Multiplicative algorithms for constructing D-optimal designs with linear constraints on weights

Radoslav Harman

Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava

Isaac Newton Institute
for Mathematical Sciences
Cambridge
8 December 2011

Overview of the talk

- $D_{\mathcal{H}}$-optimality and the multiplicative algorithm
- Constrained D-optimality viewed as $D_{\mathcal{H}}$-optimality
- The barycentric algorithm
- D-optimally augmented designs
- D-optimal stratified designs
- Concluding remarks

The problem of $D_{\mathcal{H}}$-optimality

Assume a statistical experiment with an unknown parameter $\beta \in \mathbb{R}^m$ consisting of a series of trials and an experimental domain $\mathcal{X}$ with $|\mathcal{X}| = n < \infty$ elements.

An (approximate) design is a probability measure on $\mathcal{X}$. Any design $\mathbf{w}$ can be uniquely represented by an n-dimensional vector $\mathbf{w}$ of “design weights” having components $(\mathbf{w})_x = \mathbf{w}(\{x\})$ for $x \in \mathcal{X}$. If we view designs as vectors of weights, then the set of all designs is the probability n-simplex $\mathbb{P}^n$.

For each $x \in \mathcal{X}$ let $\mathbf{H}(x)$ be a known “elementary information matrix” belonging to the set $\mathcal{S}_+^m$ of $m \times m$ nonnegative definite matrices, which represents the information gained from a single trial in the design point x.

The sequence $\mathcal{H} = (\mathbf{H}(x))_{x \in \mathcal{X}}$ will be called “the model”.

The problem of $D_{\mathcal{H}}$-optimality

Our fundamental assumption is that for a design $\mathbf{w} \in \mathbb{P}^n$ the amount of information about the unknown parameter β obtained from the entire experiment is proportional to the "information matrix"

$$\mathbf{M}_{\mathcal{H}}(\mathbf{w}) = \int_{\mathcal{X}} \mathbf{H}(x)\mathrm{d}\mathbf{w}(x) = \sum_{x \in \mathcal{X}} \mathbf{w}_x \mathbf{H}(x).$$

We will assume that $\mathbf{H}(x) \neq \mathbf{0}$ for any $x \in \mathcal{X}$ and the set of all possible information matrices $\mathcal{M}_{\mathcal{H}} = \mathrm{conv}\{\mathcal{H}\}$ contains a nonsingular matrix.

By $\mathbb{P}^n_+(\mathcal{H})$ we will denote the set of all "regular" designs for the model $\mathcal{H}$, i.e., the set of all designs $\mathbf{w} \in \mathbb{P}^n$ such that $\mathbf{M}_{\mathcal{H}}(\mathbf{w})$ is nonsingular.

The assumption of additivity of information is often satisfied for models with independent trials. The most classic example is the linear regression model with independent homoscedastic real-valued observations $Y(x)$ satisfying $E(Y(x)) = \mathbf{f}'(x)\beta$, where the elementary information matrices are $\mathbf{H}(x) = \mathbf{f}(x)\mathbf{f}'(x)$ for all $x \in \mathcal{X}$.

The problem of $D_{\mathcal{H}}$-optimality

We define the criterion of D-optimality $\Phi : \mathcal{S}_+^m \to [0, \infty)$ by $\Phi(\mathbf{M}) = \det^{1/m}(\mathbf{M})$. A design $\mathbf{w}^* \in \mathbb{P}^n$ will be called $D_{\mathcal{H}}$-optimal (on the set $\mathbb{P}^n$), iff

$$\Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w}^*)) = \sup\{\Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w})) : \mathbf{w} \in \mathbb{P}^n\}.$$

The matrix $\mathbf{M}_{\mathcal{H}}(\mathbf{w}^*)$ will be called the $D_{\mathcal{H}}$-optimal information matrix.

The problem of $D_{\mathcal{H}}$-optimality shares of many “nice” properties with the standard problem of D-optimality (see Ucinski 2005, Harman and Trnovska 2009, and elsewhere, such as in Fedorov and Hackl 1997 for important special cases):

- There exists at least one $D_{\mathcal{H}}$-optimal design.
- The $D_{\mathcal{H}}$-optimal information matrix is regular and unique.
- Simple “equivalence theorem” for $D_{\mathcal{H}}$-optimality.
- Nontrivial upper bounds on the $D_{\mathcal{H}}$-optimal weights.

The problem of $D_{\mathcal{H}}$-optimality

For any $\mathbf{w} \in \mathbb{P}_+^n(\mathcal{H})$ we define its "variance vector" to be

$$\mathbf{d}_{\mathcal{H}}(\mathbf{w}) = \left(\mathrm{tr}\left[\mathbf{H}(x)\mathbf{M}_{\mathcal{H}}^{-1}(\mathbf{w})\right]\right)'_{x\in\mathcal{X}} \in \mathbb{R}_+^n.$$

Theorem (Harman and Pronzato 2007, Harman and Trnovska 2009)

Let $\mathbf{w} \in \mathbb{P}_+^n(\mathcal{H})$, let $\mathbf{w}^*$ be $D_{\mathcal{H}}$-optimal. Then

$$\frac{\Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w}))}{\Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w}^*))} \geq \frac{m}{\max_{x\in\mathcal{X}}(\mathbf{d}_{\mathcal{H}}(\mathbf{w}))_x}.$$

For $\epsilon > 0$ let

$$h_m(\epsilon) = m\left(1 + \epsilon/2 - \sqrt{\epsilon(4+\epsilon-4/m)}/2\right).$$

If $(\mathbf{w}^*)_y > 0$ for some $y \in \mathcal{X}$, then

$$(\mathbf{d}_{\mathcal{H}}(\mathbf{w}))_y \geq h_m\left(\max_{x\in\mathcal{X}}(\mathbf{d}_{\mathcal{H}}(\mathbf{w}))_x - m\right).$$

Multiplicative algorithm for $D_{\mathcal{H}}$-optimality

The $D_{\mathcal{H}}$-optimal design can be calculated using a "multiplicative" algorithm (generalizing the classic algorithm developed by Torsney, Titterington and others) defined by the "updating rule" $T_{\mathcal{H}} : \mathbb{P}^n_+(\mathcal{H}) \to \mathbb{P}^n$, where

$$T_{\mathcal{H}}(\mathbf{w}) = \frac{1}{m}\mathrm{diag}(\mathbf{w})\mathbf{d}_{\mathcal{H}}(\mathbf{w}).$$

The transformation $T_{\mathcal{H}}$ is "monotonic" in the sense that for any $\mathbf{w} \in \mathbb{P}^n_+(\mathcal{H})$ we have (Ucinski 2005, Harman and Trnovska 2009)

$$\det(\mathbf{M}_{\mathcal{H}}(\mathbf{w})) \leq \det(\mathbf{M}_{\mathcal{H}}(T_{\mathcal{H}}(\mathbf{w}))).$$

Let $\mathbf{w}_1$ be regular and let $\mathbf{w}_t = T_{\mathcal{H}}(\mathbf{w}_{t-1})$, $t = 2, 3, \ldots$.

Theorem

$(\mathbf{M}_{\mathcal{H}}(\mathbf{w}_t))_{t=1}^{\infty}$ *converges to the $D_{\mathcal{H}}$-optimal information matrix .*

The most careful proof is in Yu 2010 (and its arxiv version).

D-optimal designs with linear constraints on weights

Assume the model $\mathcal{H} = (\mathbf{H}(x))_{x \in \mathcal{X}}$, where $\mathbf{H}(x) = \mathbf{f}(x)\mathbf{f}'(x)$, $x \in \mathcal{X}$.

Let $\mathbb{Q}^n \subseteq \mathbb{P}^n$ be a convex polytope. The aim is to find a design $\mathbf{w}^* \in \mathbb{Q}^n$ such that $\Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w}^*)) \geq \Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w}))$ for all $\mathbf{w} \in \mathbb{Q}^n$.

Note that any system of linear constraints on $\mathbf{w}$ leads to some convex polytope $\mathbb{Q}^n$ of permissible designs.

Let $\mathbf{Q} = (\mathbf{q}_{\tilde{x}})_{\tilde{x} \in \tilde{\mathcal{X}}}$ be the $n \times \tilde{n}$ matrix of extreme vectors of $\mathbb{Q}^n$.

The function $\mathbf{M}_{\mathcal{H}} : \mathbb{P}^n \to \mathcal{M}_{\mathcal{H}}$ is linear, which means that the set of all information matrices of designs $\mathbf{w} \in \mathbb{Q}^n$ is a convex polytope in $\mathcal{M}_{\mathcal{H}}$ generated by $\tilde{\mathbf{H}}(\tilde{x}) = \mathbf{M}_{\mathcal{H}}(\mathbf{q}_{\tilde{x}})$, $\tilde{x} \in \tilde{\mathcal{X}}$.

Thus, the problem of linearly constrained *D*-optimality is equivalent to the problem of $D_{\tilde{\mathcal{H}}}$-optimality, where

$$\tilde{\mathcal{H}} = (\tilde{\mathbf{H}}(\tilde{x}))_{\tilde{x} \in \tilde{\mathcal{X}}} = (\mathbf{M}_{\mathcal{H}}(\mathbf{q}_{\tilde{x}}))_{\tilde{x} \in \tilde{\mathcal{X}}}.$$

(cf. Wynn 1982, Torsney 1988, Ucinski and Patan 2007)

D-optimal designs with linear constraints on weights

We have $\mathbb{Q}^n = \mathrm{conv}\{\mathbf{q}_{\tilde{x}} : \tilde{x} \in \tilde{\mathcal{X}}\}$, which implies that $\mathbf{w} \in \mathbb{Q}^n$ if and only if there exists a "corresponding" $\tilde{\mathbf{w}} \in \mathbb{P}^{\tilde{n}}$ such that $\mathbf{w} = \mathbf{Q}\tilde{\mathbf{w}}$.

Lemma

If $\mathbf{w} \in \mathbb{Q}^n$ and $\tilde{\mathbf{w}} \in \mathbb{P}^{\tilde{n}}$ satisfy $\mathbf{w} = \mathbf{Q}\tilde{\mathbf{w}}$, then $\mathbf{M}_{\tilde{\mathcal{H}}}(\tilde{\mathbf{w}}) = \mathbf{M}_{\mathcal{H}}(\mathbf{w})$, hence $\mathbf{w}$ is regular iff $\tilde{\mathbf{w}}$ is regular. In such a case

$$\mathbf{d}_{\tilde{\mathcal{H}}}(\tilde{\mathbf{w}}) = \mathbf{Q}'\mathbf{d}_{\mathcal{H}}(\mathbf{w}).$$

Theorem

Let $\mathbf{w} \in \mathbb{Q}^n$ be a regular design and let $\mathbf{w}^*$ be D-optimal in the set $\mathbb{Q}^n$. Then

$$\frac{\Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w}))}{\Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w}^*))} \geq \frac{m}{\max_{\tilde{x} \in \tilde{\mathcal{X}}} (\mathbf{Q}'\mathbf{d}_{\mathcal{H}}(\mathbf{w}))_{\tilde{x}}}.$$

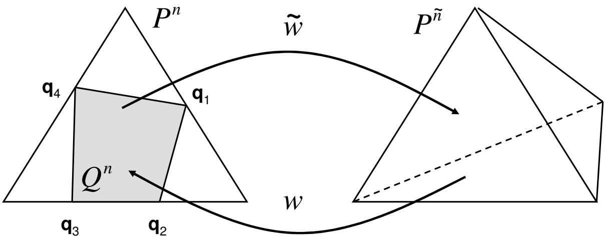
D-optimal designs with linear constraints on weights

- The D-optimum constrained design can be computed, at least in principle, by the multiplicative algorithm for $D_{\tilde{\mathcal{H}}}$-optimality.
- However, even if it is possible to find an analytic description of the extreme points of $\mathbb{Q}^n$, their number $\tilde{n} = |\tilde{\mathcal{X}}|$ is usually very large. Moreover, the direct application of the multiplicative algorithm for $D_{\tilde{\mathcal{H}}}$-optimality needs to store and operate with $\tilde{n}$ matrices $\tilde{\mathbf{H}}(\tilde{x})$ of size $m \times m$. This makes the numerical computations usually infeasible.

The key new idea is that under the linear constraints on weights we can use the multiplicative transformation $T_{\tilde{\mathcal{H}}}$ operating on $\mathbb{P}^{\tilde{n}}$, but not necessarily explicitly; we can perform all the actual computations in the much smaller space $\mathbb{P}^n$, and making them with the n vectors $\mathbf{f}(x) \in \mathbb{R}^m$.

Computing constrained D-optimal designs

- Let $w : \mathbb{P}^{\tilde{n}} \to \mathbb{Q}^n$ be defined by $w(\tilde{\mathbf{w}}) = \mathbf{Q}\tilde{\mathbf{w}}$, for $\tilde{\mathbf{w}} \in \mathbb{P}^{\tilde{n}}$. We have $\mathbf{M}_{\mathcal{H}}(w(\cdot)) = \mathbf{M}_{\tilde{\mathcal{H}}}(\cdot)$ on $\mathbb{P}^{\tilde{n}}$.
- Let $\tilde{w} : \mathbb{Q}^n \to \mathbb{P}^{\tilde{n}}$ be continuous, $\mathbf{w} = \mathbf{Q}\tilde{w}(\mathbf{w})$ for $\mathbf{w} \in \mathbb{Q}^n$. We have $\mathbf{M}_{\mathcal{H}}(\cdot) = \mathbf{M}_{\tilde{\mathcal{H}}}(\tilde{w}(\cdot))$ on $\mathbb{Q}^n$.
- $\tilde{w}(\mathbf{w})$ is the "Choquet probability measure" or the vector of "generalized barycentric coordinates" of $\mathbf{w} \in \mathbb{Q}^n$.
- The mapping $\tilde{w}$ is (usually) nonlinear and not unique.

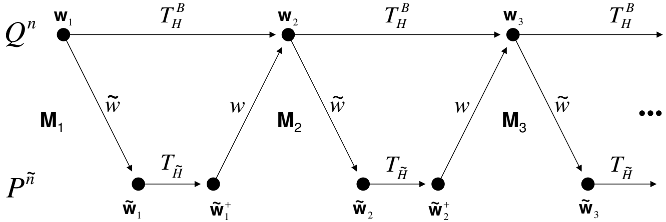


The barycentric algorithm

Using w, $\tilde{w}$, $T_{\tilde{\mathcal{H}}}$ we can define the "barycentric" updating rule:

$$T^B_{\mathcal{H}}(\cdot) = w(T_{\tilde{\mathcal{H}}}(\tilde{w}(\cdot))) = \frac{1}{m}\mathbf{Q}\mathrm{diag}(\tilde{w}(\cdot))\mathbf{Q}'\mathbf{d}_{\mathcal{H}}(\cdot).$$

Let $\tilde{w}(\mathbf{w}_1) > \mathbf{0}$, $\mathbf{w}_{t+1} = T^B_{\mathcal{H}}(\mathbf{w}_t)$, $\mathbf{M}_t = \mathbf{M}_{\mathcal{H}}(\mathbf{w}_t)$ for $t = 1, 2, \ldots$.



Theorem

The sequence $(\mathbf{M}_{\mathcal{H}}(\mathbf{w}_t))_{t=1}^{\infty}$ *converges to some* $\mathbf{M}_\infty \in \mathcal{M}_{\tilde{\mathcal{H}}}$.

D-optimal augmented designs

Let $\mathbf{b} \in \mathbb{R}^n$ have nonnegative components and $\mathbf{b}'\mathbf{1} < 1$. Let $\mathbb{Q}^n$ be the set of all design $\mathbf{w}$ bounded from below by $\mathbf{b}$, i.e.

$$\mathbb{Q}^n = \{\mathbf{w} \in \mathbb{P}^n : (\mathbf{w})_x \geq (\mathbf{b})_x \text{ for all } x \in \mathcal{X}\}.$$

- This type of constraints occurs, e.g., if we want to enlarge the existing set of trials in a D-optimal way (cf. Chapter 19 in Atkinson et al 2007).

The set of extreme vectors of $\mathbb{Q}^n$ has only n elements and the corresponding matrix is $\mathbf{Q} = \mathbf{b}\mathbf{1}' + (1 - \mathbf{b}'\mathbf{1})\mathbf{I}$.

Theorem

Let $\mathbf{w} \in \mathbb{Q}^n$ be regular, and let $\mathbf{w}^*$ be a D-optimal design in $\mathbb{Q}^n$.

$$\frac{\Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w}))}{\Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w}^*))} \geq \frac{m}{(1 - \mathbf{b}'\mathbf{1}) \max_{x \in \mathcal{X}} (\mathbf{d}_{\mathcal{H}}(\mathbf{w}))_x + \mathbf{b}'\mathbf{d}_{\mathcal{H}}(\mathbf{w})}.$$

Barycentric algorithm for D-optimal augmentation

For this very simple type of constraints, barycentric coordinates can be defined by $\tilde{w}(\mathbf{w}) = \mathbf{Q}^{-1}\mathbf{w}$ for all $\mathbf{w} \in \mathbb{Q}^n$.

The resulting barycentric updating rule has the form:

$$T^B_{\mathcal{H}}(\mathbf{w}) = \mathbf{b} + \frac{1}{m}(\mathbf{w} - \mathbf{b}) \odot ((1 - \mathbf{b}'\mathbf{1})\mathbf{d}_{\mathcal{H}}(\mathbf{w}) + (\mathbf{b}'\mathbf{d}_{\mathcal{H}}(\mathbf{w}))\mathbf{1}),$$

where $\odot$ is the componentwise multiplication. Note that implementing the resulting algorithm is extremely simple in any matrix-based programming language such as Matlab or R.

Theorem

Let $\mathbf{w}_1$ is such that $(\mathbf{w})_x > (\mathbf{b})_x$ for all $x \in \mathcal{X}$, and let $\mathbf{w}_{t+1} = T^B_{\mathcal{H}}(\mathbf{w}_t)$ for $t = 1, 2, ...$. Then the sequence $(\mathbf{M}_{\mathcal{H}}(\mathbf{w}_t))_{t=1}^{\infty}$ monotonically converges to the information matrix of a design that is D-optimal in $\mathbb{Q}^n$.

D-optimal stratified design

Let $\mathcal{X}_1, ..., \mathcal{X}_k \neq \emptyset$ be a decomposition of $\mathcal{X}$ into "strata", and let $s_1, ..., s_k > 0$, and $\sum_{j=1}^{k} s_j = 1$. Let $\mathbb{Q}^n$ be the set of all "stratified" designs, i.e.

$$\mathbb{Q}^n = \{\mathbf{w} \in \mathbb{P}^n : \sum_{x \in \mathcal{X}_j} (\mathbf{w})_x = s_j \text{ for all } j = 1, ..., k\},$$

- The most important special case is the "marginally restricted design" (Cook and Thibodeau 1980), where k is equal to the number of levels of one factor.
- Another special case is the problem of computing a D-optimal design with one or more of the values $\mathbf{w}(\{x\})$ fixed, which can be used, e.g., in branch-and-bound algorithms for calculating exact D-optimal designs (Ucinski and Patan 2007).

D-optimal stratified design

Let $\tilde{\mathcal{X}}$ be the set of all sequences $\tilde{x} = (x_1, \cdots, x_k)$ satisfying $x_j \in \mathcal{X}_j$ for all $j = 1, ..., k$. For $\tilde{x} = (x_1, \cdots, x_k) \in \tilde{\mathcal{X}}$ let $\mathbf{q}_{\tilde{x}} = \sum_{j=1}^k s_j \mathbf{e}_{x_j}$, where $\mathbf{e}_i$ is the i-th standardized unit vector.

It can be verified that $\{\mathbf{q}_{\tilde{x}} : \tilde{x} \in \tilde{\mathcal{X}}\}$ is the set of all extreme vectors of the polytope $\mathbb{Q}^n$. Note that $\tilde{n} = \prod_{j=1}^k |\mathcal{X}_j|$.

Theorem

Let $\mathbf{w}$ *be a regular stratified design, and let* $\mathbf{w}^*$ *be a D-optimal stratified design. Then*

$$\frac{\Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w}))}{\Phi(\mathbf{M}_{\mathcal{H}}(\mathbf{w}^*))} \geq \frac{m}{\sum_{j=1}^k s_j \max_{x \in \mathcal{X}_j} (\mathbf{d}_{\mathcal{H}}(\mathbf{w}))_x}.$$

(cf. Cook and Thibodeau 1980, Chapter 4 in Fedorov and Hackl 1997)

Barycentric algorithm for D-optimal stratified designs

Barycentric coordinates $\tilde{w} : \mathbb{Q}^n \to \mathbb{P}^{\tilde{n}}$ can be defined by

$$(\tilde{w}(\mathbf{w}))_{\tilde{x}} = \prod_{j=1}^{k} \frac{(\mathbf{w})_{x_j}}{s_j};\ \tilde{x} = (x_1, ..., x_k)' \in \tilde{\mathcal{X}}.$$

The resulting barycentric updating rule has the form:

$$T_{\mathcal{H}}^{B}(\mathbf{w}) = \mathbf{w} + \frac{1}{m}\mathbf{w} \odot \left[(\mathbf{Ks}) \odot \mathbf{d}_{\mathcal{H}}(\mathbf{w}) - \mathbf{K}(\mathbf{K}'(\mathbf{d}_{\mathcal{H}}(\mathbf{w}) \odot \mathbf{w})) \right],$$

where $(\mathbf{K})_{ij} = 1$ iff $x_i \in \mathcal{X}_j$ (otherwise 0), and $\mathbf{s} = (s_1, ..., s_k)'$.

Theorem

Let $\mathbf{w}_1$ have all components positive and let $\mathbf{w}_{t+1} = T_{\mathcal{H}}^{B}(\mathbf{w}_t)$ for $t = 1, 2, ...$ Then the sequence $(\mathbf{M}_{\mathcal{H}}(\mathbf{w}_t))_{t=1}^{\infty}$ monotonically converges to the information matrix of a design that is D-optimal in $\mathbb{Q}^n$.

Further notes

- Since the barycentric algorithm is based on the multiplicative algorithm, we can easily combine it with the “deletion rules” (Harman and Pronzato 2007, Harman and Trnovska 2009) which can significantly speed up the computations. It is also very probable, that additional speed up can be based on a “clustering approach” as in Martin-Martin and Torsney 2006, or a combination with other optimization methods, such as in Yu 2010.
- For the problem of stratified *D*-optimality, it is possible to suggest a different multiplicative heuristics generalizing the one in Martin-Martin, Torsney and Lopez-Fidalgo 2007. The heuristics has been tested and it is more rapid than the barycentric algorithm, but its convergence properties (monotonicity, convergence to the global optimum) are unclear. Moreover, for other types of constraints, there is no similarly simple heuristics at all.

Further notes

- Many practical design problems require linear constraints on the design weights (see, e.g., Cook and Fedorov 1995) but my real motivation to study the algorithms for computing *D*-optimal approximate constrained designs is that they are crucial for some algorithms for computing exact designs of experiments (closely related to the branch-and-bound algorithms, see Welch 1982, Ucinski and Patan 2007). In a sense, the design problems involving constraints on weights form a link between the areas of approximate and exact designs. (Use the blackboard to explain the idea for block designs with blocks of size 2.)

Further notes

- Many practical design problems require linear constraints on the design weights (see, e.g., Cook and Fedorov 1995) but my real motivation to study the algorithms for computing D-optimal approximate constrained designs is that they are crucial for some algorithms for computing exact designs of experiments (closely related to the branch-and-bound algorithms, see Welch 1982, Ucinski and Patan 2007). In a sense, the design problems involving constraints on weights form a link between the areas of approximate and exact designs. (Use the blackboard to explain the idea for block designs with blocks of size 2.)

Thank you!