



School of Mathematics



Interior Point Methods: Semidefinite Programming

Jacek Gondzio

Email: J.Gondzio@ed.ac.uk

URL: <http://www.maths.ed.ac.uk/~gondzio>

 NATCOR, June 2018

1

J. Gondzio

IPMs for SDP

Outline

- **Self-concordant Barriers**
- **Semidefinite Programming**
 - background (linear matrix inequalities)
 - duality
 - example SDP problems
 - logarithmic barrier function
 - IPM for SDP
- **Final Comments**

 NATCOR, June 2018

2

Self-concordant Barrier

Def: Let $C \in \mathcal{R}^n$ be an open nonempty convex set. Let $f : C \mapsto \mathcal{R}$ be a 3 times continuously diff'able 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 $\nabla^2 f(x)[h, h]$.

 NATCOR, June 2018

3

J. Gondzio

IPMs for SDP

Self-concordant Barrier

Lemma The barrier function $-\log x$ is self-concordant on \mathcal{R}_+ .

Proof:

Consider $f(x) = -\log x$ and compute

$$f'(x) = -x^{-1}, \quad f''(x) = x^{-2} \quad \text{and} \quad f'''(x) = -2x^{-3}$$

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 \mathcal{R}_+ .

Lemma

The barrier function $e^{1/x}$ is not self-concordant on \mathcal{R}_+ .

Use self-concordant barriers in optimization

 NATCOR, June 2018

4

Semidefinite Programming (SDP)

SDP: Semidefinite Programming

- Generalization of LP.
- Deals with symmetric positive semidefinite matrices (Linear Matrix Inequalities, LMI).
- Solved with IPMs.
- Numerous applications:
 - eigenvalue optimization problems,
 - quasi-convex programs,
 - convex quadratically constrained optimization,
 - robust mathematical programming,
 - matrix norm minimization,
 - combinatorial optimization (provides good relaxations),
 - control theory,
 - statistics.

SDP: Semidefinite Programming

This lecture is based on two survey papers:

- **L. Vandenberghe and S. Boyd**,
Semidefinite Programming,
SIAM Review 38 (1996) pp. 49-95.
- **M.J. Todd**,
Semidefinite Optimization,
Acta Numerica 10 (2001) pp. 515-560.

SDP: Background

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

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

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

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

Linear Matrix Inequalities

Def. *Linear Matrix Inequalities*

Let $U, V \in \mathcal{SR}^{n \times n}$.

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 $U, V \in \mathcal{SR}^{n \times n}$, and $Q \in \mathcal{R}^{n \times n}$ is orthogonal (i.e. $Q^T Q = I$), then $U \bullet V = (Q^T U Q) \bullet (Q^T V Q)$.
More generally, if P is nonsingular, then $U \bullet V = (P U P^T) \bullet (P^{-T} V P^{-1})$.
3. Every $U \in \mathcal{SR}^{n \times n}$ can be written as $U = Q \Lambda Q^T$, where Q is orthogonal and Λ is diagonal. Then $U Q = Q \Lambda$.
In other words the columns of Q are the eigenvectors, and the diagonal entries of Λ the corresponding eigenvalues of U .
4. If $U \in \mathcal{SR}^{n \times n}$ and $U = Q \Lambda Q^T$, then $\text{trace}(U) = \text{trace}(\Lambda) = \sum_i \lambda_i$.

Properties (cont'd)

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

- (i) $U \succeq 0$ ($U \succ 0$)
- (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 $\Lambda \succeq 0$ ($\Lambda \succ 0$).
- (iv) $U = P^T P$ for some matrix P ($U = P^T P$ for some square nonsingular matrix P).

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

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 $U^{1/2} U^{1/2} = U$.

Properties (cont'd)

7. Suppose

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

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

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

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

Proof: follows easily from the factorization:

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

8. If $U \in \mathcal{SR}^{n \times n}$ and $x \in \mathcal{R}^n$, then $x^T U x = U \bullet x x^T$.

Primal-Dual Pair of SDPs

Primal

$$\begin{aligned} \min \quad & C \bullet X \\ \text{s.t.} \quad & A_i \bullet X = b_i, \quad i = 1..m \\ & X \succeq 0; \end{aligned}$$

Dual

$$\begin{aligned} \max \quad & b^T y \\ \text{s.t.} \quad & \sum_{i=1}^m y_i A_i + S = C, \\ & S \succeq 0, \end{aligned}$$

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

Theorem: Weak Duality in SDP

If X is feasible in the primal and (y, S) in the dual, then

$$C \bullet X - b^T y = X \bullet S \geq 0.$$

Proof:

$$\begin{aligned} C \bullet X - b^T y &= \left(\sum_{i=1}^m y_i A_i + S \right) \bullet X - b^T y \\ &= \sum_{i=1}^m (A_i \bullet X) y_i + S \bullet X - b^T y \\ &= S \bullet X = X \bullet S. \end{aligned}$$

Further, since X is positive semidefinite, it has a square root $X^{1/2}$ (Property 6), and so

$$X \bullet S = \text{trace}(XS) = \text{trace}(X^{1/2} X^{1/2} S) = \text{trace}(X^{1/2} S X^{1/2}) \geq 0.$$

We use Property 1 and the fact that S and $X^{1/2}$ are positive semidefinite, hence $X^{1/2} S X^{1/2}$ is positive semidefinite and its trace is nonnegative.

SDP Example 1: Minimize the Max. Eigenvalue

We wish to choose $x \in \mathcal{R}^k$ to minimize the maximum eigenvalue of $A(x) = A_0 + x_1 A_1 + \dots + x_k A_k$, 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

$$\lambda_{\max}(A(x) - tI) \leq 0 \iff \lambda_{\min}(tI - A(x)) \geq 0.$$

This holds iff

$$tI - A(x) \succeq 0.$$

So we get the SDP in the *dual* form:

$$\begin{aligned} \max \quad & -t \\ \text{s.t.} \quad & tI - A(x) \succeq 0, \end{aligned}$$

where the variable is $y := (t, x)$.

SDP Example 2: Logarithmic Chebyshev Approx.

Suppose we wish to solve $Ax \approx b$ approximately,

where $A = [a_1 \dots a_n]^T \in \mathcal{R}^{n \times k}$ and $b \in \mathcal{R}^n$.

In Chebyshev approximation we minimize the ℓ_∞ -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 :

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & -t \leq a_i^T x - b_i \leq t, \quad i = 1..n. \end{aligned}$$

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

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

or,

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & \begin{bmatrix} t - a_i^T x / b_i & 0 & 0 \\ 0 & a_i^T x / b_i & 1 \\ 0 & 1 & t \end{bmatrix} \succeq 0, \quad i = 1..n \end{aligned}$$

which is a semidefinite program.

Logarithmic Barrier Function

Define the **logarithmic barrier function** for the cone $\mathcal{SR}_+^{n \times n}$ of positive definite matrices.

$$f : \mathcal{SR}_+^{n \times n} \mapsto \mathcal{R}$$

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

Let us evaluate its derivatives.

Let $X \succ 0, H \in \mathcal{SR}^{n \times n}$. Then

$$\begin{aligned} f(X + \alpha H) &= -\ln \det[X(I + \alpha X^{-1}H)] \\ &= -\ln \det X - \ln(1 + \alpha \text{trace}(X^{-1}H) + \mathcal{O}(\alpha^2)) \\ &= f(X) - \alpha X^{-1} \bullet H + \mathcal{O}(\alpha^2), \end{aligned}$$

so that $f'(X) = -X^{-1}$ and $Df(X)[H] = -X^{-1} \bullet H$.

Logarithmic Barrier Function (cont'd)

Similarly

$$\begin{aligned} f'(X + \alpha H) &= -[X(I + \alpha X^{-1}H)]^{-1} \\ &= -[I - \alpha X^{-1}H + \mathcal{O}(\alpha^2)]X^{-1} \\ &= f'(X) + \alpha X^{-1}HX^{-1} + \mathcal{O}(\alpha^2), \end{aligned}$$

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

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

Finally,

$$f'''(X)[H, G] = -X^{-1}HX^{-1}GX^{-1} - X^{-1}GX^{-1}HX^{-1}.$$

Logarithmic Barrier Function (cont'd)

Theorem: $f(X) = -\ln \det X$ is a convex barrier for $\mathcal{SR}_+^{n \times n}$.

Proof: Define $\phi(\alpha) = f(X + \alpha H)$. We know that f is convex if, for every $X \in \mathcal{SR}_+^{n \times n}$ and every $H \in \mathcal{SR}^{n \times n}$, $\phi(\alpha)$ is convex in α .

Consider a set of α such that $X + \alpha H \succ 0$. On this set

$$\phi''(\alpha) = D^2f(\bar{X})[H, H] = \bar{X}^{-1}H\bar{X}^{-1} \bullet H,$$

where $\bar{X} = X + \alpha H$.

Since $\bar{X} \succ 0$, so is $V = \bar{X}^{-1/2}$ (Property 6), and

$$\begin{aligned} \phi''(\alpha) &= V^2HV^2 \bullet H = \text{trace}(V^2HV^2H) \\ &= \text{trace}((VHV)(VHV)) = \|VHV\|_F^2 \geq 0. \end{aligned}$$

So ϕ is convex.

When $X \succ 0$ approaches a singular matrix, its determinant approaches zero and $f(X) \rightarrow \infty$.

Simplified Notation

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

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

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

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

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 \mathcal{R}^m to $\mathcal{SR}^{n \times n}$.

Simplified Notation (cont'd)

With this notation the **primal** SDP becomes

$$\begin{aligned} \min \quad & C \bullet X \\ \text{s.t.} \quad & \mathcal{A}X = b, \\ & X \succeq 0, \end{aligned}$$

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

The associated **dual** SDP writes

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

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

Solving SDPs with IPMs

Replace the **primal SDP**

$$\begin{aligned} \min \quad & C \bullet X \\ \text{s.t.} \quad & \mathcal{A}X = b, \\ & X \succeq 0, \end{aligned}$$

with the **primal barrier SDP**

$$\begin{aligned} \min \quad & C \bullet X + \mu f(X) \\ \text{s.t.} \quad & \mathcal{A}X = b, \end{aligned}$$

(with a barrier parameter $\mu \geq 0$).

Formulate the Lagrangian

$$L(X, y, S) = C \bullet X + \mu f(X) - y^T (\mathcal{A}X - b),$$

with $y \in \mathcal{R}^m$, and write the first order conditions (FOC) for a stationary point of L :

$$C + \mu f'(X) - \mathcal{A}^* y = 0.$$

Solving SDPs with IPMs (cont'd)

Use $f(X) = -\ln \det(X)$ and $f'(X) = -X^{-1}$.

Therefore the FOC become:

$$C - \mu X^{-1} - \mathcal{A}^* y = 0.$$

Denote $S = \mu X^{-1}$, i.e., $XS = \mu I$.

For a positive definite matrix X its inverse is also positive definite.

The FOC now become:

$$\begin{aligned} \mathcal{A}X &= b, \\ \mathcal{A}^* y + S &= C, \\ XS &= \mu I, \end{aligned}$$

with $X \succ 0$ and $S \succ 0$.

Newton direction

We derive the Newton direction for the system:

$$\begin{aligned} \mathcal{A}X &= b, \\ \mathcal{A}^*y + S &= C, \\ -\mu X^{-1} + S &= 0. \end{aligned}$$

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

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

Newton direction (cont'd)

The differentiation in the above system is a **nontrivial** operation. The direction is the solution of the system:

$$\begin{bmatrix} \mathcal{A} & 0 & 0 \\ 0 & \mathcal{A}^* & \mathcal{I} \\ \mu(X^{-1} \odot X^{-1}) & 0 & \mathcal{I} \end{bmatrix} \cdot \begin{bmatrix} \Delta X \\ \Delta y \\ \Delta S \end{bmatrix} = \begin{bmatrix} \xi_b \\ \xi_C \\ \xi_\mu \end{bmatrix}.$$

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, $f : \mathcal{SR}_+^{n \times n} \mapsto \mathcal{R}$

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

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

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

Nesterov and Nemirovskii,

Interior Point Polynomial Algorithms in Convex Programming: Theory and Applications, SIAM, Philadelphia, 1994.

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

Interior Point Methods:

- Logarithmic barrier functions for SDP and SOCP
 - Self-concordant barriers
 - polynomial complexity (predictable behaviour)
- Unified view of optimization
 - from LP via QP to NLP, SDP, SOCP
- Efficiency
 - good for SOCP
 - problematic for SDP because solving the problem of size n involves linear algebra operations in dimension n^2
 - and this requires n^6 flops!

Use IPMs in your research!

Newton Method goes far beyond IPMs

- **K. Fountoulakis, J. Gondzio and P. Zhlobich**, Matrix-free interior point method for compressed sensing problems, *Math Prog Computation* 6 (2014), pp. 1-31.
- **K. Fountoulakis and J. Gondzio**, A second-order method for strongly convex ℓ_1 -regularization problems, *Mathematical Programming* 156 (2016), pp. 189-219.
- **I. Dassios, K. Fountoulakis and J. Gondzio**, A preconditioner for a primal-dual Newton conjugate gradients method for compressed sensing problems, *SIAM Journal on Scientific Computing* 37 (2015) pp. A2783-A2812.
- **K. Fountoulakis and J. Gondzio**, Performance of first- and second-order methods for ℓ_1 -regularized least squares problems, *Comp Optimization and Appls* 65 (2016) 605–635.
(reports on the solution of a problem with $2^{40} \approx 10^{12}$ vars).

Papers available: <http://www.maths.ed.ac.uk/~gondzio/>