Mixed-integer conic optimization and MOSEK

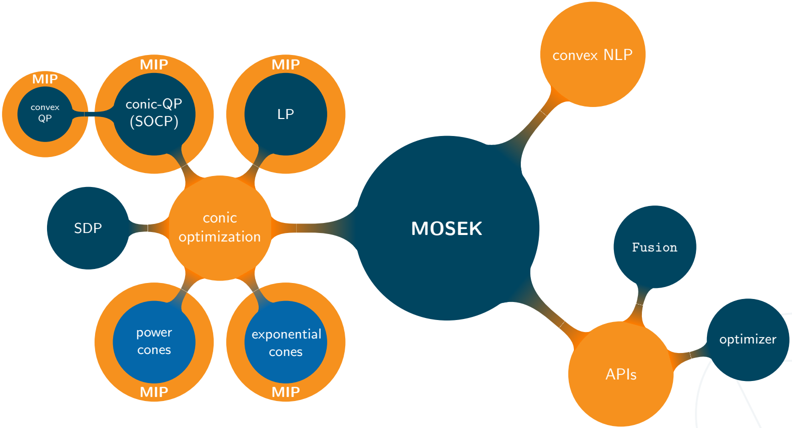
Dagstuhl seminar on MINLP, February 20th 2018

Sven Wiese

www.mosek.com

What is MOSEK ?

- MOSEK ApS is a Danish company founded in 1997.
- Creates software for mathematical optimization problems.



The classical linear optimization problem:

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

Pro:

- Structure is explicit and simple.
- Data is simple: c, A, b.
- Structure implies convexity i.e. data independent.
- Powerful duality theory including Farkas lemma.
- Smoothness, gradients, Hessians are not an issue.

Therefore, we have powerful algorithms and software.

Con:

- It is linear only.

The classical nonlinear optimization problem

The classical nonlinear optimization problem:

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & g(x) & \leq 0. \end{array}$$

Pro

- It is very general.

Con:

- Structure is hidden.
  - How to specify the problem at all in software?
  - How to compute gradients and Hessians if needed?
  - How to exploit structure?
- Convexity checking!
  - Verifying convexity is NP-hard.
  - Solution: Disciplined convex modeling by Grant, Boyd and Ye [1] to assure convexity.

A fundamental question

Is there a class of nonlinear optimization problems that preserve almost all of the good properties of the linear optimization problem?

Conic optimization

Linear cone problem:

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

with $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_K$ a product of proper cones.

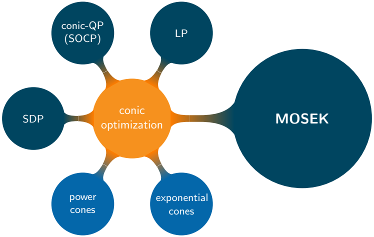
The beauty of conic optimization

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

Lubin et al. [2] show that all convex instances (333) in MINLPLIB2 are conic representable using only 4 types of cones.

Extremely disciplined convex programming

These 4 cones, including symmetric and non-symmetric ones, and extended by another popular cone, are:



Allowing for the nonsymmetric conic formulation leads to **extremely disciplined convex programming**. Simple, yet flexible for modeling, and with efficient numerical algorithms.

Symmetric cones (supported by **MOSEK** 8)

- *the nonnegative orthant*

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

- *the quadratic cone*

$$\mathcal{K}_q^n = \{x \in \mathbb{R}^n \mid x_1 \geq \left(x_2^2 + \cdots + x_n^2\right)^{1/2}\},$$

- *the rotated quadratic cone*

$$\mathcal{K}_r^n = \{x \in \mathbb{R}^n \mid 2x_1x_2 \geq x_3^2 + \dots x_n^2,\ x_1, x_2 \geq 0\}.$$

- *the semidefinite matrix cone*

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

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

Examples of quadratic cones

- Absolute value:

$$|x| \le t \quad \Longleftrightarrow \quad (t, x) \in \mathcal{K}_q^2.$$

- Euclidean norm:

$$\|x\|_2 \le t \quad \Longleftrightarrow \quad (t, x) \in \mathcal{K}_q^{n-1},$$

- Second-order cone inequality:

$$\|Ax + b\|_2 \le c^T x + d \quad \Longleftrightarrow \quad (c^T x + d, Ax + b) \in \mathcal{K}_q^{m+1}.$$

Examples of rotated quadratic cones

- Squared Euclidean norm:

$$\|x\|_2^2 \le t \quad \iff \quad (1/2, t, x) \in \mathcal{K}_r^{n+2}.$$

- Convex quadratic inequality:

$$(1/2)x^T Qx \le c^T x + d \quad \iff \quad (1/2, c^T x + d, F^T x) \in \mathcal{K}_r^{k+2}$$

with $Q = F^T F$, $F \in \mathbb{R}^{n \times k}$.

Examples of rotated quadratic cones

- Convex hyperbolic function:

$$\frac{1}{x} \leq t,\ x > 0 \quad \Longleftrightarrow \quad (x, t, \sqrt{2}) \in \mathcal{K}_r^3.$$

- Convex negative rational power:

$$\frac{1}{x^2} \leq t,\ x > 0 \quad \Longleftrightarrow \quad (t, \frac{1}{2}, s), (x, s, \sqrt{2}) \in \mathcal{K}_r^3.$$

- Square roots:

$$\sqrt{x} \geq t,\ x \geq 0 \quad \Longleftrightarrow \quad (\frac{1}{2}, x, t) \in \mathcal{K}_r^3.$$

- Convex positive rational power:

$$x^{3/2} \leq t,\ x \geq 0 \quad \Longleftrightarrow \quad (s, t, x), (x, 1/8, s) \in \mathcal{K}_r^3.$$

Nonsymmetric cones (in next **MOSEK** release)

- *the three-dimensional power cone*

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

  for $0 < \alpha < 1$.

- *the three-dimensional exponential cone*

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

IPMs for nonsymmetric cones are less studied, and less mature.

Examples of power cones

- Models many quadratic cone examples more succinctly.
- Powers:

$$t \geq |x|^p \quad \iff \quad (t, 1, x) \in \mathcal{K}_p^{1/p}$$

- p-norm cones ($p > 1$):

$$t \geq \|x\|_p \quad \iff \quad \sum r_i = t,\ (r_i, t, x_i) \in \mathcal{K}_p^{1/p},\ i = 1, \ldots, n.$$

Examples of exponential cones

- Expontial:

$$e^x \leq t \quad \Longleftrightarrow \quad (t, 1, x) \in \mathcal{K}_e.$$

- Logarithm:

$$\log x \geq t \quad \Longleftrightarrow \quad (x, 1, t) \in \mathcal{K}_e.$$

- Entropy:

$$-x \log x \geq t \quad \Longleftrightarrow \quad (1, x, t) \in \mathcal{K}_e.$$

- Softplus function:

$$\log(1+e^x) \leq t \quad \Longleftrightarrow \quad (u, 1, x-t), (v, 1, -t) \in \mathcal{K}_e,\ u+v \leq 1.$$

- Log-sum-exp:

$$\log(\sum_i e^{x_i}) \leq t \Longleftrightarrow \sum u_i \leq 1,\ (u_i, 1, x_i-t) \in \mathcal{K}_e,\ i = 1, \ldots, n.$$

The homogeneous model for conic problems

Solution to the homogenous model

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

encapsulates different duality cases:

- If $\tau > 0$, $\kappa = 0$ then $\frac{1}{\tau}(x, y, s)$ is optimal,

$$Ax = b\tau, \quad c\tau - A^T y = s, \quad c^T x - b^T y = 0.$$

- If $\tau = 0$, $\kappa > 0$ then the problem is infeasible,

$$Ax = 0, \quad -A^T y = s, \quad c^T x - b^T y < 0.$$

- If $\tau = 0$, $\kappa = 0$ then the problem is ill-posed.

1

Shifted central-path for cone problems

Let $F(\cdot)$ be a logarithmic barrier for $\mathcal{K}$. Central-path for interior point $(x^0, s^0, y^0, \tau^0, \kappa^0)$:

$$\begin{aligned} Ax_\mu - b\tau_\mu &= \mu(Ax^0 - b\tau^0) \\ s_\mu + A^T y_\mu - c\tau_\mu &= \mu(s^0 + A^T y^0 - c\tau^0) \\ c^T x_\mu - b^T y_\mu + \kappa_\mu &= \mu(c^T x^0 - b^T y^0 + \kappa^0) \\ s_\mu = -\mu F'(x_\mu), \quad x_\mu &= -\mu F'_*(s_\mu), \quad \kappa_\mu \tau_\mu = \mu, \end{aligned}$$

parametrized in μ.

For (our three) symmetric cones, we have a bilinear product $\circ$, and the barrier function satisfies

$$F'(x) = -x^{-1}$$

(using the inverse defined by the product), so the centrality condition becomes

$$x \circ s = \mu e.$$

Non-symmetric cones are more difficult to handle

- For the non-symmetric cones, there is no such bilinear product.

- The three symmetric cones are also *self-scaling*, and there exists a Nesterov-Todd scaling

$$Wx = W^{-1}s = \lambda.$$

  For the non-symmetric cones, this does not exist.

- Higher-order Mehrotra-type correctors are illusive.

A logistic regression example

Given n binary training-points $\{(x_i, y_i)\}$.

Training:

$$\begin{array}{ll} \text{minimize} & \sum_i t_i + \lambda r \\ \text{subject to} & t_i \geq \log(1 + \exp(-\theta^T x_i)), \quad y_i = 1, \\ & t_i \geq \log(1 + \exp(\theta^T x_i)), \quad y_i = 0, \\ & r \geq \|\theta\|_2, \end{array}$$

$2n$ exponential cones + 1 quadratic cone.

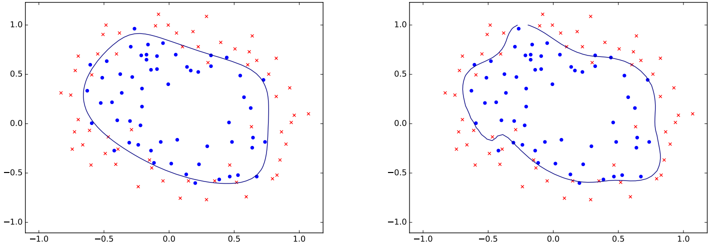
Classifier:

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

A logistic regression example

```python
from mosek.fusion import *

# t >= log( 1 + exp(u) )
def softplus(M, t, u):
    aux = M.variable(2)
    M.constraint(Expr.sum(aux), Domain.lessThan(1.0))
    M.constraint(Expr.hstack(aux, Expr.constTerm(2, 1.0),
                             Expr.vstack(Expr.sub(u,t), Expr.neg(t))),
                 Domain.inPExpCone())

# Model logistic regression (regularized with full 2-norm of theta)
# lambda - regularization parameter
def logisticRegression(X, y, lamb=1.0):
    n, d = X.shape         # num samples, dimension
    M = Model()
    theta = M.variable(d)
    t     = M.variable(n)
    reg   = M.variable()

    M.objective(ObjectiveSense.Minimize, Expr.add(Expr.sum(t), Expr.mul(lamb,reg)))
    M.constraint(Var.vstack(reg, theta), Domain.inQCone())

    for i in range(n):
        dot = Expr.dot(X[i], theta)
        if y[i]==1:
            softplus(M, t.index(i), Expr.neg(dot))
        else:
            softplus(M, t.index(i), dot)

    return M, theta
```

A logistic regression example



Decision regions for different regularizations. Data lifted to the space of degree 6 polynomials.

A logistic regression example

```
Optimizer  - threads                : 20
Optimizer  - solved problem         : the primal
Optimizer  - Constraints            : 236
Optimizer  - Cones                  : 237
Optimizer  - Scalar variables       : 855               conic                  : 737
Optimizer  - Semi-definite variables: 0                 scalarized             : 0
Factor     - setup time             : 0.00              dense det. time        : 0.00
Factor     - ML order time          : 0.00              GP order time          : 0.00
Factor     - nonzeros before factor : 7257              after factor           : 7257
Factor     - dense dim.             : 0                 flops                  : 9.66e+05
ITE PFEAS    DFEAS    GFEAS    PRSTATUS   POBJ              DOBJ              MU       TIME
0   1.6e+00  1.3e+00  9.9e+01  0.00e+00   9.768593109e+01   0.000000000e+00   1.0e+00  0.00
1   8.1e-01  6.6e-01  6.8e+01  6.60e-01   9.011469440e+01   3.552297591e+01   5.4e-01  0.01
2   3.2e-01  2.6e-01  4.3e+01  8.10e-01   7.052557003e+01   4.814005503e+01   2.2e-01  0.01
3   1.6e-01  1.3e-01  3.0e+01  9.51e-01   5.716944320e+01   4.630260918e+01   1.1e-01  0.01
4   7.2e-02  5.9e-02  2.0e+01  9.28e-01   4.754032019e+01   4.248014972e+01   5.2e-02  0.01
5   4.2e-02  3.4e-02  1.5e+01  8.70e-01   4.269747692e+01   3.971483592e+01   3.1e-02  0.01
6   2.5e-02  2.0e-02  1.1e+01  8.15e-01   3.929422749e+01   3.748825666e+01   1.9e-02  0.01
7   1.6e-02  1.3e-02  8.6e+00  7.54e-01   3.712558491e+01   3.593418437e+01   1.2e-02  0.01
8   9.3e-03  7.6e-03  6.4e+00  7.23e-01   3.535772247e+01   3.462356155e+01   7.5e-03  0.02
9   6.4e-03  5.2e-03  5.2e+00  7.14e-01   3.443934016e+01   3.391733535e+01   5.4e-03  0.02
10  5.0e-03  4.1e-03  4.6e+00  7.48e-01   3.396250049e+01   3.355009827e+01   4.3e-03  0.02
11  3.3e-03  2.7e-03  3.6e+00  7.22e-01   3.331083099e+01   3.303323369e+01   2.9e-03  0.02
12  2.7e-03  2.2e-03  3.2e+00  7.28e-01   3.302865568e+01   3.280278682e+01   2.4e-03  0.02
13  2.2e-03  1.8e-03  2.9e+00  7.56e-01   3.282977819e+01   3.264128094e+01   2.0e-03  0.02
14  1.5e-03  1.2e-03  2.3e+00  6.97e-01   3.247818711e+01   3.234470459e+01   1.5e-03  0.02
15  1.1e-03  8.9e-04  1.8e+00  6.52e-01   3.221441130e+01   3.211463097e+01   1.1e-03  0.02
16  9.3e-04  7.6e-04  1.6e+00  6.00e-01   3.210593508e+01   3.201793882e+01   9.4e-04  0.03
17  7.0e-04  5.7e-04  1.3e+00  5.24e-01   3.191120208e+01   3.184089496e+01   7.4e-04  0.03
18  5.3e-04  4.4e-04  1.1e+00  4.64e-01   3.174702006e+01   3.168994262e+01   5.8e-04  0.03
19  3.5e-04  2.9e-04  8.0e-01  4.37e-01   3.153180306e+01   3.149066395e+01   4.0e-04  0.03
20  2.4e-04  1.9e-04  6.1e-01  4.88e-01   3.136835364e+01   3.133901910e+01   2.8e-04  0.03
21  1.5e-04  1.3e-04  4.6e-01  5.95e-01   3.123979806e+01   3.122013885e+01   1.9e-04  0.03
22  8.1e-05  6.6e-05  3.1e-01  6.43e-01   3.110705011e+01   3.109619585e+01   1.0e-04  0.03
23  5.2e-05  4.3e-05  2.4e-01  7.72e-01   3.104953216e+01   3.104241448e+01   6.9e-05  0.04
24  3.3e-05  2.7e-05  1.8e-01  8.40e-01   3.100710801e+01   3.100267855e+01   4.4e-05  0.04
25  1.7e-05  1.4e-05  1.3e-01  8.85e-01   3.097269722e+01   3.097037756e+01   2.4e-05  0.04
26  4.1e-06  3.4e-06  5.8e-02  9.39e-01   3.094182106e+01   3.094128534e+01   6.0e-06  0.04
27  1.1e-06  9.1e-07  3.0e-02  9.84e-01   3.093399109e+01   3.093384551e+01   1.6e-06  0.04
28  8.5e-08  6.9e-08  8.1e-03  9.97e-01   3.093131789e+01   3.093130706e+01   1.3e-07  0.04
29  2.1e-08  1.7e-08  4.0e-03  1.00e+00   3.093115672e+01   3.093115404e+01   3.1e-08  0.04
30  5.7e-09  4.6e-09  2.1e-03  1.00e+00   3.093112004e+01   3.093111935e+01   8.8e-09  0.05
31  1.6e-09  5.4e-10  7.4e-04  1.00e+00   3.093110805e+01   3.093110797e+01   1.1e-09  0.05
```

Mixed-integer conic optimization

- **MOSEK** allows mixed-integer variables in combination with the linear, the conic-quadratic, the exponential and the power cones.
- Applies a branch-and-cut/branch-and-bound framework.
- Preliminary work in case of the non-symmetric cones.
- Tested on mixed-integer exp-cone instances from CBLIB by Miles Lubin.

Mixed-integer exponential-cone instances I

Successfully solved instances

	Time	Obj. value	# nodes
syn40m04h	6.58	-901.75	476
syn40m03h	2.31	-395.15	276
syn40m02h	0.43	-388.77	14
syn40h	0.19	-67.713	16
syn30m04h	3.27	-865.72	450
syn30m03h	1.11	-654.16	165
syn30m02m	1091.4	-399.68	348085
syn30m02h	0.44	-399.68	58
syn30m	9.98	-138.16	7849
syn30h	0.13	-138.16	11
syn20m04m	1833.48	-3532.7	534769
syn20m04h	0.55	-3532.7	27
syn20m03m	300.47	-2647	118089
syn20m03h	0.37	-2647	25
syn20m02m	28.21	-1752.1	14321
syn20m02h	0.19	-1752.1	11
syn20m	0.63	-924.26	645
syn20h	0.09	-924.26	11
syn15m04m	16.59	-4937.5	5567
syn15m04h	0.33	-4937.5	7
syn15m03m	4.77	-3850.2	1907
syn15m03h	0.19	-3850.2	5
syn15m02m	1.24	-2832.7	751
syn15m02h	0.11	-2832.7	5
syn15m	0.12	-853.28	85
syn15h	0.04	-853.28	3
syn10m04m	2.99	-4557.1	1983
syn10m04h	0.16	-4557.1	5

Mixed-integer exponential-cone instances II

Successfully solved instances

syn10m03m	1.13	-3354.7	923
syn10m03h	0.11	-3354.7	5
syn10m02m	0.36	-2310.3	409
syn10m02h	0.08	-2310.3	5
syn10m	0.05	-1267.4	31
syn10h	0	-1267.4	0
syn05m04m	0.17	-5510.4	45
syn05m04h	0.06	-5510.4	3
syn05m03m	0.09	-4027.4	33
syn05m03h	0.04	-4027.4	3
syn05m02m	0.06	-3032.7	23
syn05m02h	0.03	-3032.7	3
syn05m	0.02	-837.73	11
syn05h	0.02	-837.73	5
rsyn0840m04h	39.28	-2564.5	2197
rsyn0840m03h	15.34	-2742.6	1577
rsyn0840m02h	1.56	-734.98	149
rsyn0840h	0.27	-325.55	19
rsyn0830m04h	29.9	-2529.1	2115
rsyn0830m03h	8.3	-1543.1	935
rsyn0830m02h	2.38	-730.51	299
rsyn0830m	227.14	-510.07	99495
rsyn0830h	0.44	-510.07	117
rsyn0820m04h	10.59	-2450.8	635
rsyn0820m03h	18.16	-2028.8	2079
rsyn0820m02h	3.35	-1092.1	510
rsyn0820m	110.08	-1150.3	58607
rsyn0820h	0.46	-1150.3	145
rsyn0815m04h	5.79	-3410.9	587
rsyn0815m03h	7.37	-2827.9	866

Mixed-integer exponential-cone instances III

Successfully solved instances

rsyn0815m02m	2345.68	-1774.4	567030
rsyn0815m02h	2.08	-1774.4	365
rsyn0815m	10.47	-1269.9	7059
rsyn0815h	0.36	-1269.9	238
rsyn0810m04h	6.95	-6581.9	677
rsyn0810m03h	4.95	-2722.4	740
rsyn0810m02m	1353.22	-1741.4	425403
rsyn0810m02h	1.15	-1741.4	159
rsyn0810m	8.31	-1721.4	9041
rsyn0810h	0.21	-1721.4	134
rsyn0805m04m	578.5	-7174.2	66975
rsyn0805m04h	1.92	-7174.2	101
rsyn0805m03m	186.01	-3068.9	37908
rsyn0805m03h	1.61	-3068.9	177
rsyn0805m02m	86.81	-2238.4	34126
rsyn0805m02h	0.87	-2238.4	201
rsyn0805m	3.16	-1296.1	4639
rsyn0805h	0.19	-1296.1	120

Mixed-integer exponential-cone instances

Timed-out instances

	Time	Obj. value	# nodes
gams01	3600.0	22265	70232
rsyn0810m03m	3600.0	-2722.4	493926
rsyn0810m04m	3600.0	-6580.9	307231
rsyn0815m03m	3600.1	-2827.9	420782
rsyn0815m04m	3600.2	-3359.8	309729
rsyn0820m02m	3600.2	-1077.6	683356
rsyn0820m03m	3600.2	-1980.4	380611
rsyn0820m04m	3600.1	-2401.1	262880
rsyn0830m02m	3600.4	-705.46	568113
rsyn0830m03m	3600.2	-1456.3	368794
rsyn0830m04m	3600.1	-2395.7	206456
rsyn0840m	3600.3	-325.55	1157426
rsyn0840m02m	3600.5	-634.17	422224
rsyn0840m03m	3600.1	-2656.5	252651
rsyn0840m04m	3600.0	-2426.3	142895
syn30m03m	3600.2	-654.15	831798
syn30m04m	3600.2	-848.07	643266
syn40m02m	3600.2	-366.77	748603
syn40m03m	3600.3	-355.64	607359
syn40m04m	3600.2	-859.71	371521

Future directions

- Outer-approximation algorithms for the mixed-integer case?
- Other cones?
- How to exploit conic structure in mixed-integer optimization?

Further information

- Docs: `https://www.mosek.com/documentation/`
  - Manuals for interfaces.
  - Modeling cook-book.
  - White papers.

- Examples and tutorials:
  - `https://github.com/MOSEK/Tutorials`

References

[1] M. Grant, S. Boyd, and Y. Ye.
Disciplined convex programming.
In L. Liberti and N. Maculan, editors, *Global Optimization: From Theory to Implementation*, pages 155–210. Springer, 2006.

[2] 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.