Polyhedral Approaches for Mixed-integer Convex Optimization

GERAD

Miles Lubin

February 8, 2019

Google, New York City (This work was conducted at MIT.)

Optimization

Given: Feasible region $\mathcal{H}$ and objective function $\hat{a} : \mathcal{H} \; ! \; \mathrm{R}$

Find: $\_ \; 2 \; \mathcal{H}$ with lowest possible value of $\hat{a}(\_)$ (global solution)

Optimization

Given: Feasible region $\mathcal{H}$ and objective function $\hat{a} : \mathcal{H} \; ! \; \mathrm{R}$

Find: $\_ \; 2 \; \mathcal{H}$ with lowest possible value of $\hat{a}(\_)$ (global solution)

How to solve?

- Local search (e.g., gradient descent), genetic algorithms, ...
- Approximate or reformulate (if lucky) as another optimization problem we know how to solve to a global solution

Which optimization problems can we solve to global optimality?

Which optimization problems can we solve to global optimality?

- Integer programming
- Convex optimization

In this talk, we consider **mixed-integer convex optimization**, or mixed-integer convex programming (MICP).

These are convex optimization problems with integrality constraints.

We look for global solutions or at least global bounds.

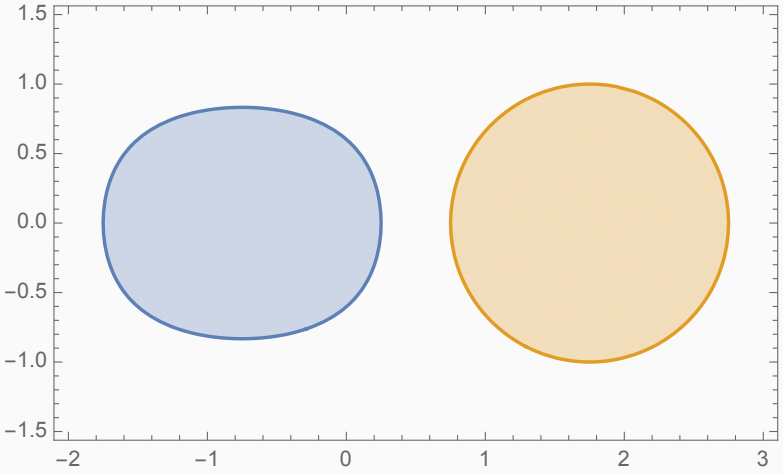
Themes

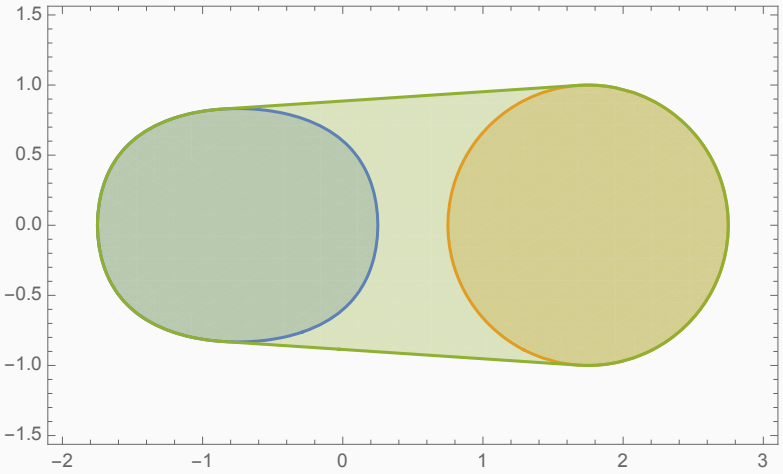
We apply extended formulations, conic duality, and branch and cut to get a state-of-the-art open-source MICP solver.

This work motivated developments that will appear in MOSEK 9 (presented at ISMP 2018).

What can you model with MICP?

It is possible to model a broad scope of problems using MILP. What ɘ ƗƵ¢Ù can you do if allowed to model with convex constraints?





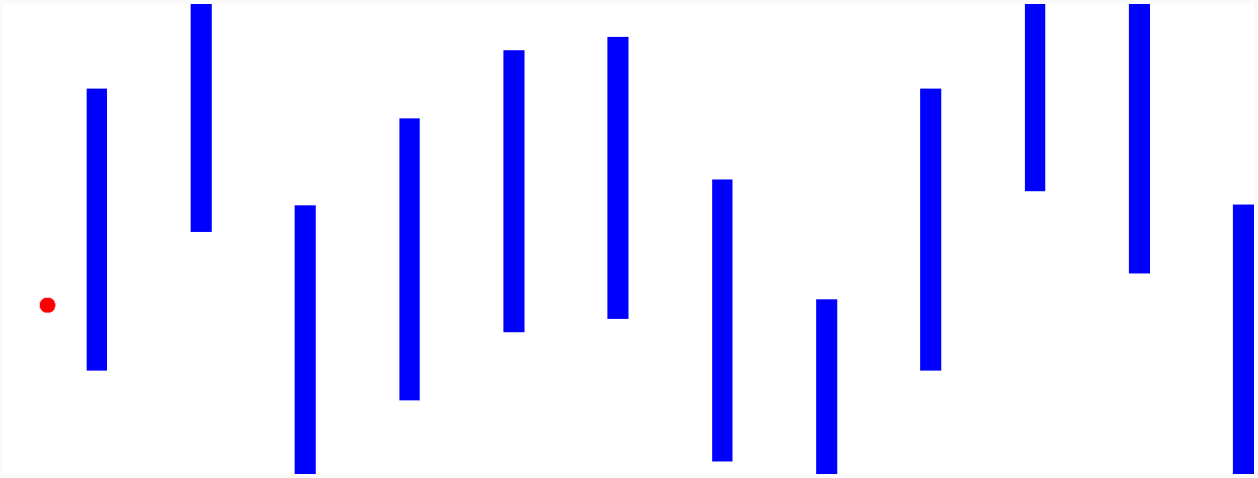
Ceria and Soares (1999) develop MICP formulations for unions of certain (e.g., bounded) convex sets.

See L., Zadik, and Vielma (IPCO 2017, arXiv 1706.05135) for much more on this.

A fun example

Mixed-integer sum-of-squares optimization

- Sum-of-squares constraints (which are convex) can be used to model polynomial constraints, including trajectories of a moving object
- Control decisions with discrete aspect (e.g., minimize number of thrusts) yield integer constraints



Credit: Joey Huchette (SIAM Conference on Optimization, 2017)

We define MICP general form as:

$$\min_{\_ \in \mathbb{R}^{\text{Þ}}} \quad \text{ß}^{1}\_ :$$

$$\_ \in X;$$

$$\__{\text{ζ}} \in \mathbb{Z}; \qquad\qquad \forall \text{ζ} \in \text{J}^{3}\text{K} = \{1; : : : ; {}^{3}g;$$

where X is a closed, convex set, and the first ^{3} of Þ variables are constrained to take integer values.

We can always assume objective is linear, via an epigraph transformation.

When X is a polyhedron, this reduces to MILP.

Solution methods

- Branch & bound – recursively partition possible assignments to integer variables and use convex relaxations to prune the search tree (Gupta and Ravindran, 1985; Bonami et al., 2013)

Solution methods

- Branch & bound – recursively partition possible assignments to integer variables and use convex relaxations to prune the search tree (Gupta and Ravindran, 1985; Bonami et al., 2013)
- **Iterative outer approximation (OA)** – based on polyhedral relaxation of convex constraints, solve a sequence of MILP relaxations and convex subproblems (Duran and Grossmann, 1986; Leyffer, 1993; Bonami et al., 2008)

Solution methods

- Branch & bound – recursively partition possible assignments to integer variables and use convex relaxations to prune the search tree (Gupta and Ravindran, 1985; Bonami et al., 2013)
- **Iterative outer approximation (OA)** – based on polyhedral relaxation of convex constraints, solve a sequence of MILP relaxations and convex subproblems (Duran and Grossmann, 1986; Leyffer, 1993; Bonami et al., 2008)
- **LP-based branch & bound** – use polyhedral relaxations within a single branch & bound (B&B) search tree (Quesada and Grossmann, 1992; Bonami et al., 2008)

Developing polyhedral relaxations

Suppose:

$$X = \{ \_ \in \mathrm{R}^{\sigma} : \hat{E}_{\phi}(\_) \leq 0; \forall \phi \in \mathrm{JÆK} \}$$

Then under assumptions on $\hat{E}_{\phi}$:

$$X = \{ \_ \in \mathrm{R}^{\sigma} : \hat{E}_{\phi}(\_^{0}) + r\hat{E}_{\phi}(\_^{0})^{\intercal}(\_ - \_^{0}) \leq 0; \forall \_^{0} \in \mathrm{R}^{\sigma}; \phi \in \mathrm{JÆK} \}$$

Any finite subset of linearization points yields a polyhedral outer approximation.

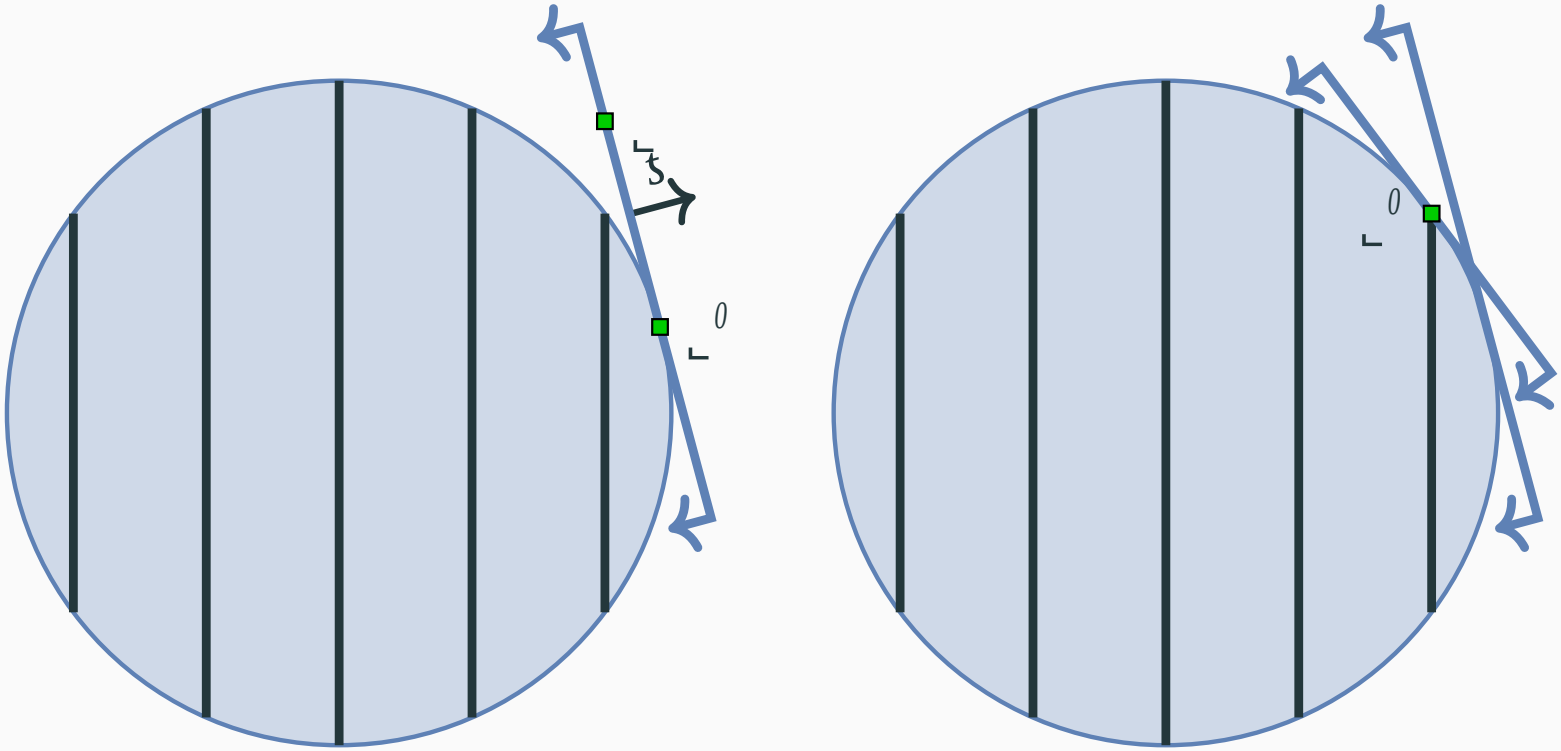
Given a polyhedral outer approximation $\dot{E}$ X, consider:

$$\begin{aligned} \min_{\llcorner} \quad & \mathbf{s}^{\prime}{}_{\llcorner} : \\ & {}_{\llcorner} \; 2 \; \dot{E}; \\ & {}_{\llcorner\zeta} \; 2 \; \mathrm{Z}; \qquad\qquad 8\zeta \; 2 \; \mathrm{J}^{3}\mathrm{K} : \end{aligned}$$

The above problem is a mixed integer linear relaxation.

Algorithmic idea: Iteratively improve the relaxation until lower bound matches feasible solution.

An illustration:



What can go wrong?

Classical approaches form polyhedral outer approximations by a finite collection of "gradient linearizations." But a good polyhedral outer approximation might need **too many linear constraints**.

Recall the ℓ_1 ball:

$$B_1 = \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\} = \left\{x \in \mathbb{R}^n : \sum_{i=1}^{n} r_i x_i \leq 1, r \in \{-1, +1\}^n\right\}$$

$$B_1 = \{ \in \mathbb{R}^n : \|\| \|_1 \quad 1\} = \left(\in \mathbb{R}^n : \sum_{=1} r_{,} \quad 1; r \in \{ 1; +1\}^n \right)$$

Standard LP extended formulation:

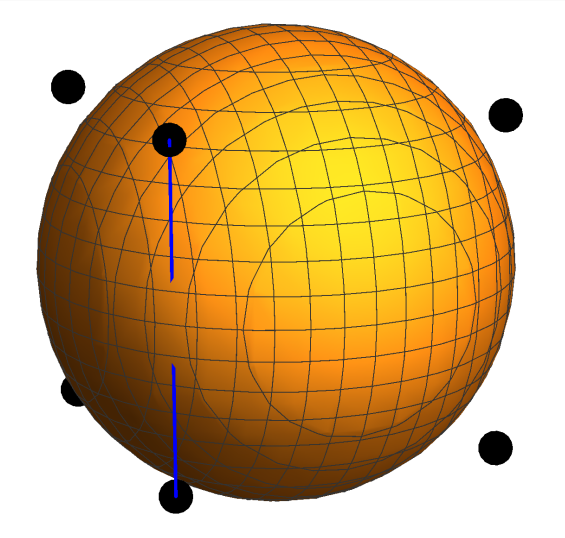
$$B_1 = \left(\in \mathbb{R}^n : \exists \in \mathbb{R}^n : \quad ; \sum_{=1} \quad 1 \right)$$

The new formulation has $2n$ variables and $2n + 1$ constraints versus n variables and 2^n constraints.

The same happens with smooth constraints

Hijazi et al. (2014) consider the set:

$$B_2 = \left(\sqsubset 2\ f0;1g^{\sigma} : \sum_{\zeta=1} \sqsubset\zeta \quad \frac{1}{2}^{2} \quad \frac{\sigma \quad 1}{4} \right)$$



How to fix it?

Consider the equivalent extended formulation ($B_2 = \text{proj}_{\_} \hat{B}_2$):

$$\hat{B}_2 = \left\{ (\_;\tilde{}) \in [0;1]^{\square} \times \mathbb{R}^{\square} : \sum_{\zeta=1}^{\square} \tilde{}_{\zeta} \geq \frac{\square - 1}{4}; \; \__{\zeta} \geq \frac{1}{2} \tilde{}_{\zeta}; \forall \zeta \in [\square] \right\}$$

How to fix it?

Consider the equivalent extended formulation ($B_2 = \text{proj}_{\_} \hat{B}_2$):

$$\hat{B}_2 = \left\{ (\_;\tilde{}) \in \{0;1\}^{\sigma} \times \mathbb{R}^{\sigma} : \sum_{\zeta=1}^{\sigma} \tilde{}_{\zeta} \left(\frac{\sigma}{4} - \frac{1}{2} \right)^{2} \__{\zeta} \tilde{}_{\zeta}; \forall \zeta \in [\sigma] \right\}$$

A polyhedral outer approximation in the space of $(\_;\tilde{})$ needs only 2σ hyperplanes to exclude all integer points.

Say a user writes down a constraint:

$$\text{“} \quad \hat{a}(\_) + \hat{E}(\_):$$

If $\hat{a}$ and $\hat{E}$ are convex, then extended formulation is obtained by introducing $?_1; ?_2$ such that:

$$\text{“} \quad ?_1 + ?_2; ?_1 \quad \hat{a}(\_); ?_2 \quad \hat{E}(\_):$$

But $\hat{a} + \hat{E}$ convex doesn't imply $\hat{a}$ and $\hat{E}$ convex (e.g., $\hat{a}(\_) = \__1^2 \quad \__2^2$ and $\hat{E}(\_) = 2\__2^2$).

Say a user writes down a constraint:

$$\text{“} \quad \hat{a}(\_) + \hat{E}(\_):$$

If $\hat{a}$ and $\hat{E}$ are convex, then extended formulation is obtained by introducing $?_1; ?_2$ such that:

$$\text{“} \quad ?_1 + ?_2; ?_1 \quad \hat{a}(\_); ?_2 \quad \hat{E}(\_):$$

But $\hat{a} + \hat{E}$ convex doesn't imply $\hat{a}$ and $\hat{E}$ convex (e.g., $\hat{a}(\_) = \__1^2 \quad \__2^2$ and $\hat{E}(\_) = 2\__2^2$).

So if we have access to the algebraic representation, in principle need to solve a subproblem of convexity detection before constructing extended formulation.

Convexity detection is hard.

Say a user writes down a constraint:

$$\text{“} \quad \hat{a}(\_) + \hat{E}(\_):$$

If $\hat{a}$ and $\hat{E}$ are convex, then extended formulation is obtained by introducing $?_1; ?_2$ such that:

$$\text{“} \quad ?_1 + ?_2; ?_1 \quad \hat{a}(\_); ?_2 \quad \hat{E}(\_):$$

But $\hat{a} + \hat{E}$ convex doesn't imply $\hat{a}$ and $\hat{E}$ convex (e.g., $\hat{a}(\_) = \__1^2 \quad \__2^2$ and $\hat{E}(\_) = 2\__2^2$).

So if we have access to the algebraic representation, in principle need to solve a subproblem of convexity detection before constructing extended formulation.

Convexity detection is hard. (But some solvers try.)

Conic optimization solves the problem

$$\min_{\square} \quad \mathfrak{s}^{l}{}_{\square} : \\ \quad \mathcal{y} \quad {}_{\square} \; 2 \; K$$

A set $K \quad \mathrm{R}^{\mathcal{e}}$ is a **closed convex cone** if it is closed and contains all conic combinations of its points, i.e.,

$$_{1\leftarrow 1} + \quad _{2\leftarrow 2} \; 2 \; K \qquad 8 \; _{1}; \quad _{2} \quad 0 \qquad 8_{\leftarrow 1}; {}_{\leftarrow 2} \; 2 \; K:$$

An example

The constraint

$$\log\left(\sum_{i=1}^{n} \exp(x_i)\right) \leq t$$

is convex, not separable. Converting it to conic form exposes the summation structure.

An example

The constraint

$$\log \left(\sum_{i=1} \exp(a_i) \right) \leq b$$

is convex, not separable. Converting it to conic form exposes the summation structure.

$$\sum_{i=1} \exp(a_i) \leq \exp(b)$$

$$\sum_{i=1} \exp(a_i - b) \leq 1$$

$$\exists c : \sum_{i=1} c_i \leq 1 \text{ and } \forall i; (c_i, 1, a_i - b) \in E,$$

where $E = \text{cone}(\{(t, x) : t \geq \exp(x)\})$.

In general...

- “Lectures on Modern Convex Optimization” by Ben-Tal and Nemirovski is the canonical source for how to write a problem in conic form.
- Disciplined convex programming (DCP) provides a systematic and easy-to-use approach to translating algebraic expressions into conic form (, ¯µ, , ¯µq¶, , ĎćκÛĺЩúĀ)
- If not convinced, read more at “Polyhedral Approximation in Mixed-integer Convex Optimization”, Math Programming B, 2017.

Which cones are needed?

- Nonnegative orthant $\mathrm{R}^{\sigma}_{+} = f^{\imath} \; 2 \; \mathrm{R}^{\sigma} : {}^{\imath} \quad \mathbf{0}g$
- Second-order cone $L^{1+\sigma} = \quad (\dot{\iota}; {}^{\imath}) \; 2 \; \mathrm{R}^{1+\sigma} : \dot{\iota} \quad k^{\imath}k_2$
- Exponential cone $E = \mathrm{cl} \quad (\dot{\iota}; {}^{r}; {}^{\imath}) \; 2 \; \mathrm{R}^{3} : {}^{r} > 0; \dot{\iota} \quad {}^{r} \exp \; \frac{{}^{\imath}}{r}$
- PSD cone $\mathrm{S}^{\sigma}_{+} = f\imath \; 2 \; \mathrm{S}^{\sigma} : \; \min(\imath) \quad 0g$
- Power cone $W \; = f(\llcorner; \lrcorner; ^{\prime\prime}) \; 2 \; \mathrm{R}^{3} : j^{\prime\prime}j \quad \llcorner \; \lrcorner^{1} \quad ; \llcorner \quad 0; \lrcorner \quad 0g$

We categorized the 333 mixed-integer convex instances in the MINLPLIB2 library according to conic representability. Excluding R^{σ}_{+}, the following cones are needed:

L only	E only	L and E	W only	Total
217	107	7	2	333

Open problems solved from MINLPLIB2

- æÄĆġ6ϑ, L+E-representable
  - 145 variables, 1268 constraints (1158 linear)
  - Using , qX6µ as MILP solver and V_I~te as convex solver, iterative OA solved in 6 iterations ($<$ 10 hours)
  - Optimal solution is 21380.20 (previous best bound was 1735.06, and best known solution was 21516.83)
- ĨĀġϕ and ĨĀġϖ, L-representable, unsolved since 2001
  - Solved in less than 24h by converting to MISOCP and then using AǐĝĎÐî

Mosek 9 adds support for the exponential and power cones (first commercial solver).

ëÎĨĚġУЯЯ×ĎÑġЩĆĎġÛýЩÑĎĆЯġĀî×ÛġЯYƃϑQЯîġĆĚYƃϑQЯ îġĆĚюÄć×ÛĝġÛćЩĚ×å

Recall:

- Iterative OA: Solve a sequence of MILP relaxations
- LP-based B&B: Use a single B&B tree with relaxations computed by solving LPs

How to generate cuts in the conic case?

A set $K \subseteq \mathbb{R}^n$ is a **closed convex cone** if it is closed and contains all conic combinations of its points, i.e.,

$$\theta_1 x_1 + \theta_2 x_2 \in K \quad \forall\, x_1, x_2 \in K \quad \forall\, \theta_1, \theta_2 \geq 0 \quad \forall x_1, x_2 \in K. \tag{1}$$

The **dual cone** of a set is defined as:

$$K^* = \{ y \in \mathbb{R}^n : y^T x \geq 0, \forall x \in K \}$$

If K is a closed convex cone, then:

$$x \in K \iff y^T x \geq 0, \forall y \in K^*$$

Any finite subset $Z \quad K$ yields a polyhedral outer approximation of K. That is:

$$\dot{E} := \quad \llcorner \ 2 \ \mathrm{R}^{\sigma} : \llcorner^{\prime\prime} \quad 0; 8^{\prime\prime} \ 2 \ Z \quad :$$

$\dot{E}$ is a polyhedron that contains K.

We say that an element $^{\prime\prime} \ 2 \ K$ corresponds to a K cut.

- The second-order cone L and PSD cone P are **self-dual**
- The exponential cone E is $\mathfrak{sk}^2$ self dual, but E is easily described
- K cuts can be validated by checking membership in dual cone
- K cuts can be safely rounded by projection onto boundary of dual cone

Compare with gradient inequalities:

$$X = f_{\_} \; 2 \; \mathrm{R}^{\sigma} : \hat{E}(\_) \quad \mathbf{0}g:$$

Then $^{\imath}\_$ is a gradient inequality iff there exists $\_^{0} \; 2 \; \mathrm{R}^{\sigma}$ such that:

$$= \; r\,\hat{E}_{\emptyset}(\_^{0});$$
$$= \; r\,\hat{E}_{\emptyset}(\_^{0})^{\imath}\_^{0} \quad \hat{E}(\_^{0}):$$

This is much harder to check.

Conic duality

Primal

$$\min_{x} \quad c^T x : \quad Ax - b \in K$$

Dual

$$\max_{y} \quad b^T y : \quad A^T y + c = 0, \quad y \in K^*$$

Reduces to LP duality when $K = \mathbb{R}^p_+$.

Conic duality can fail

The following MISOCP instance cannot be solved by the iterative OA algorithm:

$$\min_{\llcorner;\leftarrow;''} \quad '' :$$

$$\llcorner\leftarrow \quad ''^2;$$

$$\llcorner = 0;$$

$$\leftarrow \quad 0;$$

$$\llcorner \ 2 \ f0; 1g:$$

Conic duality can fail

The following MISOCP instance cannot be solved by the iterative OA algorithm:

$$\min_{\llcorner;\leftarrow;''} \quad '' :$$

$$\llcorner\leftarrow \quad ''^2;$$

$$\llcorner = 0;$$

$$\leftarrow \quad 0;$$

$$\llcorner \in \{0;1\}:$$

Why not?

- The optimal solution is zero
- The dual of the root node relaxation is infeasible
- $\sigma_{\leftarrow}$ polyhedral OA is unbounded below

MI-conic form

$$\min_{x} \quad c^T x : \quad Ax - b \in K, \qquad x_i \in \mathbb{Z} \quad \forall i \in I$$

(When applying B&B, we also assume known lower and upper bounds on the integer-constrained variables.)

Certificate K cuts

Suppose we solve:

$$\min_{\square} \quad \$^{\prime} \square \; : \quad \mathcal{Y} \quad \square \; 2 \; K \qquad (\text{''}) \quad \mathbb{E}_{\zeta} \quad \square_{\zeta} \quad \text{'}_{\zeta} \qquad 8\zeta \; 2 \; \mathrm{J}^{3}\mathrm{K} \qquad \text{(assuming } \mathbb{E}_{\zeta} \quad \text{'}_{\zeta}\text{)}$$

and obtain an optimal objective value $\imath$ and dual solution $''$ (ignoring multipliers on inequalities). Then

$$\imath \; = \; \min_{\square} \quad \$^{\prime} \square \; : \quad (\mathcal{Y} \quad \square)^{\prime\prime} \quad 0 \quad \mathbb{E}_{\zeta} \quad \square_{\zeta} \quad \text{'}_{\zeta} \qquad 8\zeta \; 2 \; \mathrm{J}^{3}\mathrm{K}$$

Iterative OA uses certificate cuts with integer variables fixed to the optimal solution returned from the MILP relaxation. ( ) finite convergence)

LP-based B&B uses conic certificate cuts at multiple points in the tree. Must use at least whenever a potential new incumbent integer feasible solution is found.

Recall ″ 2 K 8 0. Does it matter how we scale ″ ?

LP solvers almost never guarantee exact feasibility. Suppose instead that we solve,

$$\min_{\lrcorner} \quad \mathbf{s}^{l}\lrcorner :$$

$$(\mathcal{Y} \quad \lrcorner)^{l''} \qquad \hat{a}\grave{U}\mathcal{E}^{r}$$

$$\not{E}_{\zeta} \quad \lrcorner_{\zeta} \quad \tau_{\zeta} \qquad\qquad 8\zeta\ 2\ \mathrm{J}^{3}\mathrm{K}$$

This has optimal value $\imath \quad \frac{}{\hat{a}\grave{U}\mathcal{E}^{r}}$!

Recall ″ 2 K 8 0. Does it matter how we scale ″ ?

LP solvers almost never guarantee exact feasibility. Suppose instead that we solve,

$$\min_{\lrcorner} \quad \mathfrak{s}^{l}\lrcorner :$$

$$(\mathcal{y} \quad \lrcorner)^{l''} \qquad \hat{a}\grave{U}\mathcal{E}^{r}$$

$$\mathbb{E}_{\zeta} \quad \lrcorner_{\zeta} \quad \ulcorner_{\zeta} \qquad\qquad 8\zeta\ 2\ \mathrm{J}^{3}\mathrm{K}$$

This has optimal value $\imath$ $\frac{}{\hat{a}\grave{U}\mathcal{E}^{r}}$!

So choose to make $\frac{}{\hat{a}\grave{U}\mathcal{E}^{r}}$ on the scale of the optimal objective, e.g., 0:01$\imath$.

Separation K cuts

Given $\bar{x}$ such that $\bar{x} \notin K$, any hyperplane that separates $\bar{x}$ from K can be shifted (if needed) to become a K cut.

Example: If $\boldsymbol{A} \notin \mathbb{S}_+^n$ then $\lambda_{\min}(\boldsymbol{A}) < 0$. Let $\boldsymbol{v}$ be an eigenvector corresponding to the smallest eigenvalue of $\boldsymbol{A}$. Then:

$$\boldsymbol{v}^T \boldsymbol{A} \boldsymbol{v} = \lambda_{\min}(\boldsymbol{A}) < 0; \tag{2}$$

so $\boldsymbol{v}^T \boldsymbol{x} \geq 2$ corresponds to a K cut that separates $\boldsymbol{A}$ from $\mathbb{S}_+^n$.

Example: If $\mathbf{Z} \notin \mathbb{S}^n_+$ then $\lambda_{\min}(\mathbf{Z}) < 0$. Let $\mathbf{v}$ be an eigenvector corresponding to the smallest eigenvalue of $\mathbf{Z}$. Then:

$$\mathbf{v}^T \mathbf{Z} \mathbf{v} = \lambda_{\min}(\mathbf{Z}) < 0; \tag{2}$$

so $\mathbf{v}^T \geq 0$ $(\mathbb{S}^n_+)$ corresponds to a K cut that separates $\mathbf{Z}$ from $\mathbb{S}^n_+$.

x, Iqюx2q (Gally et al., 2017) uses these cuts for MISDP. x, Iq, AĭĝĎĐî, , qX6µ, and µĚĝÛġġ use separation cuts for MISOCP.

Extensions to K cuts

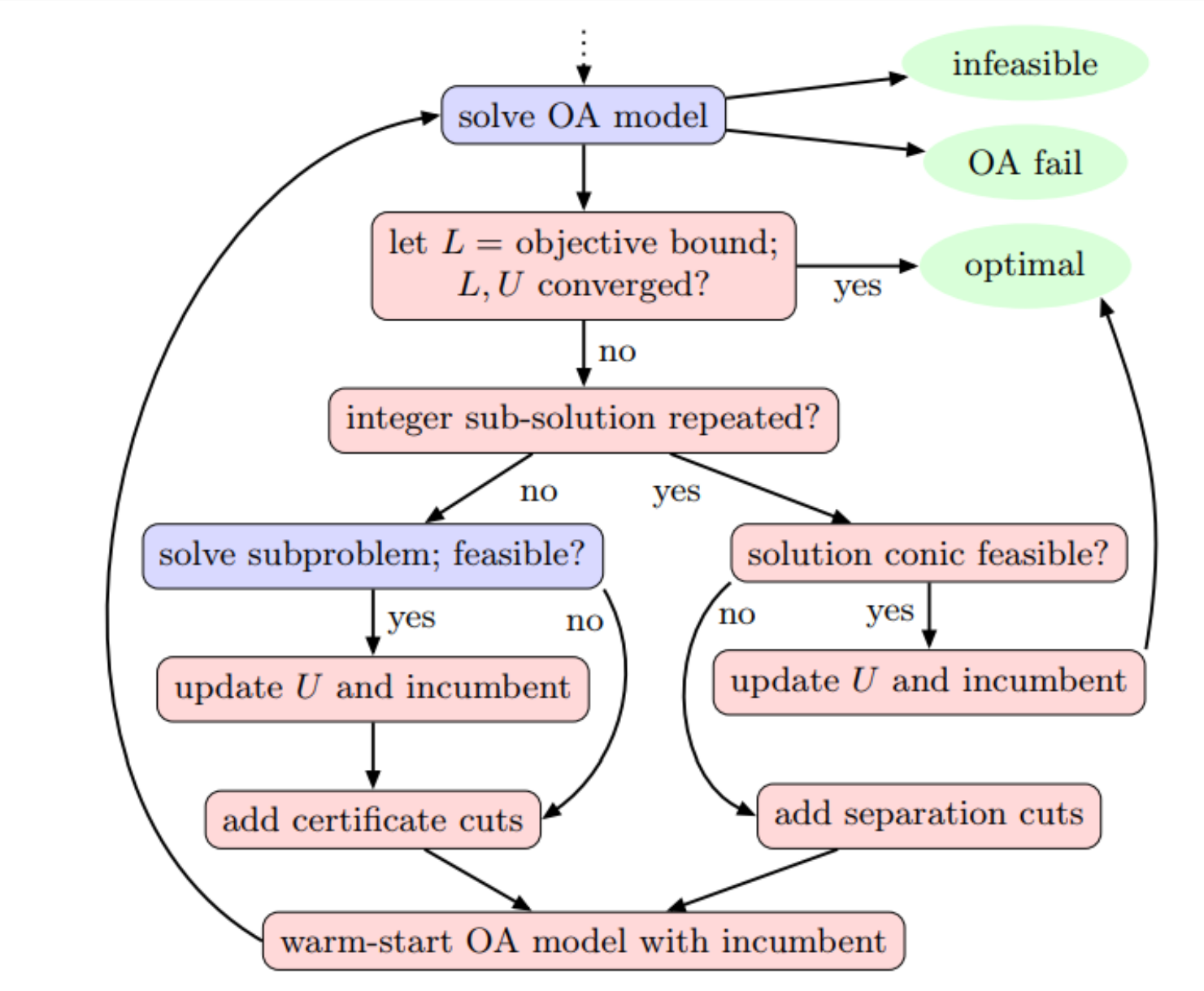
Natural extensions to the K cuts framework include:

1. Obtaining multiple K cuts from a single one by decomposing it by products of cones or into extreme rays
2. Second-order cone outer approximation of the PSD cone
3. Lifting of the second-order cone (Vielma et al., 2016)

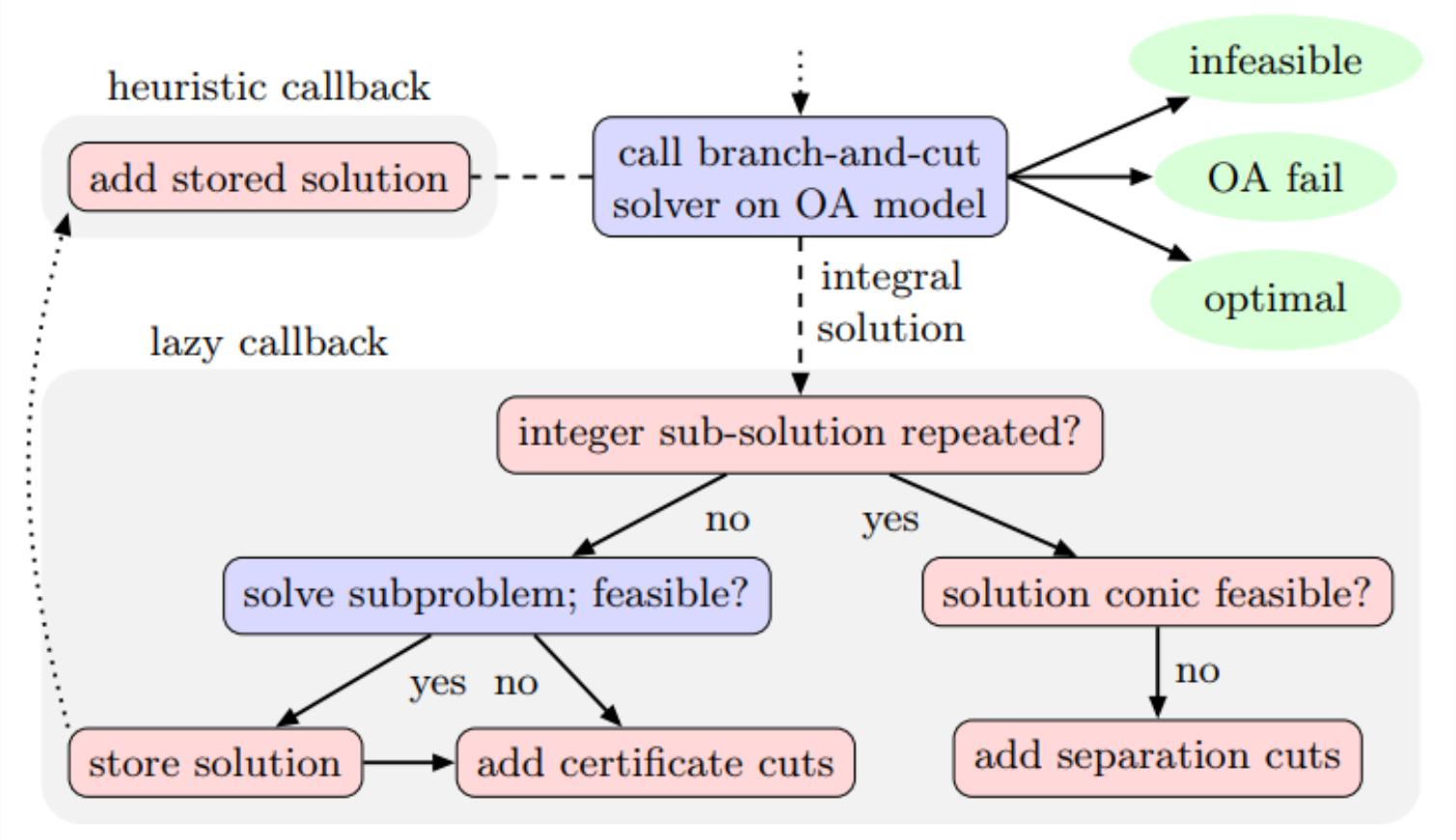
Pajarito

- An open-source MICP solver released in 2016
- Written in Julia, accessible through JuMP and via CBF files
- Can use any conic and MILP solver available to JuMP
- Implements iterative OA and LP-based B&B using callbacks

Note: For convex MINLP see Pavito. This functionality was split off last year from Pajarito.



The iterative OA algorithm in Pajarito.



The LP-based B&B algorithm in Pajarito. We call it MIP-solver-driven (MSD) because the MIP solver drives the algorithm.

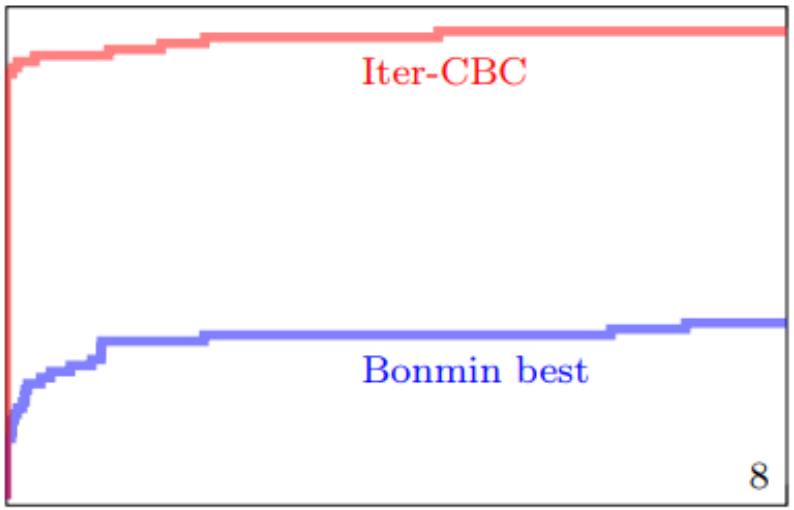
Numerical experiments

- Selection of 120 MISOCP instances from CBLIB library
- Solutions with large feasibility violations marked as ‘excluded’
- ‘conv‘ means converged, ‘error’ means failed to converge, and ‘limit’ means timeout
- Time summaries presented as shifted geometric means $(\prod_{i=1}^{n}(t_i + s))^{\frac{1}{n}} - s$ with $s = 10$ seconds

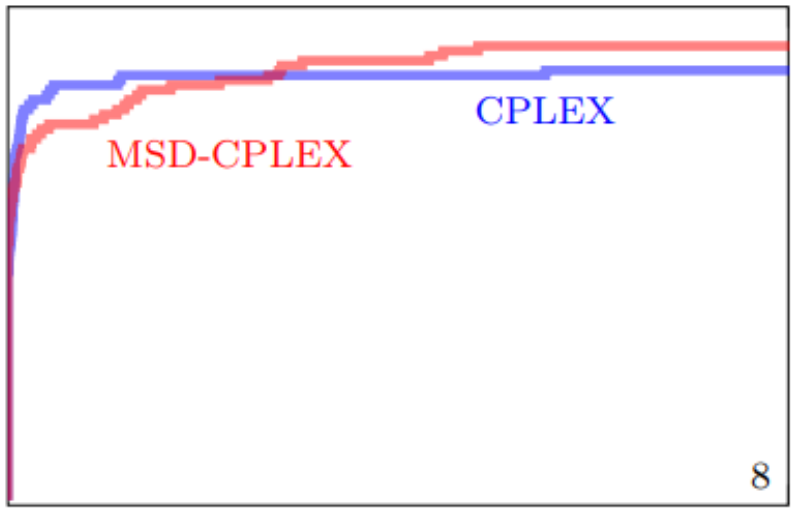
Status summary

		statuses				time (s)
	solver	conv	limit	error	excluded	all conv
open source	Bonmin-BB	34	44	11	31	38.0
	Bonmin-OA	25	53	29	13	64.2
	Bonmin-OA-D	30	48	29	13	15.1
	Iter-GLPK	56	60	3	1	2.0
	Iter-CBC	78	30	3	9	1.6
restricted	SCIP	74	35	8	3	3.2
	CPLEX	90	16	5	9	0.9
	Iter-CPLEX	86	26	0	8	0.4
	MSD-CPLEX	97	20	2	1	0.4

Performance profiles



(a) Open source Bonmin (instance-wise best of 3) and Pajarito iterative solvers.



(b) CPLEX MISOCP and Pajarito MSD solvers.

Test of algorithic variants on 95 MICP instances with a variety of cones:

- Experimental design: PSD cone and exponential cone
- Portfolio problems with risk constraints: exponential cone (entropy risk), second-order cone (norm risk), PSD cone (robust norm risk)
- Retrofit-synthesis of process networks: exponential cone, from MINLPLIB2
- Representative MISOCP instances from CBLIB

Effect of certificate cut scaling

	scale	statuses			
		conv	limit	error	excluded
Iter	off	63	1	28	3
	on	69	1	22	3
MSD	off	60	0	30	5
	on	67	0	26	2

Significant improvement in reliability. Effect on solve time (not shown) is hard to tell.