A primal-dual algorithm for expontial-cone optimization

ICCOPT Berlin, August 8th, 2019

joachim.dahl@mosek.com

www.mosek.com

Conic optimization

Linear cone problem:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \in K, \end{array}$$

with $K = K_1 \times K_2 \times \cdots \times K_p$ a product of proper cones.

Dual:

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & c - A^T y = s \\ & s \in K^*, \end{array}$$

with $K^* = K_1^* \times K_2^* \times \cdots \times K_p^*$.

Conic optimization

MOSEK 9 supports the following symmetric cones,

- linear, quadratic and semidefinite cones

and the nonsymmetric cones,

- three-dimensional power cone for $0 < \alpha < 1$,

$$K_{\text{pow}}^{\alpha} = \{x \in \mathbb{R}^3 \mid x_1^{\alpha} x_2^{(1-\alpha)} \geq |x_3|,\ x_1, x_2 > 0\},$$

- **exponential cone**

$$K_{\text{exp}} = \mathrm{cl}\{x \in \mathbb{R}^3 \mid x_1 \geq x_2 \exp(x_3/x_2),\ x_2 > 0\}.$$

Self-concordant barriers

Self-concordant barrier for K_{exp}:

$$F(x) = -\log(x_2 \log(x_1/x_2) - x_3) - \log x_1 - \log x_2.$$

Conjugate barrier:

$$F_*(s) = \max\{-\langle x, s\rangle - F(x) \; : \; x \in \textbf{int}(K)\}.$$

Standard properties:

$$F^{(k)}(\tau x) = \frac{1}{\tau^k} F^{(k)}(x) \qquad F^{(k)}(x)[x] = -kF^{(k-1)}(x)$$

$$-F'(x) \in \textbf{int}(K_*) \qquad -F'_*(s) \in \textbf{int}(K)$$

$$F'(-F'_*(s)) = -s \qquad F''(-F'_*(s)) = [F''_*(s)]^{-1}$$

Central path for conic problem

Central path for homogenous model parametrized by μ:

$$\begin{aligned} Ax_\mu - b\tau_\mu &= \mu(Ax - b\tau) \\ s_\mu + A^T y_\mu - c\tau_\mu &= \mu(s + A^T y - c\tau) \\ c^T x_\mu - b^T y_\mu + \kappa_\mu &= \mu(c^T x - b^T y + \kappa) \\ s_\mu = -\mu F'(x_\mu), \quad x_\mu &= -\mu F'_*(s_\mu), \quad \kappa_\mu \tau_\mu = \mu, \end{aligned}$$

or equivalently

$$\begin{bmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{bmatrix} \begin{bmatrix} y_\mu \\ x_\mu \\ \tau_\mu \end{bmatrix} - \begin{bmatrix} 0 \\ s_\mu \\ \kappa_\mu \end{bmatrix} = \mu \begin{bmatrix} r_p \\ r_d \\ r_g \end{bmatrix}$$

$$s_\mu = -\mu F'(x_\mu), \quad x_\mu = -\mu F'_*(s_\mu), \quad \kappa_\mu \tau_\mu = \mu,$$

$$r_p := Ax - b\tau, \quad r_d := c\tau - A^T y - s, \quad r_g := \kappa - c^T x + b^T y, \quad r_c := x^T s + \tau\kappa.$$

Scaling for nonsymmetric cones

Following Tunçel [5] we consider a scaling $W^T W \succ 0$,

$$v = Wx = W^{-T} s, \qquad \tilde{v} = W\tilde{x} = W^{-T}\tilde{s}$$

where $\tilde{x} := -F'_*(s)$ and $\tilde{s} := -F'(x)$. The centrality conditions

$$x = \mu\tilde{x}, \quad s = \mu\tilde{s}$$

can then be written symmetrically as

$$v = \mu\tilde{v},$$

and we linearize the centrality condition $v = \mu\tilde{v}$ as

$$W\Delta x + W^{-T}\Delta s = \mu\tilde{v} - v.$$

An affine search-direction

$$\begin{bmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{bmatrix} \begin{bmatrix} \Delta y_a \\ \Delta x_a \\ \Delta \tau_a \end{bmatrix} - \begin{bmatrix} 0 \\ \Delta s_a \\ \Delta \kappa_a \end{bmatrix} = - \begin{bmatrix} r_p \\ r_d \\ r_g \end{bmatrix}$$

$$\Delta s_a + W^T W \Delta x_a = -s, \qquad \tau \Delta \kappa_a + \kappa \Delta \tau_a = -\kappa \tau,$$

satisfying

$$(\Delta x_a)^T \Delta s_a + \Delta \tau_a \Delta \kappa_a = 0.$$

Let $\alpha_a \in (0, 1]$ denote largest feasible step in the affine direction.

We estimate a centering parameter as

$$\gamma := (1 - \alpha_a) \min\{(1 - \alpha_a)^2, 1/4\}.$$

A centering search-direction

Let $\mu = (x^T s + \tau\kappa)/(\nu + 1)$.

$$\begin{bmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{bmatrix} \begin{bmatrix} \Delta y_c \\ \Delta x_c \\ \Delta \tau_c \end{bmatrix} - \begin{bmatrix} 0 \\ \Delta s_c \\ \Delta \kappa_c \end{bmatrix} = (\gamma - 1) \begin{bmatrix} r_p \\ r_d \\ r_g \end{bmatrix}$$

$$W \Delta x_c + W^{-T} \Delta s_c = \gamma \mu \tilde{v} - v, \quad \tau \Delta \kappa_c + \kappa \Delta \tau_c = \gamma \mu - \kappa \tau,$$

Constant decrease of residuals and complementarity:

$$\begin{aligned} Ax^+ - b\tau^+ &= (1 - \alpha(1-\gamma)) \cdot r_p, \\ c\tau^+ - A^T y^+ - s^+ &= (1 - \alpha(1-\gamma)) \cdot r_d, \\ b^T y^+ - c^T x^+ - \kappa^+ &= (1 - \alpha(1-\gamma)) \cdot r_g, \\ (x^+)^T s^+ + \tau^+ \kappa^+ &= (1 - \alpha(1-\gamma)) \cdot r_c, \end{aligned}$$

where $z^+ := (z + \alpha \Delta z_c)$.

A higher-order corrector term

Derivatives of $s_\mu = -\mu F'(x_\mu)$:

$$\begin{aligned} \dot{s}_\mu + \mu F''(x_\mu)\dot{x}_\mu &= -F'(x_\mu), \\ \ddot{s}_\mu + \mu F''(x_\mu)\ddot{x}_\mu &= -2F''(x_\mu)\dot{x}_\mu - \mu F'''(x_\mu)[\dot{x}_\mu, \dot{x}_\mu]. \end{aligned}$$

Using $F''(x)x = -F'(x)$ and $F'''(x)[x] = -2F''(x)$ we obtain

$$\ddot{s}_\mu + \mu F''(x_\mu)\ddot{x}_\mu = F'''(x_\mu)[\dot{x}_\mu, (F''(x_\mu))^{-1}\dot{s}_\mu].$$

We interpret $\dot{s}_\mu \approx -\mu\Delta s_\mathrm{a}$ and $\dot{x}_\mu \approx -\mu\Delta x_\mathrm{a}$, *i.e.*,

$$\Delta s_\mathrm{cor} + W^T W \Delta x_\mathrm{cor} = \frac{1}{2}F'''(x)[\Delta x_\mathrm{a}, (F''(x))^{-1}\Delta s_\mathrm{a}],$$

satisfying

$$x^T \Delta s_\mathrm{cor} + s^T \Delta x_\mathrm{cor} = -(\Delta x_\mathrm{a})^T \Delta s_\mathrm{a}.$$

Combined centering-corrector direction

A combined centering-corrector direction:

$$\begin{bmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \\ \Delta \tau \end{bmatrix} - \begin{bmatrix} 0 \\ \Delta s \\ \Delta \kappa \end{bmatrix} = (\gamma - 1) \begin{bmatrix} r_p \\ r_d \\ r_g \end{bmatrix}$$

$$W \Delta x + W^{-T} \Delta s = \gamma \mu \tilde{v} - v + \frac{1}{2} W^{-T} F'''(x)[\Delta x_{\mathrm{a}}, (F''(x))^{-1} \Delta s_{\mathrm{a}}],$$

$$\tau \Delta \kappa + \kappa \Delta \tau = \gamma \mu - \tau \kappa - \Delta \tau_{\mathrm{a}} \Delta \kappa_{\mathrm{a}}.$$

All residuals and complementarity decrease by $(1 - \alpha(1 - \gamma))$.

Computing the scaling matrix

Theorem (Schnabel [4])

Let $S, Y \in \mathbb{R}^{n\times p}$ have full rank p. Then there exists $H \succ 0$ such that $HS = Y$ if and only if $Y^TS \succ 0$.

Let

$$S := \begin{pmatrix} x & \tilde{x} \end{pmatrix}, \quad Y := \begin{pmatrix} s & \tilde{s} \end{pmatrix}$$

both be full rank. As a consequence of Thm. 1 (for $n = 3$),

$$H = Y(Y^TS)^{-1}Y^T + zz^T$$

where $S^Tz = 0$, $z \neq 0$ and

$$\det(Y^TS) = \left((x^Ts)\cdot(\tilde{x}^T\tilde{s}) - \nu^2\right) > 0$$

vanishing towards the central path.

Computing the scaling matrix

Expanding the BFGS update [4]

$$\hat{H} = H_0 + Y(Y^T S)^{-1} Y^T - H_0 S(S^T H_0 S)^{-1} S^T H_0,$$

for $H_0 \succ 0$ gives the scaling by Tunçel [5] and Myklebust [2], *i.e.*,

$$\hat{z}\hat{z}^T = H_0 - H_0 S(S^T H_0 S)^{-1} S^T H_0.$$

We choose $H_0 := \mu F''(x)$.

In other words, $W^T W = \hat{H} \approx \mu F''(x)$ and satisfies

$$W^T W x = s, \qquad W^T W \tilde{x} = \tilde{s}.$$

Tunçel's scaling bounds

Let $\mu := (x^T s)/\nu$ and $\tilde{\mu} := (\tilde{x}^T \tilde{s})/\nu$. Tunçel defines

$$\mathcal{T}_2(\xi, x, s) := \Big\{ H \succ 0 \mid Hx = s,\ H\tilde{x} = \tilde{s},$$
$$\frac{\mu}{\xi(\nu(\mu\tilde{\mu} - 1) + 1)} F''(x) \preceq H \preceq \frac{\xi(\nu(\mu\tilde{\mu} - 1) + 1)}{\mu} F''(\tilde{x}) \Big\}$$

and shows polynomial convergence for a potential reduction method if

$$\inf_{\xi} \mathcal{T}_2(\xi, x, s) \leq \mathcal{O}(1), \quad \forall x \in \mathbf{int}(K), s \in \mathbf{int}(K^*).$$

For symmetric cones $\xi^\star \leq 4/3$.

Bounds for the exponential cone

Given $s \in \mathbf{int}(K^*_{\text{exp}})$ and $\mu > 0$. Let $h := (0, 0, \nu\mu/s_3)$ and

$$x_\alpha := h - \alpha(\mu F'(s) + h).$$

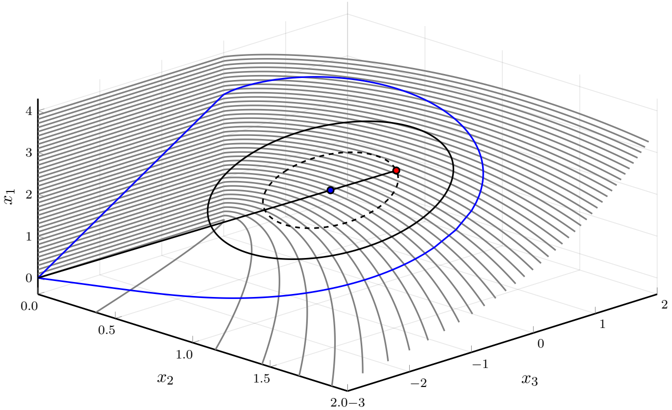
1. $x_\alpha \in K_{\text{exp}}$, $\alpha \in [0, \nu/2]$.
2. $\dfrac{\langle x_\alpha, s\rangle}{\nu} = \mu$.
3. $\mu\langle F'(x_\alpha), F'_*(s)\rangle = \dfrac{\nu - 1}{\alpha} + \dfrac{1}{\nu - (\nu - 1)\alpha}$.
4. $\|x_\alpha\|^2_{-\mu F'_*(s)} = (\alpha^2 - 2\alpha)\nu(\nu - 1) + \nu^2$.

Conjecture (Øbro [3]): For the exponential cone $\xi^\star \approx 1.2532$, *i.e.*,

$$\xi^\star = \left(\frac{2\nu}{\nu - 1} - \frac{2\sqrt{\nu}}{\sqrt{\nu - 1}}\right)^{-1} \left(\frac{(\nu - 1)^{3/2}}{\sqrt{\nu}} + \frac{1}{\nu - \sqrt{\nu(\nu - 1)}} - \nu + 1\right)^{-1}$$

attained for $x_{\alpha^\star}$ with $\alpha^\star = \nu(\nu(\nu - 1))^{-1/2}$.

Øbro's conjecture



Plot of $K_{\text{exp}} \cap \{x : x^T s = \nu\mu\}$, $D(-\mu F'_*(s), 1)$ and $x_{\alpha^\star}$ (red).

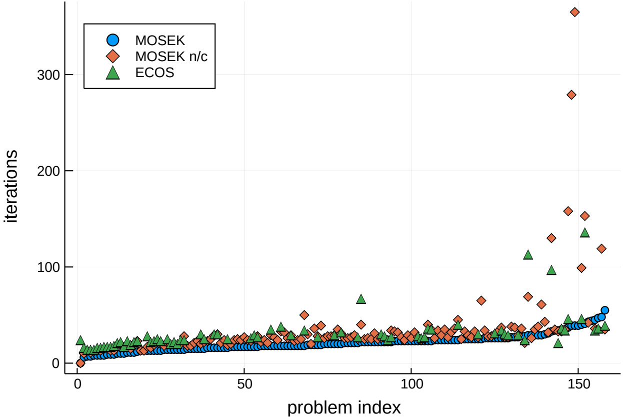
Implications for the exponential-cone

- $F(x)$ does not have negative curvature, *i.e.*,

$$F'''(x)[u] \not\preceq 0, \quad \forall x \in \textbf{int}(K_{\text{exp}}), \forall u \in K_{\text{exp}}.$$

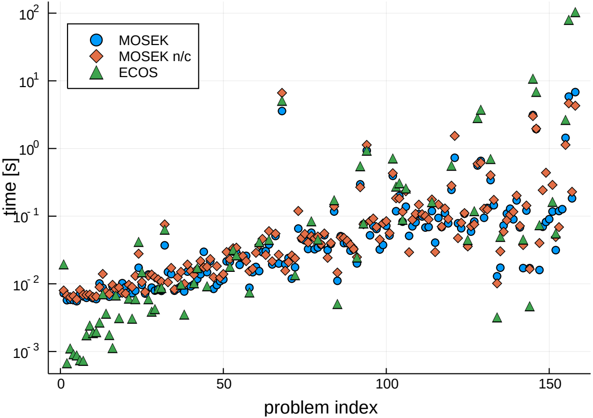
- But F'' is still bounded, for another reason.
- Tunçel's potential-reduction method for expontial-cones have polynomial-time complexity.
- No equivalent proof yet for MOSEK's algorithm, even with optimal scalings.
- The BFGS scaling appears to be bounded as well, and often coincides with the optimal scaling, leaving more to be proved.

Comparing MOSEK and ECOS conic solvers



Iteration counts for different exponential cone problems, comparing MOSEK (with and without proposed corrector) and ECOS.

Comparing MOSEK and ECOS conic solvers



Solution time for different exponential cone problems, comparing MOSEK (with and without proposed corrector) and ECOS.

Conclusions

- Exponential cone optimization included in MOSEK 9.
- Works very well in practice, especially with the proposed corrector.
- Solution-time, accuracy, number of iterations on level with symmetric cone implementation.
- No proof of polynomial-time complexity yet.
- More details can be found in [1].

References

[1] J. Dahl and E. D. Andersen.
A primal-dual interior-point algorithm for nonsymmetric exponential-cone optimization.
Technical report, MOSEK ApS., 2019.

[2] T. Myklebust and L. Tunçel.
Interior-point algorithms for convex optimization based on primal-dual metrics.
Technical report, University of Waterloo, 2014.

[3] M. Øbro.
Conic optimization with exponential cones.
Master's thesis, Technical University of Denmark, 2019.

[4] R. B. Schnabel.
Quasi-newton methods using multiple secant equations.
Technical report, Colorado Univ., Boulder, Dept. Comp. Sci., 1983.

[5] L. Tunçel.
Generalization of primal-dual interior-point methods to convex optimization problems in conic form.
Foundations of Computational Mathematics, 1:229–254, 2001.