



## **Introduction to conic optimization: Lecture 4**

June 20th 2016 - June 22th 2016

[e.d.andersen@mosek.com](mailto:e.d.andersen@mosek.com)

[www.mosek.com](http://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





- 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].





## 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

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

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

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

Define

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



## 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





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

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

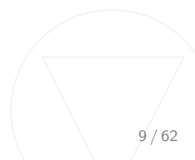
where  $q \approx \log_2(n)$ .

- 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].





- 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





The linear optimization problem:

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

Assumptions:

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



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

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



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

- $\mu$  is a given **positive** parameter.
- Clearly, any feasible solution to (2) is a feasible solution to (1).
- Claim: As  $\mu$  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

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

where



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

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  $\mu$  sufficiently small using Newton's method.
- The barrier term gets rid of the inequalities!





Primal problem:

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

- We assume ONE cone for simplicity.



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

Primal barrier problem:

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

- $\mu$  is positive parameter.
- The optimum is in the interior for  $\mu > 0$ .
- Claim: For  $\mu \rightarrow 0$  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

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



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  $x \in \text{int}(\mathcal{K}_q)$  then  $\text{arrow}(x)$  is nonsingular.

Therefore, alternatively we may write

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





- Optimality conditions perturbed by  $\mu$ .
- 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





Generalized Goldman-Tucker homogeneous model:

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

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].



## 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.





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

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

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

It holds

$$Q Q^T = I.$$



**Definition:**

$$V := \text{mat}(v) = \begin{bmatrix} v_1 & v_{2:n}^T \\ v_{2:n} & v_1 I_{n-1} \end{bmatrix}.$$

Given  $x, s \in K$  then

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

where

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

**Definition:**

$$\begin{aligned} X &:= \text{diag}(X^1, \dots, X^k), \\ S &:= \text{diag}(S^1, \dots, S^k). \end{aligned}$$



Let

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

be given such that

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

Central path definition:

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



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  $\gamma = 1$ .





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

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

- 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 ( $\mathcal{N}(\beta)$ ):

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

and

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

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





- 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.



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

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.





### Definition

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

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

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

$$\begin{aligned} W &:= \text{diag}(W^1, \dots, W^k), \\ \Theta &:= \text{diag}(\theta^1 1_{n^1}; \dots; \theta^k 1_{n^k}). \end{aligned}$$

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)  $x \in \mathcal{K} \Leftrightarrow \bar{x} \in \mathcal{K} \ (x \in \text{int}(\mathcal{K}) \Leftrightarrow \bar{x} \in \text{int}(K)).$
- v) *Given a  $\beta \in (0, 1)$  then*

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







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

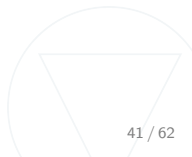
$$\Theta W x = \bar{x} = \bar{s} = (\Theta W)^{-1} s$$

or equivalently

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



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





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

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

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

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,

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



## 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

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



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

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

New iterate:

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



## Lemma

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

## 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.



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





- 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  $\varepsilon_g$  all are small user specified constants.

- If

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

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

- If

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

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





- 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.



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

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

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

This implies

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

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

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

and

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

we reach the easy solvable linear system

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



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

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

where

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

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





- 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.





- 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.



- 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).

## Section 4

### Summary





- 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.



- [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 Mathématique et de Recherche  
Opérationnelle, 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.



- [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.