



Exponential cone in MOSEK

ISMP2018, Relative Entropy Optimization, 6 July 2018

Michał Adamaszek, MOSEK ApS

www.mosek.com





- linear conic solver: SOCP, SDP, EXP, POW,
- primal/dual simplex for LPs,
- convex QPs,
- + mixed-integer,
- APIs: MATLAB, C, Python, Java, .NET, R, Julia,
- conic modeling language *Fusion*, C++, Java, .NET, Python,
- third party: AMPL, GAMS, CVX, CVXPY, YALMIP, JuMP
- version 9 (soon).





A conic problem in canonical primal form:

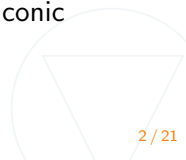
$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{s.t.} & Ax = b \\ & x \in \mathcal{K} \end{array}$$

with dual

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{s.t.} & c - A^T y \in \mathcal{K}^* \end{array}$$

where $\mathcal{K} = K_1 \times \cdots \times K_s$ is a product of cones.

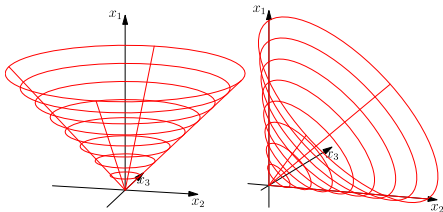
Extremely disciplined convex programming: a problem in conic form is convex by construction.





Nonlinear symmetric cones supported in MOSEK:

- quadratic (SOC) and rotated quadratic:



$$x_1 \geq (x_2^2 + \dots + x_n^2)^{1/2}, \quad 2x_1x_2 \geq x_3^2 + \dots + x_n^2$$

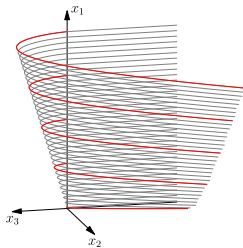
- semidefinite:

$$\mathbb{S}_+^n = \{X \in \mathbb{R}^{n \times n} : X = FF^T\}$$





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



Equivalently

$$-x_3 \geq x_2 \log x_2/x_1 = \text{rel_entr}(x_2, x_1)$$

or the perspective cone (epigraph of the perspective function $(x, y) \rightarrow xf(y/x)$) for either $f(u) = \exp(u)$ or $f(u) = u \log(u)$.



- $t \geq \exp(x) \iff (t, 1, x) \in K_{\text{exp}}$
- $t \leq \log(x) \iff (x, 1, t) \in K_{\text{exp}}$
- $t \geq a_1^{x_1} \cdots a_k^{x_k} \iff (t, 1, \sum x_i \log a_i) \in K_{\text{exp}}, a_i \in \mathbb{R}_+$
- $t \geq x \exp(x)$

$$\begin{array}{ll} t \geq x \exp(y/x) & (t, x, y) \in K_{\text{exp}} \\ y \geq x^2 & (0.5, y, x) \in \mathcal{Q}_r \end{array}$$





What is (SOC,EXP,POW,SDP) — representable? Probably a lot.

From ask.cvxr.com:

Hi guys,

I met a problem while using `cvx`. My objective function is $f(x,y) = x \log(1+x/y)$, $x > 0, y > 0$. The function is convex, but I don't know how to express which in `cvx`. Can you help me?

I believe $x \log(1+x/y)$ is convex for $x > 0, y > 0$.

I don't see how to get it into CVX. Perhaps someone else knows how, although I don't know whether it can be done.

$x \log(1+x/y) = \text{rel_entr}(x+y, y) + \text{rel_entr}(y, x+y)$





- Product of variables in the objective

$$\max(x_1 x_2 \cdots x_n) \iff \max\left(\sum \log x_i\right)$$

Appears in maximum likelihood optimization.

- Log-sum-exp

$$t \geq \log(e^{x_1} + \cdots + e^{x_n})$$

is equivalent to

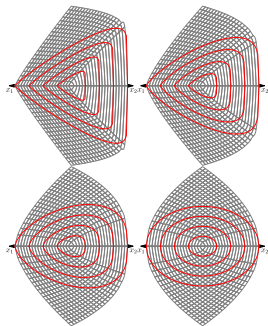
$$e^{x_1-t} + \cdots + e^{x_n-t} \leq 1.$$





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

- generalizes the Lorentz cone ($p = 2$),
- is also a perspective cone (of $f(u) = |u|^p$),
- allows modeling of x^p , $\|x\|_p$, etc.





A geometric program (GP) has the form

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{s.t.} && f_j(x) \leq 1, \quad j = 1, \dots, m \\ & && x_i > 0, \quad i = 1, \dots, n. \end{aligned}$$

where each f is a *posynomial*:

$$f(x) = \sum_j c_k x^{\alpha_k}, \quad c_k > 0, \alpha_k \in \mathbb{R}^n,$$

e.g. $2\sqrt{x} + 0.1x^{-1}z^3 \leq 1$.

For $x_i = \exp(y_i)$ constraints take a convex (conic) form

$$\sum_k c_k \exp(\alpha_k^T y_k) \leq 1.$$

Applications: circuit design, chemical engineering, mechanical engineering, wireless networks, ...



Training data: $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^d \times \{0, 1\}$.

Classify new data using

$$h_{\theta}(x) = \frac{1}{1 + \exp(-\theta^T x)} \sim P[y = 1].$$

Cost function

$$J(\theta) = \sum_i -y_i \log(h_{\theta}(x_i)) - (1 - y_i) \log(1 - h_{\theta}(x_i)).$$

Regularized optimization problem

$$\text{minimize}_{\theta \in \mathbb{R}^d} J(\theta) + \lambda \|\theta\|_2.$$





$$\text{minimize}_{\theta \in \mathbb{R}^d} \sum_i -y_i \log(h_\theta(x_i)) - (1 - y_i) \log(1 - h_\theta(x_i)) + \lambda \|\theta\|_2.$$

Formulate as:

$$\begin{aligned} \text{minimize} \quad & \mathbf{1}^T t_i + \lambda r \\ \text{s.t} \quad & t_i \geq -\log(h_\theta(x)) = \log(1 + \exp(-\theta^T x_i)) \quad \text{if } y_i = 1, \\ & t_i \geq -\log(1 - h_\theta(x)) = \log(1 + \exp(\theta^T x_i)) \quad \text{if } y_i = 0, \\ & r \geq \|\theta\|_2, \end{aligned}$$

Each constraint is conic-representable:

- $r \geq \|\theta\|_2 \iff (r, \theta) \in \mathcal{Q}$
- $t \geq \log(1 + \exp(u)) \iff \exp(-t) + \exp(u - t) \leq 1 \iff$

$$\begin{aligned} y_1 + y_2 &\leq 1, \\ (y_1, 1, u - t) &\in K_{\text{exp}}, \\ (y_2, 1, -t) &\in K_{\text{exp}}. \end{aligned}$$



```
# t >= log( 1 + exp(u) )
def softplus(M, t, u):
    y = M.variable(2)
    # y_1 + y_2 <= 1
    M.constraint(Expr.sum(y), Domain.lessThan(1.0))
    # [ y_1  1  u-t ]
    # [ y_2  1  -t ] in ExpCone
    M.constraint(Expr.hstack(y, Expr.constTerm(2, 1.0),
                             Expr.vstack(Expr.sub(u,t), Expr.neg(t))),
                 Domain.inPExpCone())

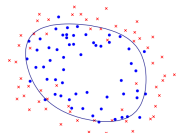
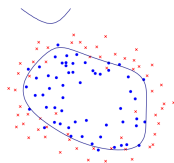
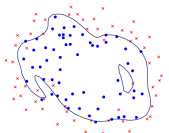
def logisticRegression(X, y, lamb=1.0):
    n, d = X.shape          # num samples, dimension
    M = Model()
    theta = M.variable(d)
    t      = M.variable(n)
    reg    = M.variable()

    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)

    M.solve()
```





Logistic regression with increasing regularization. Every point lifted through 28 degree ≤ 6 monomials.

Remark: logistic regression is a (log-)likelihood maximization problem:

$$J(\theta) = \log \prod_i h_{\theta}(x_i)^{y_i} (1 - h_{\theta}(x_i))^{1-y_i}.$$



Dirk Lorenz <https://regularize.wordpress.com/2018/05/24/>

building-norms-from-increasing-and-convex-functions-the-luxemburg-norm/

$\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ — increasing, convex with $\varphi(0) = 0$. Then the following is a norm on \mathbb{R}^n :

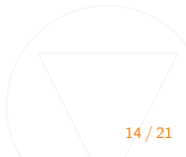
$$\|x\|_\varphi = \inf \left\{ \lambda > 0 : \sum_i \varphi \left(\frac{|x_i|}{\lambda} \right) \leq 1 \right\}.$$

Example: $\varphi(x) = x^p$:

$$\sum_i \left(\frac{|x_i|}{\lambda} \right)^p \leq 1 \iff \lambda \geq \left(\sum_i |x_i|^p \right)^{1/p},$$

so

$$\|x\|_\varphi = \|x\|_p.$$





Observation. The epigraph of the φ -Luxemburg-norm

$$t \geq \|x\|_\varphi$$

is conic representable if the perspective function of φ is.

Proof.

$$\begin{aligned}w_i &\geq |x_i| \\s_i &\geq t\varphi(w_i/t) \\ \sum s_i &= t\end{aligned}$$

add up to

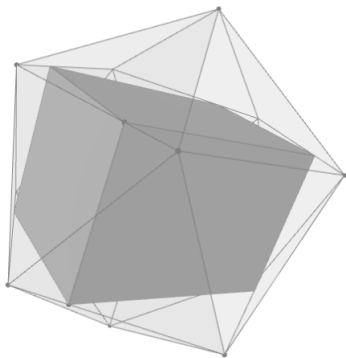
$$1 \geq \sum \varphi(|x_i|/t) \iff t \geq \|x\|_\varphi.$$

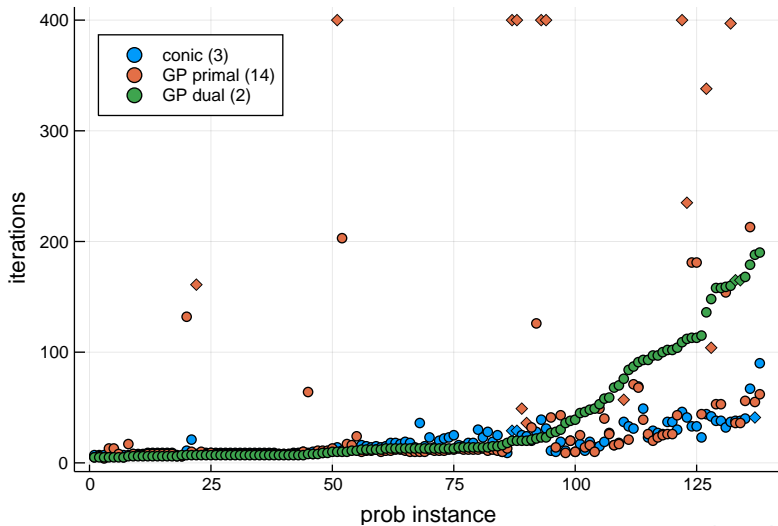
Corollary. We can compute with balls in Luxemburg norms for x^p ,
 $x \cdot \log(1 + x)$, $\exp(x) - 1$.



Find the maximal volume axis-parallel cuboid inscribed in a given convex (conic-representable) set $K \subseteq \mathbb{R}^n$.

$$\begin{aligned} & \text{maximize} && \sum \log d_i \\ & \text{s.t.} && x + \varepsilon \circ d \in K, \text{ for all } \varepsilon \in \{0, 1\}^n \\ & && x, d \in \mathbb{R}^n. \end{aligned}$$







Log-exponential convex risk measure, (Vinel, Krokmal, 2017).

$$\begin{aligned} \text{minimize} \quad & \eta + (1 - \alpha)^{-1} f^{-1} \left(\sum_{j=1}^m p_j f(-r_j^T x - \eta) \right) \\ \text{s.t.} \quad & 1^T x \leq 1 \\ & x^T \sum r_j p_j \geq \bar{r} \\ & x \in \mathbb{R}^n, \eta \in \mathbb{R} \end{aligned}$$

- generalization of CVaR (Rockafellar, Uryasev, 2002),
- f — vanishing on \mathbb{R}_- , $f(0) = 0$, convex on \mathbb{R}_+ . Here:

$$f(u) = \exp([u]_+) - 1.$$

- n — number of assets.
- m — number of historical scenarios $r_1, \dots, r_m \in \mathbb{R}^n$ with probabilities p_1, \dots, p_m .



Easy instances

n	m	8	9
200	100	0.08 (20)	0.05 (22)
200	200	0.17 (21)	0.19 (25)
200	500	0.91 (31)	0.35 (27)
200	1000	4.08 (28)	0.57 (27)
200	2000	3.32 (39)	0.99 (28)
500	100	0.13 (20)	0.11 (23)
500	200	0.28 (20)	0.36 (27)
500	500	1.61 (34)	1.41 (31)
500	1000	5.92 (29)	1.56 (30)
500	2000	25.25 (34)	2.44 (30)
1000	100	0.21 (22)	0.21 (29)
1000	200	0.42 (20)	0.59 (30)
1000	500	3.03 (34)	2.53 (31)
1000	1000	9.43 (31)	6.87 (35)
1000	2000	35.26 (32)	8.66 (32)
1500	100	0.24 (18)	0.20 (23)
1500	200	0.62 (20)	0.82 (31)
1500	500	4.11 (35)	3.99 (33)
1500	1000	16.39 (33)	10.42 (37)
1500	2000	45.67 (31)	12.15 (34)

Numerically harder instances

n	m	8	9
200	100	0.12 (23)	0.06 (29)
200	200	0.42 (67)	0.29 (37)
200	500	1.12 (43)	0.77 (59)
200	1000	6.01 (51)	1.83 (71)
200	2000		3.44 (87)
500	100		0.09 (24)
500	200	0.35 (27)	0.37 (31)
500	500		2.08 (44)
500	1000	8.12 (46)	4.45 (80)
500	2000		5.84 (64)
1000	100	0.31 (38)	0.13 (22)
1000	200	0.51 (27)	0.58 (28)
1000	500	3.66 (43)	3.23 (40)
1000	1000	12.32 (44)	12.83 (66)
1000	2000		16.78 (70)
1500	100	0.31 (24)	0.18 (22)
1500	200	2.08 (83)	0.70 (28)
1500	500		6.04 (51)
1500	1000		11.65 (42)
1500	2000	73.21 (52)	24.77 (67)

time in sec. (intpnt. iterations)



Software:

- CVXPY has a K_{exp} -capable MOSEK interface (Riley Murray).
- Also YALMIP.
- MOSEK Version 9 release this year.

Links:

- WWW www.mosek.com
- Demos github.com/MOSEK/Tutorials
- Blog themosekblog.blogspot.com/
- I found a bug! / MOSEK is too slow! support@mosek.com
- Twitter @mosektw
- Modeling Cookbook www.mosek.com/documentation
- Slides: www.mosek.com/resources/presentations

Reading:

- V.Chandrasekaran, P.Shah, *Relative entropy optimization and its applications*, Math. Program., Ser. A (2017) 161:1-32



Smallest enclosing ball of a random point set in \mathbb{R}^2 in the $(\exp(x) - 1)$ -Luxemburg norm.

