Conic Programming

Michael J. Todd

School of Operations Research and Industrial Engineering,

Cornell University

www.orie.cornell.edu/ ˜miketodd/todd.html

ICCOPT I Summer School

August 1, 2004

Outline

- Conic programming problems
- Weak duality
- Examples and applications
- Strong duality
- Algorithms

I. Conic programming problems

- Linear programming (LP)
- Semidefinite programming (SDP)
- Second-order cone programming (SOCP)
- General conic programming problem
- Hyperbolic, nonnegative polynomial cones

LP:

Given $A \in \Re^{m \times n}$, $b \in \Re^m$, $c \in \Re^n$, consider:

$$
\text{(P)} \qquad \begin{array}{rrcl} \min_x & c^T x & & \\ & Ax & = & b, \\ & x & \geq & 0. \end{array}
$$

Using the same data, we can construct the dual problem:

$$
\text{(D)} \qquad \begin{array}{rrcl} \max_y & b^T y & & \\ & A^T y & \leq & c. \end{array}
$$

LP, cont'd:

We will see that it is useful to explicitly introduce slack variables, to get

$$
\text{(D)} \qquad \begin{array}{rrcrcl}
\max_{y,s} & b^T y & & & & \\
 & A^T y & + & s & = & c, \\
 & & & s & \geq & 0.
\end{array}
$$

SDP:

Given $A_i \in \mathcal{SR}^{p \times p}$ (symmetric real matrices of order p), $i = 1, \ldots, m$, $b \in \Re^m$, $C \in \mathcal{SR}^{p \times p}$, consider:

$$
\text{(P)} \qquad \begin{array}{lrcl}
\min_X & C \bullet X & & \\
& A_i \bullet X & = & b_i, \ i = 1, \ldots, m, \\
& X & \succeq & 0,
\end{array}
$$

where $S \bullet Z :=$ Trace $S^T Z = \sum_i \sum_j s_{ij} z_{ij}$ for matrices of the same dimensions, and $X \succeq 0$ means X is symmetric and positive semidefinite (psd). (We'll also write $A \succeq B$ and $B \preceq A$ for $A - B \succeq 0$.) We'll write $\mathcal{SR}^{p \times p}_+$ for the cone of psd real matrices of order p. Note that, instead of the components of the vector x being nonnegative, now the p eigenvalues of the symmetric matrix X are nonnegative.

SDP, cont'd

Using the same data, we can construct another SDP in dual form:

$$
\text{(D)} \qquad \begin{array}{lrcl} \max_y & b^T y & & \\ & \sum_i y_i A_i & \preceq & C, \end{array}
$$

or with an explicit slack matrix,

$$
\text{(D)} \qquad \begin{array}{lrcrcl} \max_{y,S} & b^T y & & & & \\ & \sum_i y_i A_i & + & S & = & C, \\ & & & S & \succeq & 0. \end{array}
$$

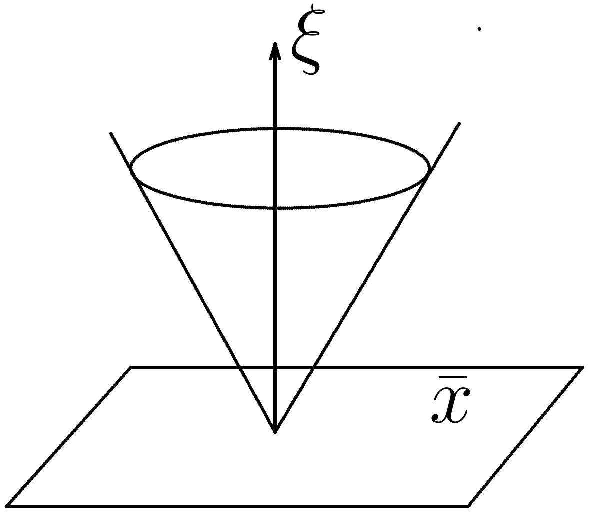
SOCP:

Given $A_j \in \Re^{m\times(1+n_j)}$, $c_j \in \Re^{1+n_j}$, $j = 1,\ldots,k$, and $b \in \Re^m$, consider:

$$
\text{(P)} \qquad \begin{array}{lrcl}
\min_{x_1,\ldots,x_k} & c_1^T x_1 + \ldots + c_k^T x_k & & \\
 & A_1 x_1 + \ldots + A_k x_k & = & b, \\
 & x_j & \in & \mathcal{S}_2^{1+n_j},\ j = 1,\ldots,k,
\end{array}
$$

where $\mathcal{S}_2^{1+q}$ is the second-order cone:

$$\{x := (\xi; \bar{x}) \in \Re^{1+q} :\ \xi \geq \|\bar{x}\|_2\}$$



Again using the same data, we can construct a problem in dual form:

$$
\text{(D)} \quad
\begin{array}{lrcl}
\max_y & b^T y & & \\
 & c_j - A_j^T y & \in & \mathcal{S}_2^{1+n_j},\ j = 1, \ldots, k,
\end{array}
$$

or

$$
\text{(D)} \quad
\begin{array}{lrcccl}
\max_{y, s_1, \ldots, s_k} & b^T y & & & & \\
 & A_1^T y & + & s_1 & = & c_1 \\
 & \vdots & & & & \\
 & A_k^T y & + & s_k & = & c_k \\
 & & & s_j & \in & \mathcal{S}^{1+n_j},\ j = 1, \ldots, k.
\end{array}
$$

General conic programming problem:

Given again $A \in \Re^{m\times n}$, $b \in \Re^m$, $c \in \Re^n$, and a closed convex cone $\mathcal{K} \subset \Re^n$,

$$(\mathrm{P}) \qquad \begin{array}{lrcl} \min_x & \langle c, x\rangle & & \\ & Ax & = & b, \\ & x & \in & K, \end{array}$$

where we have written $\langle c, x\rangle$ instead of $c^T x$ to emphasize that this can be thought of as a general scalar/inner product. E.g., if our original problem is an SDP involving $X \in \mathcal{SR}^{p\times p}$, we need to embed it into $\Re^n$ for some n.

Even though our problem (P) looks very much like LP, it is important to note that every convex programming problem can be written in the form (P).

Standard embedding for matrices ($X \in \Re^{p\times q}$):

$$X \longleftrightarrow x = \mathrm{vec}(X) := \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{p1} \\ x_{p2} \\ \vdots \\ x_{pq} \end{pmatrix} \in \Re^{pq} \text{ and then } S \bullet Z = s^T z =: \langle s, z \rangle.$$

Our matrices are symmetric. For $X \in \mathcal{SR}^{p \times p}$, we could define

$$X \longleftrightarrow \tilde{x} = \widetilde{\mathrm{svec}}(X) := \begin{pmatrix} x_{11} \\ x_{21} \\ x_{22} \\ x_{31} \\ \vdots \\ x_{pp} \end{pmatrix} \in \Re^{p(p+1)/2}$$ and then

$$S \bullet Z = \tilde{s}_1 \tilde{z}_1 + 2\tilde{s}_2 \tilde{z}_2 + \tilde{s}_3 \tilde{z}_3 + \ldots =: \langle \tilde{s}, \tilde{z} \rangle,$$

or better $X \longleftrightarrow x = \mathrm{svec}(X) = \begin{pmatrix} x_{11} \\ \sqrt{2}x_{21} \\ x_{22} \\ \sqrt{2}x_{31} \\ \vdots \\ x_{pp} \end{pmatrix} \in \Re^{p(p+1)/2}$

and then $S \bullet Z = s^T z =: \langle s, z \rangle$.

Conic problem in dual form

How do we construct the corresponding problem in dual form? We need the dual cone:

$$K^* = \{s \in \Re^n : \ \langle s, x \rangle \geq 0 \text{ for all } x \in K\}.$$

Then we define

$$\text{(D)} \qquad \begin{array}{rrcrcl} \max_{y,s} & \langle b, y \rangle & & & & \\ & A^* y & + & s & = & c, \\ & & & s & \in & K^*. \end{array}$$

What is A^*? The operator adjoint to A, so that for all x, y, $\langle A^* y, x \rangle = \langle Ax, y \rangle$.

If $\langle \cdot, \cdot \rangle$ is the usual dot product, $A^* = A^T$.

Two other cones of interest:

Let $p : \Re^n \mapsto \Re$ be a polynomial, and fix some $e \in \Re^n$. We say p is hyperbolic in direction e if for every $x \in \Re^n$, $p(\lambda e - x)$ has all roots λ real. These roots are called the eigenvalues of x. Such a p defines a cone K via

$$K := \{x \in \Re^n : \text{ all eigenvalues of } x \text{ are nonnegative}\}.$$

Surprisingly, this is a closed convex cone, called (the closure of) the hyperbolicity cone for p in the direction e.

(Work by Güler, Bauschke, Lewis, Sendov, and Renegar: see, e.g.,

www.optimization-online.org/DB_HTML/2004/03/844.html.)

Next, consider the vector space of all polynomials of total degree d in q variables (think of the coefficients as components in some large-dimensional $\Re^n$), and within it the cone of those polynomials that are always nonnegative.

This allows you to model the problem of global minimization of a nonconvex polynomial as a convex problem! Must be hard ...

The dual cone is the cone of moments.

(Work by Shor, Parrilo, Lasserre, Bertsimas, Pena, ...: see, e.g., the Gloptipoly home page at www.laas.fr/~henrion/software/gloptipoly/gloptipoly.html.)

II. Weak duality

Above we have seen problems “in primal form” and “in dual form” constructed from the same data. Here we note that weak duality holds for these pairs of problems, so we are justified in calling them dual problems.

We start with the well-known one-line proof for LP:

If x is feasible for (P) and (y, s) for (D), then

$$c^T x - b^T y = (A^T y + s)^T x - (Ax)^T y \overset{(i)}{=} s^T x \overset{(ii)}{\geq} 0.$$

Here the key ingredients are:

(i) $(A^T y)^T x = (Ax)^T y$, and

(ii) $s^T x \geq 0$,

both trivial.

SDP weak duality

Now for SDP as we have written it above:

For X feasible in (P) and (y, S) in (D), we have

$$\begin{array}{rcl} C \bullet X - b^T y & = & (\sum y_i A_i + S) \bullet X - ((A_i \bullet X)_{i=1}^m)^T y \\ & = & (\sum y_i A_i) \bullet X + S \bullet X - \sum y_i (A_i \bullet X) \\ & \stackrel{(i)}{=} & S \bullet X \stackrel{(ii)}{\geq} 0. \end{array}$$

Here the key facts are:

(i) $(\sum y_i A_i) \bullet X = \sum y_i (A_i \bullet X)$ by linearity of the trace; and

(ii) $S \bullet X \geq 0$, i.e., if K denotes the cone $\mathcal{SR}_+^{p \times p}$ of psd matrices, then $K \subseteq K^*$;

indeed we'll see below that $K = K^*$.

SOCP weak duality

Next for SOCP:

For $(x_1, \ldots, x_k)$ feasible for (P) and $(y, (s_1, \ldots, s_k))$ for (D), we have

$$\begin{aligned} \sum c_j^T x_j - b^T y &= \sum (A_j^T y + s_j)^T x_j - (\sum A_j x_j)^T y \\ &\stackrel{(i)}{=} \sum s_j^T x_j \stackrel{(ii)}{\geq} 0. \end{aligned}$$

Here we have used

(i) $\sum (A_j^T y)^T x_j = (\sum A_j x_j)^T y$ and

(ii) If $K = \mathcal{S}_2^{1+q}$, then $K \subseteq K^*$; indeed we'll see that again $K = K^*$ for this cone.

Weak duality for general conic problems

These are all special cases of weak duality for general conic programming:

If x is feasible for (P) and (y, s) for (D), then

$$\langle c, x\rangle - \langle b, y\rangle = \langle A^* y + s, x\rangle - \langle Ax, y\rangle \stackrel{(i)}{=} \langle s, x\rangle \stackrel{(ii)}{\geq} 0,$$

where (i) follows by definition of the adjoint operator A^* and (ii) by definition of the dual cone K^*.

So in all cases we have weak duality, which suggests that it is worthwhile to consider (P) and (D) together. In many cases, strong duality holds, and then it is very worthwhile!

Weak duality, cont'd

In the cases above, our proofs of (i) indicate that we have the correct adjoint operator A^* for LP, SDP, and SOCP. We need to show that, if K is the second-order cone or the cone of psd matrices, then $K^* = K$, i.e., K is self-dual. It is easy to see that

$$(K_1 \times \ldots \times K_k)^* = K_1^* \times \ldots \times K_k^*,$$

so we will also have covered general SOCP.

The SO cone is self-dual

$$(\mathcal{S}_2^{1+q})^* = \mathcal{S}_2^{1+q}.$$

First, $\supseteq$: if $s := (\sigma; \bar{s})$ and $x := (\xi; \bar{x})$ lie in $\mathcal{S}_2^{1+q}$, then

$$s^T x = \sigma\xi + \bar{s}^T \bar{x} \geq \sigma\xi - \|\bar{s}\|_2 \|\bar{x}\|_2$$

by Cauchy-Schwarz, and this is nonnegative.

Next, $\subseteq$: Suppose $s := (\sigma; \bar{s})$ has $s^T x \geq 0$ for all x in $\mathcal{S}_2^{1+q}$. If $\bar{s} = 0$, take $x := (1; 0)$ to get $\sigma \geq 0 = \|\bar{s}\|_2$. Else choose $x := (\|\bar{s}\|_2; -\bar{s})$ to get

$$0 \leq s^T x = \sigma\|\bar{s}\|_2 - \bar{s}^T \bar{s} = \sigma\|\bar{s}\|_2 - \|\bar{s}\|_2^2$$

and hence conclude that $\sigma \geq \|\bar{s}\|_2$.

The psd cone is self-dual

$$(\mathcal{SR}_+^{p\times p})^* = \mathcal{SR}_+^{p\times p}.$$

First, $\supseteq$: Suppose S and X are psd. We use

- S has a psd square root $S^{1/2}$.

(Proof: $S = Q\Lambda Q^T$ with Q orthogonal and Λ diagonal with nonnegative diagonal entries λ_j. Define $\Lambda^{1/2} :=$ Diag$(\lambda_j^{1/2})$, and note that $\Lambda^{1/2}\Lambda^{1/2} = \Lambda$. Then define $S^{1/2} := Q\Lambda^{1/2}Q^T$. This is psd (its eigenvalues are $\lambda_j^{1/2} \geq 0$), and

$$S^{1/2}S^{1/2} = Q\Lambda^{1/2}Q^TQ\Lambda^{1/2}Q^T = Q\Lambda^{1/2}\Lambda^{1/2}Q^T = S.)$$

Also,

- For any $r \times s$ P and $s \times r$ Q, Trace $(PQ) =$ Trace (QP).
  (Proof: Both are $\sum_{i,j} p_{ij}q_{ji}$.)

The psd cone is self-dual, cont'd

Putting these facts together, we get

$$S \bullet X = \text{Trace }(SX) = \text{Trace }(S^{1/2}S^{1/2}X) = \text{Trace }(S^{1/2}XS^{1/2}).$$

Now $S^{1/2}XS^{1/2}$ is psd, and hence its trace

($=$ the sum of its eigenvalues $=$ the sum of its diagonal entries) is nonnegative.

An alternative proof writes

$$S \bullet X = \text{Trace }(SX) = \text{Trace }(Q\Lambda Q^T X) = \text{Trace }(\Lambda(Q^T XQ)).$$

Now $Q^T XQ$ is psd, so its diagonal entries are nonnegative, and premultiplying by Λ just multiplies these by nonnegative numbers.

The psd cone is self-dual, cont'd

Next, $\subseteq$: This uses another key fact,

- For any $u \in \Re^p$, $uu^T \in \mathcal{SR}_+^{p \times p}$.

Indeed, for any $v \in \Re^p$, $v^T(uu^T)v = (u^Tv)^2 \geq 0$.

So if $S \in (\mathcal{SR}_+^{p \times p})^*$, we have

$$u^T S u = \text{Trace } (u^T S u) = \text{Trace } (S u u^T) = S \bullet uu^T \geq 0$$

for any $u \in \Re^p$, so $S \in \mathcal{SR}_+^{p \times p}$.

Optimizing ...

James Branch Cabell:

The optimist proclaims that we live in the best of all possible worlds; and the pessimist fears this is true.

Antonio Gramsci:

I’m a pessimist because of intelligence, but an optimist because of will.

III. Examples and applications

- matrix optimization
- quadratically constrained quadratic programming (QCQP)
- control theory
- relaxations in combinatorial optimization
- extensions of Chebyshev's inequality
- Fermat-Weber problem
- global optimization of polynomials

More applications

Many other interesting applications are to be discussed in ICCOPT. See sessions MM2, MM4, MM5, MA5, MA6, MS (Scherer), TA6, TM1, TM2, WM2, WS (Tseng), WA3, and WA4.

In addition, survey papers/books/articles can be found at the following sites:

www.stanford.edu/~boyd/sdp-apps.html

www.stanford.edu/~boyd/socp.html

rutcor.rutgers.edu/~alizadeh/Sdppage/PAPER3/papers.ps.gz

www-math.mit.edu/~goemans/semidef-survey.ps

www-fp.mcs.anl.gov/otc/Guide/OptWeb/continuous/constrained/sdp/

www.ec-securehost.com/SIAM/MP02.html

www.gams.com/conic/.

Finally, the field of robust optimization gives rise to SDPs and SOCPs.

Matrix optimization

Suppose we have a symmetric matrix

$$A(y) := A_0 + \sum_{i=1}^{m} y_i A_i$$

depending affinely on $y \in \Re^m$. We wish to choose y to

minimize the maximum eigenvalue of $A(y)$.

Note: $\lambda_{\max}(A(y)) \le \eta$ iff all e-values of $\eta I - A(y)$ are nonnegative iff $A(y) \preceq \eta I$.

This gives

$$\begin{array}{lcccccc} \max_{\eta,y} & -\eta & & & & & \\ & -\eta I & + & \sum_{i=1}^m y_i A_i & \preceq & -A_0, \end{array}$$

an SDP problem of form (D).

QCQP

Proposition (Schur complements) Suppose $B \succ 0$. Then

$$\begin{pmatrix} B & P \\ P^T & C \end{pmatrix} \succeq 0 \Leftrightarrow C - P^T B^{-1} P \succeq 0.$$

Hence the convex quadratic constraint $(Ay+b)^T(Ay+b) - c^T y - d \leq 0$ holds iff

$$\begin{pmatrix} I & Ay+b \\ (Ay+b)^T & c^T y + d \end{pmatrix} \succeq 0,$$

or alternatively iff $\sigma \geq \|\bar{s}\|_2$, $\sigma := c^T y + d + \frac{1}{4}$, $\bar{s} := (c^T y + d - \frac{1}{4}; Ay + b)$.

This allows us to model the QCQP of minimizing a convex quadratic function subject to convex quadratic inequalities as either an SDP or an SOCP.

Control theory

Suppose the state of a system is defined by $\dot{x} \in \text{conv}\{P_1, P_2, \ldots, P_m\}\, x$.

A sufficient condition that $x(t)$ is bounded for all time is that there is $Y \succ 0$ with $V(x) := \frac{1}{2} x^T Y x$ nonincreasing, i.e.,

$$\dot{V}(x) = \frac{1}{2} x^T (YP + P^T Y) x \leq 0$$

for all $P \in \text{conv}\{P_1, P_2, \ldots, P_m\}$. This leads to

$$\begin{array}{llcrcl}
\max_{\eta, Y} & -\eta & & & & \\
 & -\eta I & + & Y & \preceq & 0, \\
 & & & -Y & \preceq & -I, \\
 & YP_i & + & P_i^T Y & \preceq & 0,\ i = 1, \ldots, m.
\end{array}$$

(Note the block diagonal structure.)

Relaxations in combinatorial optim'n

The Maximum Cut Problem: given an undirected (wlog complete) graph on $V = \{1, \ldots, n\}$ with nonnegative edge weights $W = (w_{ij})$, find a cut $\delta(S) := \{\{i, j\} : i \in S, j \notin S\}$ with maximum weight.

(IP): $\max\{\frac{1}{4}\sum_i \sum_j w_{ij}(1 - x_i x_j) : x_i \in \{-1, +1\}, i = 1, \ldots, n\}$.

The constraint is the same as $x_i^2 = 1$ all i. Now

$\{X : x_{ii} = 1, i = 1, \ldots, n,\ X \succeq 0, \text{rank}(X) = 1\} = \{xx^T : x_i^2 = 1, i = 1, \ldots, n\}$.

So a relaxation is:

$$\begin{array}{rcl} \frac{1}{4}\sum\sum w_{ij} - \frac{1}{4}\min_X W \bullet X & & \\ e_i e_i^T \bullet X & = & 1,\ i = 1, \ldots, n, \\ X & \succeq & 0. \end{array}$$

This gives a good bound and a good feasible solution (within 14%) (Goemans and Williamson).

Extension of Chebyshev's inequality

Suppose we have a random vector $X \in \Re^n$ and we know $E(X) = \bar{x}$, $E(XX^T) = \Sigma$. We wish to bound the probability that $X \in C$, with

$$C := \{x \in \Re^n : x^T A_i x + 2b_i^T x + c_i < 0,\ i = 1, \ldots, m\}.$$

A tight bound is given by the solution to the SDP

$$\begin{array}{lrcl}
\max_{Y,y,\eta,\zeta} & 1 - \Sigma \bullet Y - 2\bar{x}^T y - \eta & & \\
 & \begin{bmatrix} Y - \zeta_i A_i & y - \zeta_i b_i \\ (y - \zeta_i b_i)^T & \eta - 1 - \zeta_i c_i \end{bmatrix} & \succeq & 0,\ i = 1, \ldots, m \\
 & \begin{bmatrix} Y & y \\ y^T & \eta \end{bmatrix} & \succeq & 0, \\
 & \zeta_i & \geq & 0,\ i = 1, \ldots, m.
\end{array}$$

Chebyshev's inequality, cont'd

Suppose we have a feasible solution. Then, for any $x \in \Re^n$,

$$x^T Y x + 2y^T x + \eta \geq 1 + \zeta_i (x^T A_i x + 2b_i^T x + c_i),\ i = 1, \ldots, m,$$

and $x^T Y x + 2y^T x + \eta \geq 0$. So this quantity is at least 1 if $x \notin C$, and at least 0 for $x \in C$. Hence the expectation of $X^T Y X + 2y^T X + \eta$ is at least $1 - P(X \in C)$, but this expectation is exactly $\Sigma \bullet Y + 2y^T \bar{x} + \eta$.

To show that it is tight we use SDP duality (Vandenberghe, Boyd, and Comanor).

The Fermat-Weber location problem

We want to choose $y \in \Re^2$ to minimize the sum of its distances to the given points $p_i \in \Re^2$, $i = 1, \ldots, m$. This becomes

$$\begin{array}{ll} \min_{y,\eta} & \eta_1 + \cdots + \eta_m \\ & \eta_i \geq \|y - p_i\|_2, \quad i = 1, \ldots, m, \end{array}$$

an SOCP problem in dual form. Note that here

all the second-order cones have dimension 3.

The dual is also interesting: it can be written as

$$\begin{array}{lll} \max_{x_1,\ldots,x_m} & p_1^T x_1 + \cdots + p_m^T x_m & \\ & x_1 + \cdots + x_m & = \quad 0, \\ & \|x_i\|_2 & \leq \quad 1, \; i = 1, \ldots, m. \end{array}$$

Global optimization of polynomials

Lastly, we just indicate the approach to global optimization of polynomials using conic programming.

Given a polynomial function θ of q variables, the globally optimal value of minimizing $\theta(x)$ over all $x \in \Re^q$ is the maximum value of η such that the polynomial $p(x) \equiv \theta(x) - \eta$ is nonnegative for all x, and this is a convex set of polynomials (described say by all their coefficients).

This equivalence indicates that the convex cone of nonnegative polynomials must be hard to deal with. It can be approximated using SDPs; clearly if p is the sum of squares of polynomials then it is nonnegative (but not conversely); however, using extensions of these ideas we can approximate the optimal value as closely as desired.

IV. Strong duality

Consider

$$\min \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \bullet X, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \bullet X = 0, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} \bullet X = 2,\ X \succeq 0,$$

with optimal solution $X = \text{Diag}(0;0;1)$ and optimal value 1, while its dual

$$\max\ 2y_2, \qquad y_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + y_2 \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} \preceq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

has optimal solution $y = (0;0)$ and optimal value 0.

Strong duality, cont'd

Hence strong duality, by which we mean that both (P) and (D) have optimal solutions and there is no duality gap, doesn't hold in general in conic programming. We need to add a regularity condition.

We say x is a strictly feasible solution for (P) if it is feasible and $x \in \text{int } K$; similarly (y, s) is a strictly feasible solution for (D) if it is feasible and $s \in \text{int } K^*$.

Theorem Suppose (P) has a feasible solution and (D) a strictly feasible solution. Then (P) has a nonempty bounded set of optimal solutions, and there is no duality gap.

Corollary If both (P) and (D) have strictly feasible solutions, strong duality holds.

Notation: $\mathcal{F}(\text{P}) := \{$feasible solutions of (P)$\}$ and similarly for (D).
$\mathcal{F}^0(\text{P}) := \{$strictly feasible solutions of (P)$\}$ and similarly for (D).

Strong duality, cont'd

Proof sketch The set of optimal solutions to (P) is unchanged if we add the constraint $\langle c, x\rangle \leq \langle c, \hat{x}\rangle$ for an arbitrary feasible solution $\hat{x}$. But this constraint is equivalent to $\langle \hat{s}, x\rangle \leq \langle \hat{s}, \hat{x}\rangle$, where $(\hat{y}, \hat{s})$ is an arbitrary strictly feasible solution for (D), and the set of $x \in K$ satisfying this is bounded. Hence we are minimizing a continuous function on a compact set, giving the first part.

If ζ is the optimal value of (P), we can apply a separating hyperplane argument to K and $\{x \in \Re^n : Ax = b, \langle c, x\rangle \leq \zeta - \epsilon\}$ for an arbitrary positive ϵ to get a feasible dual solution within ϵ of the optimal value of (P).

Henceforth, assume that both (P) and (D) have strictly feasible solutions, and (wlog) that A has full row rank.

Barriers ...

Thomas Jefferson:

To draw around the whole nation the strength of the General Government, as a barrier against foreign foes ...

Mary Wollstonecraft:

What a weak barrier is truth when it stands in the way of an hypothesis!

Robert Frost:

My apple trees will never get across
And eat the cones under his pines, I tell him.
He only says, “Good fences make good neighbors.”

V. Algorithms

We will concentrate on interior-point methods, which have the theoretical advantage of polynomial-time complexity, while also performing very well in practice on medium-scale problems.

F : int $K \rightarrow \Re$ is a barrier function for K if

- F is strictly convex; and
- $x_k \rightarrow \bar{x} \in \partial K \Rightarrow F(x_k) \rightarrow +\infty$.

It is helpful to think of F as defined on $\Re^n$: set $F(x) = +\infty$ for $x \notin$ int K.

Similarly, let F_* be a barrier function for int K^*.

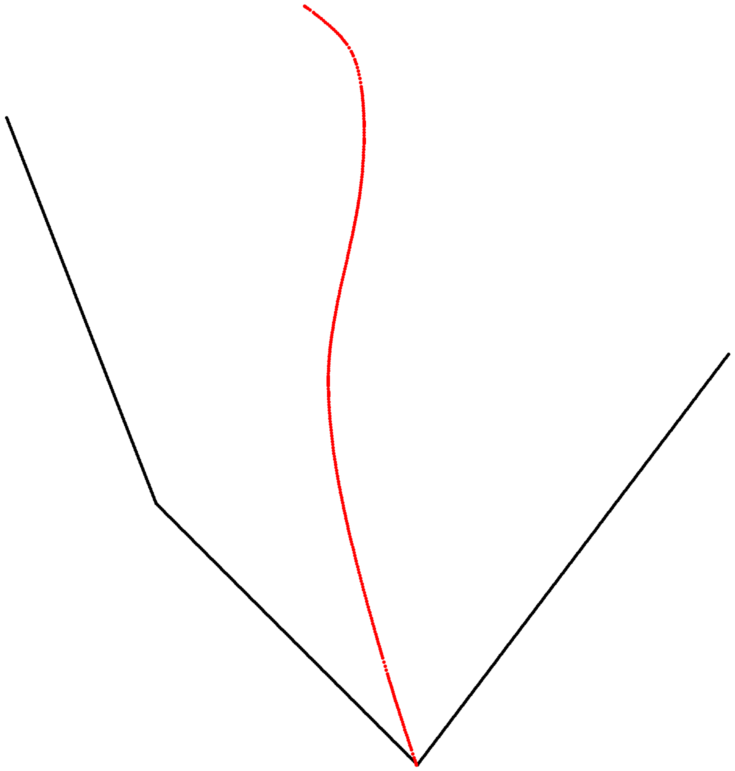
Barrier Problems: Choose $\mu > 0$ and consider

$$(\mathrm{BP}_\mu) \quad \min \langle c, x \rangle + \mu F(x), \quad Ax = b \quad (x \in \text{int } K),$$

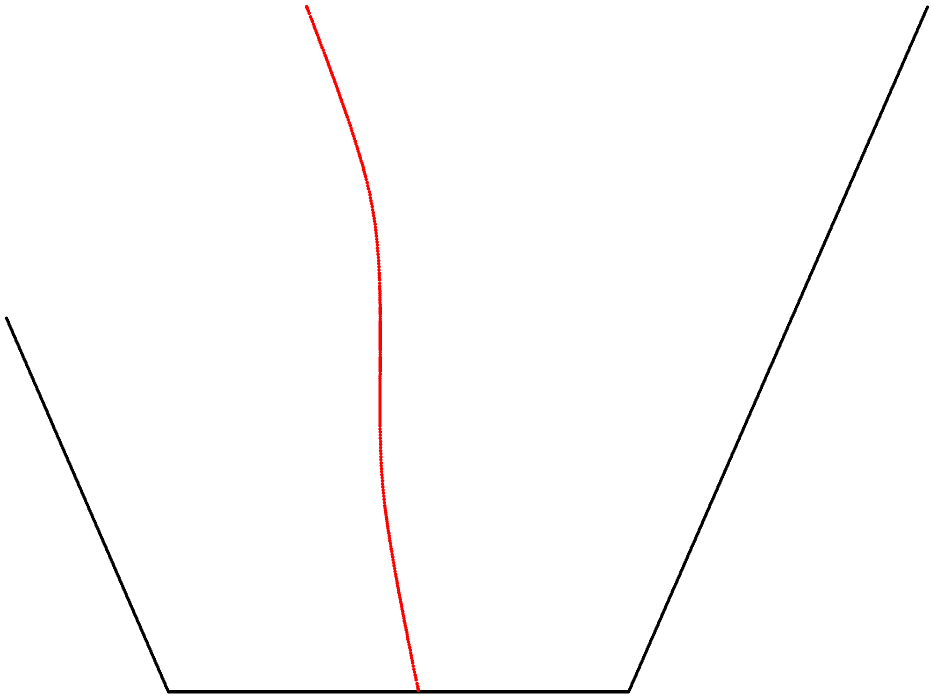
$$(\mathrm{BD}_\mu) \quad \max \langle b, y \rangle - \mu F_*(s), \quad A^* y + s = c \quad (s \in \text{int } K^*).$$

Central paths

These have unique solutions $x(\mu)$ and $(y(\mu), s(\mu))$ varying smoothly with μ, forming trajectories in the feasible regions, the so-called central paths:



$\mathcal{F}$(P)



$\mathcal{F}$(D)

Self-concordant barriers

F is a ν-self-concordant barrier for K (Nesterov and Nemirovski) if

- F is a C^3 barrier for K;
- For all $x \in \text{int } K$, $D^2 F(x)$ is pd; and
- For all $x \in \text{int } K$, $d \in \Re^n$,

  (i) $|D^3 F(x)[d,d,d]| \leq 2(D^2 F(x)[d,d])^{3/2}$;

  (ii) $|DF(x)[d]| \leq \sqrt{\nu}(D^2 F(x)[d,d])^{1/2}$.

F is ν-logarithmically homogeneous if

- For all $x \in \text{int } K$, $\tau > 0$, $F(\tau x) = F(x) - \nu \ln \tau$ ($\Rightarrow$ (ii)).

Examples: for $K = \Re^n_+$: $F(x) := -\ln(x) := -\sum \ln(x^{(j)})$ with $\nu = n$;

for $K = \mathcal{SR}^{p \times p}_+$: $F(X) := -\ln \det X = -\sum \ln(\lambda_j(X))$ with $\nu = p$;

for $K = \mathcal{S}^{1+q}_2$: $F(\xi; \bar{x}) := -\ln(\xi^2 - \|\bar{x}\|_2^2)$ with $\nu = 2$.

Properties

Henceforth, F is a ν-LHSCB for K.

Define the dual barrier: $F_*(s) := \sup\{-\langle s, x\rangle - F(x)\}$.

Then F_* is a ν-LHSCB for K^*.

$$F(x) = -\ln(x) \Rightarrow F_*(s) = -\ln(s) - n;$$

$$F(X) = -\ln\det X \Rightarrow F_*(S) = -\ln\det S - p.$$

Properties: For all $x \in \text{int } K$, $\tau > 0$, $s \in \text{int } K^*$,

- $F'(\tau x) = \tau^{-1} F'(x), \quad F''(\tau x) = \tau^{-2} F''(x), \quad F''(x)x = -F'(x)$.
- $x \in \text{int } K \Rightarrow -F'(x) \in \text{int } K^*$.
- $\langle -F'(x), x\rangle = \langle s, -F'_*(s)\rangle = \nu$.
- $s = -F'(x) \Leftrightarrow x = -F'_*(s)$.
- $F''_*(-F'(x)) = [F''(x)]^{-1}$.
- $\nu \ln\langle s, x\rangle + F(x) + F_*(s) \geq \nu\ln\nu - \nu$, with equality iff $s = -\mu F'(x)$ (or $x = -\mu F'_*(s)$) for some $\mu > 0$.

Central path equations

Optimality conditions for barrier problems:

x is optimal for (BP$_\mu$) iff $\exists(y,s)$ with

$$\begin{array}{ccccccl} & A^*y & + & s & = & c, & s \in \text{int } K^*, \\ Ax & & & & = & b, & x \in \text{int } K, \\ \mu F'(x) & & + & s & = & 0. & \end{array}$$

Similarly, (y,s) is optimal for (BD$_\mu$) iff $\exists x$ with the same first two equations and

$$x + \mu F_*'(s) = 0.$$

These two sets of equations are equivalent if F and F_* are as above!

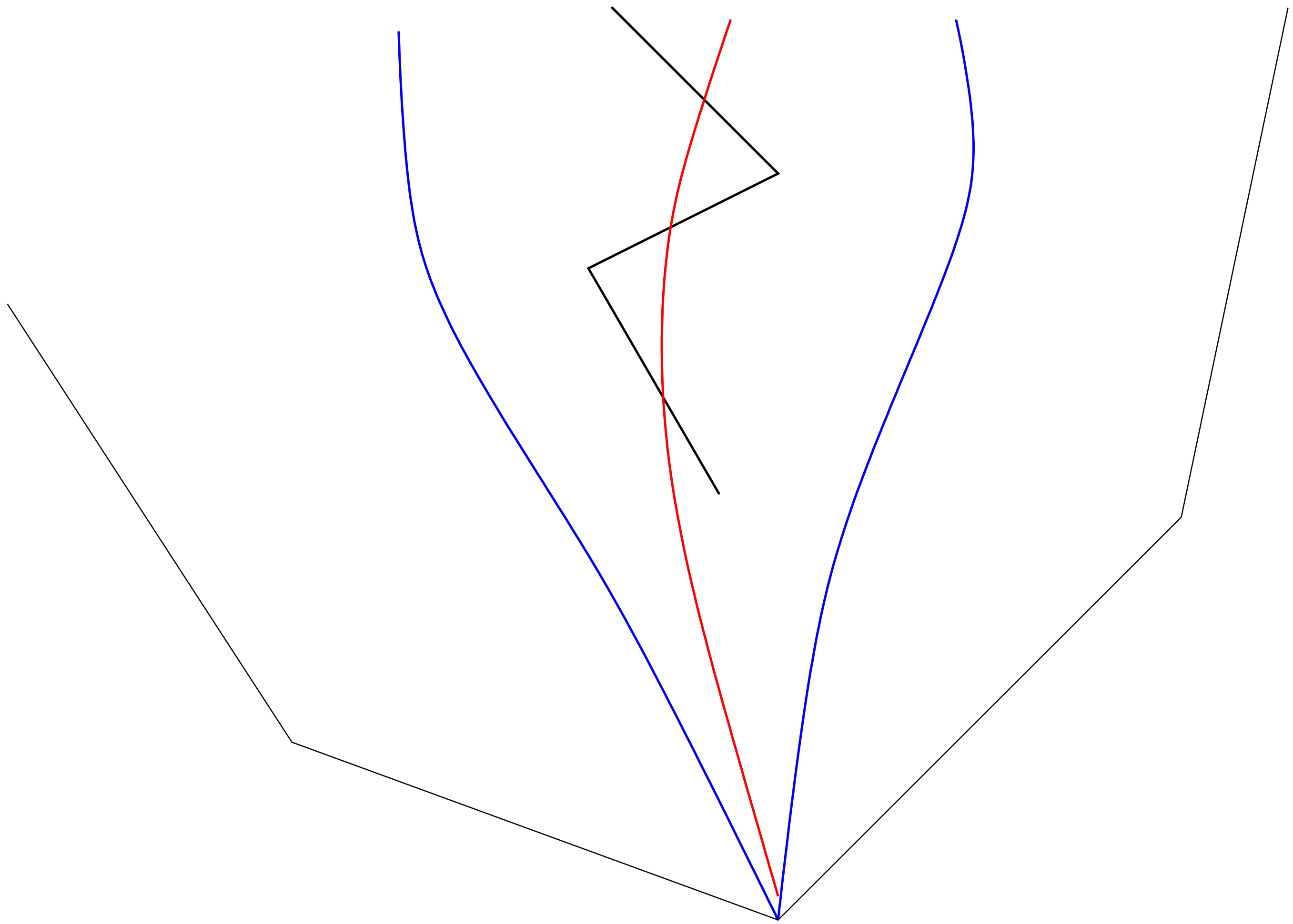
Also, if we have $x(\mu)$ solving (BP$_\mu$), we can easily get $(y(\mu), s(\mu))$ with duality gap

$$\langle s(\mu), x(\mu)\rangle = \mu \left\langle -F'(x(\mu), x(\mu)\right\rangle = \nu\mu,$$

which tends to zero as $\mu \downarrow 0$ (this provides an alternative proof of strong duality).

Path-following algorithms

This leads to theoretically efficient path-following algorithms which use Newton's method to approximately follow the paths:



Complexity

Given a strictly feasible (x_0, y_0, s_0) close to the central path, we can produce a strictly feasible (x_k, y_k, s_k) close to the central path with

$$\langle c, x_k \rangle - \langle b, y_k \rangle = \langle s_k, x_k \rangle \leq \epsilon \langle s_0, x_0 \rangle$$

within

$$O(\nu \ln(1/\epsilon)) \quad \text{or} \quad O(\sqrt{\nu} \ln(1/\epsilon))$$

iterations. This is a primal or dual algorithm, unlike the primal-dual algorithms typically used for LP.

Major work per iteration: forming and factoring the sparse or dense Schur complement matrix $A[F''(x)]^{-1}A^T$ or $AF''_*(s)A^T$.

For LP, $A\,\text{Diag(x)}^2 A^T$ or $A\,\text{Diag(s)}^{-2} A^T$;

for SDP, $(A_i \bullet (XA_jX))$ or $(A_i \bullet (S^{-1}A_jS^{-1}))$.

Can we devise symmetric primal-dual algorithms?

Self-scaled cones

Yes, for certain cones K and barriers F. We need to find, for every $x \in \text{int } K$ and $s \in \text{int } K^*$, a scaling point $w \in \text{int } K$ with

$$F''(w)x = s.$$

Then $F''(w)$ approximates $\mu F''(x)$ and simultaneously $F''_*(t) := F''_*(-F'(w)) = [F''(w)]^{-1}$ approximates $\mu F''_*(s)$. Hence we find our search direction $(\Delta x, \Delta y, \Delta s)$ from

$$\begin{array}{rcccl} & A^*\Delta y & + & \Delta s & = & r_d, \\ A\Delta x & & & & = & r_p, \\ F''(w)\Delta x & & + & \Delta s & = & r_c. \end{array}$$

This generalizes standard primal-dual methods for LP.

Self-scaled cones, cont'd

For what cones can we find such barriers? So-called self-scaled cones (Nesterov-Todd), also the same as symmetric (homogeneous and self-dual) cones (Güler), which have been completely characterized. Includes LP, SDP, SOCP (and not much else).

There is another approach to defining central paths and hence algorithms, with no barrier functions. The idea is to generalize the characterization of LP optimality using complementary slackness, and the definition of the central path using perturbed complementary slackness conditions $x_j s_j = \mu$ for each j. The corresponding general structure is a Euclidean Jordan algebra and its cone of squares. These give precisely the same class of cones as above! (Faybusovich and Güler.)

The corresponding perturbed complementary slackness conditions for SDP are

$$\frac{1}{2}(XS + SX) = \mu I.$$

LP and NLP approaches

There are a variety of other methods for conic programming problems, which typically sacrifice the polynomial time complexity of interior-point methods to get improved efficiency for certain large-scale problems.

There are active-set-based or simplex-like methods (Anderson and Nash, Pataki, Goldfarb, Muramatsu).

There are methods that treat $\min \lambda_{\max}(A(y))$ and related problems as nonsmooth convex minimization problems, and exploit their special structure (the spectral bundle method of Helmberg and Rendl).

And there are methods that derive a smooth but nonconvex nonlinear programming problem (e.g., by substituting $S = LL^T$, and replacing the constrained variable S by the unconstrained variable L) (Burer, Monteiro, and Zhang).

Punch line

The wealth of applications of conic programming problems and the availability of efficient algorithms for solving medium- to large-scale instances has revolutionized optimization in the last ten years!

Resources

Books:

The “bible” is Nesterov and Nemirovski’s

“Interior Point Polynomial Algorithms in Convex Programming”

(www.ec-securehost.com/SIAM/AM13.html), but it is very hard to read.

Easier is Renegar’s

“A Mathematical View of Interior-Point Methods in Convex Optimization”

(www.ec-securehost.com/SIAM/MP03.html).

Ben-Tal and Nemirovski’s “Lectures on Modern Convex Optimization”

(www.ec-securehost.com/SIAM/MP02.html).

Nesterov’s “Introductory Lectures on Convex Optimization”

(www.wkap.nl/prod/b/1-4020-7553-7).

Resources, cont'd

Of the general books on interior-point methods for mainly LP I recommend Wright's "Primal-Dual Interior-Point Methods" (www.ec-securehost.com/SIAM/ot54.html).

For information on symmetric cones see Faraut and Koranyi's "Analysis on Symmetric Cones" (www.oup.co.uk/isbn/0-19-853477-9).

A lecture series and a survey talk: Nemirovski's "Five Lectures on Convex Optimization" (www.core.ucl.ac.be/SumSch/COO_A.PDF) and Wright's "The Ongoing Impact of Interior-Point Methods" (www.cs.wisc.edu/~swright/papers/siopt_talk_may02.pdf).

Survey papers by Boyd and his collaborators on applications of SDP and SOCP (www.stanford.edu/~boyd/sdp-apps.html, www.stanford.edu/~boyd/socp.html).

A paper by Goemans on the use of SDP in combinatorial optimization (www-math.mit.edu/~goemans/semidef-survey.ps).

Resources, cont'd

Papers by Lewis and Overton and by Todd on SDP (cs.nyu.edu/cs/faculty/overton/papers/psfiles/acta.ps, www.orie.cornell.edu/~miketodd/soa5.ps).

Handbook of SDP: see www.wkap.nl/prod/b/0-7923-7771-0.

Useful web sites: the Interior-Point Methods Online site of Wright (www-unix.mcs.anl.gov/otc/InteriorPoint/) and the SDP pages of Helmberg and Alizadeh (www-user.tu-chemnitz.de/~helmberg/semidef.html, rutcor.rutgers.edu/~alizadeh/Sdppage/index.html).

Sites for software: See Helmberg's site above and also www-neos.mcs.anl.gov/neos/server-solvers.html#SDP, www.gamsworld.org/cone/solvers.htm.