Exploiting Convexity in Experimental Design

Michel Baes

Niels Haverbeke
What can you do with Convex Optimization?
Obviously, solve convex optimization problems

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad Ax = b \\
& \quad x \in X
\end{align*}
\]

\(f\) 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

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad Ax = b \\
& \quad x \in X
\end{align*}
\]

\[
\min \quad t = \min \quad t \\
\text{s.t.} \quad Ax = b \
& \quad x \in X \\
& \quad f(x) \leq t
\]

\[
\min \quad \langle \bar{c}, \bar{x} \rangle \\
\text{s.t.} \quad \bar{A}\bar{x} = \bar{b} \\
& \quad \bar{x} \in \bar{X}
\]

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

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

\(X\) a closed \textbf{convex set} with nonempty interior

▷ Equivalent reformulation:

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

▷ Barrier function approximation:

\[ p^* \simeq \min \{ \langle c, x \rangle + \mu \phi(x) : Ax = b \}, \]
where \( \phi \) is a \textit{barrier} of \( X \):
\( \phi \) is \textit{strongly convex} (at least convex + quadratic),
\( \text{dom} \phi = \text{int} X \), 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 $\phi$ 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 $\mu$ to $\mu_-$, so that $\hat{x}(\mu)$ is a ”good” starting point
But there are some barriers for which **miracles** happen

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

If $\phi$ is a **self-concordant barrier**, then

- we **know** the quadratic convergence zone of Newton’s method
- $\mu$ can decrease **linearly**
- only **one** Newton step per $\mu$

**Examples:**

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

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

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 \( X = S^N_+ \)
- Guaranteed accuracy of the computed solution

**Software:**

- **Sedumi** (in C, for Matlab), for symmetric programming (a.o. \( X = \mathcal{R}^n_+ \times S^N_+ \)) — 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

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

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

\[
\min_{S^n_+} \left( \sum_{j=1}^{p} m_j v_j v_j^T \right)^{-1}
\text{ s.t. } \sum_{j=1}^{p} m_j c_j = B
\]
\[ m_j \in \mathbb{N} \]
Two things to simplify:
integrality and non-scalar objective

\[
\begin{align*}
\min_{S^n_+} & \quad \left( \sum_{j=1}^p m_j v_j v_j^T \right)^{-1} \\
\text{s.t.} & \quad \sum_{j=1}^p m_j c_j = B \\
& \quad m_j \in \mathbb{N}
\end{align*}
\]

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

\[
\begin{align*}
\min_{S^n_+} & \quad \left( \sum_{j=1}^p \lambda_j v_j v_j^T \right)^{-1} \\
\text{s.t.} & \quad \sum_{j=1}^p \lambda_j c_j = 1 \\
& \quad \lambda_j \geq 0
\end{align*}
\]
A first scalarization: D-design

\[
\begin{align*}
\min & \quad \det \left( \left( \sum_{j=1}^{p} \lambda_j v_j v_j^T \right)^{-1} \right) \\
\text{s.t.} & \quad \sum_{j=1}^{p} \lambda_j c_j = 1 \\
& \quad \lambda_j \geq 0
\end{align*}
\]

Observation:
- \( -\ln \det(X) \) is the "miraculous" barrier of Interior-point methods. Thus, feed \textbf{Sedumi} with:

\[
\begin{align*}
\min & \quad -\ln \det(X) \\
\text{s.t.} & \quad \sum_{j=1}^{p} \lambda_j c_j = 1 \\
& \quad X = \sum_{j=1}^{p} \lambda_j v_j v_j^T \\
& \quad \lambda_j \geq 0, X \in S^n_+ 
\end{align*}
\]

\[
\begin{align*}
\min & \quad 0 \\
\text{s.t.} & \quad \sum_{j=1}^{p} \lambda_j c_j = 1 \\
& \quad X = \sum_{j=1}^{p} \lambda_j v_j v_j^T \\
& \quad \lambda_j \geq 0, X \in S^n_+
\end{align*}
\]
A second scalarization: E-design

\[
\begin{align*}
\min & \quad \lambda_{\text{max}} \left( \left( \sum_{j=1}^{p} \lambda_j v_j v_j^T \right)^{-1} \right) \\
\text{s.t.} & \quad \sum_{j=1}^{p} \lambda_j c_j = 1 \\
& \quad \lambda_j \geq 0
\end{align*}
\]

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

\[
\begin{align*}
\max & \quad \lambda_{\text{min}}(X) \\
\text{s.t.} & \quad \sum_{j=1}^{p} \lambda_j c_j = 1 \\
& \quad X = \sum_{j=1}^{p} \lambda_j v_j v_j^T \\
& \quad \lambda_j \geq 0
\end{align*}
\]  →  

\[
\begin{align*}
\min & \quad t \\
\text{s.t.} & \quad \sum_{j=1}^{p} \lambda_j c_j = 1 \\
& \quad X = \sum_{j=1}^{p} \lambda_j v_j v_j^T \\
& \quad \lambda_j \geq 0, X - tI \in S_+^n
\end{align*}
\]
A third scalarization: M-design

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

Observation: Using Schur Lemma,

\[
0 \leq \left( \left( \sum_{j=1}^{p} \lambda_j v_j^T v_j^T \right)^{-1} \right)_{kk} \leq t_k \iff \left( \sum_{j=1}^{p} \lambda_j v_j^T v_j^T e_k \begin{pmatrix} e_k \\ t_k \end{pmatrix} \right) \in S_{+}^{n+1}
\]

\[
\begin{align*}
\min \ & \ t \\
\text{s.t.} \ & \ \sum_{j=1}^{p} \lambda_j c_j = 1 \\
& \ X = \sum_{j=1}^{p} \lambda_j v_j^T v_j^T \\
& \lambda_j \geq 0, \ t_k \leq t \\
& \begin{pmatrix} X \\ e_k^T \\ e_k \\ t_k \end{pmatrix} \in S_{+}^{n+1} \text{ for } 1 \leq k \leq n
\end{align*}
\]
A fourth scalarization: A-design

\[
\begin{align*}
\min & \quad \text{tr} \left( \left( \sum_{j=1}^{p} \lambda_j v_j v_j^T \right)^{-1} \right) \\
\text{s.t.} & \quad \sum_{j=1}^{p} \lambda_j c_j = 1 \\
& \quad \lambda_j \geq 0
\end{align*}
\]

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

\[
\begin{align*}
\min & \quad \sum_{k} t_k \\
\text{s.t.} & \quad \sum_{j=1}^{p} \lambda_j c_j = 1 \\
& \quad X = \sum_{j=1}^{p} \lambda_j v_j v_j^T \\
& \quad \lambda_j \geq 0 \\
& \quad \left( \begin{array}{cc}
X & e_k \\
e_k^T & t_k
\end{array} \right) \in S^{n+1}_+ \text{ for } 1 \leq k \leq n
\end{align*}
\]
In Conclusion:

**Do It Convex**

*Interior-Point Methods are a technology that deserves the practitioner’s attention***
Some references

