Exploiting Convexity in Experimental Design

Michel Baes

Niels Haverbeke

OPTEC
K.U.Leuven Optimization in Engineering Center

What can you do
with Convex Optimization?

Obviously,

solve convex optimization problems

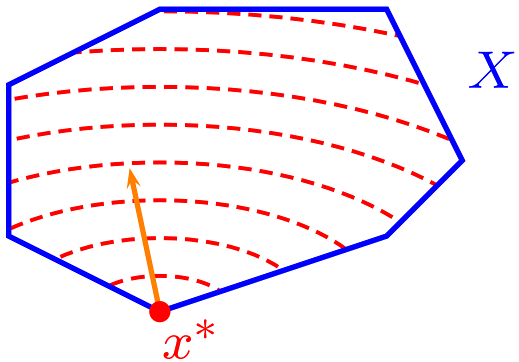
$$\begin{array}{ll} \text{min} & f(x) \\ \text{s.t.} & Ax = b \\ & x \in X \end{array}$$

f is a **convex** function
X a **convex** set
with nonempty interior

Advantage: no local optima

Kuhn-Tucker conditions are necessary and sufficient:

$$x^* \text{ optimal} \Leftrightarrow \langle f'(x^*), x - x^* \rangle \geq 0 \quad \forall x \in X \text{ with } Ax = b$$



A linear objective:
Separate the wheat from the chaff

$$\begin{array}{rl} \min & f(x) \\ \text{s.t.} & Ax = b \\ & x \in X \end{array}$$

$$\begin{array}{rl} \min & t \\ \text{s.t.} & Ax = b \\ & x \in X \\ & f(x) \le t \end{array} \quad = \quad \begin{array}{rl} \min & t \\ \text{s.t.} & Ax = b \\ & (t, x) \in Y := \{(t, x) \in \text{epi} f : x \in X\} \end{array}$$

$$\begin{array}{rl} \min & \langle \bar{c}, \bar{x} \rangle \\ \text{s.t.} & \bar{A}\bar{x} = \bar{b} \\ & \bar{x} \in \bar{X} \end{array}$$

Note: one can further make the nonlinear feasible set conic with $K := \text{cl}\{(s, y) : s > 0, y/s \in Y\}$

Get rid of "difficult" constraints with a barrier

$$p^* = \min \ \langle c, x \rangle \quad \text{s.t.} \quad Ax = b, \quad x \in X$$

X a closed **convex** set with nonempty interior

▷ **Equivalent reformulation**:
$p^* = \min\{\langle c, x\rangle + \mu\chi_X(x) : Ax = b\}$,
where χ_X is the characteristic function of Q.

▷ **Barrier function approximation**:
$p^* \simeq \min\{\langle c, x\rangle + \mu\phi(x) : Ax = b\}$, where ϕ is a *barrier* of X:
ϕ is *strongly convex* (at least convex + quadratic),
$\mathrm{dom}\phi = \mathrm{int}\, X$, and $\lim_{x\to\partial X} \phi(x) = +\infty$. We let $\mu \to 0$.

That's the strategy of Interior-Point Methods

In principle, this barrier idea works

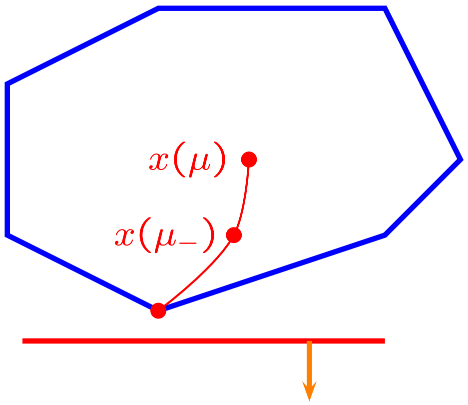
Theorem 1 *If the barrier ϕ is bounded below, then the sequence x_k of solutions for $\mu_k \to 0$ satisfies $\langle c, x_k\rangle + \mu_k \phi(x_k) \to p^*$.* ■

▷ Fix $\mu > 0$

▷ Solve by Newton's method

$$x(\mu) = \arg\min_{Ax=b} \langle c, x\rangle + \mu\phi(x)$$

▷ Decrease μ to μ_-, so that $\widehat{x}(\mu)$ is a "good" starting point



But there are some barriers for which miracles happen

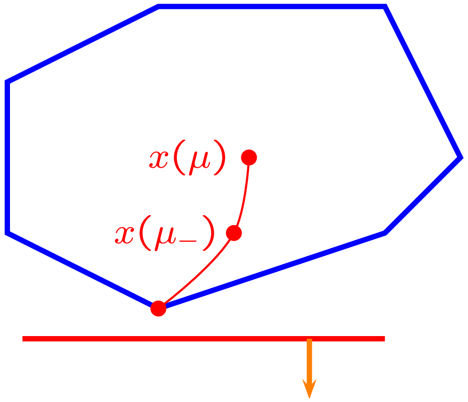
(From Yu. Nesterov and A. Nemirovski, [3])

If ϕ is a *self-concordant barrier*, then

▷ we **know** the quadratic convergence zone of Newton's method

▷ μ can decrease **linearly**

▷ only **one** Newton step per μ



Examples:

for $X = \mathcal{R}^n_+$, $\phi(x) = -\sum_i \ln(x_i)$

for $X = \mathcal{S}^n_+$, $\phi(X) = -\ln(\det(X))$

for $X = X_1 \times X_2$, $\phi(x_1, x_2) = \phi_1(x_1) + \phi(x_2)$

Interior-Point Methods
Guarantees and software

Guarantees: nothing is heuristic

▷ Theoretically guaranteed overall complexity:
$\mathcal{O}(\sqrt{N}\,\ln(1/\epsilon))$ it. for $X = \mathcal{S}_+^N$

▷ Guaranteed accuracy of the computed solution

Software:

▶ **Sedumi** (in C, for Matlab), for symmetric programming
(a.o. $X = \mathcal{R}_+^n \times \mathcal{S}_+^N$) — The best free software

▶ **SDPT3** (Matlab), for symmetric programming

▶ **Mosek**, up-to-date software, licensed

solve easily problems with n up to 10000 and N up to 500
in **few minutes** ($\sim$ 20-30 least-square problems)

How does this connect to experimental design?

Quick recall on linear experimental design 1

$$y_i = a_i^T x + w_i$$

$y_1, \ldots, y_N$ are the measured quantities

$w_1, \ldots, w_N$ are iid noises $\sim N(0, 1)$

$a_1, \ldots, a_N$ correspond to performed experiments

Least-square solution is

$$\hat{x} = \left(\sum_{i=1}^{N} a_i a_i^T \right)^{-1} \sum_{i=1}^{N} y_i a_i, \text{ with properties:}$$

$$\mathbf{E}(\hat{x}) = \mathbf{E}(x), \text{ and } \mathbf{Var}(x - \hat{x}) = \left(\sum_{i=1}^{N} a_i a_i^T \right)^{-1}$$

Quick recall on linear experimental design 2

$$y_i = a_i^T x + w_i$$

Suppose we have a catalog of p different experiments $v_1, \ldots, v_p$.
Question: which one should we conduct to have the most relevant least-square solution?

Let m_j be the number of time the experiment v_j is conducted
c_j be the cost of v_j
B be the total budget

$$\begin{array}{rl} \min_{\mathcal{S}_+^n} & \left(\sum_{j=1}^p m_j v_j v_j^T\right)^{-1} \\ \text{s.t.} & \sum_{j=1}^p m_j c_j = B \\ & m_j \in \mathbb{N} \end{array}$$

Two things to simplify: integrality and non-scalar objective

$$
\begin{aligned}
\min_{\mathcal{S}_+^n} \quad & \left(\sum_{j=1}^p m_j v_j v_j^T\right)^{-1} \\
\text{s.t.} \quad & \sum_{j=1}^p m_j c_j = B \\
& m_j \in \mathbb{N}
\end{aligned}
$$

Let $\lambda_j := m_j/B$. Relax integrality constraint, and round the solution to recover $m_j \in \mathbb{N}$

$$
\begin{aligned}
\min_{\mathcal{S}_+^n} \quad & \left(\sum_{j=1}^p \lambda_j v_j v_j^T\right)^{-1} \\
\text{s.t.} \quad & \sum_{j=1}^p \lambda_j c_j = 1 \\
& \lambda_j \geq 0
\end{aligned}
$$

A first scalarization: D-design

$$\begin{array}{ll} \min & \det\left(\left(\sum_{j=1}^{p} \lambda_j v_j v_j^T\right)^{-1}\right) \\ \text{s.t.} & \sum_{j=1}^{p} \lambda_j c_j = 1 \\ & \lambda_j \geq 0 \end{array}$$

Observation:

$-\ln\det(X)$ is the "miraculous" barrier of Interior-point methods. Thus, feed **Sedumi** with:

$$\begin{array}{ll} \min & -\ln\det(X) \\ \text{s.t.} & \sum_{j=1}^{p} \lambda_j c_j = 1 \\ & X = \sum_{j=1}^{p} \lambda_j v_j v_j^T \\ & \lambda_j \geq 0, X \in \mathcal{S}_+^n \end{array} \quad \rightarrow \quad \boxed{\begin{array}{ll} \min & 0 \\ \text{s.t.} & \sum_{j=1}^{p} \lambda_j c_j = 1 \\ & X = \sum_{j=1}^{p} \lambda_j v_j v_j^T \\ & \lambda_j \geq 0, X \in \mathcal{S}_+^n \end{array}}$$

A second scalarization: E-design

$$\begin{array}{ll} \min & \lambda_{\max}\left(\left(\sum_{j=1}^{p} \lambda_j v_j v_j^T\right)^{-1}\right) \\ \text{s.t.} & \sum_{j=1}^{p} \lambda_j c_j = 1 \\ & \lambda_j \geq 0 \end{array}$$

Observation: $\lambda_{\max}(X^{-1}) = 1/\lambda_{\min}(X)$.
Thus, the problem can be rewritten as:

$$\begin{array}{ll} \max & \lambda_{\min}(X) \\ \text{s.t.} & \sum_{j=1}^{p} \lambda_j c_j = 1 \\ & X = \sum_{j=1}^{p} \lambda_j v_j v_j^T \\ & \lambda_j \geq 0 \end{array} \quad \rightarrow \quad \begin{array}{ll} \min & t \\ \text{s.t.} & \sum_{j=1}^{p} \lambda_j c_j = 1 \\ & X = \sum_{j=1}^{p} \lambda_j v_j v_j^T \\ & \lambda_j \geq 0,\ X - tI \in \mathcal{S}_+^n \end{array}$$

A third scalarization: M-design

$$\begin{array}{ll} \text{min} & \max_k \left(\left(\sum_{j=1}^p \lambda_j v_j v_j^T \right)^{-1} \right)_{kk} \\ \text{s.t.} & \sum_{j=1}^p \lambda_j c_j = 1 \\ & \lambda_j \geq 0 \end{array}$$

Observation: *Using Schur Lemma,*

$$0 \leq \left(\left(\sum_{j=1}^p \lambda_j v_j v_j^T \right)^{-1} \right)_{kk} \leq t_k \text{ iff } \begin{pmatrix} \sum_{j=1}^p \lambda_j v_j v_j^T & e_k \\ e_k^T & t_k \end{pmatrix} \in \mathcal{S}_+^{n+1}$$

$$\begin{array}{ll} \text{min} & t \\ \text{s.t.} & \sum_{j=1}^p \lambda_j c_j = 1 \\ & X = \sum_{j=1}^p \lambda_j v_j v_j^T \\ & \lambda_j \geq 0, \; t_k \leq t \\ & \begin{pmatrix} X & e_k \\ e_k^T & t_k \end{pmatrix} \in \mathcal{S}_+^{n+1} \text{ for } 1 \leq k \leq n \end{array}$$

A fourth scalarization: A-design

$$\begin{array}{ll} \min & \mathrm{tr}\left(\left(\sum_{j=1}^{p} \lambda_j v_j v_j^T\right)^{-1}\right) \\ \text{s.t.} & \sum_{j=1}^{p} \lambda_j c_j = 1 \\ & \lambda_j \geq 0 \end{array}$$

Very similar to M-design,
except we take the sum of diagonal elements

$$\begin{array}{ll} \min & \sum_k t_k \\ \text{s.t.} & \sum_{j=1}^{p} \lambda_j c_j = 1 \\ & X = \sum_{j=1}^{p} \lambda_j v_j v_j^T \\ & \lambda_j \geq 0 \\ & \begin{pmatrix} X & e_k \\ e_k^T & t_k \end{pmatrix} \in \mathcal{S}_+^{n+1} \text{ for } 1 \leq k \leq n \end{array}$$

In Conclusion:

Do It Convex

Interior-Point Methods

are a technology

that deserves

the practitioner's attention

Some references

[1] - S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004

[2] - Y. Nesterov, *Introductory lectures on convex optimization: a basic course*, Kluwer, 2003

[3] - Yu. Nesterov, A. Nemirovski, *Interior Point Polynomial Algorithms in Convex Programming*, SIAM, 1994