Lecture #5 Notes Summary

$\boldsymbol{c}-$optimality, Elfving's theorem

Scalar optimality

Definition 1 (Feasibility cone)**.** Let K be an $m \times r-$matrix. The *Feasibility cone* $\Xi(K)$ of K is the set of all designs ξ such that the estimability condition $\operatorname{im} K \subseteq \operatorname{im} A(\xi)^T = \operatorname{im} M(\xi)$ is satisfied.

The quantity $\zeta = K^T\boldsymbol{\theta}$ is said to be *estimable* if the feasibility cone is nonempty.

Recall our first informal statement of the approximate optimal design problem: *Find a design* $\xi = \{\boldsymbol{x}, \boldsymbol{w}\} \in \Xi(K)$ *that minimizes (in a certain sense) the matrix*

$$K^T \left(\sum_{i=1}^{s} w_i\, \boldsymbol{a}(\boldsymbol{x}_i)\boldsymbol{a}(\boldsymbol{x}_i)^T \right)^{-} K$$

There is one case where this problem is well stated. Namely, when $K \in m \times r$ is a column vector (i.e., $r = 1$). In this case, we write $K = \boldsymbol{c}$, and we want to estimate a single scalar quantity, $\zeta = \boldsymbol{c}^T\boldsymbol{\theta}$. That is, we are interested by a particular linear combination of the parameters. Particular cases are:

- estimation of a single parameter ($\zeta = \theta_1$);
- estimation of the difference between two parameters ($\zeta = \theta_i - \theta_j$);
- estimation of the average of the parameters ($\zeta = \sum_{i=1}^{m} \frac{1}{m}\theta_i$).

The approximate optimal design problem can be rewritten as:

$$\begin{aligned} \min_{\xi=\{\boldsymbol{x},\boldsymbol{w}\}} \quad & \boldsymbol{c}^T \left(\sum_{i=1}^{s} w_i\, \boldsymbol{a}(\boldsymbol{x}_i)\boldsymbol{a}(\boldsymbol{x}_i)^T \right)^{-} \boldsymbol{c} \\ \text{s.t.} \quad & \xi \in \Xi(\boldsymbol{c}). \end{aligned} \tag{1}$$

We say that a design solving Problem (1) is $\boldsymbol{c}-$*optimal*. But in fact, we can almost ignore the constraint $\xi \in \Xi(\boldsymbol{c})$, because we can set $\boldsymbol{c}^T M(\xi)^- \boldsymbol{c} := +\infty$ when $\boldsymbol{c} \notin M(\xi)$.

We are next going to see that this problem has a solution that has a nice geometric construction.

Definition 2 (Elfving's set)**.** The *Elfving' set* for the design of experiments (in the linear model) is

$$E = \text{convex-hull}\left(\left\{ \pm\, \boldsymbol{a}(\boldsymbol{x}) :\ \boldsymbol{x} \in \mathcal{X} \right\}\right).$$

We denote its boundary by ∂E.

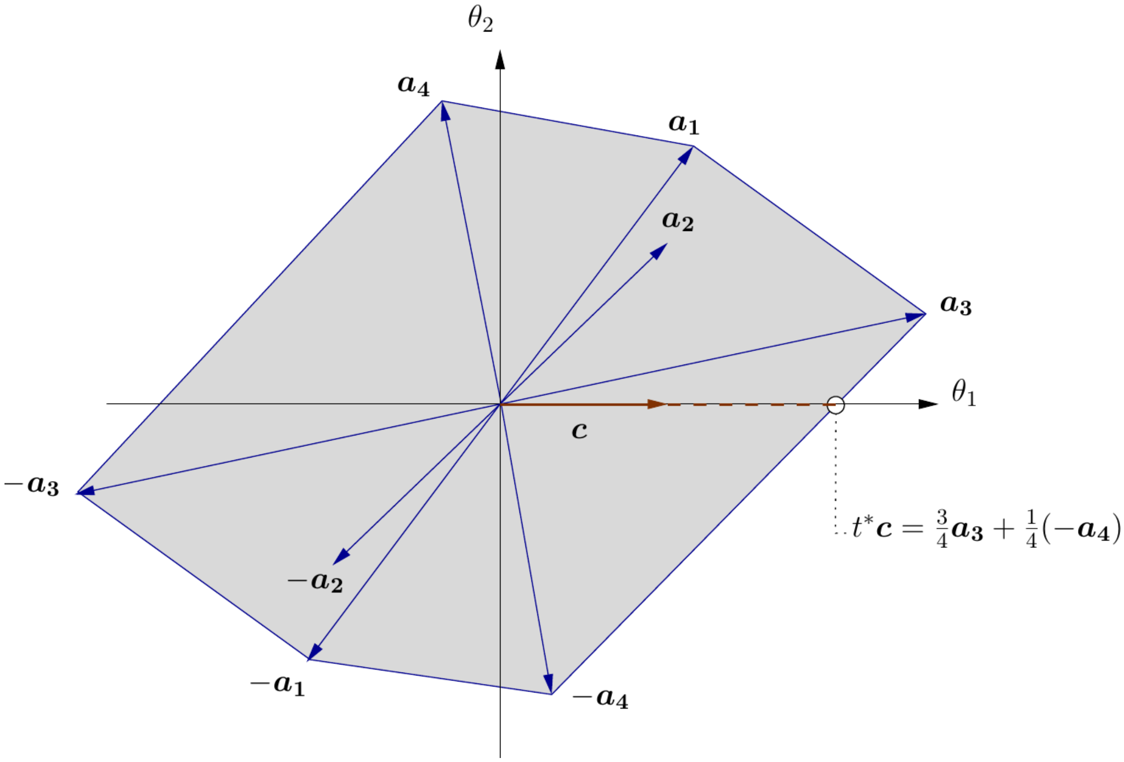
Theorem 1 (Elfving)**.** *An approximate design* $\xi = \{\boldsymbol{x}, \boldsymbol{w}\}$ *is* $\boldsymbol{c}-$*optimal if and only if there exists scalars* $\varepsilon_i = \pm 1$ *and a positive real* t *such that*

$$t\boldsymbol{c} = \sum_{i=1}^{s} \varepsilon_i\, w_i\, \boldsymbol{a}(\boldsymbol{x}_i) \in \partial E.$$

Moreover, $t^{-2} = \boldsymbol{c}^T M(\xi)^- \boldsymbol{c}$ *is the minimal variance.*

This theorem has a nice geometrical interpretation: take the intersection of the straight line directed by $\boldsymbol{c}$ and the boundary of the Elfving's set. This point can be expressed as a convex combination of extreme points of E, and the weights of this combination represent the $\boldsymbol{c}-$optimal design. In the next example, $\boldsymbol{c}$ is directed along θ_1, i.e., we just want to estimate θ_1 and we do not care about the accuracy of the estimation of θ_2. The intersection point (between $\boldsymbol{c}\mathbb{R}$ and ∂E) is $\frac{3}{4}\boldsymbol{a}_3 + \frac{1}{4}(-\boldsymbol{a}_4)$, and so a $\boldsymbol{c}-$optimal design is

$$\xi = \begin{pmatrix} \boldsymbol{x}_3 & \boldsymbol{x}_4 \\ 3 & 1 \end{pmatrix}.$$



We will sketch a geometrical proof of this result. But we first need the following lemma:

Lemma 2. *Let $\xi = \{\boldsymbol{x}, \boldsymbol{w}\} \in \Xi(\boldsymbol{c})$, and assume w.l.o.g. that $\boldsymbol{w} > 0$ (we can remove the points with a zero weight from the design). Then,* $\mathbf{c}^T M(\xi)^- \mathbf{c} = \min \left\{ \sum_{i=1}^s \frac{u_i^2}{w_i} :\ \boldsymbol{u} \in \mathbb{R}^s,\ \sum_i u_i \boldsymbol{a}(\boldsymbol{x_i}) = \boldsymbol{c} \right\}$.

Proof. As a direct consequence of Gauss markov theorem:

$$K^T (A^T A)^- K = \min\nolimits_{\preceq} \{H^T H :\ A^T H = K\},$$

we have

$$\boldsymbol{c}^T M(\xi)^- \boldsymbol{c} = \min \left\{ \sum_{i=1}^s v_i^2 :\ \boldsymbol{v} \in \mathbb{R}^s,\ \sum_i v_i \sqrt{w_i} \boldsymbol{a}(\boldsymbol{x_i}) = \boldsymbol{c} \right\}.$$

(Set $K := \boldsymbol{c}$ and $A := \left[\sqrt{w_1}\boldsymbol{a}(\boldsymbol{x}_1), \cdots, \sqrt{w_s}\boldsymbol{a}(\boldsymbol{x}_s)\right]^T$, so that $M(\xi) = A^T A$.) The result of the lemma is obtained after the change of variable $u_i = \sqrt{w_i} v_i$.

□

Proof of Elfving's theorem. Let $\xi = \{\boldsymbol{x}, \boldsymbol{w}\}$ be a design satisfying the estimability condition $\xi \in \Xi(\boldsymbol{c})$. We define the set $\mathcal{A}(\xi) := \{\boldsymbol{x} :\ \boldsymbol{x} \in \operatorname{im} M(\xi),\ \boldsymbol{x}^T M(\xi)^- \boldsymbol{x} \leq 1\}$, which is an ellipsoid (it can be a *flat* ellipsoid if $M(\xi)$ is singular). Denote by t the largest scalar $\alpha \geq 0$ such that $\alpha\boldsymbol{c} \in \mathcal{A}(\xi)$ and by t^* the largest $\alpha \geq 0$ such that $\alpha\boldsymbol{c} \in E$.

The proof is in 3 steps:

1. We show that $\mathcal{A}(\xi) \subseteq E$
2. This implies $\boldsymbol{c}^T M(\xi)^- \boldsymbol{c} = t^{-2} \geq (t^*)^{-2}$
3. We show that this bound is attained iff ξ is a design satisfying the condition of the theorem.

1. Let $\boldsymbol{x}$ be a point in $\mathcal{A}(\xi)$. By Lemma 2 we know that

$$\boldsymbol{x}^T M(\xi)^- \boldsymbol{x} = \min \left\{ \sum_{i=1}^{s} \frac{u_i^2}{w_i} :\ \boldsymbol{u} \in \mathbb{R}^s,\ \sum_i u_i \boldsymbol{a}(\boldsymbol{x_i}) = \boldsymbol{x} \right\}.$$

This implies the existence of a vector $\boldsymbol{u}$ so that $\sum_{i=1}^{s} \frac{u_i^2}{w_i} \leq 1$ and $\sum_i u_i \boldsymbol{a}(\boldsymbol{x_i}) = \boldsymbol{x}$. Now, we apply Cauchy Schwarz:

$$1 \geq \sum_{i=1}^{s} \frac{u_i^2}{w_i} = \sum_{i=1}^{s} \left(\frac{|u_i|}{\sqrt{w_i}} \right)^2 \underbrace{\sum_{i=1}^{s} (\sqrt{w_i})^2}_{=1} \geq \left(\sum_{i=1}^{s} |u_i| \right)^2$$

So for $\epsilon_i = \text{sign}(u_i)$ we have $\boldsymbol{x} = \sum_i \epsilon_i |u_i| \boldsymbol{a}(\boldsymbol{x_i})$, and $\sum_i |u_i| \leq 1$. This shows that $\boldsymbol{x} \in E = \text{convex-hull}(\pm a(\mathcal{X}))$.

2. In particular, we have $\alpha\boldsymbol{c} \in \operatorname{im} M(\xi)$ for all $\alpha \geq 0$, so $\alpha\boldsymbol{c} \in \mathcal{A}(\xi)$ iff

$$(\alpha\boldsymbol{c})^T M(\xi)^- (\alpha\boldsymbol{c}) \leq 1 \iff \alpha^2 \boldsymbol{c}^T M(\xi)^- \boldsymbol{c} \leq 1 \iff \alpha^{-2} \geq \boldsymbol{c}^T M(\xi)^- \boldsymbol{c}.$$

So $t^{-2} = \boldsymbol{c}^T M(\xi)^- \boldsymbol{c}$, and from point 1. we know that $t \leq t^*$, hence the bound $\boldsymbol{c}^T M(\xi)^- \boldsymbol{c} \geq (t^*)^{-2}$.

3. Let $t^*\boldsymbol{c} = \sum_i \epsilon_i w_i^* \boldsymbol{a}(\boldsymbol{x}_i^*) \in \partial E$ as in the theorem, and define $\xi^* = \{\boldsymbol{x}^*, \boldsymbol{w}^*\}$. By Lemma 2 we have

$$\boldsymbol{c}^T M(\xi^*)^- \boldsymbol{c} = \min \left\{ \sum_{i=1}^{s} \frac{u_i^2}{w_i^*} :\ \boldsymbol{u} \in \mathbb{R}^s,\ \sum_i u_i \boldsymbol{a}(\boldsymbol{x_i^*}) = \boldsymbol{c} \right\}.$$

In partucular, if we set $u_i = (t^*)^{-1} \epsilon_i w_i^*$ we have $\boldsymbol{c}^T M(\xi^*)^- \boldsymbol{c} \leq \sum_{i=1}^{s} (t^*)^{-2} w_i^* = (t^*)^{-2}$. So $\boldsymbol{c}^T M(\xi^*)^- \boldsymbol{c} = (t^*)^{-2}$.

Conversely, assume that $t\boldsymbol{c} \in \partial E$. It implies that the Cauchy Schwarz inequality used above must be an equality for $\boldsymbol{x} = t\boldsymbol{c} = \sum_i u_i \boldsymbol{a}(\boldsymbol{x_i})$, otherwise the weights $|u_i|$ do not sum to 1 and we have $t'\boldsymbol{c} \in E$ for $t' = \frac{t}{\sum_i |u_i|} > t$. So the $|u_i|/\sqrt{w_i}$ are proportional to $\sqrt{w_i}$, which implies $w_i = \pm u_i$ and $t\boldsymbol{c} = \sum_i \epsilon_i w_i \boldsymbol{a}(\boldsymbol{x_i}) \in \partial E$. That is, ξ satisfies the condition of the theorem. ☐

Corollary 3 (LP-formulation)**.** *Assume that the experimental region $\mathcal{X}$ is finite, or that the candidate design points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ are given (recall that in that case, we write $\boldsymbol{a_i}$ instead of $\boldsymbol{a}(\boldsymbol{x}_i)$). Then, the optimal weights of a $\mathbf{c}-$optimal design and the quantity t such that t^{-2} is the minimal variance can be computed by the following linear program:*

$$\begin{aligned} \max_{t,\boldsymbol{\lambda}} \quad & t \\ \text{s.t.} \quad & t\boldsymbol{c} = \sum_{i=1}^{N} \lambda_i \boldsymbol{a_i} \\ & \sum_{i=1}^{N} \underbrace{|\lambda_i|}_{w_i} \leq 1, \end{aligned}$$

(we have left absolute values for more readibility).

Exercises

1. On the finite experimental region $\mathcal{X} = \{1,2,3\}$, $a_1 = [1,0]^T,\ a_2 = [4,1]^T,\ a_3 = [4,2]^T$, find the optimal design for θ_1 (that is, the $(1,0)-$optimal design).

2. Consider the line fit model $y(x) = ax + b$ over $x \in \mathcal{X} = [-1,1]$, where the unknown parameter is $\boldsymbol{\theta} = [a,b]$. Show that the optimal design for a is unique, but that there are an infinity of optimal designs for b.

3. In the quadratic fit model $y(x) = \theta_1 + \theta_2 x + \theta_3 x^2$ over $x \in \mathcal{X} = [-1,1]$, find the optimal designs for

   - $\theta_1 + \theta_2 + \theta_3$

   - $\theta_1 - \theta_3$

   - $2\theta_2 + \theta_3$ [*Hint:* try to write the vector $[1, 2t, t]$ as a barycenter of 2 vectors of the form $[1, x, x^2]$ and $[-1, -y, -y^2]$].