Conic Optimization In MOSEK 9: New Cones And Algorithms

INFORMS 2019 Annual Meeting
October 20th, Seattle, WA

Sven Wiese

www.mosek.com

(Mixed-Integer) Conic Optimization

We consider problems of the form

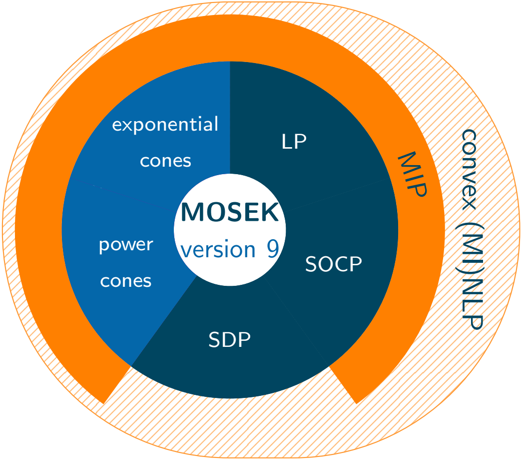
$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \in \mathcal{K} \cap \left(\mathbb{Z}^p \times \mathbb{R}^{n-p}\right), \end{array}$$

where $\mathcal{K}$ is a convex cone.

- Typically, $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_K$ is a product of lower-dimensional cones.
- How can these so-called conic building blocks look like?

What is **MOSEK** ?

The software package **MOSEK** can be employed to solve such problems with the following building blocks:



MOSEK 9.0 released January 2019, 9.1 released October 2019

Symmetric cones - supported by **MOSEK** 8

- the nonnegative orthant

$$\mathbb{R}^n_+ := \{x \in \mathbb{R}^n \mid x_j \geq 0,\ j = 1, \dots, n\},$$

- the quadratic cone

$$\mathcal{Q}^n = \{x \in \mathbb{R}^n \mid x_1 \geq \left(x_2^2 + \cdots + x_n^2\right)^{1/2} = \|x_{2:n}\|_2\},$$

- the rotated quadratic cone

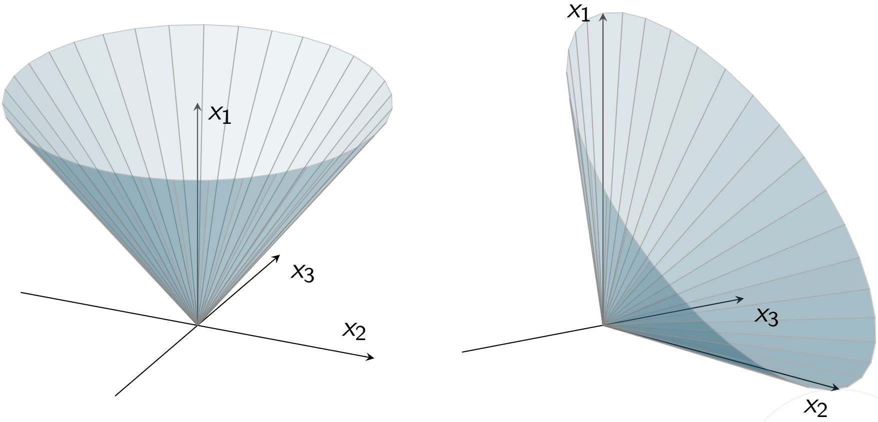
$$\mathcal{Q}^n_r = \{x \in \mathbb{R}^n \mid 2x_1 x_2 \geq x_3^2 + \cdots + x_n^2 = \|x_{3:n}\|_2^2,\ x_1, x_2 \geq 0\}.$$

- the semidefinite matrix cone

$$\mathcal{S}^n = \{x \in \mathbb{R}^{n(n+1)/2} \mid z^T \mathbf{mat}(x) z \geq 0,\ \forall z\},$$

with $\mathbf{mat}(x) := \begin{bmatrix} x_1 & x_2/\sqrt{2} & \dots & x_n/\sqrt{2} \\ x_2/\sqrt{2} & x_{n+1} & \dots & x_{2n-1}/\sqrt{2} \\ \vdots & \vdots & & \vdots \\ x_n/\sqrt{2} & x_{2n-1}/\sqrt{2} & \dots & x_{n(n+1)/2} \end{bmatrix}.$

Quadratic cones in dimension 3



Quadratic-cone use case: unit commitment

We want to schedule the power production of a set of n generators over T periods.

$$
\begin{array}{ll}
\text{minimize} & \sum_i a_i \sum_t p_{i,t}^2 + \sum_i b_i \sum_t p_{i,t} \\
\text{subject to} & \sum_i p_{i,t} \geq d_t, \\
& u_{i,t} p_i^{min} \leq p_{i,t} \leq u_{i,t} p_i^{max}, \\
& u \in U \\
& u_{i,t} \in \{0,1\}.
\end{array}
$$

First introduce $s_i \geq \sum_t p_{i,t}^2$ and rewrite the objective as

$$\sum_i a_i s_i + \sum_i b_i \sum_t p_{i,t}.$$

Now $s_i \geq \sum_t p_{i,t}^2 = \|p_{i,\cdot}\|_2^2 \iff (1/2, s_i, p_{i,\cdot}) \in \mathcal{Q}_r^{n+2}$.

Quadratic-cone use case: convex quadratics

- Every convex (MI)QCP can be reformulated as a (MI)SOCP:

  $$t \geq x^T Q x \text{ with } Q \text{ p.s.d.} \iff t \geq \|Fx\|_2^2 \text{ with } Q = F^T F.$$

  This reformulation can be performed automatically, but F may as well be known explicitly to the modeler.

- In some applications, like least-squares regression, a SOC-formulation is more direct than a QP-formulation.

- The symmetric cones in **MOSEK** are thus enough to tackle LP, MILP, SDP, QCP and MIQCP.

Non-symmetric cones - supported by **MOSEK** 9!

- the three-dimensional exponential cone

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

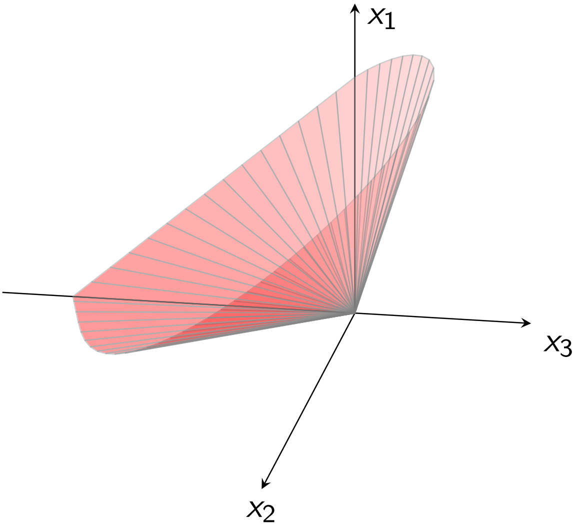
- the three-dimensional power cone

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

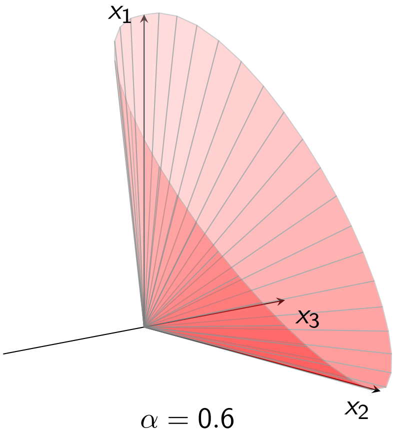
for $0 < \alpha < 1$.

Symmetric cones are homogeneous and self-dual by definition, and the above lack at least one of these properties.

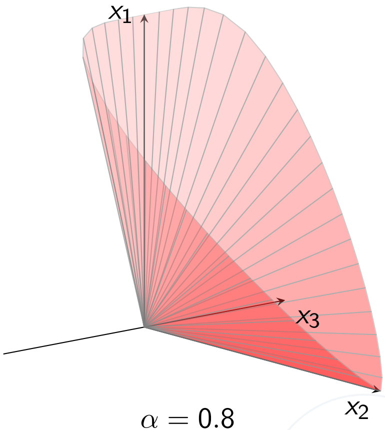
The exponential cone



The power cone



$\alpha = 0.6$



$\alpha = 0.8$

Exponential-cone use case: logistic regression

Given n binary training-points $\{(x_i, y_i)\}$ in $\mathbb{R}^{d+1}$, we want to determine the classifier

$$h_\theta(x) = \frac{1}{1 + \exp(-\theta^T x)}.$$

Training with $2n$ exponential cones:

$$\begin{array}{lll} \text{minimize} & \sum_i t_i \;+\; F \cdot |\{j \mid \theta_j \neq 0\}| & \\ \text{subject to} & t_i \geq \log(1 + \exp(-\theta^T x_i)), & y_i = 1, \\ & t_i \geq \log(1 + \exp(\theta^T x_i)), & y_i = 0. \end{array}$$

We may also consider simultaneous feature selection [10], giving rise to additional d binary variables!

Exponential-cone use case: logistic regression (cont.)

We need to model the so-called softplus function:

$$t \geq \log(1 + e^x) \quad \iff \quad 1 \geq e^{-t} + e^{x-t}$$

$$\iff \quad 1 \geq u + v,\ \ u \geq e^{-t},\ \ v \geq e^{x-t}$$

$$\iff \quad 1 \geq u + v,\ \ (u, 1, -t), (v, 1, x - t) \in \mathcal{K}_{exp}.$$

- Other use cases of the exponential cone arise in Geometric Programming, log-exponential convex risk measuring, power allocation in mobile networks, ...
- Use cases of the power cone arise, e.g., in connection with p-norms.

Non-symmetric cones - why?

- The exponential- and power-cone inequalities are tractable with general convex methods, e.g., the convex interface of **MOSEK** 8.
- Yet, the theoretical foundations for conic interior-point methods are stronger as compared to nonlinear programming.
- Until now, there had simply not been a satisfactory algorithm handling the non-symmetric cones.

A breakthrough!

- Performance and stability are improved, often on level with symmetric-cone implementation.

What about the generality?

- The 5 cones - linear, quadratic, exponential, power and semidefinite- together are highly versatile for modeling.

Continuous Optimization Folklore

"Almost all convex constraints which arise in practice are representable using these cones."

- Lubin et al. [8] show that all convex instances (333) in MINLPLIB2 are conic representable using only 4 types of cones.
- We call modeling with the aforementioned 5 cones **extremely disciplined convex programming**.

Extremely disciplined convex programming in action

mosek

Conic Modeling Cheatsheet

Cones

Quadratic cone $\mathcal{Q}^n$

$$x_1 \geq \sqrt{x_2^2 + \cdots + x_n^2}$$

Rotated quadratic cone $\mathcal{Q}_r^n$

$$2x_1x_2 \geq x_3^2 + \cdots + x_n^2,\ x_1, x_2 \geq 0$$

Power cone $\mathcal{P}_3^{\alpha,1-\alpha}$, $\alpha \in (0,1)$

$$x_1^{\alpha} x_2^{1-\alpha} \geq |x_3|,\ x_1, x_2 \geq 0$$

Exponential cone $K_{\exp}$

$$x_1 \geq x_2 e^{x_3/x_2},\ x_2 \geq 0$$

Simple bounds

$t \geq x^2$	$(0.5, t, x) \in \mathcal{Q}_r^3$
$\lvert t\rvert \leq \sqrt{x}$	$(0.5, x, t) \in \mathcal{Q}_r^3$
$t \geq \lvert x\rvert$	$(t, x) \in \mathcal{Q}^2$
$t \geq 1/x,\ x > 0$	$(x, t, \sqrt{2}) \in \mathcal{Q}_r^3$
$t \geq \lvert x\rvert^p,\ p > 1$	$(t, 1, x) \in \mathcal{P}_3^{1/p,1-1/p}$
$t \geq 1/x^p,\ x > 0,\ p > 0$	$(t, x, 1) \in \mathcal{P}_3^{1/(1+p),p/(1+p)}$
$\lvert t\rvert \leq x^p,\ x > 0,\ p \in (0,1)$	$(x, 1, t) \in \mathcal{P}_3^{p,1-p}$
$t \geq \lvert x\rvert^p/y^{p-1},\ y \geq 0$ $p > 1$	$(t, y, x) \in \mathcal{P}_3^{1/p,1-1/p}$
$t \geq x^Tx/y,\ y \geq 0$	$(0.5t, y, x) \in \mathcal{Q}_r^{n+2}$
$t \geq e^x$	$(t, 1, x) \in K_{\exp}$
$t \leq \log x$	$(x, 1, t) \in K_{\exp}$
$t \geq 1/\log x,\ x > 1$	$(u, t, \sqrt{2}) \in \mathcal{Q}_r^3$ $(x, 1, u) \in K_{\exp}$
$t \geq a_1^{x_1} \cdots a_n^{x_n},\ a_i > 0$	$(t, 1, \sum x_i \log a_i) \in K_{\exp}$
$t \geq xe^x,\ x \geq 0$	$(t, x, u) \in K_{\exp}$ $(0.5, u, x) \in \mathcal{Q}_r^3$
$t \geq \log(1 + e^x)$	$u + v \leq 1$ $(u, 1, x - t) \in K_{\exp}$ $(v, 1, -t) \in K_{\exp}$
$t \geq \lvert x\rvert^{3/2}$	$(t, 1, x) \in \mathcal{P}_3^{2/3,1/3}$
$t \geq x^{3/2},\ x \geq 0$	$(s, t, x), (x, 1/8, s) \in \mathcal{Q}_r^3$
$t \geq 1/x^3,\ x > 0$	$(t, x, 1) \in \mathcal{P}_3^{3/4,1/4}$
$0 \leq t \leq x^{2/5},\ x \geq 0$	$(x, 1, t) \in \mathcal{P}_3^{2/5,3/5},\ t \geq 0$

Means and averaging

Log-sum-exp $t \geq \log(\sum e^{x_i})$	$(z_i, 1, x_i - t) \in K_{\exp}$ $i = 1, \ldots, n$ $\sum z_i \leq 1$
Harmonic mean $0 \leq t \leq n(\sum x_i^{-1})^{-1}$ $x_i > 0$	$(z_i, x_i, t) \in \mathcal{Q}_r^3$ $i = 1, \ldots, n$ $\sum z_i = nt/2$
Geometric mean $\lvert t\rvert \leq (x_1 \cdots x_n)^{1/n}$ $x_i > 0$	$(z_i, x_i, z_{i+1}) \in \mathcal{P}_3^{1-1/i,1/i}$ $i = 2, \ldots, n$ $z_2 = x_1,\ z_{n+1} = t$
$\lvert t\rvert \leq \sqrt{xy},\ x, y > 0$	$(x, y, \sqrt{2}t) \in \mathcal{Q}_r^3$
Weighted geom. mean $\lvert t\rvert \leq x_1^{\alpha_1} \cdots x_n^{\alpha_n},\ x_i > 0$ $\alpha_i > 0, \sum \alpha_i = 1$	$(z_i, x_i, z_{i+1}) \in \mathcal{P}_3^{1-\beta_i,\beta_i}$ $\beta_i = \alpha_i/(\alpha_1 + \cdots + \alpha_i)$ $i = 2, \ldots, n$ $z_2 = x_1,\ z_{n+1} = t$
$\lvert t\rvert \leq x^{1/4}y^{5/12}z^{1/3}$ $x, y, z \geq 0$	$(s, z, t) \in \mathcal{P}_3^{2/3,1/3}$ $(x, y, s) \in \mathcal{P}_3^{3/8,5/8}$

Entropy

$t \leq -x \log x$	$(1, x, t) \in K_{\exp}$
$t \geq x \log(x/y)$	$(y, x, -t) \in K_{\exp}$
$t \geq \log(1 + 1/x)$ $x > 0$	$(x + 1, u, \sqrt{2}) \in \mathcal{Q}_r^3$ $(1 - u, 1, -t) \in K_{\exp}$
$t \leq \log(1 - 1/x)$ $x > 1$	$(x, u, \sqrt{2}) \in \mathcal{Q}_r^3$ $(1 - u, 1, t) \in K_{\exp}$
$t \geq x \log(1 + x/y)$ $x, y > 0$	$(y, x + y, u) \in K_{\exp}$ $(x + y, y, v) \in K_{\exp}$ $t + u + v = 0$

Convex quadratic problems

Let $\Sigma \in \mathbb{R}^{n \times n}$, symmetric, p.s.d.
Find $\Sigma = LL^T$, $L \in \mathbb{R}^{n \times k}$ (Cholesky factor).
Then $x^T\Sigma x = \|L^Tx\|_2^2$.

$t \geq \frac{1}{2}x^T\Sigma x$	$(1, t, L^Tx) \in \mathcal{Q}_r^{k+2}$
$t \geq \sqrt{x^T\Sigma x}$	$(t, L^Tx) \in \mathcal{Q}^{k+1}$
$\frac{1}{2}x^T\Sigma x + p^Tx + q \leq 0$	$(1, -p^Tx - q, L^Tx) \in \mathcal{Q}_r^{k+2}$
$\max_x c^Tx - \frac{1}{2}x^T\Sigma x$	$\max c^Tx - r$ $(1, r, L^Tx) \in \mathcal{Q}_r^{k+2}$
$c^Tx + d \geq \|Ax + b\|_2$	$(c^Tx + d, Ax + b) \in \mathcal{Q}^{m+1}$

Norms, $x \in \mathbb{R}^n$

$\|\cdot\|_1,\ t \geq \sum \lvert x_i\rvert$	$(z_i, x_i) \in \mathcal{Q}^2,\ t = \sum z_i$
$\|\cdot\|_2,\ t \geq (\sum x_i^2)^{1/2}$	$(t, x) \in \mathcal{Q}^{n+1}$
$\|\cdot\|_p,\ p > 1$ $t \geq (\sum \lvert x_i\rvert^p)^{1/p}$	$(z_i, t, x_i) \in \mathcal{P}_3^{1/p,1-1/p}$ $i = 1, \ldots, n$ $\sum z_i = t$

Geometry

Bounding ball $\min_x \max_i \|x - x_i\|_2$	$\min r$ $(r, x - x_i) \in \mathcal{Q}^{n+1}$
Geometric median $\min_x \sum \|x - x_i\|_2$	$\min \sum t_i$ $(t_i, x - x_i) \in \mathcal{Q}^{n+1}$
Analytic center $\max_x \sum \log(b_i - a_i^Tx)$	$\max \sum t_i$ $(b_i - a_i^Tx, 1, t_i) \in K_{\exp}$

Regression and fitting

Regularized least squares $\min_w \|Xw - y\|_2^2 + \lambda\|w\|_2^2$	$\min t + \lambda r$ $(0.5, t, Xw - y) \in \mathcal{Q}_r^{m+2}$ $(0.5, r, w) \in \mathcal{Q}_r^{n+2}$
Max likelihood $\max_p p_1^{a_1} \cdots p_n^{a_n}$	$\max \sum a_i t_i$ $(p_i, 1, t_i) \in K_{\exp}$
Logistic cost function $t \geq -\log(1/(1 + e^{-\theta^Tx}))$	$u + v \leq 1$ $(u, 1, -\theta^Tx - t) \in K_{\exp}$ $(v, 1, -t) \in K_{\exp}$

Risk-return

$\Sigma \in \mathbb{R}^{n \times n}$ – covariance, $\Sigma = LL^T$, $L \in \mathbb{R}^{n \times k}$

$\max_x \alpha^Tx$ s.t. $x^T\Sigma x \leq \gamma$	$\max_x \alpha^Tx$ $(\sqrt{\gamma}, L^Tx) \in \mathcal{Q}^{k+1}$
$\max_x \alpha^Tx - \delta x^T\Sigma x$	$\max_x \alpha^Tx - \delta r$ $(0.5, r, L^Tx) \in \mathcal{Q}^{k+2}$
Risk plus $x^{1.5}$ impact cost $t \geq \delta x^T\Sigma x + \beta \sum \lvert x_i\rvert^{3/2}$	$t \geq \delta r + \beta \sum u_i$ $(0.5, r, L^Tx) \in \mathcal{Q}^{k+2}$ $(u_i, 1, x_i) \in \mathcal{P}_3^{2/3,1/3}$
Risk in factor model $\gamma \geq x^T(D + FSF^T)x$ D – specific risk (diag.) $F \in \mathbb{R}^{n \times k}$ – factor loads $S = UU^T$ – factor cov.	$\gamma \geq t + s$ $(0.5, t, \sqrt{D}x) \in \mathcal{Q}_r^{n+2}$ $(0.5, s, U^TF^Tx) \in \mathcal{Q}_r^{k+2}$

The computational beauty of Conic Optimization

In continuous optimization, conic (re-)formulations have been highly advocated for quite some time, e.g., by Nemirovski [9].

- Separation of data and structure:
  - Data: c, A and b.
  - Structure: $\mathcal{K}$.
- Structural convexity.
- Duality (almost...).
- No issues with smoothness and differentiability.

Extending **MOSEK** with non-symmetric cones

MOSEK solves the homogenous model

$$
\begin{aligned}
Ax - b\tau &= 0 \\
c\tau - A^T y - s &= 0 \\
c^T x - b^T y + \kappa &= 0 \\
x \in \mathcal{K},\ s \in \mathcal{K}^*,\ \tau, \kappa &\geq 0.
\end{aligned}
$$

The challenges of its extension to non-symmetric cones include:

- The symmetric cones are equipped with a bilinear product that simplifies the centrality condition of the shifted central path problem. For non-symmetric cones there is no such bilinear product.
- On the symmetric cones, the Nesterov-Todd scaling can be employed, but not on non-symmetric cones.
- Making corrector terms work requires more effort.

Cones in Mixed-Integer Optimization

The exploitation of conic structures in the mixed-integer case is slightly newer, but nonetheless an active research area:

- Outer approximation: Coey et al. [7].
- Lift-and-project cuts: Tanneau and Vielma [11].
- MISOCP:
  - Extended Formulations: Vielma et al. [12].
  - Cutting planes: Andersen and Jensen [1], Kılınç-Karzan and Yıldız [6], Belotti et al. [2], ...
  - Primal heuristics: Çay et al. [3].

Limited structure facilitates the development of various ingredients of modern MINLP-solvers.

Mixed-Integer optimization in **MOSEK**

MOSEK implements conic ($\approx$ nonlinear) branch-and-cut and conic outer-approximation frameworks.

Conic outer approximation is new in **MOSEK** 9!

- For a cone $\mathcal{K} = \{x \mid a^T x \leq 0 \; \forall a \in \mathcal{K}^\circ\}$, any point $a \in \mathcal{K}^\circ$ separates $\hat{x} \notin \mathcal{K}$: $a^T \hat{x} > 0$.

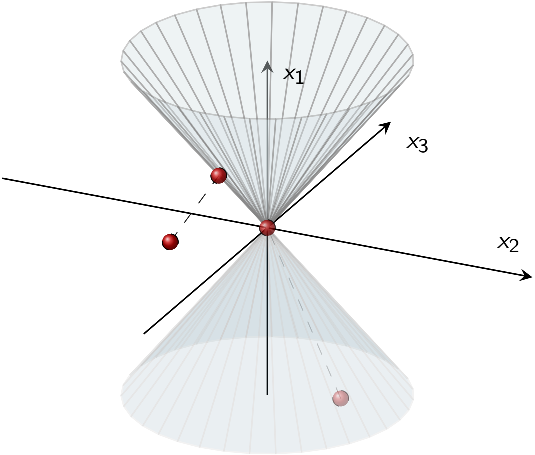
- If $\mathcal{K} = \{x \mid f(x) \leq 0\}$, then $a = \nabla f(\hat{x})$ is a separator [7].

- In **MOSEK** instead, we solve the maximal separation problem

$$\max_{a \in \mathcal{K}^\circ, \|a\|_2 \leq 1} a^T \hat{x}.$$

- This is the dual of the projection problem $\min_{x \in \mathcal{K}} \|x - \hat{x}\|_2$.

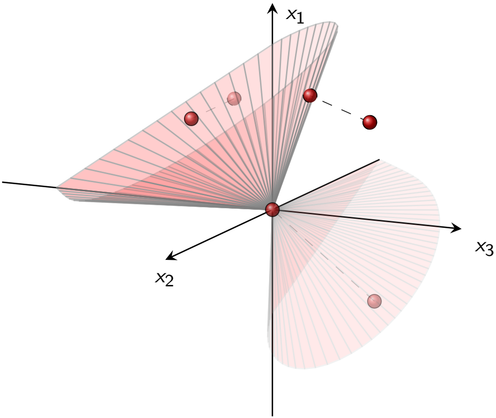
Cone projections



For the symmetric cones, the projection problem can be solved algebraically!

Cone projections (cont.)

For the exponential and power cones, the projection problem is at most a univariate root-finding problem [5, 4].



Wrap-up

- **MOSEK** 9 adds new modeling power, via the exponential and power cones, to the already existing symmetric cones, giving the possibility to tackle most convex (MI)NLP problems.
- Robust numerical algorithms are available for solving these problems in the continuous and mixed-integer case.
- Consider (Mixed-Integer) Conic Optimization as a research area - there are some fruits to pick!

References I

[1] Kent Andersen and Anders Nedergaard Jensen.
Intersection cuts for mixed integer conic quadratic sets.
In Michel Goemans and José Correa, editors, *Integer Programming and Combinatorial Optimization*, pages 37–48, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[2] Pietro Belotti, Julio C. Góez, Imre Pólik, Ted K. Ralphs, and Tamás Terlaky.
On families of quadratic surfaces having fixed intersections with two hyperplanes.
Discrete Appl. Math., 161(16-17):2778–2793, November 2013.

[3] Sertalp Çay, Imre Pólik, and Tamás Terlaky.
The first heuristic specifically for mixed-integer second-order cone optimization.
Technical Report 18T-002, Lehigh University, January 2018.

References II

[4] Henrik Alsing Friberg.
Projection onto the exponential cone: a univariate root-finding problem.
Technical report, Optimization Online, January 2018.

[5] Le Thi Khanh Hien.
Differential properties of euclidean projection onto power cone.
Mathematical Methods of Operations Research, 82(3):265–284, Dec 2015.

[6] F. Kılınç-Karzan and S. Yıldız.
Two-term disjunctions on the second-order cone.
Mathematical Programming, 154(1):463–491, April 2015.

[7] Miles Lubin.
Mixed-integer convex optimization: outer approximation algorithms and modeling power.
PhD thesis, Massachusetts Institute of Technology, 2017.

References III

[8] M. Lubin and E. Yamangil and R. Bent and J. P. Vielma.
Extended Formulations in Mixed-integer Convex Programming.
In Q. Louveaux and M. Skutella, editors, *Integer Programming and Combinatorial Optimization. IPCO 2016. Lecture Notes in Computer Science, Volume 9682*, pages 102–113. Springer, Cham, 2016.

[9] A. Nemirovski.
Advances in convex optimization: Conic programming.
In Marta Sanz-Sol, Javier Soria, Juan L. Varona, and Joan Verdera, editors, *Proceedings of International Congress of Mathematicians, Madrid, August 22-30, 2006, Volume 1*, pages 413–444. EMS - European Mathematical Society Publishing House, April 2007.

[10] Yuichi Takano, Toshiki Sato, Ryuhei Miyashiro, and Akiko Yoshishe.
Feature Subset Selection for Logistic Regression via Mixed Integer Optimization.
Talk at Workshop on Advances in Optimization (WAO2016).
`www.me.titech.ac.jp/~mizu_lab/KAKEN2014/WAO2016/takano.pdf`, 2016.

References IV

[11] Mathieu Tanneau and Juan Pablo Vielma.
Cutting Planes for Mixed-Integer Conic Programming.
Poster session at the 2019 YinzOR Student Conference, 2019.

[12] J. P. Vielma, I. Dunning, J. Huchette, and M. Lubin.
Extended Formulations in Mixed Integer Conic Quadratic Programming.
Mathematical Programming Computation, 9:369–418, 2017.

Further information on **MOSEK**

- Documentation at
  `https://www.mosek.com/documentation/`
  - Manuals for interfaces.
  - Modeling cook book.
  - White papers.
- Tutorials and more at
  `https://github.com/MOSEK/`