Lecture #9 Notes Summary

SDP formulations of some optimal experimental design problems

The SDP approach to compute optimal designs

In this lecture, we consider the Φ_K-optimal design problem over a finite regression range $\Xi \equiv \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s\}$. Recall that in this situation, a design is simply a vector of weights $\boldsymbol{w} \in \Delta_s := \{\boldsymbol{w} \geq 0 : \sum_{i=1}^s w_i = 1\}$, and we write $\boldsymbol{a}_i$ instead of $\boldsymbol{a}(\boldsymbol{x}_i)$:

$$\max_{\boldsymbol{w} \in \Xi(K)} \quad \Phi\Big(M_K(\boldsymbol{w})\Big), \tag{1}$$

where $M_K(\boldsymbol{w}) := \Big(K^T \left(\sum_{i=1}^s w_i \boldsymbol{a}_i \boldsymbol{a}_i^T\right)^- K\Big)^{-1}$ and $\boldsymbol{w} \in \Xi_K$ means that $M(\boldsymbol{w}) = \sum_i w_i \boldsymbol{a}_i \boldsymbol{a}_i^T$ contains the columns of K in its range, so $M_K(\boldsymbol{w})$ is well defined.

In fact, it is possible to extend the definition of M_K by continuity to designs $\boldsymbol{w} \notin \Xi(K)$, in which case $M_K(\xi)$ is singular.

Theorem 1. *Let K be a $m \times k-$matrix of full column rank, and let $p < 1$. Assume that the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s$ span the columns of K. Then, a $(\Phi_p)_K-$optimal design exists and must be solution of the following concave maximization problem:*

$$\max_{\boldsymbol{w} \in \Delta_s} \quad \Phi_p\Big(M_K(\boldsymbol{w})\Big). \tag{2}$$

Moreover if $p \neq -\infty$ and $\boldsymbol{w}$ and $\boldsymbol{w'}$ are optimal, then $M_K(\boldsymbol{w}) = M_K(\boldsymbol{w'}) \succ 0$.

Proof. Let $\boldsymbol{w}$ be a solution of Problem (2). We show that that $M := M_K(\boldsymbol{w}) \succ 0$, which implies $\boldsymbol{w} \in \Xi(K)$, and hence $\boldsymbol{w}$ is a $(\Phi_p)_K-$optimal design. This is clear for $p \leq 0$, because $\Phi_p(X) = 0$ for all singular matrices X, and by assumption a non-singular design exists. For $0 < p < 1$, we use the duality theorem: there exists a matrix $N \succeq 0$ such that $D = K^T N K$ satisfies

$$\Phi_p(M)\Phi_p^\star(D) = \langle M, D\rangle = 1.$$

This implies $\Phi_p^\star(D) > 0$, and hence $D \succ 0$ because $\Phi_p^\star = m\Phi_q$ for a $-\infty < q < 0$, and $\Phi_q(X) = 0$ for singular matrices X. Now, let $\boldsymbol{z}$ be a vector such that $\boldsymbol{z}^T M \boldsymbol{z} = 0$. We show by contradition that $\boldsymbol{z} = 0$, from which the conclusion $M \succ 0$ follows. If $\boldsymbol{z} \neq \boldsymbol{0}$ we have:

$$\Phi_p(M)\Phi_p^\star(D + \boldsymbol{z}\boldsymbol{z}^T) \leq \langle M, D + \boldsymbol{z}\boldsymbol{z}^T\rangle = \langle M, D\rangle = \Phi_p(M)\Phi_p^\star(D) < \Phi_p(M)\Phi_p^\star(D + \boldsymbol{z}\boldsymbol{z}^T),$$

where the first inequality follows from the definition of $\Phi_p^\star$, and the second inequality comes from the strict monotonicity (w.r.t. $\preceq$) of $\Phi_p^\star = m\Phi_q$ over $\mathbb{S}_{++}^m$ (which we admit). Finaly, the unicity of the optimal information matrix $M_k(\boldsymbol{w})$ is a consequence of the strict concavity of Φ_p over $\mathbb{S}_{++}^m$ for all $p \in (-\infty, 1)$. $\square$

In this lecture, we show that Problem (2) can be reformulated as an SDP for $p = -\infty$ (E-optimality) and $p = -1$ (A-optimality). The case $p = 0$ (D-optimality) is treated in exercises. In fact, it is possible to give an SDP formulation of Problem (2) for all rational values of $p \leq 1$.

A-optimality

Recall that an A_K-optimal design must minimize trace $K^T M(\boldsymbol{w})^- K$ over $\Xi(K)$.

If $\boldsymbol{w} \in \Xi(K)$, the (extended) Schur complement lemma gives

$$\begin{pmatrix} M(\boldsymbol{w}) & K \\ K^T & U \end{pmatrix} \succeq 0 \Longleftrightarrow U \succeq K^T M(\boldsymbol{w})^- K.$$

This suggests to consider the following SDP:

$$\begin{aligned} \min_{\boldsymbol{w},U} \quad & \text{trace } U \\ s.t. \quad & \begin{pmatrix} M(\boldsymbol{w}) & K \\ K^T & U \end{pmatrix} \succeq 0 \\ & \boldsymbol{w} \in \Delta_s \end{aligned} \tag{3}$$

The $(m+k) \times (m+k)$ matrix inequality is linear in $\boldsymbol{w}$ and in the entries of $U \in \mathbb{S}^k$, so (3) is an SDP indeed.

Proposition 2. *$\boldsymbol{w}$ is A_K-optimal if and only if there is a matrix U such that $(U, \boldsymbol{w})$ is an optimal solution of (3).*

Proof. We first show that $\boldsymbol{w} \in \Xi(K)$ for all feasible $(\boldsymbol{w}, U)$. Indeed, if the LMI is satisfied, the big $(m+k) \times (m+k)$ can be expressed as $[A, B]^T [A, B]$ for some matrices A, B of respective sizes $(m+k) \times m$ and $(m+k) \times k$. So we have $\operatorname{im} K = \operatorname{im} A^T B \subseteq \operatorname{im} A^T = \operatorname{im} A^T A = \operatorname{im} M(\boldsymbol{w})$, which shows $\boldsymbol{w} \in \Xi(K)$. So we have $U \succeq K^T M(\boldsymbol{w})^- K$, which implies trace $U \geq$ trace $K^T M(\boldsymbol{w})^- K$.

Conversely, let $\boldsymbol{w}^*$ be A_K-optimal, and set $U^* = K^T M(\boldsymbol{w}^*)^- K$. Note that the pair $(\boldsymbol{w}^*, U^*)$ is feasible, so the optimal value of the problem must be trace $K^T M(\boldsymbol{w}^*)^- K$. This concludes the proof. □

E-optimality

Proposition 3. *Let $M \in \mathbb{S}^m$. Then,*

$$M \succeq \lambda I_m \Longleftrightarrow \lambda \leq \lambda_{min}(M).$$

Proof. Recall the following characterization of the smallest eigenvalue of a symmetric matrix M:

$$\lambda_{min}(M) := \inf_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\boldsymbol{x}^T M \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

Now we write

$$\begin{aligned} M \succeq \lambda I_m \Longleftrightarrow & \quad \forall \boldsymbol{x} \in \mathbb{R}^m, \quad \boldsymbol{x}^T M \boldsymbol{x} \geq \lambda \boldsymbol{x}^T \boldsymbol{x} \\ \Longleftrightarrow & \quad \lambda \leq \inf_{\boldsymbol{x} \neq 0} \frac{\boldsymbol{x}^T M \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \lambda_{min}(M). \end{aligned}$$

□

This suggests the following SDP for $E-$optimality:

$$\begin{aligned} \max_{\boldsymbol{w},\lambda} \quad & \lambda \\ s.t. \quad & M(\boldsymbol{w}) \succeq \lambda I_m \\ & \boldsymbol{w} \in \Delta_s \end{aligned} \tag{4}$$

Proposition 4. $\boldsymbol{w}$ *is* $E-$*optimal if and only if there is a scalar* λ *such that* $(\boldsymbol{w}, \lambda)$ *is an optimal solution of* (4).

Proof. We recall that an $E-$optimal design maximizes $\lambda_{min}\big(M(\boldsymbol{w})\big)$ over Δ_s. If $(\boldsymbol{w}, \lambda)$ is feasible, then we know from the previous proposition that $\lambda \leq \lambda_{min}\big(M(\boldsymbol{w})\big)$. Conversely, let $\boldsymbol{w^*}$ be $E-$optimal and set $\lambda^* = \lambda_{min}\big(M(\boldsymbol{w^*})\big)$. The pair $(\boldsymbol{w^*}, \lambda^*)$ is feasible, and so it must be optimal for the SDP (4). □

Remark 5. The SDP (4) can actually be extended to an SDP-formualtion for E_K-optimality, cf. Exercise 3.

Exercises

1. Recall the Elfving theorem for $\boldsymbol{c}-$optimality, and show that a $\boldsymbol{c}-$optimal design can be computed by *linear programming* (for the case of a finite design space $\mathcal{X}$).

2. Show that the dual problem (with respect to the polar information function $\Phi^\star$) for $D-$optimality is equivalent to

$$\begin{aligned} \max_N \quad & \det N \\ s.t. \quad & \boldsymbol{a}_i^T N \boldsymbol{a}_i \leq 1 \qquad (\forall i \in \leq 1,\ldots,m) \\ & N \succeq 0 \end{aligned}$$

   Give a geometrical interpretation to this problem

3. Show that the dual problem (with respect to the polar information function $\Phi^\star$) for E_K-optimality can be written as an SDP. Form the Lagrangian dual of this SDP and make a change of variables to find an SDP formulation of the E_K-optimal design problem.

4. SDP-representation of a geometric mean.

   In this exercise, we show (by some examples) that inequalities of the form $t \leq \prod_{i=1}^m x_i^{1/m}$ are equivalent to an LMI (with respect to the variables $t, u_1, \ldots, u_m$).

   - Let $u_1, u_2 \geq 0$. Show that the inequality $t^2 \leq u_1 u_2$ can be rewritten as a $2 \times 2-$LMI.
   - Let $\boldsymbol{x} \in \mathbb{R}_+^5$. Show that

$$\begin{aligned} t^5 \leq x_1x_2x_3x_4x_5 &\iff t^8 \leq x_1x_2x_3x_4x_5t^3 \\ &\iff \exists \boldsymbol{u} \in \mathbb{R}_+^5 : \begin{cases} u_1^2 & \leq x_1x_2, \qquad u_4^2 \leq u_1u_2, \\ u_2^2 & \leq x_3x_4, \qquad u_5^2 \leq u_3 t, \\ u_3^2 & \leq x_5 t, \qquad\ \ t^2 \leq u_4u_5, \end{cases} \end{aligned}$$

   - Conclude that $t \leq \prod_{i=1}^5 x_i^{1/5}$ can be rewritten as a big $12 \times 12-$LMI.
   - By using a similar construction, rewrite $t \leq \prod_{i=1}^9 x_i^{1/9}$ as a big LMI.

5. Let $(\boldsymbol{w}, \boldsymbol{u}, L) \in \Delta_s \times \mathbb{R}^m \times \mathbb{R}^{m \times m}$ be such that

$$\begin{aligned} (i) \quad & \begin{pmatrix} M(\boldsymbol{w}) & L \\ L^T & \mathrm{Diag}(\boldsymbol{u}) \end{pmatrix} \succeq 0 \\ (ii) \quad & L_{ii} = u_i \quad (\forall i = 1,\ldots,m) \\ (iii) \quad & L_{ij} = 0 \quad (\forall\, 1 \leq i < j \leq m) \end{aligned}$$

   - We assume that $\boldsymbol{u} > \boldsymbol{0}$ for simplicity. Set $J = L\,\mathrm{Diag}(\boldsymbol{u})^{-1/2}$. Show that $M(\boldsymbol{w}) \succeq JJ^T$.
   - Deduce that $\det M(\boldsymbol{w}) \geq (\det J)^2 = \prod_{i=1}^m u_i$ (even if some $u_i = 0$).
   - Conclude with an SDP formulation for $D-$optimality (*hint*: use the construction of exercise 4).