The value of conic optimization for analytics practitioners

December 15, 2020

Erling D. Andersen
MOSEK ApS,
Email: e.d.andersen@mosek.com
Personal WWW: https://erling.andersens.name
Company WWW: https://mosek.com

Outline

Section 1

Introduction

The generic problem

$$\begin{array}{ll} \min & f(x), \\ \text{st} & x \in X. \end{array}$$

- x is the decision variables.
- f is a function.
- X is a set of feasible solutions.
- Think about minimizing cost subject to various constraints.
- Fact: Very hard to solve in general.
- Too generic to be practical useful!!

A linear problem

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{st} \quad & Ax = b, \\ & x \geq 0. \end{aligned}$$

- Explicit structure that is linear.
- Simple data (2 vectors, 1 matrix).
- Powerful algorithms e.g. simplex and interior-point methods.
- Convex (local optima $=$ global optima).
- Easy to verify optimality via duality.
- But cannot deal with nonlinearities!

The question

Is there a comprise problem class that allows some nonlinearities while keeping almost all the good properties of linear optimization?

Section 2

Conic optimization

$$\begin{array}{lll} \text{minimize} & \sum_k (c^k)^T x^k & \\ \text{subject to} & \sum_k A^k x^k & = \quad b, \\ & x^k \in \mathcal{K}^k, & \quad \forall k, \end{array}$$

where

- $c^k \in \mathbb{R}^{n^k}$,
- $A^k \in \mathbb{R}^{m \times n^k}$,
- $b \in \mathbb{R}^m$,
- $\mathcal{K}^k$ are convex cones.

Convex cone definition

$\mathcal{K}^k$ is a nonempty pointed convex cone i.e.

- (Convexity) $\mathcal{K}$ is a convex set.
- (Conic) $x \in \mathcal{K} \Rightarrow \lambda x \in \mathcal{K}, \ \forall \lambda \geq 0.$
- (Pointed) $x \in \mathcal{K}$ and $-x \in \mathcal{K} \Rightarrow x = 0.$

Comments:

- Wikipedia reference: `https://en.wikipedia.org/wiki/Convex_cone`.

$$
\begin{array}{ll}
\text{maximize} & b^T y \\
\text{subject to} & c^k - (A^k)^T y \;\in\; (\mathcal{K}^k)^*, \forall k.
\end{array}
$$

- $(\mathcal{K}^k)^*$ are the corresponding dual cones i.e.

$$(\mathcal{K}^k)^* = \{s \mid s^T x \geq 0, \; \forall x \in \mathcal{K}^k\}.$$

- Equally general.
- The primal and dual problem are both convex!
- The objective sense is not important.

Beauty of conic optimization

- Separation of data and structure:
  - Data: c^k, A^k and b.
  - Structure: $\mathcal{K}$.
- See Nemirovski [5] for more details.
- Very similar to linear optimization.

The 5 cones

Almost all practical convex optimization models can be formulated with **5** convex cone types:

- Linear.
- Quadratic.
- Semidefinite.
- Exponential.
- Power.

Evidence:

- Lubin [4] shows all convex instances(85) in the benchmark library MINLPLIB2 is conic representable using the 5 cones.
- See discussion on the Cvx and Cvxpy forums.

The linear and quadratic cones

The linear cone:

$$\{x \in \mathbb{R} : \; x \geq 0\}.$$

The quadratic cones:

$$\begin{aligned} \mathcal{K}_q &:= \left\{ x \in \mathbb{R}^n : \; x_1 \geq \sqrt{\sum_{j=2}^{n} x_j^2} \right\}, \\ \mathcal{K}_r &:= \left\{ x \in \mathbb{R}^n : \; 2 x_1 x_2 \geq \sum_{j=3}^{n} x_j^2, \; x_1, x_2 \geq 0 \right\}. \end{aligned}$$

Examples:

$$
\begin{aligned}
(t,x)\in\mathcal{K}_q \quad &\Leftrightarrow \quad t\geq \|x\|\,,\\
(1/2,t,V^Tx)\in\mathcal{K}_r \quad &\Leftrightarrow \quad t\geq x^TVV^Tx,\\
\left\{(t,x)\mid t\geq \frac{1}{x},\, x\geq 0\right\} \quad &\Leftrightarrow \quad \left\{(t,x)\mid (x,t,\sqrt{2})\in\mathcal{K}_r^3\right\},\\
\left\{(t,x)\mid t\geq x^{3/2},\, x\geq 0\right\} \quad &\Leftrightarrow \quad \left\{(t,x)\mid (s,t,x),(x,1/8,s)\in\mathcal{K}_r^3\right\},\\
\left\{(t,x)\mid t\geq \frac{1}{x^2},\, x\geq 0\right\} \quad &\Leftrightarrow \quad \left\{(t,x)\mid (t,1/2,s),(x,s,\sqrt{2})\in\mathcal{K}_r^3\right\},
\end{aligned}
$$

The Harmonic mean

$$
t\leq n\left(\sum_{j=1}^{n} x_j^{-1}\right)^{-1}
$$

for $x_j > 0$ has the representation

$$
(z_j,x_j,t)\in\mathcal{K}_r^3 \text{ and } \sum_j z_j = \frac{nt}{2}.
$$

(Proof is left as an exercise!)

Robust optimization

A normal inequality

$$a^T x \leq b$$

and a robust inequality

$$a^T x \leq b, \quad \forall a \in \mathcal{E} := \{z : \; z = \bar{a}^T + Hy, \; \|y\| \leq 1\}.$$

Equivalent formulation

$$\bar{a}^T x + \|Hy\| \leq b$$

or the CQO form

$$\bar{a}^T x + t = b,$$
$$(t, Hy) \in \mathcal{K}_q.$$

The semi-definite cone

The cone of symmetric positive semi-definite matrices:

$$\mathcal{K}_s \quad := \quad \{X \in \mathbb{R}^{n \times n} \mid \quad X = X^T, \quad \lambda_{\min}(X) \geq 0\}$$

Examples:

- Approximation of nonconvex problems e.g graph partitioning.
- Nearest correlation matrix.

Symmetric cones

Facts:

- The first 3 cones belong to the class of symmetric cones.
  - Are **SELF-DUAL**.
  - Are homogeneous.
- Only 5 different symmetric cones.
  - Esoteric: The p.s.d. cone of Hermitian matrices with complex, or quaternion entries.
  - Highly esoteric: 3 by 3 positive semidefinite matrices with octonion entries.
- Closely related to Jordan algebra.

The primal power cone:

$$\mathcal{K}_{pow}(\alpha) := \left\{ (x, z) : \prod_{j=1}^{n} x_j^{|\alpha_j|} \geq \|z\|^{\sum_{j=1}^{n} |\alpha_j|}, \; x \geq 0 \right\}.$$

Examples ($\alpha \in (0,1)$):

$$\begin{aligned} (t, 1, x) \in \mathcal{K}_{pow}(\alpha, 1-\alpha) &\Leftrightarrow t \geq |x|^{1/\alpha}, \; t \geq 0, \\ (x, 1, t) \in \mathcal{K}_{pow}(\alpha, 1-\alpha) &\Leftrightarrow x^{\alpha} \geq |t|, x \geq 0, \\ (x, t) \in \mathcal{K}_{pow}(e) &\Leftrightarrow \underbrace{\left(\prod_{j=1}^{n} x_j \right)^{1/n}}_{\text{geometric mean}} \geq |t|, \; x \geq 0. \end{aligned}$$

An illustrative example

Given the convex function

$$y^{-0.5}$$

then the set

$$x(y/x)^{-0.5} \leq t$$

is convex for $x, y, t \geq 0$. This implies

$$|x|^3 \leq t^2 y$$

or

$$|x| \leq t^{2/3} y^{1/3} = t^{1/3} t^{1/3} y^{1/3}.$$

Note the geometric mean pops up!

Dual power cone

The dual power cone:

$$\mathcal{K}_{pow}(\alpha)^* := \left\{ (x, z) : \prod_{j=1}^{n} \left(\frac{x_j}{\frac{|\alpha_j|}{\sum_{j=1}^{n} |\alpha_j|}} \right)^{|\alpha_j|} \geq \|z\|^{\sum_{j=1}^{n} |\alpha_j|},\ x \geq 0 \right\}.$$

Comment

- Appears naturally when working with a sum of nonnegative circuit polynomials [6].

Facts:

- Is self-dual using a redefined inner-product.
- But is not homogeneous.
- Hence not symmetric.

The exponential cone

The perspective function of e^{x_3} is

$$x_2 e^{\frac{x_3}{x_2}}$$

and next the exponential cone

$$\begin{aligned}\mathcal{K}_{\exp} \quad := \quad & \{(x_1, x_2, x_3) : \; x_1 \geq x_2 e^{\frac{x_3}{x_2}}, \; x_2 > 0\} \\ & \cup \{(x_1, x_2, x_3) : \; x_1 \geq 0, x_2 = 0, \; x_3 \leq 0\}\end{aligned}$$

Applications:

$$\begin{aligned}(t, 1, x) \in \mathcal{K}_{\exp} \quad & \Leftrightarrow \quad t \geq e^x \\ (x, 1, t) \in \mathcal{K}_{\exp} \quad & \Leftrightarrow \quad t \leq \ln(x), \\ (1, x, t) \in \mathcal{K}_{\exp} \quad & \Leftrightarrow \quad t \leq -x \ln(x), \\ (y, x, -t) \in \mathcal{K}_{\exp} \quad & \Leftrightarrow \quad t \geq x \ln(x/y), \text{(relative entropy)}.\end{aligned}$$

Analytic center:

$$\begin{array}{ll} \text{maximize} & \sum_j \ln x_j \\ \text{subject to} & Ax = b, \\ & x \geq 0, \end{array}$$

is conic representable

$$\begin{array}{lll} \text{maximize} & \sum_j t_j & \\ \text{subject to} & Ax = b, & \\ & (x_j, 1, t_j) \in \mathcal{K}_{\exp}, & \forall j. \end{array}$$

The set

$$\log\left(\sum_j e^{x_j}\right) \leq t$$

is equivalent to

$$\begin{aligned} (z_j, 1, x_j - t) &\in \mathcal{K}_{\exp}, \\ \sum_j z_j &\leq 1. \end{aligned}$$

A plot shows

$$\log\left(1+\frac{1}{xy}\right)$$

is convex for $x, y \geq 0$.
Equivalent

$$\begin{array}{rcl} \log\left(e^r+e^s\right) & \leq & t, \\ r & = & 0, \\ e^s & \geq & \dfrac{1}{xy} \end{array}$$

leading to

$$\begin{array}{rcl} \log\left(e^r+e^s\right) & \leq & t, \\ r & = & 0, \\ s+\log x+\log y & \geq & 0 \end{array}$$

which is conic representable.

More examples:

- Geometric programming (Duffin, Kortanek, Peterson, ...)
- Lambert function [2, 1].

Dual exponential cone

$$(\mathcal{K}_{\exp})^* := \{(s_1, s_2, s_3) : \; s_1 \geq (-s_3)e^{-1}e^{\frac{s_3}{s_2}}, \; s_3 < 0\}$$
$$\cup\{(s_1, s_2, s_3) : \; s_1 \geq 0, s_2 \geq 0, \; s_3 = 0\}$$

- Not self-dual.
- Hence not a symmetric cone.

A counter example
Logarithmic mean

The inverse of the logarithmic mean of temperature difference

$$\frac{x - y}{\ln(x/y)}$$

where $x, y \geq 0$ is not convex but the inverse

$$\frac{\ln(x/y)}{x - y}$$

is. Conic representable using the 5 cones? Maybe! See `https://themosekblog.blogspot.com/2019/06/logarithmic-mean-temperature-difference.html`.

Related to the convex function

$$\frac{x}{e^x - 1}.$$

Conic representable using the 5 cones? Maybe!

The primal problem

$$\begin{array}{lll} \text{minimize} & \sum_k (c^k)^T x^k & \\ \text{subject to} & \sum_k A^k x^k = b, & \\ & x^k \in \mathcal{K}^k, & \forall k, \end{array}$$

and

$$\begin{array}{llll} \text{maximize} & b^T y & & \\ \text{subject to} & (A^k)^T y + s^k & = & c^k, \forall k, \\ & s^k & \in & (\mathcal{K}^k)^*, \forall k. \end{array}$$

Weak duality

For any primal dual feasible solution we have weak duality:

$$\begin{aligned} \text{duality gap} &= \sum_k (c^k)^T x^k - b^T y \\ &= \sum_k (x^k)^T s^k \\ &\geq 0. \end{aligned}$$

Strong duality

Under some regularity conditions there is an optimal primal dual solution such that

$$(c^k)^T x^k - b^T y = 0.$$

Some example of failure

$$\begin{array}{ll} \text{minimize} & x^{-1} \\ \text{subject to} & x \geq 0 \end{array}$$

or

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & e^{-x} \leq 0. \end{array}$$

Primal infeasibility certificate

The primal problem is infeasible if (y, s^k) satisfies:

$$\begin{array}{ccl} b^T y & > & 0 \\ (A^k)^T y + s^k & = & 0, \forall k, \\ s^k & \in & (\mathcal{K}^k)^*, \forall k. \end{array}$$

Dual infeasibility certificate

The dual problem is infeasible if (x^k) satisfies:

$$\begin{array}{rcl} \sum(c^k)^T x^k & < & 0, \\ \sum\limits_k Ax^k & = & 0, \forall k, \\ x^k & \in & \mathcal{K}^k, \forall k. \end{array}$$

Section 3

Benchmark results

Conic quadratic optimization

- Hardware: AMD EPYC 7402P 2.80GHz, 24C/48T, 128M Cache (180W) (virum).
- MOSEK: Version 9.2.30.
- Threads: 8 threads is used in test to simulate a typical user environment.
- All timing results t are in wall clock seconds.
- Test problems: Public (e.g `cblib.zib.de`) and customer supplied.

Name	# con.	# cone	# var.	# mat. var.
2D_43_dual	10645	10081	50965	0
2D_43_primal	10645	10081	50965	0
Barcelona_p4	83388	5042	245019	0
a_case118	614	188	3096	0
beam30	113627	30	115817	0
c-nql180	53074	32400	117396	0
c-qssp180	64799	65338	260278	0
c-traffic-36	2740	1331	5401	0
chainsing-1000000	2499995	2499995	8499985	0
frozenpizza_2_cap100_conic_mc	894	670	5795	0
igl-eth-schuller-optProblemSmall	20000	1	76560	0
igl_n	160	122	521	0
multibody-3	12000	1	215012	0
paper-large-scale-co-for-wifi-net-50	10000	101	25151	0
pp-n100000-d10000	1	100000	300001	0
primal_minball_100000	3	100000	300000	0
soccer-trajectory-20160118-ticket11770	55322	82980	359576	0
sssd-strong-30-8	110	24	368	0
swiss_quant_challenge_100_100000_0_primal_quad	100	1	100102	0
tv_employee_200x200_sigma10	79600	40000	159600	0
yuriv2cm	951	421	4832	0
yuriy4	908	421	3496	0
2013_firL2a	0	0	0	0
2013_firLinf	2001	19952	59855	0
2013_wbNRL	1041	9	39417	0
igl-eth-schuller-optProblem2	297493	1	1080539	0

Name	P. obj.	# sig. fig.	# iter	time(s)
2D_43_dual	-8.7829000553e-01	12	22	1.9
2D_43_primal	8.7829000553e-01	12	22	1.9
Barcelona_p4	1.4838294461e+00	7	38	36.8
a_case118	1.2542405625e+02	10	23	0.1
beam30	1.0196771737e+01	10	35	116.3
c-nql180	-9.2772861998e-01	9	21	20.8
c-qssp180	-6.6396108490e+00	10	18	11.7
c-traffic-36	-5.3902456339e+03	9	28	0.7
chainsing-1000000	6.0504878176e+05	9	17	78.0
frozenpizza_2_cap100_conic_mc	1.0381568386e+00	12	24	0.3
igl-eth-schuller-optProblemSmall	7.8481434655e+01	11	30	7.2
igl_n	-7.8587308813e+03	9	20	0.0
multibody-3	-3.0447631860e-01	9	22	9.1
paper-large-scale-co-for-wifi-net-50	-2.7585512927e+00	9	11	11.3
pp-n100000-d10000	2.1957042399e+07	9	19	2.6
primal_minball_100000	4.7675622219e+00	11	23	1.9
soccer-trajectory-20160118-ticket11770	3.6231654500e+01	10	26	8.6
sssd-strong-30-8	3.5848245069e+05	11	20	0.0
swiss_quant_challenge_100_100000_0_primal_quad	4.2175370289e+03	9	12	4.4
tv_employee_200x200_sigma10	1.3368471313e+05	9	22	10.4
yuriv2cm	2.2897899929e-07	9	25	4.1
yuriy4	3.2029655992e-05	7	24	4.0
2013_firL2a	-1.4367664238e-01	15	0	5.2
2013_firLinf	-1.0022267372e-02	13	17	145.0
2013_wbNRL	-3.8821826904e-05	9	14	11.5
igl-eth-schuller-optProblem2	2.2954986269e+01	8	21	89.5

Exponential/power cone optimization

- Threads: 8 threads is used in test to simulate a typical user environment.
- All timing results t are in wall clock seconds.
- Test problems: Public (e.g `cblib.zib.de`) and customer supplied.

Name	# con.	# cone	# var.	# mat. var.
C_Table5_POW_T12	1086456	1328602	3991646	0
WassersteinReg_2img	972	1229312	3687936	0
dump_Prostate_VMAT_308	26872	14649	126320	0
logistic_mip-8-1000-bayes	326	237	1007	0
c-diaz_test_c47	164404	160000	519810	0
x20565-3over2pow	7932	2482	20320	0
z17926	2	75000	225000	0
task_dopt16	1600	26	376	2
entolib_ento2	26	4695	14085	0
entolib_a_56	37	9702	29106	0
exp-ml-scaled-20000	19999	20000	79998	0
exp-ml-20000	19999	20000	79998	0
patil3_conv	418681	413547	1264340	0
c-diaz_test_c47	164404	160000	519810	0
utkarsh_robust_29012019	1228800	819201	2867301	0
varun_conv	333	328	1009	0
z19841	160767	160766	483856	0
z19502	2354679	524286	6546535	0
udomsak	97653	97653	294519	0
relentr25000	1	25000	75000	0
cbf_mra02	3739	3620	11105	0
log-utility-200-5000	10003	5001	25206	0
cbf_cx02-100	5247	5149	15645	0
elmore_delay_16_conv	830	762	2375	0
gp_dave_3_conv	568	373	2052	0
fsparc_6_075_10	942	432	1806	0
c-260209-1	2238	1424	10079	0

Exponential/power cone optimization

Result

Name	P. obj.	# sig. fig.	# iter	time(s)
C_Table5_POW_T12	5.3366628233e-02	8	29	215.8
WassersteinReg_2img	-1.9199612101e+01	6	112	338.2
dump_Prostate_VMAT_308	5.6832289442e+04	7	56	681.0
logistic_mip-8-1000-bayes	6.6082723990e+01	9	24	0.1
c-diaz_test_c47	1.8880304374e-02	9	69	79.1
x20565-3over2pow	2.3137375000e+02	8	44	2.1
z17926	1.7942114183e-01	9	45	7.7
task_dopt16	1.3214504598e+01	9	12	0.6
entolib_ento2	-1.1354764143e+01	9	31	0.3
entolib_a_56	-8.2834835217e+00	9	89	3.6
exp-ml-scaled-20000	-3.3125927467e+00	9	146	11.7
exp-ml-20000	-1.9796344233e+04	10	115	5.2
patil3_conv	-1.0539216058e+00	9	90	134.2
c-diaz_test_c47	1.8880304374e-02	9	69	79.1
utkarsh_robust_29012019	1.6172995191e+00	7	83	717.2
varun_conv	-2.3527295666e+01	9	34	0.1
z19841	-2.6100559671e+00	9	84	294.5
z19502	5.1527245209e+06	10	66	403.8
udomsak	7.6475916495e-02	10	212	704.9
relentr25000	6.3511190583e-02	9	22	1.3
cbf_mra02	4.3179836817e+00	10	160	3.1
log-utility-200-5000	-1.8520357724e+03	9	27	1.6
cbf_cx02-100	7.7292741084e+00	9	16	0.5
elmore_delay_16_conv	4.6571223457e+00	8	27	0.2
gp_dave_3_conv	6.1849202244e+00	9	24	0.1
fsparc_6_075_10	4.6989895082e+02	9	20	0.0
c-260209-1	-6.8419290856e-02	7	37	3.8

Section 4

Summary

Extremely disciplined modelling

- Conic modelling using the linear, quadratic, semidefinite, power, and exponential cones. (More cones in the future?).
- Inspired by disciplined modeling of Grant, Boyd and Ye [3].
- Models can be solved efficiently using a primal-dual interior-point method.
- Preserves all the good properties of linear optimization:
  - Simple and explicit data.
  - Structural convexity.
  - Duality (**almost**).
  - No issues with smoothness and differentiability.

Further information

- Mosek `https://mosek.com`
  - Trial and free academic license.
  - Solves linear and conic mixed problems.
  - Interfaces C, Java, Julia, Matlab, R, Python, ...
- Documentation at `https://www.mosek.com/documentation/`
  - Modelling cook book.
  - Modeling cheat sheet.
- Examples
  - Tutorials at Github: `https://github.com/MOSEK/Tutorials`

[1] J. Borwein and S. Lindstrom.
Meetings with Lambert W and Other Special Functions in Optimization and Analysis.
Technical report, University of Newcastle, 2016.

[2] Peter Robert Chares.
Cones and interior-point algorithms for structed convex optimization involving powers and exponentials.
PhD thesis, Ecole polytechnique de Louvain, Universitet catholique de Louvain, 2009.

[3] 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.

[4] M. Lubin and E. Yamangil and R. Bent and J. P. Vielma.
Extended Formulations in Mixed-integer Convex Programming.
Technical report, 2016.
arXiv:1511.06710v3.

[5] 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.

[6] Dvid Papp.
Duality of sum of nonnegative circuit polynomials and optimal sonc bounds, 2019.