The Mixed-integer Conic Optimizer in MOSEK

23rd International Symposium on Mathematical Programming,
July 2nd 2018, Bordeaux

Sven Wiese

www.mosek.com

Mixed-Integer Conic Optimization

We consider problems of the form

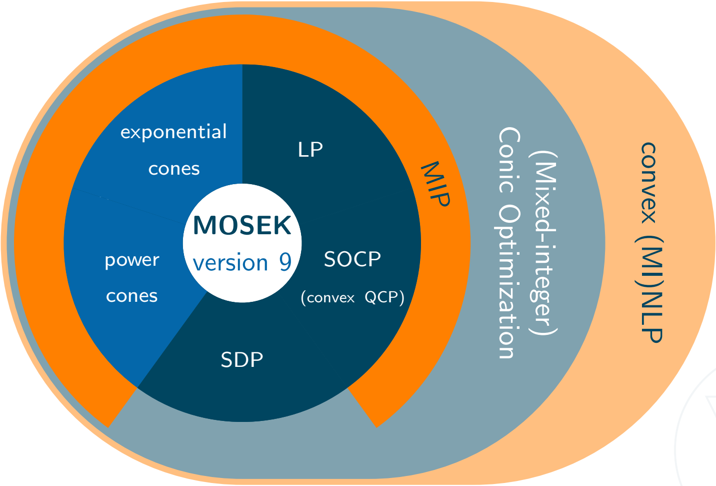
$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \in \mathcal{K} \cap \left(\mathbb{Z}^p \times \mathbb{R}^{n-p}\right), \end{array}$$

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 - so-called conic building blocks.

What is **MOSEK** ?

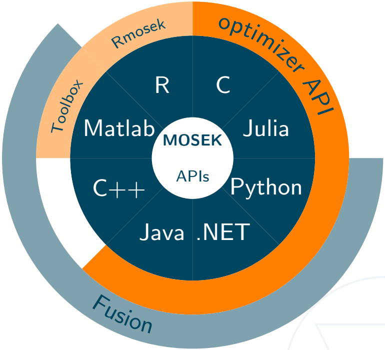
MOSEK is a Copenhagen-based company developing the homonymous software package since 1997.



What is **MOSEK** ? (cont.)

MOSEK at ISMP 2018:

- Henrik A. Friberg,
  Projection and presolve in MOSEK: exponential and power cones,
  Tue, 8:30AM
- Joachim Dahl,
  Extending MOSEK with exponential cones,
  Wed, 8:30AM
- Erling D. Andersen,
  MOSEK version 9,
  Wed, 3:15PM
- Michał Adamaszek,
  Exponential cone in MOSEK: overview and applications,
  Fri, 3:15PM



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

- *the nonnegative orthant*

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

- *the quadratic cone*

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

- *the rotated quadratic cone*

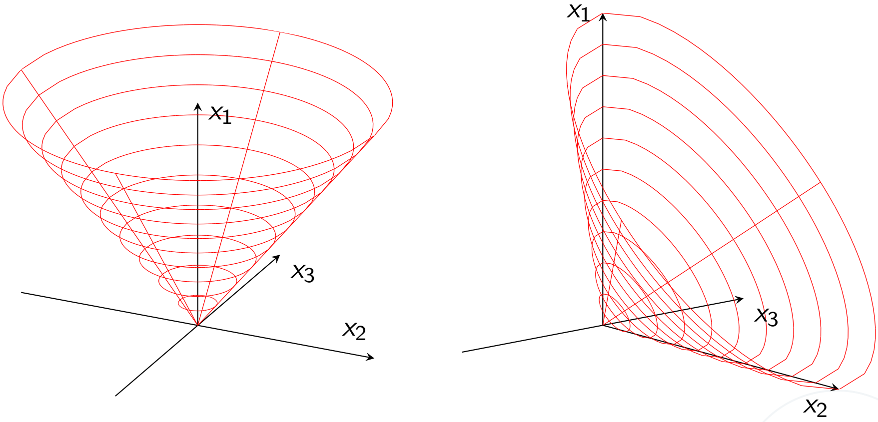
$$\mathcal{Q}^n_r = \{x \in \mathbb{R}^n \mid 2x_1x_2 \geq x_3^2 + \ldots x_n^2,\ x_1, x_2 \geq 0\}.$$

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

Quadratic cones in dimension 3



Examples of quadratic cones

- Absolute value:

$$|x| \leq t \quad \iff \quad (t, x) \in \mathcal{Q}^2.$$

- Euclidean norm:

$$\|x\|_2 \leq t \quad \iff \quad (t, x) \in \mathcal{Q}^{n+1},$$

- Second-order cone inequality:

$$\|Ax + b\|_2 \leq c^T x + d \quad \iff \quad (c^T x + d, Ax + b) \in \mathcal{Q}^{m+1}.$$

Examples of rotated quadratic cones

- Squared Euclidean norm:

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

- Convex quadratic inequality:

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

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

Examples of rotated quadratic cones (cont.)

- Convex hyperbolic function:

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

- Convex negative rational power:

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

- Square roots:

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

- Convex positive rational power:

$$x^{3/2} \le t,\ x \ge 0 \quad \Longleftrightarrow \quad (s, t, x), (x, 1/8, s) \in \mathcal{Q}_r^3.$$

Non-symmetric cones (in next **MOSEK** release)

- *the three-dimensional exponential cone*

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

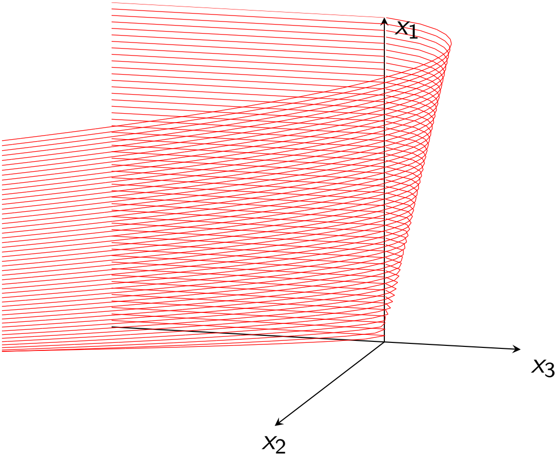
- *the three-dimensional power cone*

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

for $0 < \alpha < 1$.

Interior-point methods for non-symmetric cones are less studied, and less mature.

The exponential cone



Examples of exponential cones

- Expontial:

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

- Logarithm:

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

- Entropy:

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

- Softplus function:

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

- Log-sum-exp:

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

Examples of power cones

The power cone models many quadratic cone examples more succinctly.

- Powers:

$$t \geq |x|^p \quad \Longleftrightarrow \quad (t, 1, x) \in \mathcal{P}^{1/p}$$

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

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

A logistic regression example

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:

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

Some authors consider simultaneous Feature selection [9], giving rise to additional d binary variables!

A logistic regression example (cont.)

```
# t >= log(1 + exp(x))
def softplus(M, t, x):
    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(x,t), Expr.neg(t))),
                 Domain.inPExpCone())

# Model logistic regression
def logisticRegression(X, y, F=1.0, bigM=100):
    n, d = X.shape
    M = Model()
    theta = M.variable(d)
    t     = M.variable(n)
    z     = M.variable(d, Domain.binary())

    # objective
    M.objective(ObjectiveSense.Minimize, Expr.add(Expr.sum(t), Expr.mul(F, Expr.sum(z))))

    for i in range(n): # 2n cone constraints
        dot = Expr.dot(X[i], theta)
        softplus(M, t.index(i), Expr.neg(dot)) if y[i] == 1 else softplus(M, t.index(i), dot)

    for j in range(d): # 2d bigM constraints
        M.constraint(Expr.dot([1.0, bigM], Expr.vstack(theta.index(j), z.index(j))),
                     Domain.greaterThan(0.0))
        M.constraint(Expr.dot([-1.0, bigM], Expr.vstack(theta.index(j), z.index(j))),
                     Domain.greaterThan(0.0))

    return M, theta, z
```

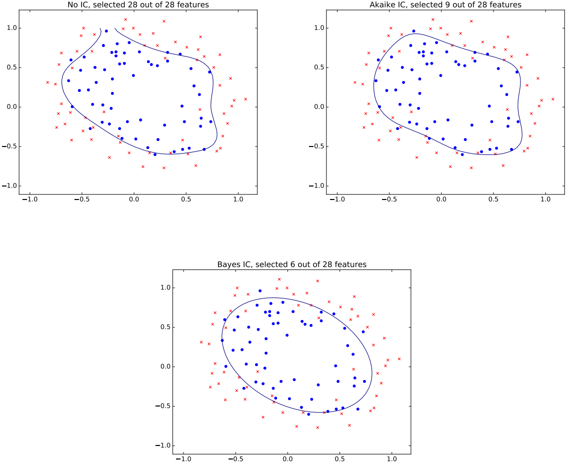
A logistic regression example (cont.)

```
Problem
Objective sense        : min
Type                   : CONIC (conic optimization problem)
Constraints            : 882
Cones                  : 236
Scalar variables       : 1118
Matrix variables       : 0
Integer variables      : 28

Optimizer started.
Mixed integer optimizer started.
Threads used: 20
Presolve started.
Presolve terminated. Time = 0.02
Presolved problem: 764 variables, 292 constraints, 3885 non-zeros
Presolved problem: 0 general integer, 28 binary, 736 continuous
Clique table size: 0
BRANCHES RELAXS   ACT_NDS  DEPTH    BEST_INT_OBJ         BEST_RELAX_OBJ       REL_GAP(%)  TIME
0        1        1        0        1.2123260449e+02     9.8928494362e+01     18.40       0.1
0        1        1        0        1.1848950471e+02     9.8928494362e+01     16.51       0.4
8        12       7        3        1.1848950471e+02     1.0134750080e+02     14.47       0.9
13       17       10       4        1.1669250047e+02     1.0195462270e+02     12.63       1.0
24       28       17       5        1.1669250047e+02     1.0510431665e+02     9.93        1.1
37       41       26       7        1.1669250047e+02     1.0510431665e+02     9.93        1.2
57       61       34       6        1.1669250047e+02     1.0510431665e+02     9.93        1.4
71       75       28       3        1.1669250047e+02     1.0604068619e+02     9.13        1.5
84       88       33       7        1.1606141255e+02     1.0604068619e+02     8.63        1.5
110      109      25       9        1.1606141255e+02     1.0604068619e+02     8.63        1.6
122      121      19       8        1.1589020619e+02     1.0604068619e+02     8.50        1.7
131      130      14       9        1.1428084164e+02     1.0604068619e+02     7.21        1.7
144      137      7        4        1.1370049054e+02     1.0963644001e+02     3.57        1.8
152      144      3        5        1.1174946324e+02     1.1131570072e+02     0.39        1.9
An optimal solution satisfying the relative gap tolerance of 1.00e-02(%) has been located.
The relative gap is 0.00e+00(%).
An optimal solution satisfying the absolute gap tolerance of 0.00e+00 has been located.
The absolute gap is 0.00e+00.

Objective of best integer solution : 1.117494632384e+02
Best objective bound               : 1.117494632384e+02
Construct solution objective       : Not employed
Construct solution # roundings     : 0
User objective cut value           : 0
Number of cuts generated           : 0
Number of branches                 : 155
Number of relaxations solved       : 145
Number of interior point iterations: 2268
Number of simplex iterations       : 0
Time spend presolving the root     : 0.02
Mixed integer optimizer terminated. Time: 1.98
```

A logistic regression example (cont.)



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

The beauty of Conic Optimization

In continuous optimization, conic (re-)formulations have been highly advocated for quite some time, e.g., by Nemirovski [8].

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

We call modeling with the aforementioned 5 cones **extremely disciplined convex programming**: "Almost all convex constraints which arise in practice are representable by using these cones."

Cones in Mixed-Integer Optimization

Lubin et al. [6] show that all convex instances (333) in MINLPLIB2 are conic representable using only 4 types of cones.

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

- MISOCP:
  - Extended Formulations: Vielma et al. [10].
  - Cutting planes: Andersen and Jensen [1], Kılınç-Karzan and Yıldız [4], Belotti et al. [2], ...
  - Primal heuristics: Çay, Pólik and Terlaky [3].
- Duality: Morán, Dey and Vielma [7].
- Outer approximation: Lubin [5].
- ...

Cones in Mixed-Integer Optimization (cont.)

Aspects that can be exploited (computationally) when dealing with (specific) cones include:

- Limited structure facilitates the development of various ingredients of modern MINLP-solvers:
  - Preprocessing.
  - Primal heuristics.
  - Cutting planes.
  - ...
- Continuous relaxations have a rich duality theory.
- Projecting onto cones is usually relatively easy. This comes in handy, e.g., in outer-approximation.

Mixed-integer optimization in **MOSEK**

- **MOSEK** allows mixed-integer variables in combination with the linear, the quadratic, the exponential and the power cones.
- Implements branch-and-bound/cut and outer-approximation frameworks.
- Preliminary work in case of outer-approximation and/or non-symmetric cones.
- Tested on mixed-integer exp-cone instances from CBLIB.

Mixed-integer exponential-cone instances I
Successfully solved instances with branch-and-bound

	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 with branch-and-bound

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 with branch-and-bound

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

Wrap-up

- **MOSEK** version 9 will be extended with new, non-symmetric cones.
- This will allow users to tackle most (if not all) convex MINLP problems.
- We are working hard on increasing the Mixed-integer Conic Optimizer's performance more and more.

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

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

[6] 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.

[7] D. Morán, S. S. Dey, and J. P. Vielma.
Strong dual for conic mixed-integer programs.
SIAM Journal on Optimization, 22:1136–1150, 2012.

References III

[8] 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.

[9] 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.

[10] 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.
- Examples
  - Tutorials at GitHub:
    `https://github.com/MOSEK/Tutorials`