Self-concordant Barrier

Def: Let $C \in R^n$ be an open nonempty convex set. Let $f : C \mapsto R$ be a 3 times continuously differentiable convex function. A function $f$ is called self-concordant if there exists a constant $p > 0$ such that

$$|\nabla^3 f(x)[h, h, h]| \leq 2p^{-1/2}(\nabla^2 f(x)[h, h])^{3/2},$$

$\forall x \in C, \forall h : x + h \in C$. (We then say that $f$ is $p$-self-concordant).

Note that a self-concordant function is always well approximated by the quadratic model because the error of such an approximation can be bounded by the $3/2$ power of $\nabla^2 f(x)[h, h]$.

Lemma The barrier function $-\log x$ is self-concordant on $R_+$.  
Proof: Consider $f(x) = -\log x$ and compute $f'(x) = -x^{-1}, f''(x) = x^{-2}$ and $f'''(x) = -2x^{-3}$ and check that the self-concordance condition is satisfied for $p = 1$.

Lemma The barrier function $1/x^\alpha$, with $\alpha \in (0, \infty)$ is not self-concordant on $R_+$.

Lemma The barrier function $e^{1/x}$ is not self-concordant on $R_+$.

Use self-concordant barriers in optimization
Part 1:

Semidefinite Programming (SDP)

SDP: Background

Def. A matrix $H \in \mathcal{R}^{n\times n}$ is positive semidefinite if $x^T H x \geq 0$ for an $x \neq 0$. We write $H \succeq 0$.

Def. A matrix $H \in \mathcal{R}^{n\times n}$ is positive definite if $x^T H x > 0$ for an $x \neq 0$. We write $H \succ 0$.

We denote with $\mathcal{S}\mathcal{R}^{n\times n}$ ($\mathcal{S}\mathcal{R}^{n\times n}_+$) the set of symmetric and symmetric positive semidefinite matrices.

Let $U, V \in \mathcal{S}\mathcal{R}^{n\times n}$. We define the inner product between $U$ and $V$ as $U \cdot V = \text{trace}(U^T V)$, where $\text{trace}(H) = \sum_{i=1}^{n} h_{ii}$.

The associated norm is the Frobenius norm, written $\|U\|_F = (U \cdot U)^{1/2}$ (or just $\|U\|$).
Linear Matrix Inequalities

**Def. Linear Matrix Inequalities**

Let $U, V \in \mathcal{S}\mathcal{R}^{n \times n}$. We write $U \succeq V$ iff $U - V \succeq 0$.

We write $U \succ V$ iff $U - V \succ 0$.

We write $U \preceq V$ iff $U - V \preceq 0$.

We write $U \prec V$ iff $U - V \prec 0$.

---

**Properties**

1. If $P \in \mathcal{R}^{m \times n}$ and $Q \in \mathcal{R}^{n \times m}$, then $\text{trace}(PQ) = \text{trace}(QP)$.

2. If $U, V \in \mathcal{S}\mathcal{R}^{n \times n}$, and $Q \in \mathcal{R}^{n \times m}$ is orthogonal (i.e. $Q^TQ = I$), then $U \bullet V = (Q^T U Q) \bullet (Q^T V Q)$.

   More generally, if $P$ is nonsingular, then $U \bullet V = (PUP^T) \bullet (P^{-T} V P^{-1})$.

3. Every $U \in \mathcal{S}\mathcal{R}^{n \times n}$ can be written as $U = Q\Lambda Q^T$, where $Q$ is orthogonal and $\Lambda$ is diagonal. Then $UQ = Q\Lambda$.

   In other words the columns of $Q$ are the eigenvectors, and the diagonal entries of $\Lambda$ the corresponding eigenvalues of $U$.

4. If $U \in \mathcal{S}\mathcal{R}^{n \times n}$ and $U = Q\Lambda Q^T$, then $\text{trace}(U) = \text{trace}(\Lambda) = \sum_i \lambda_i$.

5. For $U \in \mathcal{S}\mathcal{R}^{n \times n}$, the following are equivalent:

   (i) $U \geq 0$ ($U > 0$)

   (ii) $x^T U x \geq 0, \forall x \in \mathcal{R}^n$ ($x^T U x > 0, \forall 0 \neq x \in \mathcal{R}^n$).

   (iii) If $U = Q\Lambda Q^T$, then $\Lambda \geq 0$ ($\Lambda > 0$).

   (iv) $U = P^T P$ for some matrix $P$ ($U = P^T P$ for some square nonsingular matrix $P$).

6. Every $U \in \mathcal{S}\mathcal{R}^{n \times n}$ has a square root $U^{1/2} \in \mathcal{S}\mathcal{R}^{n \times n}$.

   **Proof:** From Property 5 (ii) we get $U = Q\Lambda Q^T$.

   Take $U^{1/2} = Q\Lambda^{1/2} Q^T$, where $\Lambda^{1/2}$ is the diagonal matrix whose diagonal contains the (nonnegative) square roots of the eigenvalues of $U$, and verify that $U^{1/2^2} = U$.

7. Suppose

   $$U = \begin{bmatrix} A & B^T \\ B & C \end{bmatrix},$$

   where $A$ and $C$ are symmetric and $A \succ 0$.

   Then $U \succeq 0$ ($U > 0$) iff $C - BA^{-1}B^T \succeq 0$ ($> 0$).

   The matrix $C - BA^{-1}B^T$ is called the Schur complement of $A$ in $U$.

   **Proof:** follows easily from the factorization:

   $$\begin{bmatrix} A & B^T \\ B & C \end{bmatrix} = \begin{bmatrix} I & 0 \\ BA^{-1}I & C-BA^{-1}B^T \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

8. If $U \in \mathcal{S}\mathcal{R}^{n \times n}$ and $x \in \mathcal{R}^n$, then $x^T U x = U \bullet xx^T$. 
Primal-Dual Pair of SDPs

Primal  
\[
\min C \cdot X \\
\text{s.t. } A_i \cdot X = b_i, \ i = 1..m \\
\quad X \succeq 0;
\]

Dual  
\[
\max b^T y \\
\text{s.t. } \sum_{i=1}^m y_i A_i + S = C, \\
\quad S \succeq 0,
\]

where \( A_i \in \mathcal{S}\mathcal{R}^{n \times n}, b \in \mathcal{R}^m, C \in \mathcal{S}\mathcal{R}^{n \times n} \) are given; and \( X, S \in \mathcal{S}\mathcal{R}^{n \times n}, y \in \mathcal{R}^m \) are the variables.

SDP Example 1: Minimize the Maximum Eigenvalue

We wish to choose \( x \in \mathcal{R}^k \) to minimize the maximum eigenvalue
\[
\lambda_{\max}(A(x)) = \lambda_{\max}(A_0 + x_1 A_1 + \ldots + x_k A_k),
\]
where \( A_i \in \mathcal{R}^{n \times n} \) and \( A_i = A_i^T \).

Observe that
\[
\lambda_{\max}(A(x)) \leq t
\]
if and only if
\[
\lambda_{\max}(A(x) - t I) \leq 0 \iff \lambda_{\min}(t I - A(x)) \geq 0.
\]

This holds iff
\[
t I - A(x) \succeq 0.
\]
So we get the SDP in the dual form:
\[
\max -t \\
\text{s.t. } t I - A(x) \succeq 0,
\]
where the variable is \( y := (t, x) \).

SDP Example 2: Logarithmic Chebyshev Approx.

Suppose we wish to solve
\[
A x \approx b
\]
aapproximately, where \( A = [a_1 \ldots a_n]^T \in \mathcal{R}^{n \times k} \) and \( b \in \mathcal{R}^n \).

In Chebyshev approximation we minimize the \( \ell_\infty \)-norm of the residual, i.e., we solve
\[
\min \max_i |a_i^T x - b_i|.
\]
This can be cast as an LP, with \( x \) and an auxiliary variable \( t \):
\[
\min \quad t \\
\text{s.t. } -t \leq a_i^T x - b_i \leq t, \ i = 1..n.
\]
In some applications \( b_i \) has a dimension of a power of intensity, and it is typically expressed on a logarithmic scale. In such cases the more natural optimization problem is
\[
\min \max_i |\log(a_i^T x) - \log(b_i)|
\]
(assuming \( a_i^T x > 0 \) and \( b_i > 0 \)).
Logarithmic Chebyshev Approximation (cont’d)
The logarithmic Chebyshev approximation problem can be cast as a semidefinite program. To see this, note that

\[ | \log(a_i^T x) - \log(b_i) | = \log \max(a_i^T x / b_i, b_i / a_i^T x). \]

Hence the problem can be rewritten as the following nonlinear program

\[
\begin{align*}
\min & \quad t \\
\text{s.t.} & \quad 1 / t \leq a_i^T x / b_i \leq t, \quad i = 1..n.
\end{align*}
\]

or,

\[
\begin{align*}
\min & \quad t \\
\text{s.t.} & \quad \begin{bmatrix} t - a_i^T x / b_i & 0 & 0 \\
0 & a_i^T x / b_i & 1 \\
0 & 1 & t \end{bmatrix} \geq 0, \quad i = 1..n
\end{align*}
\]

which is a semidefinite program.

Logarithmic Barrier Function

Define the **logarithmic barrier function** for the cone \( SR_{++}^{n \times n} \) of positive definite matrices.

\[ f : SR_{++}^{n \times n} \rightarrow \mathbb{R} \]

\[
f(X) = \begin{cases} 
- \ln \det X & \text{if } X \succ 0 \\
+ \infty & \text{otherwise.}
\end{cases}
\]

Let us evaluate its derivatives.

Let \( X \succ 0, H \in SR_{++}^{n \times n} \). Then

\[
f(X + \alpha H) = - \ln \det [X(I + \alpha X^{-1} H)]
\]

\[
= - \ln \det X - \ln(1 + \alpha \text{trace}(X^{-1} H) + O(\alpha^2))
\]

\[
= f(X) - \alpha X^{-1} \cdot H + O(\alpha^2),
\]

so that \( f'(X) = -X^{-1} \) and \( D f(X)[H] = -X^{-1} \cdot H \).

Logarithmic Barrier Function (cont’d)

Similarly

\[
f'(X + \alpha H) = -[X(I + \alpha X^{-1} H)]^{-1}
\]

\[
= -[I - \alpha X^{-1} H + O(\alpha^2)]X^{-1}
\]

\[
f'(X) + \alpha X^{-1} H X^{-1} + O(\alpha^2),
\]

so that \( f''(X)[H] = X^{-1} H X^{-1} \)

and \( D^2 f(X)[H, G] = X^{-1} H X^{-1} \boldsymbol{\cdot} G \).

Finally,

\[
f'''(X)[H, G] = -X^{-1} H X^{-1} G X^{-1} - X^{-1} G X^{-1} H X^{-1}.
\]
Simplified Notation

Define $\mathcal{A} : S\mathcal{R}^{n \times n} \mapsto \mathbb{R}^m$

$$\mathcal{A}X = (A_i \cdot X)_{i=1}^m \in \mathbb{R}^m.$$ 

Note that, for any $X \in S\mathcal{R}^{n \times n}$ and $y \in \mathbb{R}^m$,

$$(\mathcal{A}X)^T y = \sum_{i=1}^m (A_i \cdot X) y_i = (\sum_{i=1}^m y_i A_i) \cdot X,$$

so the adjoint of $\mathcal{A}$ is given by

$$\mathcal{A}^* y = \sum_{i=1}^m y_i A_i.$$ 

$\mathcal{A}^*$ is a mapping from $\mathbb{R}^m$ to $S\mathcal{R}^{n \times n}$.

Solving SDPs with IPMs (cont’d)

With this notation the primal SDP becomes

$$\begin{align*}
\min \ C \cdot X \\
\text{s.t.} \quad \mathcal{A}X = b, \\
\quad X \succeq 0,
\end{align*}$$

where $X \in S\mathcal{R}^{n \times n}$ is the variable.

The associated dual SDP writes

$$\begin{align*}
\max \ b^T y \\
\text{s.t.} \quad \mathcal{A}^* y + S = C \\
\quad S \succeq 0,
\end{align*}$$

where $y \in \mathbb{R}^m$ and $S \in S\mathcal{R}^{n \times n}$ are the variables.

NATCOR, Edinburgh, June 2014
Newton direction

We derive the Newton direction for the system:
\[
AX = b, \\
A^*y + S = C, \\
-\mu X^{-1} + S = 0.
\]

Recall that the variables in FOC are \((X, y, S)\), where \(X, S \in \mathcal{SR}^{n \times n}_+\) and \(y \in \mathbb{R}^m\).

Hence we look for a direction \((\Delta X, \Delta y, \Delta S)\), where \(\Delta X, \Delta S \in \mathcal{SR}^{n \times n}_+\) and \(\Delta y \in \mathbb{R}^m\).

Newton direction (cont’d)

The differentiation in the above system is a nontrivial operation. The direction is the solution of the system:
\[
\begin{bmatrix}
A & 0 & 0 \\
0 & A^* & I \\
\mu(X^{-1} \odot X^{-1}) & 0 & I
\end{bmatrix}
\begin{bmatrix}
\Delta X \\
\Delta y \\
\Delta S
\end{bmatrix}
= \begin{bmatrix}
\xi_b \\
\xi_C \\
\xi_{\mu}
\end{bmatrix},
\]

We introduce a useful notation \(P \odot Q\) for \(n \times n\) matrices \(P\) and \(Q\). This is an operator from \(\mathcal{SR}^{n \times n}\) to \(\mathcal{SR}^{n \times n}\) defined by
\[
(P \odot Q)U = \frac{1}{2}(PUQ^T + QUP^T).
\]

Logarithmic Barrier Function

for the cone \(\mathcal{SR}^{n \times n}_+\) of positive definite matrices, \(f : \mathcal{SR}^{n \times n}_+ \mapsto \mathbb{R}\)
\[
f(X) = \begin{cases} 
-\ln \det X & \text{if } X \succ 0 \\
+\infty & \text{otherwise.}
\end{cases}
\]

LP: Replace \(x \geq 0\) with \(-\mu \sum_{j=1}^n \ln x_j\).

SDP: Replace \(X \succeq 0\) with \(-\mu \sum_{j=1}^n \ln \lambda_j = -\mu \ln(\prod_{j=1}^n \lambda_j)\).


Lemma The barrier function \(f(X)\) is self-concordant on \(\mathcal{SR}^{n \times n}_+\).

Second-Order Cone Programming (SOCP)
SOCP: Second-Order Cone Programming

- Generalization of QP.
- Deals with conic constraints.
- Solved with IPMs.
- Numerous applications: quadratically constrained quadratic programs, problems involving sums and maxima/minima of norms, SOC-representable functions and sets, matrix-fractional problems, problems with hyperbolic constraints, robust LP/QP, robust least-squares.

Cones: Background

Def. A set $K \in \mathbb{R}^n$ is called a cone if for any $x \in K$ and for any $\lambda \geq 0$, $\lambda x \in K$.

Convex Cone:

Example:

$$K = \{ x \in \mathbb{R}^n : x_1^2 \geq \sum_{j=2}^{n} x_j^2, x_1 \geq 0 \}.$$ 

Example: Three Cones

$R_+$:

$$R_+ = \{ x \in \mathbb{R} : x \geq 0 \}.$$ 

Quadratic Cone:

$$K_q = \{ x \in \mathbb{R}^n : x_1^2 \geq \sum_{j=2}^{n} x_j^2, x_1 \geq 0 \}.$$ 

Rotated Quadratic Cone:

$$K_r = \{ x \in \mathbb{R}^n : 2x_1x_2 \geq \sum_{j=3}^{n} x_j^2, x_1, x_2 \geq 0 \}.$$
Matrix Representation of Cones

Each of the three most common cones has a matrix representation using orthogonal matrices $T$ and/or $Q$. (Orthogonal matrix: $Q^TQ = I$).

**Quadratic Cone $K_q$.** Define

$$Q = \begin{bmatrix}
1 & -1 \\
-1 & -1 \\
& \ddots \\
& & 1
\end{bmatrix}$$

and write:

$$K_q = \{x \in \mathbb{R}^n : x^TQx \geq 0, x_1 \geq 0\}.$$  

Example: $x_1^2 \geq x_2^2 + x_3^2 + \cdots + x_n^2$.

---

**Rotated Quadratic Cone $K_r$.** Define

$$Q = \begin{bmatrix}
0 & 1 \\
1 & 0 \\
& -1 \\
& \ddots \\
& & 1
\end{bmatrix}$$

and write:

$$K_r = \{x \in \mathbb{R}^n : x^TQx \geq 0, x_1, x_2 \geq 0\}.$$  

Example: $2x_1x_2 \geq x_3^2 + x_4^2 + \cdots + x_n^2$.

---

Consider a linear transformation $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$:

$$T_2 = \begin{bmatrix}
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}
\end{bmatrix}.$$  

It corresponds to a rotation by $\pi/4$. Indeed, write:

$$\begin{bmatrix} z \\ y \end{bmatrix} = T_2 \begin{bmatrix} u \\ v \end{bmatrix}$$

that is

$$z = \frac{u + v}{\sqrt{2}}, \quad y = \frac{u - v}{\sqrt{2}}$$

to get

$$2yz = u^2 - v^2.$$
### Example: Conic constraint

Consider a constraint: 
\[
\frac{1}{2}\|x\|^2 + ax \leq b.
\]

Observe that \(g(x) = \frac{1}{2}x^T x + ax - b\) is convex hence the constraint defines a convex set.

The constraint may be reformulated as an intersection of an affine (linear) constraint and a quadratic one:
\[
\begin{align*}
ax + z &= b \\
y &= 1 \\
\|x\|^2 &\leq 2yz, \quad y, z \geq 0.
\end{align*}
\]

### Example: Conic constraint (cont’d)

Now, substitute:
\[
z = \frac{u + v}{\sqrt{2}}, \quad y = \frac{u - v}{\sqrt{2}}
\]
to get
\[
\begin{align*}
ax + \frac{u + v}{\sqrt{2}} &= b \\
u - v &= \sqrt{2} \\
\|x\|^2 + v^2 &\leq u^2.
\end{align*}
\]

### Conic Optimization

Consider an optimization problem:
\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad Ax = b, \\
x &\in K,
\end{align*}
\]
where \(K\) is a convex closed cone.

We assume that 
\[K = K^1 \times K^2 \times \cdots \times K^k,\]
that is, cone \(K\) is a product of several individual cones each of which one of the three cones defined earlier.
Primal and Dual SOCPs

Consider a \textbf{primal} SOCP
\[
\begin{align*}
\min \quad & c^T x \\
\text{s.t.} \quad & Ax = b, \\
& x \in K,
\end{align*}
\]
where $K$ is a convex closed cone.

The associated \textbf{dual} SOCP
\[
\begin{align*}
\max \quad & b^T y \\
\text{s.t.} \quad & A^T y + s = c, \\
& s \in K^*.
\end{align*}
\]

\textbf{Weak Duality:}
If $(x, y, s)$ is a primal-dual feasible solution, then
\[c^T x - b^T y = x^T s \geq 0.\]

Logarithmic Barrier Fctn for Quadratic Cone

Its derivatives are given by:
\[
\nabla f(x, t) = \frac{2}{t^2 - x^T x} \begin{bmatrix} x \\ -t \end{bmatrix},
\]
and
\[
\nabla^2 f(x, t) = \frac{2}{(t^2 - x^T x)^2} \begin{bmatrix} (t^2 - x^T x)I + 2xx^T & -2tx \\ -2tx^T & t^2 + x^T x \end{bmatrix}.
\]

\textbf{Theorem:}
$f(x, t)$ is a self-concordant barrier on $K_q$.

Exercise: Prove it in case $n = 2$.

Examples of SOCP

\textbf{LP, QP} use the cone $\mathcal{R}_+$ (positive orthant).

\textbf{SDP} uses the cone $\mathcal{S}\mathcal{R}^n_{++}$ (symmetric positive definite matrices).

\textbf{SOCP} uses two quadratic cones $K_q$ and $K_r$.

Quadratically Constrained Quadratic Programming (QCQP) is a particular example of SOCP.

Typical trick to replace a quadratic constraint as a conic one!!!

Consider a constraint:
\[
\frac{1}{2}\|x\|^2 + ax \leq b.
\]
Rewrite it as:
\[
\|x\|^2 + v^2 \leq u^2.
\]
**QCQP and SOCP**

Let $P_i \in \mathbb{R}^{n \times n}$ be a symmetric positive definite matrix and $q_i \in \mathbb{R}^n$. Define a quadratic function $f_i(x) = x^T P_i x + 2 q_i^T x + r_i$ and an associated ellipsoid $E_i = \{ x \mid f_i(x) \leq 0 \}$.

The set of constraints $f_i(x) \leq 0$, $i = 1, 2, \ldots, m$, defines an intersection of (convex) ellipsoids and of course defines a convex set.

The optimization problem

$$\min f_0(x)$$

s.t. $f_i(x) \leq 0$, $i = 1, 2, \ldots, m$,

is an example of quadratically constrained quadratic program (QCQP).

QCQP can be reformulated as SOCP.

QCQP can be also reformulated as SDP.

---

**SOCP Example: Robust LP**

Consider an LP:

$$\min c^T x$$

s.t. $a_i^T x \leq b_i$, $i = 1, 2, \ldots, m$,

and assume that the values of $a_i$ are uncertain.

Suppose that $a_i \in E_i$, $i = 1, 2, \ldots, m$, where $E_i$ are given ellipsoids

$$E_i = \{ \bar{a}_i + P_i u : \| u \| \leq 1 \},$$

where $P_i$ is a symmetric positive definite matrix.

---

**SOCP Example: Linear Regression**

The least squares solution of a linear system of equations $Ax = b$ is the solution of the following optimization problem

$$\min_{x} \| Ax - b \|$$

and it can be recast as:

$$\min_{x} \quad t$$

s.t. $\| Ax - b \| \leq t$.

---

**SOCP Example: Robust LP (cont’d)**

Observe that

$$a_i^T x \leq b_i \forall a_i \in E_i \iff \bar{a}_i^T x + \| P_i x \| \leq b_i,$$

because for any $x \in \mathbb{R}^n$

$$\max \{ a^T x : a \in E \} = \bar{a}^T x + \max \{ u^T P x : \| u \| \leq 1 \} = \bar{a}^T x + \| P x \|.$$

Hence robust LP formulated as SOCP is:

$$\min_{x} \quad c^T x$$

s.t. $\bar{a}_i^T x + \| P_i x \| \leq b_i$, $i = 1, 2, \ldots, m$. 

---

NATCOR, Edinburgh, June 2014
SOCP Example: Robust QP

Consider a QP with “uncertain” objective:
\[
\min_x \max_{P \in \mathcal{E}} x^T P x + 2 q^T x + r
\]
subject to linear constraints. “Uncertain” symmetric positive definite matrix \( P \) belongs to the ellipsoid:
\[
P \in \mathcal{E} = \{ P_0 + \sum_{i=1}^m P_i u_i : \| u \| \leq 1 \},
\]
where \( P_i \) are symmetric positive semidefinite matrices.

The definition of ellipsoid \( \mathcal{E} \) implies that
\[
\max_{P \in \mathcal{E}} x^T P x = x^T P_0 x + \max_{\| u \| \leq 1} \sum_{i=1}^m (x^T P_i x) u_i.
\]

SOCP Example: Robust QP (cont’d)

From Cauchy-Schwartz inequality:
\[
\sum_{i=1}^m (x^T P_i x) u_i \leq \left( \sum_{i=1}^m (x^T P_i x)^2 \right)^{1/2} \| u \|
\]
hence
\[
\max_{\| u \| \leq 1} \sum_{i=1}^m (x^T P_i x) u_i \leq \left( \sum_{i=1}^m (x^T P_i x)^2 \right)^{1/2}.
\]

We get a reformulation of robust QP:
\[
\min_x x^T P_0 x + \left( \sum_{i=1}^m (x^T P_i x)^2 \right)^{1/2} + 2 q^T x + r.
\]