Introduction to conic optimization

June 20th 2016 - June 22th 2016

e.d.andersen@mosek.com

www.mosek.com

How to solve conic optimization problems

Topics:

- Linearizing the quadratic cone.
- The primal and primal-dual interior-point methods.
- The homogeneous primal-dual interior-point method.

Purpose:

- Learn what affects solution time.

Section 1

Linearization of conic quadratic cone

The Ben-Tal and Nemirovski linearization

- Can the quadratic cone be approximated by a polynomial number ($O(n)$) linear inequalities?
- The answer is **yes** as proved by Ben-Tal and Nemirovski [1].
  - Using an appropriate definition of approximated.
- See also the ph.d. thesis of Francois Glineur [3].

What is meant by approximation

Definition

A set $\mathcal{U} \in \mathbb{R}^n$ is said to be an $\varepsilon-$approximation of the second-order cone $\mathcal{K}_q$ if and only if we have

$$K_q \subseteq \mathcal{U} \subseteq \mathcal{K}_q^{\varepsilon} = \{x \in R^n : \ x_1(1+\varepsilon) \geq \|x_{2:n}\|\}$$

Main idea:

- Prove that there is $\varepsilon-$ approximation for the 3 dimensional quadratic cone.
- Prove that any quadratic cone can be written using a number of 3 dimensional cones.

Define the set

$$
\begin{aligned}
\mathcal{F}^k := \{(r, \alpha_0, \ldots, \alpha_k, \beta_0, \ldots \beta_k) \in \mathbb{R}^{2k+3} :& \\
\alpha_{i+1} &= \alpha_i \cos\left(\frac{\pi}{2^i}\right) + \beta_i \sin\left(\frac{\pi}{2^i}\right), \quad i = 0, \ldots, k-1, \\
\beta_{i+1} &\geq \beta_i \cos\left(\frac{\pi}{2^i}\right) - \alpha_i \sin\left(\frac{\pi}{2^i}\right), \quad i = 0, \ldots, k-1, \\
-\beta_{i+1} &\geq \beta_i \cos\left(\frac{\pi}{2^i}\right) - \alpha_i \sin\left(\frac{\pi}{2^i}\right), \quad i = 0, \ldots, k-1, \\
r &= \alpha_k \cos\left(\frac{\pi}{2^k}\right) + \beta_k \sin\left(\frac{\pi}{2^k}\right)\}
\end{aligned}
$$

Define

$$
\mathcal{G}^k := \{x \in \mathbb{R}^3 : \ (x_1, x_2, \alpha_1, \ldots, \alpha_k, x_3, \beta_1, \ldots \beta_k) \in \mathcal{F}\}.
$$

Lemma

$\mathcal{G}^k$ is an $\varepsilon-$approximation for K_q^3 where

$$\varepsilon = \cos\left(\frac{\pi}{2^k}\right)^{-1} - 1$$

For a proof see [3].
Quite good result because

k	$\varepsilon \leq$
2	0.5
4	0.02
8	1.0e-4
16	2.0e-9

The decomposition

Let

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

Note

$$\begin{aligned} \sum_{j=2}^{\lfloor \frac{n}{2} \rfloor} x_j^2 &\leq y_l^2, \\ \sum_{j=\lceil \frac{n}{2} \rceil}^{n} x_j^2 &\leq y_r^2, \\ y_l^2 + y_r^2 &\leq x_1^2, \\ 0 &\leq x_1, y_l, y_r. \end{aligned}$$

is another representation of the quadratic cone.

- The largest cone has about $\frac{1}{2}n$ variables in the new representation.
- Had to introduce 2 cones and 2 variables.

- Apply the idea recursively.
- If each 3 dimensional cone is $\varepsilon-$approximated then the approximation for the big cone is

$$\prod_{l=1}^{q}(1+\varepsilon)-1.$$

where $q \approx \log_2(n)$.
- Hence using $O(1)n$ variables and $O(1)n$ linear constraints it is possible to build a $\varepsilon-$ approximation to the quadratic cone.
- Warning: No absolute bound can be given on the quality of the objective value of an approximated quadratic problem. [3].

Computational results

- The results of Glineur suggests:
  - The linearized problems are very hard for the simplex algorithm.
  - The linearized problem can be solved using an interior-point reasonably well.
  - The primal-dual conic interior-point algorithm is much better.
- An interesting application in mixed integer conic optimization is reported in [7].
- The cone decomposition approach is used extensively in linearization of B&B algorithms for conic quadratic mixed-integer problems.

Section 2

Interior-point methods

Interior point methods for linear optimization

The linear optimization problem:

$$\begin{array}{lll} \text{minimize} & c^T x & \\ \text{subject to} & Ax & = & b, \\ & x \geq 0. & \end{array} \tag{1}$$

Assumptions:

- $A \in \mathbb{R}^{m \times n}$ is of full row rank.
- $\exists x^0$ such that $Ax^0 = b$ and $x^0 > 0$.

A barrier function

The function

$$B(x) = -\log(x)$$

is called a *barrier* function for the cone

$$\mathcal{K}_l = \{x \in \mathbb{R} : \; x \geq 0\}.$$

A barrier function is any function such that

$$\lim_{x \to 0^+} B(x) = +\infty.$$

The barrier problem

$$\begin{array}{lll} \text{minimize} & c^T x - \mu \sum_{j=1}^{n} \ln(x_j) & \\ \text{subject to} & Ax & = \; b. \end{array} \tag{2}$$

- μ is a given **positive** parameter.
- Clearly, any feasible solution to (2) is a feasible solution to (1).
- Claim: As μ goes to 0 the optimal solution to (2) converge to the true optimal solution.

Define the Lagrange function

$$L(x, y) := c^T x - \mu \sum_{j=1}^{n} \ln(x_j) - y^T (Ax - b)$$

then the optimality conditions to (2) are

$$\begin{array}{rcccl} \nabla_x L(x, y) & = & c - \mu X^{-1} e - A^T y & = & 0, \\ \nabla_y L(x, y) & = & -Ax + b & = & 0, \\ & & x > 0 & & \end{array}$$

where

$$X^{-1} = \mathsf{diag}(x_1^{-1}, \dots, x_n^{-1})) = \begin{bmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & & \\ \vdots & & \ddots & 0 \\ 0 & & & x_n \end{bmatrix}^{-1}$$

and

$$e = (1, \dots, 1)^T.$$

Now define

$$s := \mu X^{-1} e.$$

then the optimality conditions can be written as

$$\begin{aligned} c - A^T y - s &= 0, \\ -Ax + b &= 0, \\ s - \mu X^{-1} &= 0, \end{aligned}$$

or equivalently

$$\begin{aligned} c - A^T y - s &= 0, \\ -Ax + b &= 0, \\ Xs &= \mu e. \end{aligned} \tag{3}$$

Observe

$$Xs = \mu e$$

is equivalent to

$$x_j s_j = \mu.$$

The optimality conditions (3) says:

- Dual feasibility.
- Primal feasibility.
- Perturbed complementarity.

Observe that

$$\begin{aligned}
c^T x - b^T y &= c^T x - (Ax)^T y \\
&= (c - A^T y)^T x \\
&= s^T x \\
&= e^T X s \\
&= \mu e^T e \\
&= \mu n.
\end{aligned}$$

- Conclusion: Find a solution to the barrier problem (2) for μ sufficiently small using Newton's method.
- The barrier term gets rid of the inequalities!

Primal problem:

$$\begin{array}{lccc} \min & c^T x & & \\ \text{subject to} & Ax & = & b, \\ & x \in \mathcal{K}_q. & & \end{array} \tag{4}$$

- We assume ONE cone for simplicity.

The barrier

$$
\begin{aligned}
B(x) &= \frac{1}{2}\left(\ln\left(x_1 - \frac{\|x_{2:n}\|^2}{x_1}\right) + \ln(x_1)\right) \\
&= \frac{1}{2}\ln\left(x_1^2 - \|x_{2:n}\|^2\right).
\end{aligned}
$$

Primal barrier problem:

$$
\begin{array}{lll}
\text{minimize} & c^T x - \mu B(x) & \\
\text{subject to} & Ax & = b.
\end{array}
$$

- μ is positive parameter.
- The optimum is in the interior for $\mu > 0$.
- Claim: For $\mu \to 0$ the optimum converges to the true optimum.

Optimality conditions:

$$\begin{aligned} Ax &= b, \\ c_1 - \frac{\mu}{x_1^2 - \|x_{2:n}\|^2} x_1 - a_{:1}^T y &= 0, \\ c_{2:n} + \frac{\mu}{x_1^2 - \|x_{2:n}\|^2} x_{2:n} - A_{:(2:n)}^T y &= 0. \end{aligned}$$

Define

$$\begin{aligned} s_1 - \frac{\mu}{x_1^2 - \|x_{2:n}\|^2} x_1 &= 0, \\ s_2 + \frac{\mu}{x_1^2 - \|x_{2:n}\|^2} x_{2:n} &= 0 \end{aligned}$$

and

$$\mathsf{arrow}(x) := \begin{bmatrix} x_1 & x_{2:n} & \cdots & x_n \\ x_2 & x_1 & & \\ \vdots & & \ddots & \\ x_n & & & x_1 \end{bmatrix}.$$

Therefore,

$$\begin{aligned} s_1 &= \frac{\mu}{x_1^2 - \|x_{2:n}\|^2} x_1, \\ s_2 &= -\frac{\mu}{x_1^2 - \|x_{2:n}\|^2} x_{2:n}. \end{aligned}$$

and so it can be verified

$$\text{arrow}(x)s = \mu e_1.$$

If $x \in \text{int}(k_q)$ then $\text{arrow}(x)$ s nonsingular.
Therefore, alternatively we may write

$$\begin{aligned} Ax &= b, \\ A^T y + s &= 0, \\ \text{arrow}(x)s &= \mu e_1 \end{aligned}$$

Observation

- Optimality conditions perturbed by μ.
- Just like in the linear case.
- Same algorithm as for linear case is applicable.
- Serves as a motivation for the highly efficient primal-dual algorithm.

Section 3

The homogeneous primal-dual algorithm

A homogeneous model

Generalized Goldman-Tucker homogeneous model:

$$
(H) \qquad \begin{array}{rcl}
Ax - b\tau & = & 0, \\
A^T y + s - c\tau & = & 0, \\
-c^T x + b^T y - \kappa & = & 0, \\
(x;\tau) \in \bar{\mathcal{K}}, (s;\kappa) \in \bar{\mathcal{K}}^* & &
\end{array}
$$

where

$$
\bar{\mathcal{K}} := \mathcal{K} \times \mathbb{R}_+ \quad \text{and} \quad \bar{\mathcal{K}}^* := \mathcal{K}^* \times \mathbb{R}_+.
$$

- $\mathcal{K}$ is Cartesian product of k convex cones.
- The homogeneous model always has a solution.
- Partial list of references:
  - Linear case: [5], [4], [8].
  - Nonlinear case: [6].

Investigating the homogeneous model

Lemma

Let $(x^*, \tau^*, y^*, s^*, \kappa^*)$ *be any feasible solution to (H), then*

i)
$$(x^*)^T s^* + \tau^* \kappa^* = 0.$$

ii) *If* $\tau^* > 0$, *then*
$$(x^*, y^*, s^*)/\tau^*$$
is an optimal solution.

iii) *If* $\kappa^* > 0$, *then at least one of the strict inequalities*
$$b^T y^* > 0 \tag{5}$$
and
$$c^T x^* < 0 \tag{6}$$
holds. If the first inequality holds, then (P) *is infeasible. If the second inequality holds, then* (D) *is infeasible.*

Summary:

- Compute a nontrivial solution to (H).
- Provides required information.
- Illposed case:

$$\tau^* = \kappa^* = 0.$$

- Illposed case cannot occur for linear problems.

Only the quadratic cone

$$\mathcal{K}_q := \{x \in \mathbb{R}^n : \; x_1^2 \geq \|x_{2:n}\|^2, \; x_1 \geq 0\}.$$

Notes:

- $\mathbb{R}_+ = \{x \in \mathbb{R} : \; x \geq 0\} = \mathcal{K}_q^1$.
    - Linear cone is a spacial case.
- For simplicity the rotated and semidefinite cones are ignored here.

Alternatively definition

$$\mathcal{K}_q := \{x \in \mathbb{R}^n : \; x^T Q x, \; x_1 \geq 0\}.$$

where

$$Q := \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}.$$

It holds:

$$QQ^T = I.$$

Complementarity

Definition:

$$V := \mathsf{mat}(v) = \left[\begin{array}{cc} v_1 & v_{2:n}^T \\ v_{2:n} & v_1 I_{n-1} \end{array} \right].$$

Given $x, s \in K$ then

$$x^T s = 0 \Leftrightarrow X^k S^k e^k = S^k X^k e^k = 0, \quad i = 1, \ldots, k,$$

where

$$X^k := \mathsf{mat}(x^k) \text{ and } S^k := \mathsf{mat}(s^k).$$

Definition:

$$\begin{array}{rcl} X & := & \mathsf{diag}(X^1, \ldots, X^k), \\ S & := & \mathsf{diag}(S^1, \ldots, S^k). \end{array}$$

The central path

Let

$$(x^{(0)}, \tau^{(0)}, y^{(0)}, s^{(0)}, \kappa^{(0)})$$

be given such that

$$(x^{(0)}; \tau^{(0)}), (s^{(0)}; \kappa^{(0)}) \in \mathsf{int}(\bar{\mathcal{K}}).$$

Central path definition:

$$\begin{array}{rcl} Ax - b\tau & = & \gamma(Ax^{(0)} - b\tau^{(0)}), \\ A^T y + s - c\tau & = & \gamma(A^T y^{(0)} + s^{(0)} - c\tau^{(0)}), \\ -c^T x + b^T y - \kappa & = & \gamma(-c^T x^{(0)} + b^T y^{(0)} - \kappa^{(0)}), \\ XSe & = & \gamma\mu^{(0)} e, \\ \tau\kappa & = & \gamma\mu^{(0)}, \end{array} \qquad (7)$$

where $\gamma \in [0, 1]$ and

$$\mu^{(0)} := \frac{(x^{(0)})^T s^{(0)} + \tau^{(0)} \kappa^{(0)}}{r+1} \quad \text{and} \quad e := \begin{bmatrix} e^1 \\ \vdots \\ e^k \end{bmatrix}.$$

Observe:

- For instance choose

$$(x^{(0)}, \tau^{(0)}, y^{(0)}, s^{(0)}, \kappa^{(0)}) = (e, 1, 0, e, 1).$$

- That point is on the central path for $\gamma = 1$.

Neighborhood definition

Nesterov and Todd proves for $(x;\tau),(s,\kappa) \in \mathrm{int}(\mathcal{K})$:

$$(x^T s + \tau\kappa)\left(\left(\sum_{k=1}^{r} \frac{(x^k)^T s^k}{(x^k)^T Q^k x^k (s^k)^T Q^k s^k}\right) + \frac{1}{\tau\kappa}\right) \geq (r+1)^2$$

- If the inequality holds as equality if the point is on the central path.

If $\beta \in (0,1]$ and

$$\begin{aligned} \frac{(x^k)^T Q^k x^k (s^k)^T Q^k s^k}{(x^k)^T s^k} &\geq \beta \frac{x^T s + \tau\kappa}{r+1}, \quad \forall k \\ \tau\kappa &\geq \beta \frac{x^T s + \tau\kappa}{r+1} \end{aligned}$$

then

$$(x^T s + \tau\kappa)\left(\left(\sum_{k=1}^{r} \frac{(x^k)^T s^k}{(x^k)^T Q^k x^k (s^k)^T Q^k s^k}\right) + \frac{1}{\tau\kappa}\right) \leq \frac{1}{\beta}(r+1)^2.$$

Central path neighborhood ($\mathcal{N}(\beta)$):

$$\min \begin{pmatrix} \frac{(x^1)^T Q^1 x^1 (s^1)^T Q^1 s^1}{(x^1)^T s^1} \\ \vdots \\ \frac{(x^k)^T Q^k x^k (s^k)^T Q^k s^k}{(x^k)^T s^k} \\ \tau\kappa \end{pmatrix} \geq \beta\mu$$

and

$$\mu := \frac{x^T s + \tau\kappa}{r + 1}$$

where $\beta \in [0, 1]$.

Algorithm outline

- Follow the central path to the optimum.
  - I.e. stay in the neighborhood of the central path.
- Use Newton's method to compute points in the neighborhood.

The (unscaled) Newton direction

$$
\begin{array}{rcl}
Ad_x - bd_\tau & = & \eta(Ax^{(0)} - b\tau^{(0)}), \\
A^T d_y + d_s - cd_\tau & = & \eta(A^T y^{(0)} + s^{(0)} - c\tau^{(0)}), \\
-c^T d_x + b^T d_y - d_\kappa & = & \eta(-c^T x^{(0)} + b^T y^{(0)} - \kappa), \\
X^{(0)} d_s + S^{(0)} d_x & = & -X^{(0)} S^{(0)} e + \gamma\mu^{(0)} e, \\
\tau^{(0)} d_\kappa + \kappa^{(0)} d_\tau & = & -\tau^{(0)}\kappa^{(0)} + \gamma\mu^{(0)}.
\end{array}
$$

where $\eta = \gamma - 1$.

Problems:

- The search direction is not well-defined everywhere.
- Hard to prove polynomial convergence.
- Symmetry issue.

Observe

$$X^{(0)}d_s + S^{(0)}d_x = -X^{(0)}S^{(0)}e + \gamma\mu^{(0)}e$$

implies

$$d_s = (X^{(0)})^{-1}(-S^{(0)}d_x - X^{(0)}S^{(0)}e + \gamma\mu^{(0)}e).$$

But

$$(X^{(0)})^{-1}S^{(0)}$$

is not symmetric in general.

Solution to the symmetry problem
Scaling

Definition

$W^k \in \mathbb{R}^{n^k \times n^k}$ is a scaling matrix if it satisfies the conditions

$$\begin{array}{rcl} W^k & \succ & 0, \\ W^k Q^k W^k & = & Q^k. \end{array}$$

A scaled point $\bar{x}, \bar{s}$ is obtained by the transformation

$$\bar{x} := \Theta W x \quad \text{and} \quad \bar{s} := (\Theta W)^{-1} s,$$

where

$$\begin{array}{rcl} W & := & \mathrm{diag}(W^1, \ldots, W^k), \\ \Theta & := & \mathrm{diag}(\theta^1 1_{n^1}; \ldots; \theta^k 1_{n^k}). \end{array}$$

and $\theta^k > 0$.

Lemma

i) $(x^k)^T s^k = (\bar{x}^k)^T \bar{s}^k$.

ii) $\theta_k^2 (x^k)^T Q^k x^k = (\bar{x}^k)^T Q^k \bar{x}^k$.

iii) $\theta_k^{-2} (s^k)^T Q^k s^k = (\bar{s}^k)^T Q^k \bar{s}^k$.

iv) $x \in \mathcal{K} \Leftrightarrow \bar{x} \in \mathcal{K}$ ($x \in$ *int*$(\mathcal{K}) \Leftrightarrow \bar{x} \in$ *int*(K)).

v) *Given a* $\beta \in (0, 1)$ *then*

$$(x, \tau, s, \kappa) \in \mathcal{N}(\beta) \Rightarrow (\bar{x}, \tau, \bar{s}, \kappa) \in \mathcal{N}(\beta).$$

Nesterov-Todd scaling

Comments:

- Many choices for a scaling has been suggested.
- Many of them leads polynomial complexity.
- The most satisfactory one is the Nesterov-Todd scaling which chooses the scaling such that

$$\Theta W x = \bar{x} = \bar{s} = (\Theta W)^{-1} s$$

or equivalently

$$s = W \Theta^2 W x.$$

Scaling and symmetric cones

- In the scaled space the primal and dual points are identical!
- Nesterov-Todd scaling is only available for symmetric cones in general!

How to compute the scaling

Assume that $x^k, s^k \in \mathrm{int}(\mathcal{K}^k)$ then

$$\theta_k^2 = \sqrt{\frac{(s^k)^T Q^k s^k}{(x^k)^T Q^k x^k}}. \tag{8}$$

$$
\begin{aligned}
W^k &= \begin{bmatrix} w_1^k & \left(w_{2:n^k}^k\right)^T \\ w_{2:n^k}^k & I + \dfrac{w_{2:n^k}^k \left(w_{2:n^k}^k\right)^T}{1+w_1^k} \end{bmatrix} \\
&= -Q^k + \frac{(e_1^k + w^k)(e_1^k + w^k)^T}{1+(e_1^k)^T w^k}
\end{aligned} \tag{9}
$$

where

$$
w^k = \frac{\theta_k^{-1} s^k + \theta_k Q^k x^k}{\sqrt{2}\sqrt{(x^k)^T s^k + \sqrt{(x^k)^T Q^k x^k (s^k)^T Q^k s^k}}}. \tag{10}
$$

Furthermore,

$$
(W^k)^2 = -Q^k + 2w^k(w^k)^T. \tag{11}
$$

Important facts for computations

Lemma

$$(\theta_k W^k)^{-2} = \theta_k^{-2} Q^k (W^k)^2 Q^k.$$

Notes:

- W^k can be stored using a n^k dimensional vector.
- Multiplications with W^k and $(W^k)^{-1}$ can be carried out in $O(n^k)$ complexity.
- (W^k) has the simple structure

$$-Q^k + 2w^k(w^k)^T.$$

The Nesterov-Todd search direction

$$
\begin{array}{rcl}
Ad_x - bd_\tau & = & \eta(Ax^{(0)} - b\tau^{(0)}), \\
A^T d_y + d_s - cd_\tau & = & \eta(A^T y^{(0)} + s^{(0)} - c\tau^{(0)}), \\
-c^T d_x + b^T d_y - d_\kappa & = & \eta(-c^T x^{(0)} + b^T y^{(0)} - \kappa), \\
\bar{X}^{(0)}(\Theta W)^{-1} d_s + \bar{S}^{(0)} \Theta W d_x & = & -\bar{X}^{(0)} \bar{S}^{(0)} e + \gamma \mu^{(0)} e, \\
\tau^{(0)} d_\kappa + \kappa^{(0)} d_\tau & = & -\tau^{(0)} \kappa^{(0)} + \gamma \mu^{(0)}.
\end{array}
$$

where $\eta := \gamma - 1$.

New iterate:

$$
\begin{bmatrix} x^{(1)} \\ \tau^{(1)} \\ y^{(1)} \\ s^{(1)} \\ \kappa^{(1)} \end{bmatrix} = \begin{bmatrix} x^{(0)} \\ \tau^{(0)} \\ y^{(0)} \\ s^{(0)} \\ \kappa^{(0)} \end{bmatrix} + \alpha \begin{bmatrix} d_x \\ d_\tau \\ d_y \\ d_s \\ d_\kappa \end{bmatrix}.
$$

Properties of the search direction

Lemma

$$\begin{array}{rcl} Ax^{(1)} - b\tau^{(1)} & = & (1+\alpha\eta)(Ax^{(0)} - b\tau^{(0)}), \\ A^T y^{(1)} + s^{(1)} - c\tau^{(1)} & = & (1+\alpha\eta)(A^T y^{(0)} + s^{(0)} - c\tau^{(0)}), \\ -c^T x^{(1)} + b^T y^{(1)} - \kappa^{(1)} & = & (1+\alpha\eta)(-c^T x^{(0)} + b^T y^{(0)} - \kappa^{(0)}), \\ d_x^T d_s^T + d_\tau d_\kappa & = & 0, \\ (x^{(1)})^T s^{(1)} + \tau^{(1)}\kappa^{(1)} & = & (1+\alpha\eta)((x^{(0)})^T s^{(0)} + \tau^{(0)}\kappa^{(0)}). \end{array}$$

Observations:

- The complementarity gap is reduced by a factor of $(1+\alpha\eta) \in [0,1)$.
- The infeasibility is reduced by the same factor.
- Highly advantageous property.
- Implies convergence.

Practical issues

- Step-size computation
  - Back-tracking line search type.
  - Computational cheap.
- Mehrotra predictor-corrector extension.
  - Estimate γ.
  - High-order correction.

Practical stopping criteria

- A solution
$$(x, y, s) = (x^{(k)}, y^{(k)}, s^{(k)})/\tau^{(k)}$$
is said to be primal-dual optimal solution if
$$\begin{aligned} \left\| Ax^{(k)} - b\tau^{(k)} \right\|_\infty &\leq \varepsilon_p (1 + \|b\|_\infty)\tau^{(k)}, \\ \left\| A^T y^{(k)} + s^{(k)} - c\tau^{(k)} \right\|_\infty &\leq \varepsilon_d (1 + \|c\|_\infty)\tau^{(k)}, \\ \frac{|c^T x^{(k)} - b^T y^{(k)}|}{\tau^{(k)} + \max(|c^T x^{(k)}|, |b^T y^{(k)}|)} &\leq \varepsilon_g \end{aligned}$$
where $\varepsilon_p, \varepsilon_d$ and ε_g all are small user specified constants.

- If

$$b^T y^{(k)} > 0 \text{ and } b^T y^{(k)} \varepsilon_p \geq \frac{\|b\|_\infty \left\| A^T y^{(k)} + s^{(k)} \right\|_\infty}{\max(1, \|c\|_\infty, |a_{ij}|)}$$

  the problem is denoted to be primal infeasible and the certificate is $(y^{(k)}, s^{(k)})$ is reported.
- If

$$-c^T x^{(k)} > 0 \text{ and } -c^T x^{(k)} \varepsilon_d \geq \frac{\|c\|_\infty \left\| A x^{(k)} \right\|_\infty}{\max(1, \|b\|_\infty, |a_{ij}|)}$$

  is said denoted to be dual infeasible and the certificate is $x^{(k)}$ is reported.

Observations about the stopping criterion

- Only an approximate solution is computed. (We work in finite precision anyway.)
- Stopping criterion is not god given but observed to work well in practice.
- Primal accuracy is proportional to $\|b\|_\infty$.
- Dual accuracy is proportional to $\|c\|_\infty$.
- Do and don'ts.
  - Scale the problem nicely.
  - Do not add large bounds.
  - Do not use large penalties in the objective.

Computation of the search direction

The computational most expensive operation in the algorithm is the search direction computation:

$$
\begin{array}{rcl}
Ad_x - bd_\tau & = & f^1, \\
A^T d_y + d_s - cd_\tau & = & f^2, \\
-c^T d_x + b^T d_y - d_\kappa & = & f^3, \\
\bar{X}^{(0)}(\Theta W)^{-1} d_s + \bar{S}^{(0)} \Theta W d_x & = & f^4, \\
\tau^{(0)} d_\kappa + \kappa^{(0)} d_\tau & = & f^5
\end{array}
$$

where f^i represents an arbitrary right-hand side.
This implies

$$
\begin{array}{rcl}
d_s & = & (\bar{X}^{(0)}(\Theta W)^{-1})^{-1}(f^4 - \bar{S}^{(0)} \Theta W d_x) \\
 & = & (\bar{X}^{(0)}(\Theta W)^{-1})^{-1} f^4 - W \Theta^2 W d_x, \\
d_\kappa & = & (\tau^{(0)})^{-1}(f^5 - \kappa^{(0)} d_\tau).
\end{array}
$$

Hence,

$$\begin{aligned} Ad_x - bd_\tau &= f^1, \\ A^T d_y - W\Theta^2 W d_x - cd_\tau &= \hat{f}^2, \\ -c^T d_x + b^T d_y + (\tau^{(0)})^{-1}\kappa^{(0)} d_\tau &= \hat{f}^3, \end{aligned}$$

and

$$d_x = -(W\Theta^2 W)^{-1}(\hat{f}^2 - A^T d_y + cd_\tau).$$

Thus

$$\begin{aligned} A(W\Theta^2 W)^{-1}A^T d_y - (b + A(W\Theta^2 W)^{-1}c)d_\tau &= \hat{f}^1, \\ (b - A(W\Theta^2 W)^{-1}c)^T d_y + (c^T(W\Theta^2 W)^{-1}c + (\tau^{(0)})^{-1}\kappa^{(0)})d_\tau &= \tilde{f}^3. \end{aligned}$$

Given

$$M = A(W\Theta^2 W)A^T = \sum_{k=1}^{r} \theta_k^{-2} A^k (W^k)^{-2} (A^k)^T,$$

and

$$\begin{aligned} Mv^1 &= (b + A(W\Theta^2 W)^{-1} c), \\ Mv^2 &= \hat{f}^1 \end{aligned}$$

we reach the easy solvable linear system

$$\begin{aligned} d_y - v^1 d_\tau &= v^2, \\ (b - A(W\Theta^2 W)^{-1} c)^T d_y + (c^T (W\Theta^2 W)^{-1} c + (\tau^{(0)})^{-1} \kappa^{(0)}) d_\tau &= \tilde{f}^3. \end{aligned}$$

- The hard part is the linear equation systems involving M.
- Observe that:

$$M = A(W\Theta^2 W)^{-1}A^T = \sum_{k=1}^{r} \theta_k^{-2} A^k (W^k)^{-2} (A^k)^T,$$

where

$$\begin{array}{rcl} A^k(W^k)^{-2}(A^k)^T & = & A^k Q^k(-Q^k + 2w^k(w^k)^T)Q^k(A^k)^T \\ & = & -A^k Q^k (A^k)^T \\ & & +2(A^k Q^k w^k)(A^k Q^k w^k)^T, \end{array}$$

- $M = M^T$.
- M is positive definite.
- Use Cholesky factorization $M = LL^T$.

Can we use sparse computations?

- Is M sparse? Yes, if

$$-A^k Q^k (A^k)^T$$

  and

$$(A^k Q^k w^k)(A^k Q^k w^k)^T$$

  is sparse. Likely to be the case if
  - A^k is sparse.
  - A^k contains no dense columns.
  - w^k is not high dimensional.
- M is usually very sparse in the linear case.

Observations

- Fewer rows in A tends to better.
  - Does the primal or the dual has fewest rows.
- Big cones and/or dense columns in A are trouble some.
  - Dense rows is not problematic.
  - It is possible to deal with dense columns and large cones see [2] for details.

Practical implementation

- Employs presolve to reduce problem size.
- Exploit problem structure:
  - Upper bounds on linear variables: $x_j \leq u_j$.
  - Fixed variables: $x_j = u_j$.
- Sparse Cholesky (minimum degree or GP ordering).

Computational results

Section 4

Summary

Recap.

- Demonstrated how the quadratic cone can be linearized.
- Outlined an interior-point based on using a barrier function for the conic constraint.
- Discussed the homogeneous primal-dual algorithm.
  - Considered the state of the art for conic problems.
  - Showed that the algorithm exploits a deep mathematical fact about symmetric cones.
  - Discussed issues affecting the solution quality and time.

References I

[1] A. Ben-Tal and A. Nemirovski.
On polyhedral approximations of the second-order cone.
Math. Oper. Res., 26(2):193–205, 2001.

[2] F. Alizadeh and D. Goldfarb.
Second-order cone programming.
Math. Programming, 95(1):3–51, 2003.

[3] F. Glineur.
Topics in Convex Optimization: Interior-Point Methods, Conic Duality and Approximations.
PhD thesis, Service de Mathematiqué et de Recherche Opérationelle, Faculté Polytechnique de Mons, Rue de Houdain, 9, B-7000 Mons, Belgium, 2001.

[4] A. J. Goldman and A. W. Tucker.
Polyhedral convex cones.
In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and related Systems*, pages 19–40, Princeton, New Jersey, 1956. Princeton University Press.

[5] A. J. Goldman and A. W. Tucker.
Theory of linear programming.
In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and related Systems*, pages 53–97, Princeton, New Jersey, 1956. Princeton University Press.

[6] Yu. Nesterov, M. J. Todd, and Y. Ye.
Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems.
Math. Programming, 84(2):227–267, February 1999.

References III

[7] Juan Pablo Vielma, Shabbir Ahmed, and George L. Nemhauser.
A lifted linear programming branch-and-bound algorithm for mixed integer conic quadratic programs.
Technical report, Georgia Institute of Technology, 2007.

[8] Y. Ye, M. J. Todd, and S. Mizuno.
An $O(\sqrt{n}L)$ - iteration homogeneous and self-dual linear programming algorithm.
Math. Oper. Res., 19:53–67, 1994.