Perturbation-Duality Scheme in Combinatorial Optimization and Algorithms in Generalized Convexity

Seta Rakotomandimby, Michel De Lara, Jean-Philippe Chancelier

First part: Perturbation-duality scheme in combinatorial optimization

- Rewriting of Jeroslow's result

Perturbation-duality scheme
+
generalized conjugacy

- Linking

Perturbation-duality scheme
and
"Lagrangian" relaxation

- Proposing a quasi-affine dual problems for Pure Integer Linear Programming

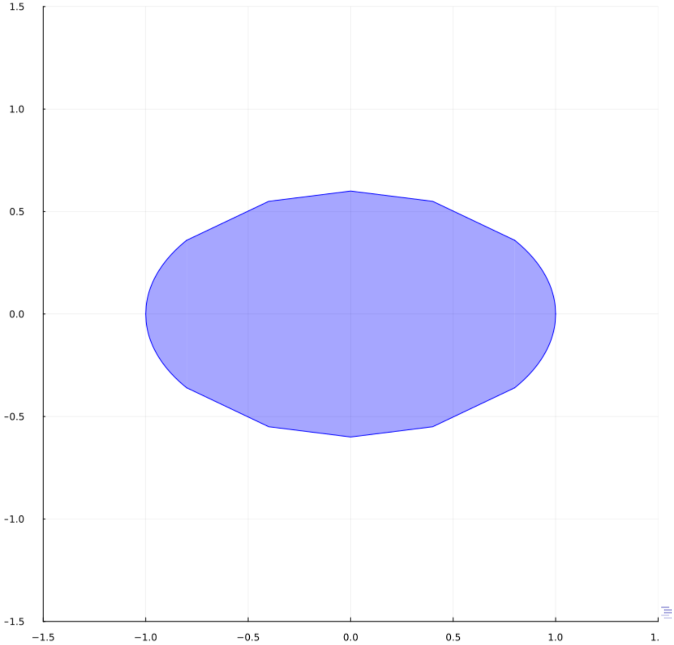
Second part: Cutting plane methods for sparse optimization

- Implementation of cutting plane methods using
  - results on CAPRA-convexity of ℓ_0 [Chancelier and De Lara, 2020, 2021]
  - and the calculation of its CAPRA-subdifferentials [Le Franc, 2021]
- Numerical tests on instances we generated in low dimension

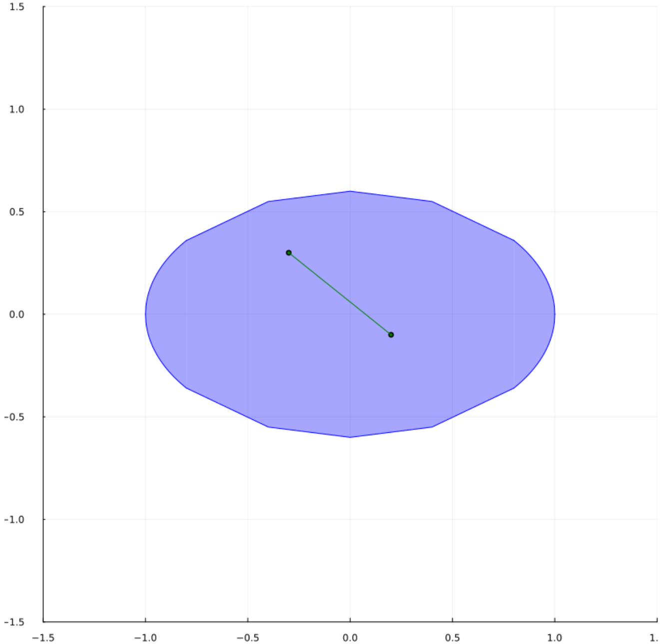
Outline

Outline

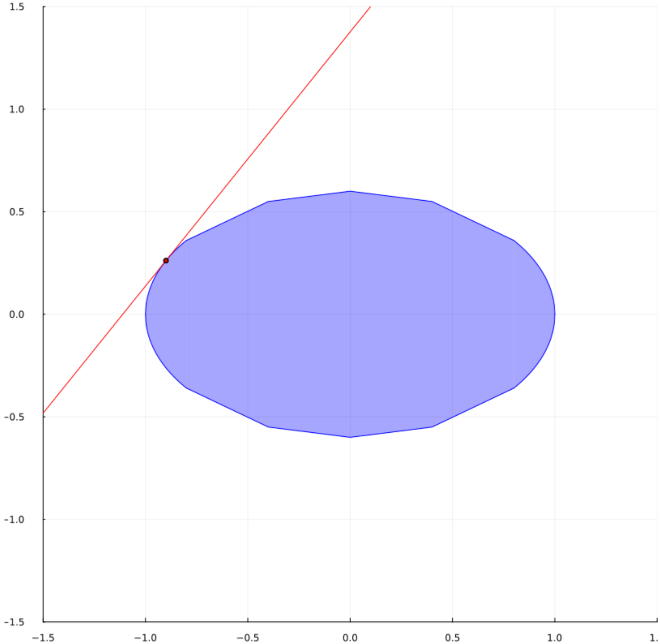
A closed convex set



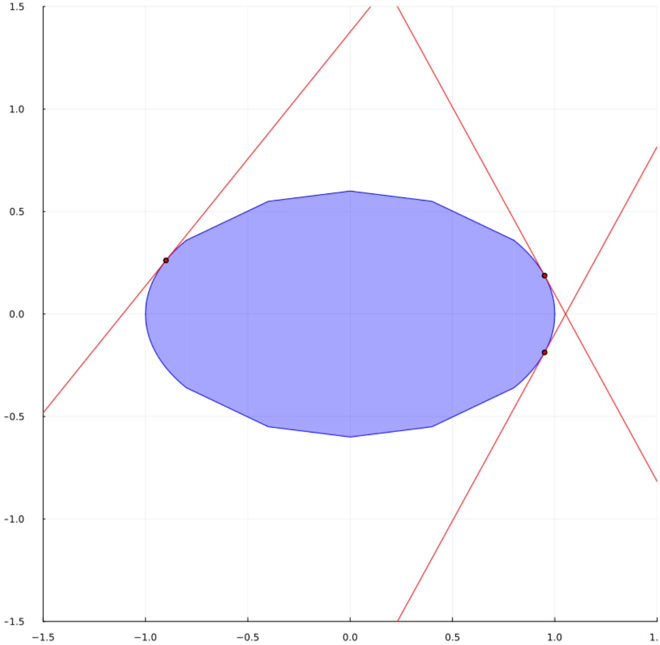
Usual definition of convexity by the interior



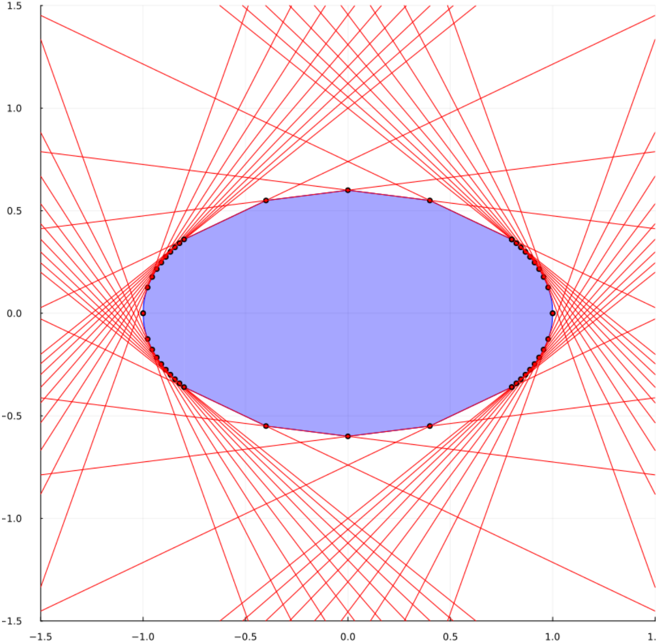
Equivalent definition for closed-convexity by the exterior



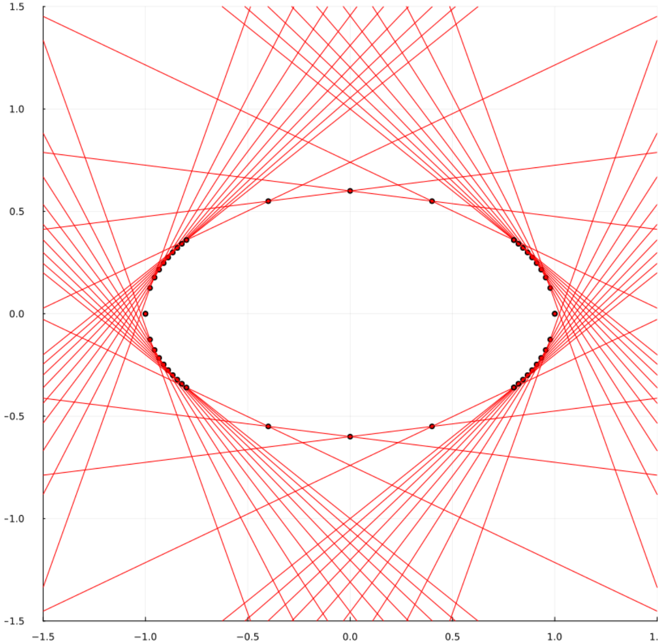
Equivalent definition for closed-convexity by the exterior



Equivalent definition for closed-convexity by the exterior

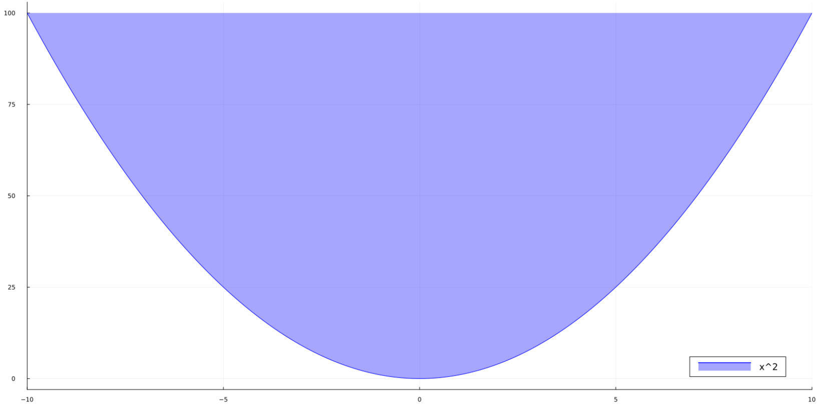


Approximation by finite number of cuts



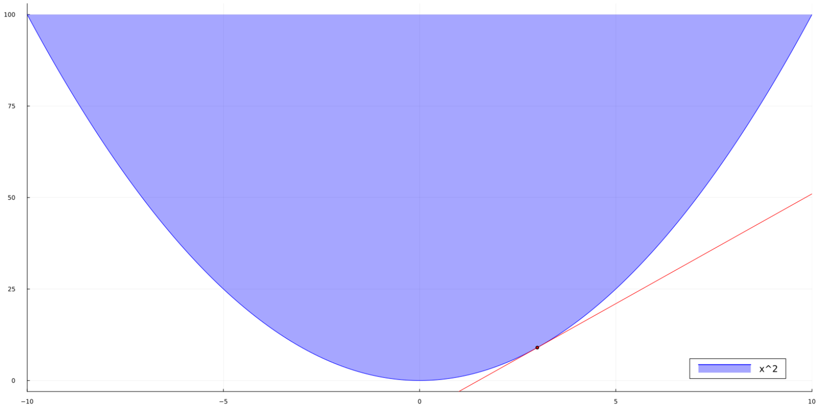
Epigraph of a closed-convex function

$$y = x^2$$



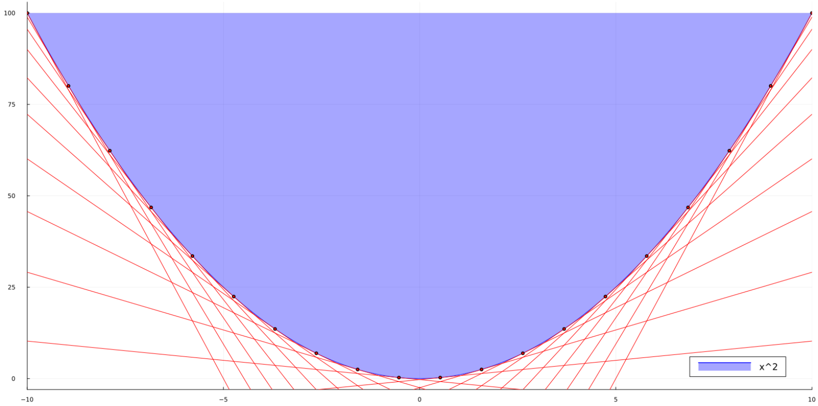
Epigraph of a closed-convex function

$$y = x^2$$



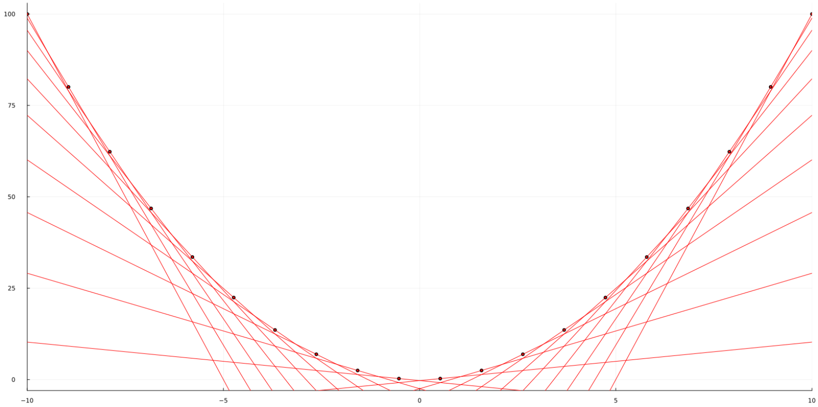
The epigraph is above its tangents

$y = x^2$

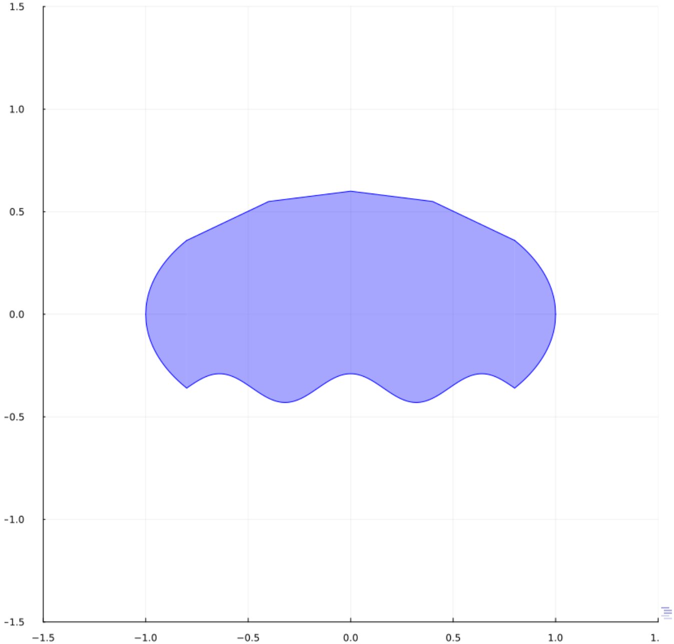


Approximation by a finite number of cuts

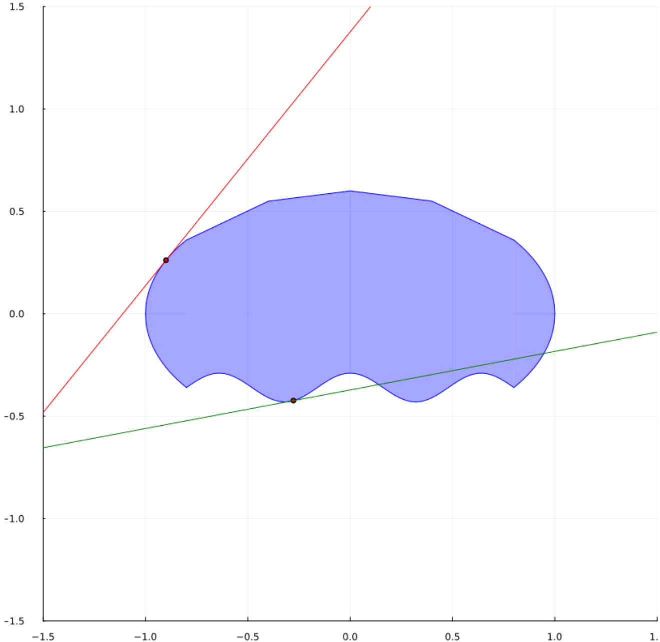
$y = x^2$



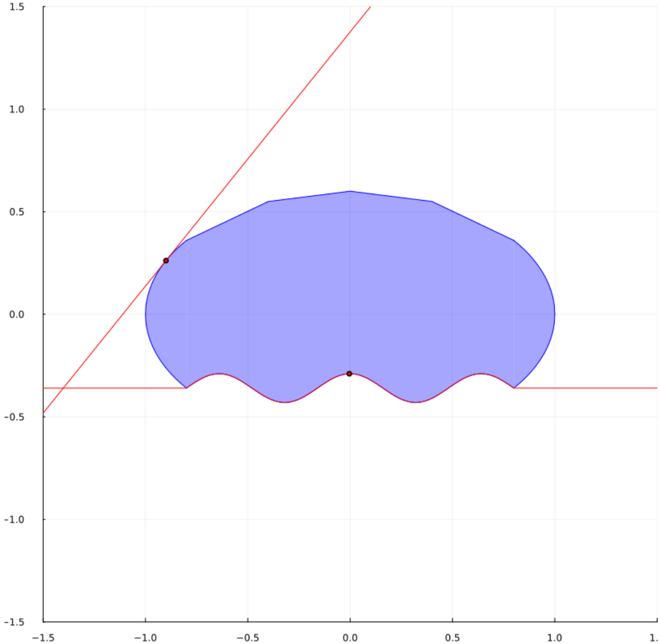
Example of a nonconvex set



Some tangents won't stay outside!



Generalized convexity: we change the shape of the tangents!



Generalized convexity: we change the shape of the tangents!

$$T(x) = \langle x, \alpha \rangle + \beta \, , \quad \forall x \in \mathbb{R}^n$$

Generalized convexity: we change the shape of the tangents!

$$T(x) = \langle x, \alpha \rangle + \beta \ , \ \ \forall x \in \mathbb{R}^n$$

Scalar product $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n$

Slope: $\alpha \in \mathbb{R}^n$

Intercept: $\beta \in \mathbb{R}$

Generalized convexity: we change the shape of the tangents!

$$T(x) = \langle x, \alpha \rangle + \beta \ , \ \ \forall x \in \mathbb{R}^n$$

Scalar product $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n$

Slope: $\alpha \in \mathbb{R}^n$

Intercept: $\beta \in \mathbb{R}$

$$T(u) = c(u, v) + \beta$$

Generalized convexity: we change the shape of the tangents!

$$T(x) = \langle x, \alpha \rangle + \beta \,, \quad \forall x \in \mathbb{R}^n$$

Scalar product $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n$

Slope: $\alpha \in \mathbb{R}^n$

Intercept: $\beta \in \mathbb{R}$

$$T(u) = c(u, v) + \beta$$

Coupling $c : U \times V \to \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$

Slope: $v \in V$

Intercept: $\beta \in \mathbb{R}$

Generalized convexity: we change the shape of the tangents!

$$T(x) = \langle x, \alpha \rangle + \beta \ , \ \ \forall x \in \mathbb{R}^n$$

Scalar product $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n$

Slope: $\alpha \in \mathbb{R}^n$

Intercept: $\beta \in \mathbb{R}$

$$T(u) = c(u, v) + \beta$$

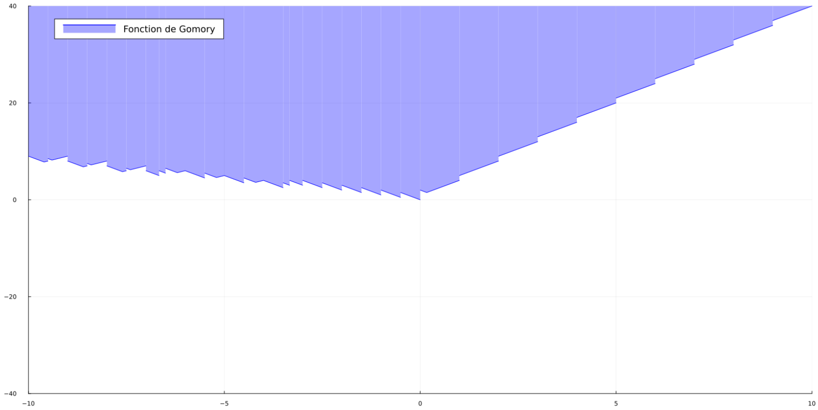
Coupling $c : U \times V \to \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$

Slope: $v \in V$

Intercept: $\beta \in \mathbb{R}$

Coming next: generalized convexity of Gomory function

$$y = \max\{3b + \lceil b \rceil, 2b + -3\lceil b \rceil, -3b + \lceil 2b \rceil + \lceil \tfrac{3}{10} b \rceil\}$$



Outline

Application of the scheme to Linear Programming

Initial minimization problem

$$\inf_{\substack{x \\ Ax = b_0 \\ x \in \mathbb{Q}_+^n}} \quad \langle x,\, k \rangle$$

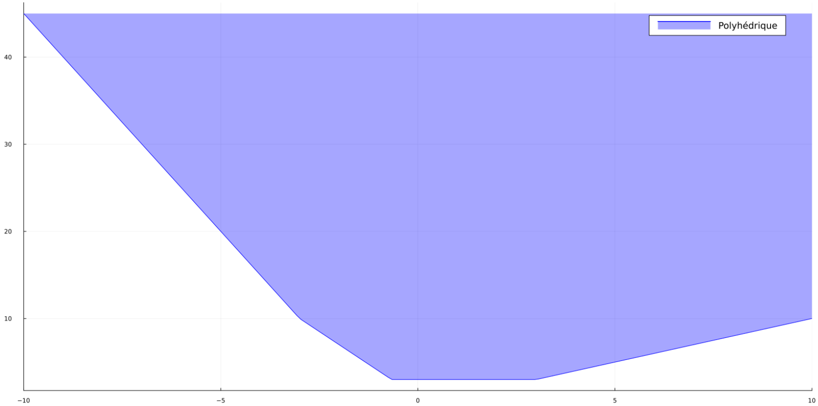
Step 1. Perturbation of the initial minimization problem

$$\forall b \in \mathbb{Q}^m\ , \quad \varphi(b) = \inf_{\substack{x \\ Ax = b \\ x \in \mathbb{Q}_+^n}} \langle x,\, k \rangle$$

- Perturbation space: $\mathbb{Q}^m$
- Perturbation function $\varphi : \mathbb{Q}^m \to \overline{\mathbb{R}}$
- Value of the initial problem: $\varphi(b_0)$

Epigraph of the perturbation function

$$\varphi(b) = \max\{-5b - 5, -3b + 1, 3, b\}$$



Step 2. Coupling and conjugate function

- Perturbation function

$$\forall b \in \mathbb{Q}^m \ , \ \ \varphi(b) = \inf_{\substack{x \\ Ax = b \\ x \in \mathbb{Q}^n_+}} \langle x, \ k \rangle$$

- Coupling $\langle \cdot, \ \cdot \rangle : \mathbb{Q}^m \times \mathbb{Q}^m \to \mathbb{R}$
- Conjugate function $\varphi^\star : \mathbb{Q}^m \to \overline{\mathbb{R}}$

$$\forall p \in \mathbb{Q}^m \ , \ \ \varphi^\star(p) = \sup_{b \in \mathbb{Q}^m} \left\{ \langle b, \ p \rangle - \varphi(b) \right\}$$

Step 3. Biconjugate and weak duality

- Biconjugate function $\varphi^{\star\star'} : \mathbb{Q}^m \to \overline{\mathbb{R}}$

$$\forall b \in \mathbb{Q}^m \ , \quad \varphi^{\star\star'}(b) = \sup_{p \in \mathbb{Q}^m} \left\{ \star(b,p) \dot{+} \left(-\varphi^{\star}(p) \right) \right\}$$

- Weak duality

$$\varphi^{\star\star'}(b) \leq \varphi(b)$$

Step 3. Biconjugate and weak duality

- Biconjugate function $\varphi^{\star\star'} : \mathbb{Q}^m \to \overline{\mathbb{R}}$

$$\forall b \in \mathbb{Q}^m , \quad \varphi^{\star\star'}(b) = \sup_{p \in \mathbb{Q}^m} \Big\{ \star(b,p) \dot{+} \big(-\varphi^{\star}(p)\big) \Big\}$$

- Property of biconjugacy

$$\varphi^{\star\star'}(b) \leq \varphi(b)$$

- Weak duality

$$\varphi^{\star\star'}(b) \leq \varphi(b) = \begin{array}{ll} \inf_x & \langle x, k \rangle \\ Ax = b & \\ x \in \mathbb{Q}^n_+ & \end{array}$$

Step 3. Biconjugate and weak duality

- Biconjugate function $\varphi^{\star\star'} : \mathbb{Q}^m \to \overline{\mathbb{R}}$

$$\forall b \in \mathbb{Q}^m \,, \quad \varphi^{\star\star'}(b) = \sup_{p \in \mathbb{Q}^m} \left\{ \star(b,p) \dotplus \left(-\varphi^{\star}(p)\right) \right\}$$

- Property of biconjugacy

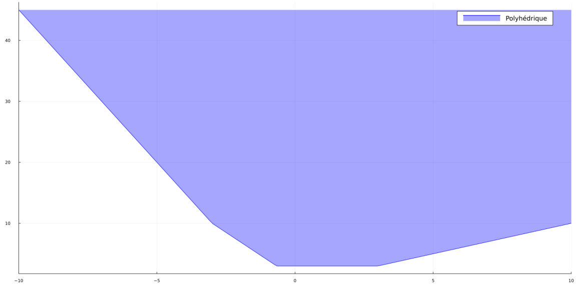
$$\varphi^{\star\star'}(b) \leq \varphi(b)$$

- Weak duality

$$\begin{array}{cc} \sup\limits_{p} & \langle p,\, b \rangle \\ p^T A \leq k & \\ p \in \mathbb{Q}^m & \end{array} = \varphi^{\star\star'}(b) \leq \varphi(b) = \begin{array}{cc} \inf\limits_{x} & \langle x,\, k \rangle \\ Ax = b & \\ x \in \mathbb{Q}^n_+ & \end{array}$$

Step 4. Closed convexity and strong duality

- φ is lower-semi-continuous convex

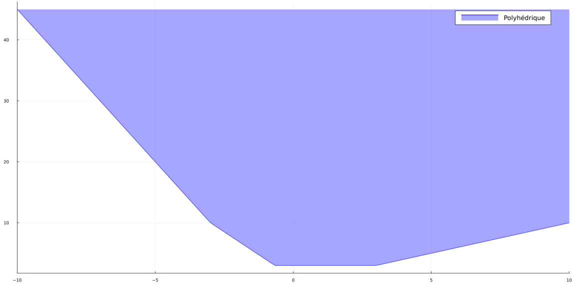


- So we have strong duality

$$\varphi^{\star\star\prime}(b) = \varphi(b)$$

Step 4. Closed convexity and strong duality

- φ is lower-semi-continuous convex



- So we have strong duality

$$\sup_{\substack{p \\ p^T A \leq k \\ p \in \mathbb{Q}^m}} \langle p,\, b_0 \rangle = \varphi^{\star\star\prime}(b) = \varphi(b) = \inf_{\substack{x \\ Ax = b \\ x \in \mathbb{Q}^n_+}} \langle x,\, k \rangle$$

Summary of the perturbation-duality scheme

[Rockafellar, 1974]

1. We perturb a minimization problem

$$\forall b \in \mathbb{Q}^m \,, \quad \varphi(b) = \inf_{\substack{x \\ Ax = b \\ x \in \mathbb{Q}^n_+}} \langle x, \, k \rangle$$

Summary of the perturbation-duality scheme

[Rockafellar, 1974]

1. We perturb a minimization problem

$$\forall b \in \mathbb{Q}^m \,, \quad \varphi(b) = \inf_{\substack{x \\ Ax = b \\ x \in \mathbb{Q}^n_+}} \langle x \,, \, k \rangle$$

2. We pair a primal space $\mathbb{Q}^m$ and a dual space $\mathbb{Q}^m$

$$\langle \cdot \,, \, \cdot \rangle : \mathbb{Q}^m \times \mathbb{Q}^m \to \mathbb{R}$$

Summary of the perturbation-duality scheme

[Rockafellar, 1974]

1. We perturb a minimization problem

$$\forall b \in \mathbb{Q}^m \,, \;\; \varphi(b) = \inf_{\substack{x \\ Ax = b \\ x \in \mathbb{Q}^n_+}} \langle x, \, k \rangle$$

2. We pair a primal space $\mathbb{Q}^m$ and a dual space $\mathbb{Q}^m$

$$\langle \cdot, \, \cdot \rangle : \mathbb{Q}^m \times \mathbb{Q}^m \to \mathbb{R}$$

3. We biconjugate the perturbation function φ

$$\underbrace{\varphi^{\star\star'}(b) \leq \varphi(b) \,, \;\; \forall b \in \mathbb{Q}^m}_{\text{Weak duality is guaranteed!}}$$

Summary of the perturbation-duality scheme

[Rockafellar, 1974]

1. We perturb a minimization problem

$$\forall b \in \mathbb{Q}^m \, , \quad \varphi(b) = \inf_{\substack{x \\ Ax = b \\ x \in \mathbb{Q}^n_+}} \langle x \, , \, k \rangle$$

2. We pair a primal space $\mathbb{Q}^m$ and a dual space $\mathbb{Q}^m$

$$\langle \cdot \, , \, \cdot \rangle : \mathbb{Q}^m \times \mathbb{Q}^m \to \mathbb{R}$$

3. We biconjugate the perturbation function φ

$$\underbrace{\varphi^{\star\star'}(b) \leq \varphi(b) \, , \quad \forall b \in \mathbb{Q}^m}_{\text{Weak duality is guaranteed!}}$$

4. Strong duality when φ is lsc convex

Introducing generalized convexity

[Balder, 1977]

Fenchel conjugate	c-conjugate
$f^{\star}(v) = \sup_{u \in \mathbb{R}^m} \langle u, v \rangle - f(u)$	$g^{c}(v) = \sup_{u \in U} c(u,v) \dot{+} (-g(u))$
Fenchel biconjugate	**c-biconjugate**
$f^{\star\star'}(u) = \sup_{v \in \mathbb{R}^m} \langle u, v \rangle - f^{\star}(v)$	$g^{cc'}(u) = \sup_{v \in V} c(u,v) \dot{+} (-g^{c}(v))$
lsc convex functions	**c-convex functions**
$\iff f = f^{\star\star'}$	$: \iff g = g^{cc'}$

Perturbation-duality scheme with generalized convexity

1. We perturb a minimization problem

$$\varphi : \mathbb{R}^m \to \overline{\mathbb{R}}$$

Perturbation-duality scheme with generalized convexity

1. We perturb a minimization problem

$$\varphi : \mathbb{R}^m \to \overline{\mathbb{R}}$$

2. We pair a primal space $\mathbb{R}^m$ and a dual space V

$$c : \mathbb{R}^m \times V \to \overline{\mathbb{R}}$$

Perturbation-duality scheme with generalized convexity

1. We perturb a minimization problem

$$\varphi : \mathbb{R}^m \to \overline{\mathbb{R}}$$

2. We pair a primal space $\mathbb{R}^m$ and a dual space V

$$c : \mathbb{R}^m \times V \to \overline{\mathbb{R}}$$

3. We biconjugate the perturbation function φ

$$\underbrace{\varphi^{cc'}(b) \leq \varphi(b) \; , \quad \forall b \in \mathbb{R}^m}_{\text{Weak duality is guaranteed!}}$$

Perturbation-duality scheme with generalized convexity

1. We perturb a minimization problem

$$\varphi : \mathbb{R}^m \to \overline{\mathbb{R}}$$

2. We pair a primal space $\mathbb{R}^m$ and a dual space V

$$c : \mathbb{R}^m \times V \to \overline{\mathbb{R}}$$

3. We biconjugate the perturbation function φ

$$\underbrace{\varphi^{cc'}(b) \leq \varphi(b) \ , \quad \forall b \in \mathbb{R}^m}_{\text{Weak duality is guaranteed!}}$$

4. Strong duality when φ is c-convex

Outline

Outline

Strong duality in LP

- Dual problem $\qquad$ "Primal" problem

$$\sup_{p} \langle p,\, b_0 \rangle \quad p^T A \le k \quad p \in \mathbb{Q}^m \qquad \underbrace{=}_{\text{strong duality}} \qquad \inf_{x} \langle x,\, k \rangle \quad Ax = b_0 \quad x \in \mathbb{Q}^n_+$$

- Complementary slackness

$$\widehat{x}_j \big(k_j - \widehat{p}^T A_j \big) = 0 \,, \;\; \forall j \in = 1, \dots, n$$

Strong duality in LP

- Dual problem

$$\sup_{p} \quad \langle p,\ b_0\rangle \qquad p^T A \le k \qquad p \in \mathbb{Q}^m$$

$$\underbrace{=}_{\text{strong duality}}$$

"Primal" problem

$$\inf_{x} \quad \langle x,\ k\rangle \qquad Ax = b_0 \qquad x \in \mathbb{Q}^n_+$$

- Complementary slackness

$$\widehat{x}_j\big(k_j - \widehat{p}^T A_j\big) = 0\ ,\ \ \forall j \in= 1, \ldots, n$$

Weak duality in PILP

- Dual problem

$$\sup_p \quad \langle p,\ b_0\rangle \qquad p^T A \le k \qquad p \in \mathbb{Q}^m$$

$\underbrace{\le}_{\text{weak duality}}$

"Primal" problem

$$\inf_x \quad \langle x,\ k\rangle \qquad Ax = b_0 \qquad x \in \mathbb{Z}^n_+$$

- Complementary slackness

???

Subadditive dual problem of Jeroslow

- [Jeroslow, 1979]

dual problem | "primal" problem

$$\sup_{F} \quad F(b_0) \qquad\qquad \inf_{x} \quad \langle x,\, k\rangle$$

$$F(A_j) \le k_j \qquad\qquad Ax = b_0$$

$$F(0) \le 0 \qquad\qquad x \in \mathbb{Z}_+^n$$

F is subadditive

$=$ (strong duality)

- Complementary slackness

$$x_j\big(k_j - F(a_j)\big) = 0\ ,\quad \forall j = 1, \ldots, n$$

$$\sum_{j=1}^{n} F(A_j)x_j = F(b_0)$$

Link between Jeroslow's result and perturbation-duality scheme?

Which scheme for PILP duality?

- We define a perturbation function $G : \mathbb{Q}^m \to \overline{\mathbb{R}}$

$$\forall b \in \mathbb{Q}^m \,, \quad G(b) = \inf_{\substack{x \\ Ax = b \\ x \in \mathbb{Z}^n_+}} \langle x, \, k \rangle$$

- We define a coupling between primal and dual space

$$c : \mathbb{Q}^m \times ?? \to \mathbb{R}$$

- We biconjugate the perturbation function

$$\underbrace{G^{cc'}(b) \leq G(b) \,, \quad \forall b \in \mathbb{Q}^m}_{\text{weak duality}}$$

Outline

Definition of Chvátal functions

Definition

The class of Chvátal functions $\mathcal{C}^m$
is the smallest class of functions $D \subset \{f | f : \mathbb{Q}^m \to \mathbb{Q}\}$ such that

$$b \in \mathbb{Q}^m \mapsto \lambda b \in D \; , \; \forall b \in \mathbb{Q}^m \qquad \text{(linear functions)}$$

$$\alpha F_1 + \beta F_2 \in D \; , \; \forall F_1, F_2 \in D \; , \; \alpha, \beta \in \mathbb{Q}_+ \qquad \text{(conic combination)}$$

$$\lceil F \rceil \in D \; , \; \forall F \in D \qquad \text{(round-up)}$$

Examples in 1D

- $b \mapsto \frac{3}{4} b$
- $b \mapsto \lceil b \rceil$
- $b \mapsto \frac{3}{4} b + \frac{7}{10} \lceil b \rceil$
- $b \mapsto 15b + \frac{39}{22} \left\lceil \frac{3}{4} b + \frac{7}{10} \lceil b \rceil \right\rceil + \lceil 16b \rceil$

Jeroslow's dual problem with Chvátal functions

Chvátal function class: $\mathcal{C}^m$

[Jeroslow, 1979] [Blair and Jeroslow, 1982]

$$\begin{array}{cc} \sup\limits_{F} & F(b_0) \\ F(A_j) \le k_j & \\ F(0) \le 0 & \\ F \text{ est sous-add.} & \end{array} \quad = \quad \begin{array}{cc} \sup\limits_{F} & F(b_0) \\ F(A_j) \le k_j & \\ F(0) \le 0 & \\ F \in \mathcal{C}^m & \end{array}$$

strong duality with initial PILP is achieved for both!

Outline

Chvátal perturbation-duality scheme

- We define a perturbation function

$$\forall b \in \mathbb{Q}^m \ , \quad G(b) = \inf_{\substack{x \\ Ax = b \\ x \in \mathbb{Z}^n_+}} \langle x, \ k \rangle$$

- We define a coupling between primal and dual space

$$c_{\mathcal{C}} : \mathbb{Q}^m \times \mathcal{C}^m \to \mathbb{R}$$
$$c_{\mathcal{C}}(b, F) = F(b) \ , \ \ \forall b \in \mathbb{Q}^m \ , \ \ \forall F \in \mathcal{C}^m$$

- We biconjugate the perturbation functions

$$\underbrace{G^{c_{\mathcal{C}} c_{\mathcal{C}}'}(b) \leq G(b) \ , \ \ \forall b \in \mathbb{Q}^m}_{\text{weak duality}}$$

- We get strong duality $G^{c_{\mathcal{C}} c_{\mathcal{C}}'}(b_0) = G(b_0)$

Obtained dual problems

Formulation 1:

$$G^{\mathcal{C}\mathcal{C}'\mathcal{C}\mathcal{C}'}(b_0) = \sup_{F \in \mathcal{C}^m} \left\{ F(b_0) + \inf_{b \in \mathbb{Q}^m} \{ G(b) - F(b) \} \right\}$$

Formulation 2:

$$G^{\mathcal{C}\mathcal{C}'\mathcal{C}\mathcal{C}'}(b_0) = \sup_{F \in \mathcal{C}^m} \left\{ F(b_0) + \inf_{x \in \mathbb{Z}^n_+} \{ \langle x, k \rangle - F(Ax) \} \right\}$$

Obtained dual problems

Formulation 1:

$$G^{c_{\mathcal{C}} c_{\mathcal{C}}'}(b_0) = \sup_{F \in \mathcal{C}^m} \left\{ F(b_0) + \inf_{b \in \mathbb{Q}^m} \{ G(b) - F(b) \} \right\}$$

Formulation 2:

$$G^{c_{\mathcal{C}} c_{\mathcal{C}}'}(b_0) = \sup_{F \in \mathcal{C}^m} \left\{ F(b_0) + \inf_{x \in \mathbb{Z}^n_+} \{ \langle x, k \rangle - F(Ax) \} \right\}$$

Reminder Jeroslow's dual problem

$$\begin{aligned} \sup_{F} \quad & F(b_0) \\ F(A_j) &\leq k_j \\ F(0) &\leq 0 \\ F &\in \mathcal{C}^m \end{aligned}$$

Generalized subdifferential and complementary slackness

Proposition

- G: bounded perturbation function of a MILP
- $A = \left(A_j\right)_{j=1,\dots,n} \in \mathbb{Q}^{m\times n}$ constraint matrix
- $b_0 \in \mathbb{Q}^n$ anchor

If $\hat{x} \in \{x \in \mathbb{Z}^n_+ | Ax = b_0\}$ and $\widehat{F} \in \mathcal{C}^m$ are "primal"-dual optimal solutions then we have the equivalence

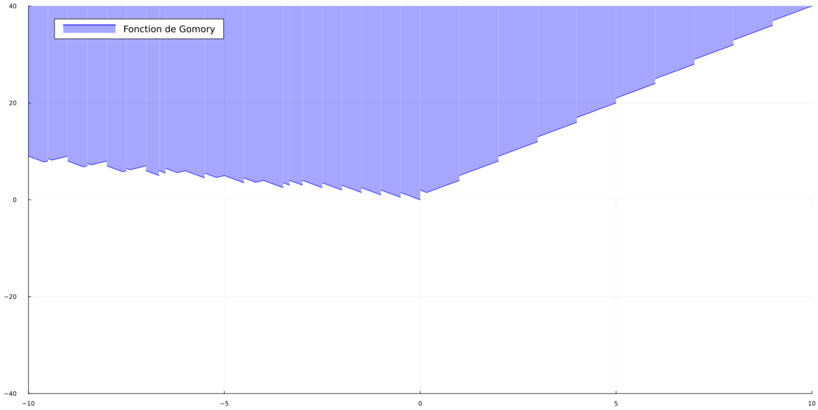
$$\widehat{F} \in \partial^{cc} G(b_0)$$
$$\iff -k \in \partial\left(-\widehat{F} \circ A \dot{+} \delta_{\mathbb{Z}^n_+}\right)(\hat{x})$$

Furthermore, if $\widehat{F}(A_j) \leq k_j, \forall j = 1, \dots, n$, then the following assertion is also equivalent

$$\widehat{F}(0) \leq 0 \;, \;\; \widehat{F}(b_0) = G(b_0) \text{ and } \left(k_j - \widehat{F}(A_j)\right)\hat{x}_j = 0 \;, \;\; \forall j = 1, \dots, n \;.$$

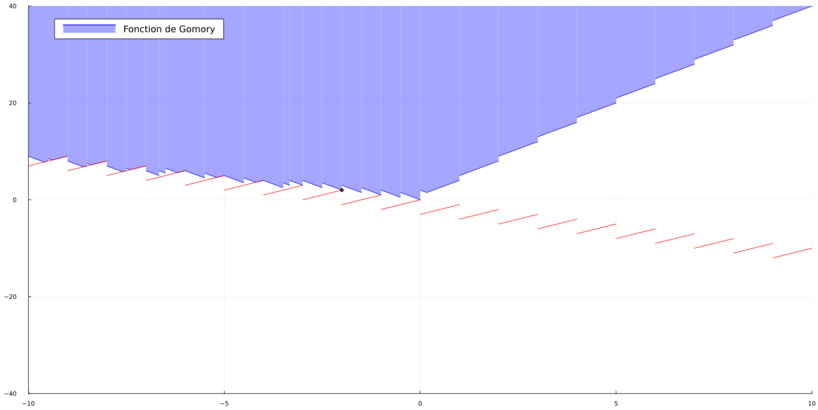
Epigraph of a perturbation function for a PILP

$$G(b) = \max\{3b + \lceil b \rceil, 2b + -3\lceil b \rceil, -3b + \lceil 2b \rceil + \lceil \tfrac{3}{10} b \rceil\}$$



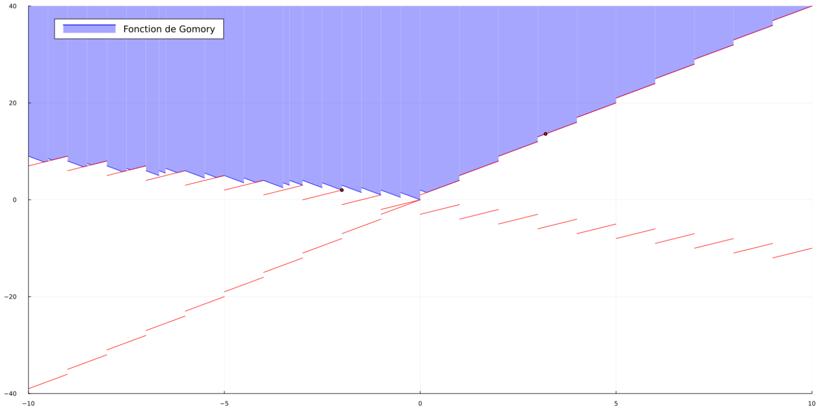
Epigraph of a perturbation function for a PILP

$$G(b) = \max\{3b + \lceil b \rceil, 2b + -3\lceil b \rceil, -3b + \lceil 2b \rceil + \lceil \tfrac{3}{10} b \rceil\}$$



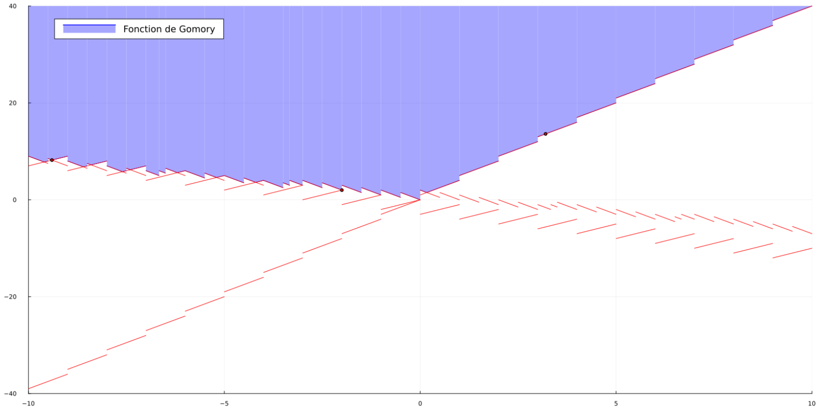
Epigraph of a perturbation function for a PILP

$$G(b) = \max\{3b + \lceil b \rceil, 2b + -3\lceil b \rceil, -3b + \lceil 2b \rceil + \lceil \tfrac{3}{10} b \rceil\}$$



Epigraph of a perturbation function for a PILP

$$G(b) = \max\{3b + \lceil b \rceil, 2b + -3\lceil b \rceil, -3b + \lceil 2b \rceil + \lceil \tfrac{3}{10} b \rceil\}$$



Outline

Limitations of Chvátal functions

- Solve the dual problem of Jeroslow: which algorithm? ([Klabjan, 2007])
- Expression of a Chvátal function $F \in \mathcal{C}^m$: no limit on the number of $\lceil \cdot \rceil$

Proposed relaxation: quasiaffine program

- Relaxation : considering a subclass of Chvátal functions

Example

$\alpha \in \mathbb{Q}_+$

$$\begin{aligned} \sup_{\lambda \in \mathbb{Q}^m} \quad & \langle \lambda, b_0 \rangle + \alpha \lceil \langle \lambda, b_0 \rangle \rceil \\ \langle \lambda, A_j \rangle + \alpha \lceil \langle \lambda, A_j \rangle \rceil \leq k_j \ , & \\ \forall j \in \{1, \ldots, n\} & \end{aligned} \tag{1}$$

- This program is quasiaffine! [Martínez-Legaz, 2005]

SECOND PART

Outline

ℓ_0 pseudonorm and sparse optimization

Definition

The pseudonorm $\ell_0 : \mathbb{R}^d \to \{0, \dots, d\}$

$$\boxed{\ell_0(x) = \#\text{nonnull components of } x} \ , \quad \forall x \in \mathbb{R}^d$$

- Examples: $\ell_0 \begin{pmatrix} 1 \\ 0 \\ -50 \end{pmatrix} = 2, \ \ell_0 \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} = 1, \ \ell_0 \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0.$
- Application in compressive sensing, image recovery, minimum description length

E-Capra conjugacy and E-Capra convex sets

The norm $\ell_2 : \mathbb{R}^d \to \mathbb{R}_+$ $\ell_2(x) = \sqrt{\sum_{i=1}^d x_i^2}$

Definition

Normalization mapping $n : \mathbb{R}^d \to \mathbb{R}^d$

$$\forall x \in \mathbb{R}^d \ , \ \ n(x) = \begin{cases} x/\ell_2(x), & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

Coupling Euclidean Constant Along PRimal RAy (E-CAPRA) $¢ : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$

$$¢(x, y) = \langle n(x), y \rangle \ , \ \ \forall x, y \in \mathbb{R}^d$$

[Chancelier and De Lara, 2022]

Proposition

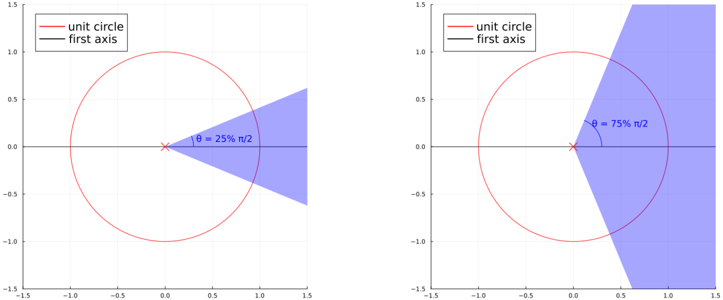
The pseudonorm ℓ_0 is E-*Capra* convex, meaning $\ell_0 = \ell_0^{¢¢'}$

Three considered problems

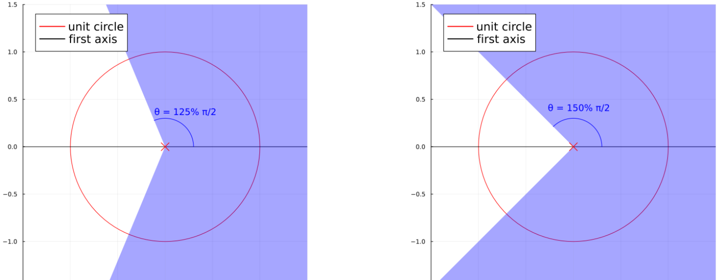
Problems	Min of norms ratio	Min of ℓ_0	Matrix spark
Objective fun.	ℓ_1/ℓ_2	ℓ_0	ℓ_0
E-Capra convex	✓	✓	✓
Feasible set	$\overline{\text{cone}}(g_1, \ldots g_r) \setminus \{0\}$	$\overline{\text{cone}}(g_1, \ldots g_r) \setminus \{0\}$	$\{x \in \mathbb{R}^d \setminus \{0\} : Ax = 0\}$
E-Capra convex	✓	✓	

Examples and counterexamples of E-Capra convex sets

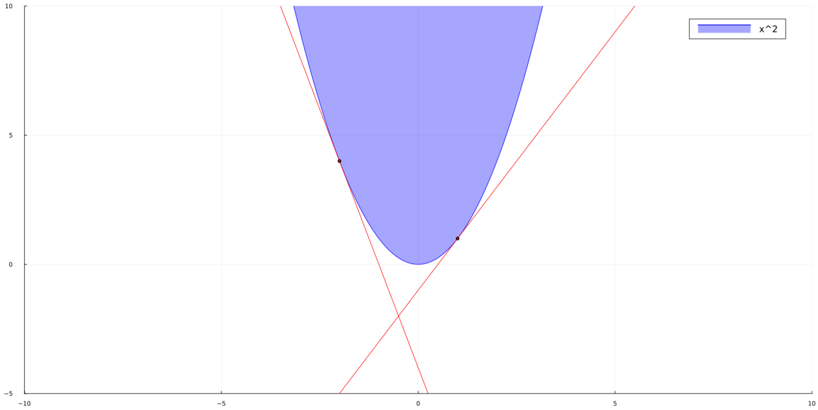
Examples



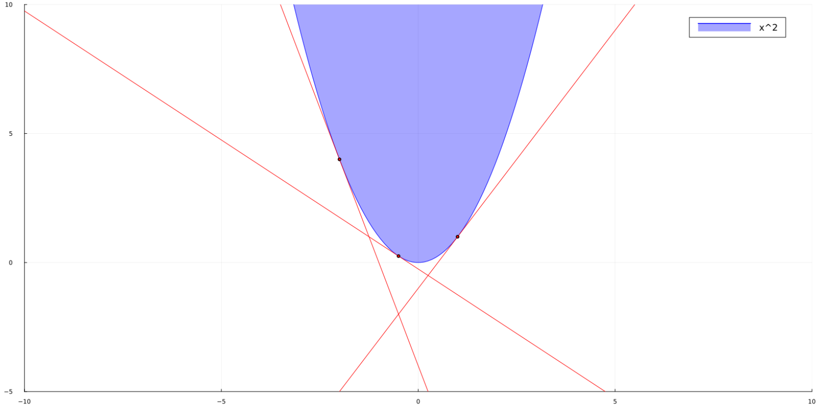
Counterexamples



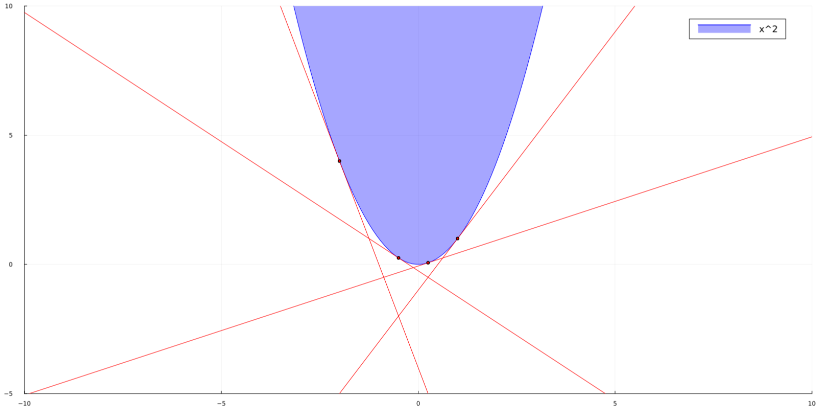
Cutting plane method in action



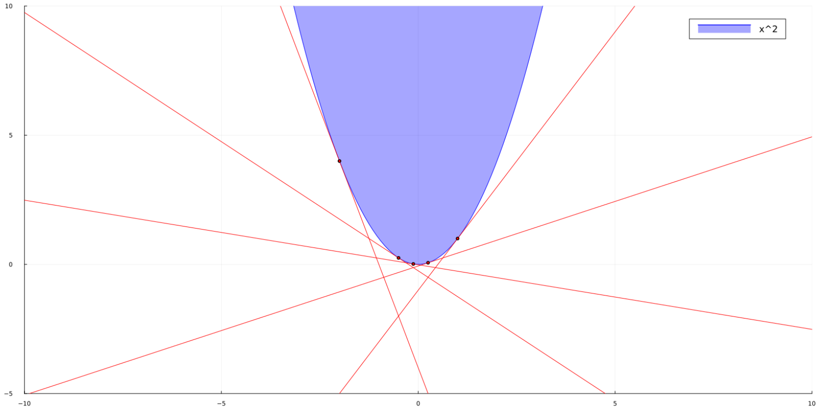
Cutting plane method in action



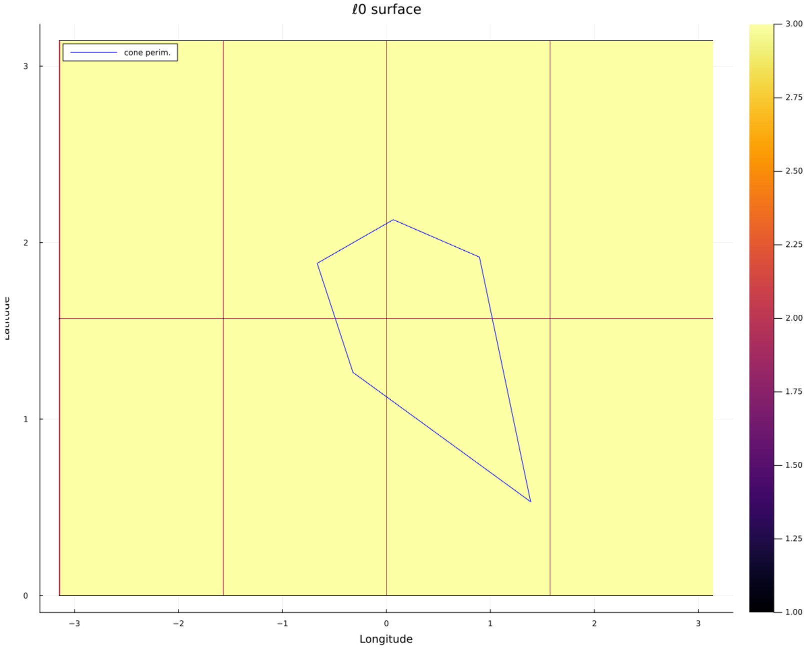
Cutting plane method in action



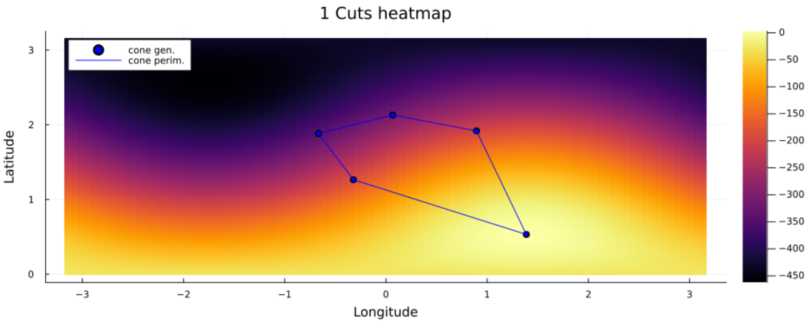
Cutting plane method in action

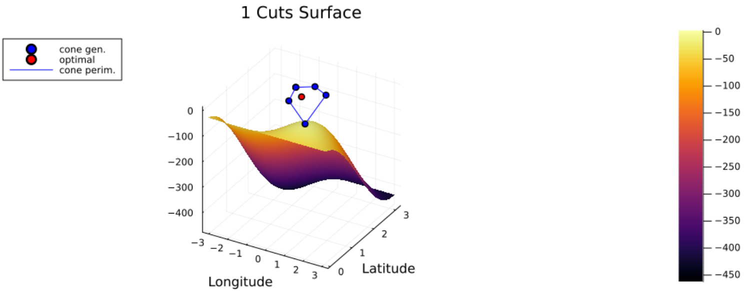


ℓ_0 graph on the sphere in $\mathbb{R}^3$

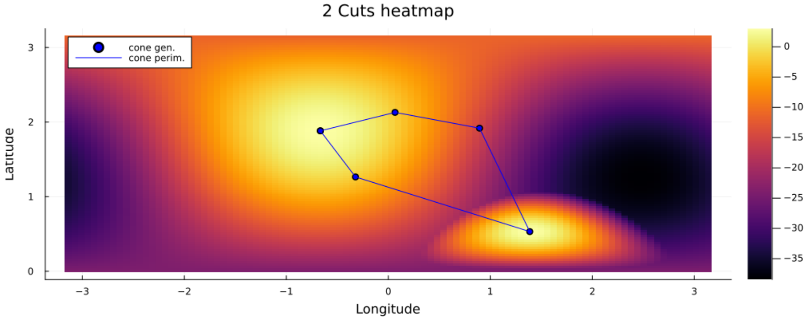


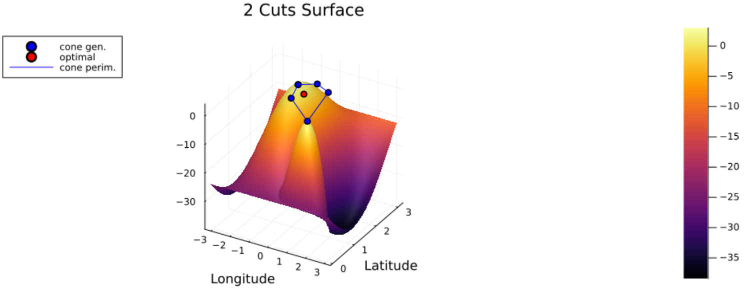
E-Capra cuts


1 Cuts heatmap
cone gen.
cone perim.
Latitude
Longitude
3
2
1
0
−3
−2
−1
0
1
2
3
0
−50
−100
−150
−200
−250
−300
−350
−400
−450

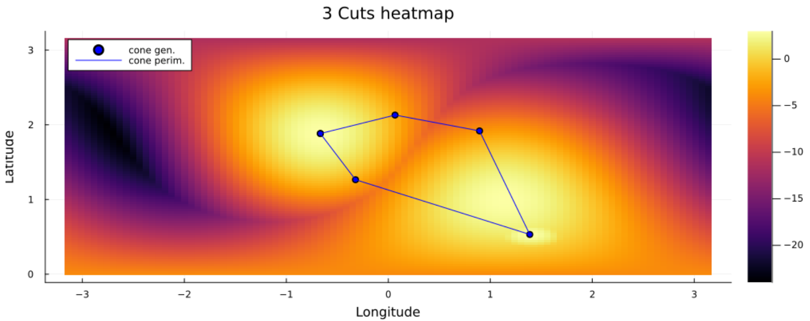

1 Cuts Surface
cone gen.
optimal
cone perim.
0
−100
−200
−300
−400
Longitude
Latitude
0
−50
−100
−150
−200
−250
−300
−350
−400
−450

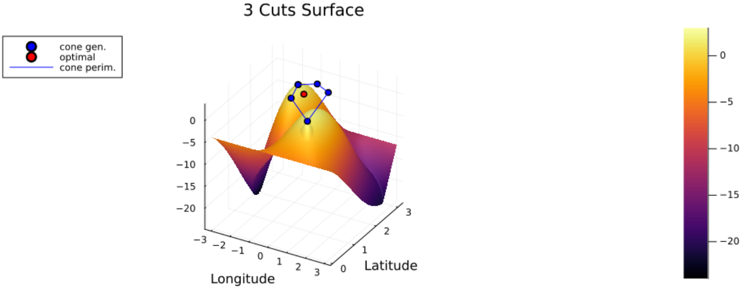
E-Capra cuts


2 Cuts heatmap
cone gen.
cone perim.
Latitude
Longitude
3
2
1
0
−3
−2
−1
0
1
2
3
0
−5
−10
−15
−20
−25
−30
−35

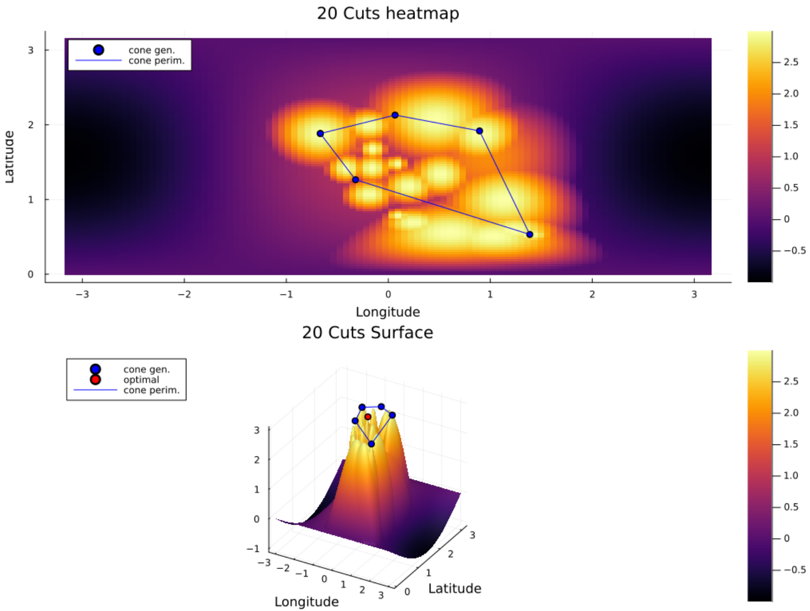

2 Cuts Surface
cone gen.
optimal
cone perim.
0
−10
−20
−30
Longitude
Latitude
0
−5
−10
−15
−20
−25
−30
−35

E-Capra cuts


3 Cuts heatmap
cone gen.
cone perim.
Latitude
Longitude
3
2
1
0
−3
−2
−1
0
1
2
3
0
−5
−10
−15
−20

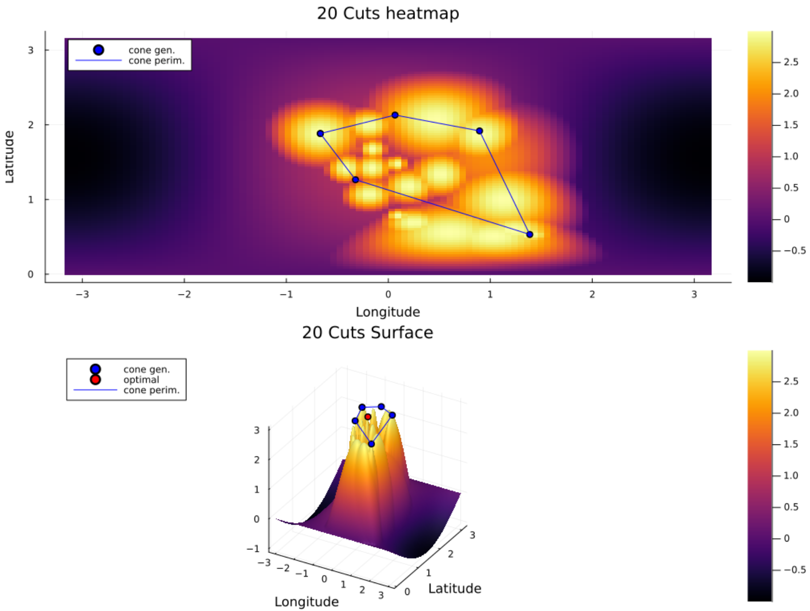

3 Cuts Surface
cone gen.
optimal
cone perim.
0
−5
−10
−15
−20
Longitude
Latitude
0
−5
−10
−15
−20

E-Capra cuts



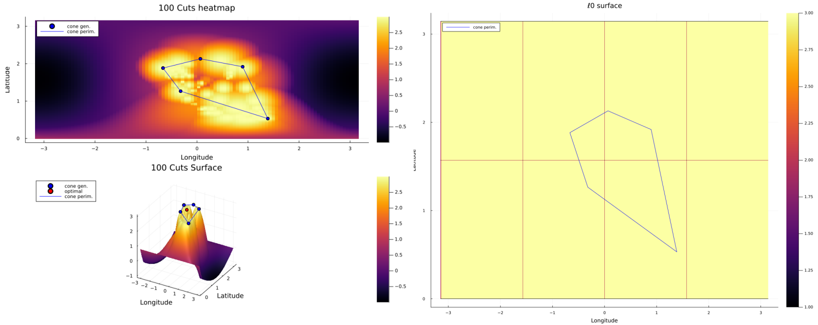
20 Cuts heatmap
cone gen.
cone perim.
Latitude
Longitude
20 Cuts Surface
cone gen.
optimal
cone perim.
Longitude
Latitude

E-Capra cuts



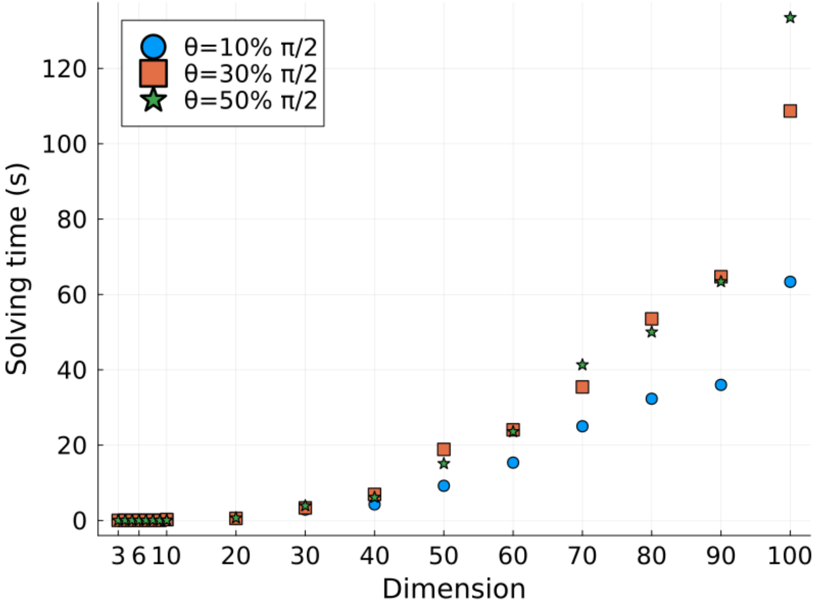
20 Cuts heatmap
cone gen.
cone perim.
Latitude
Longitude
20 Cuts Surface
cone gen.
optimal
cone perim.
Longitude
Latitude

E-Capra cuts



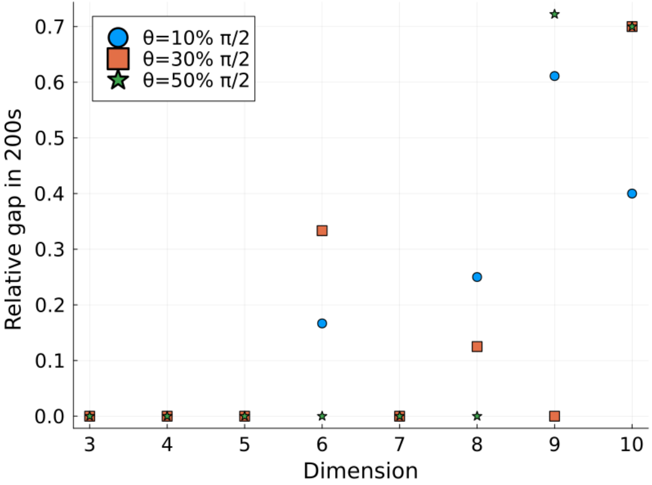
100 Cuts heatmap
cone gen.
cone perim.
Latitude
Longitude
100 Cuts Surface
cone gen.
optimal
cone perim.
Longitude
Latitude
ℓ0 surface
cone perim.
Longitude

Solving time for the ratio of norms

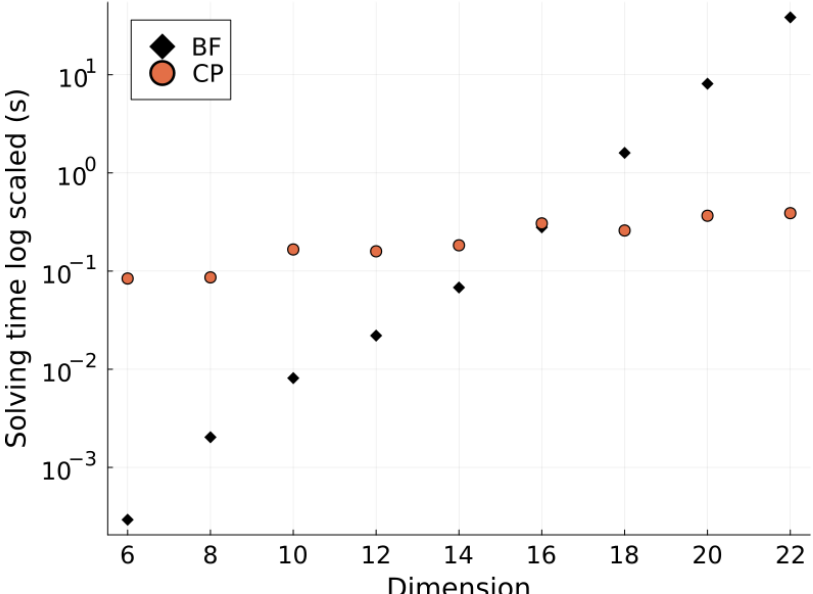


Relative gap for ℓ_0 minimization

$$\text{Relative gap} = \frac{\text{best found value} - \text{optimal value}}{\text{dimension}}$$



Solving time comparison between Brute Force and cutting plane



Outline

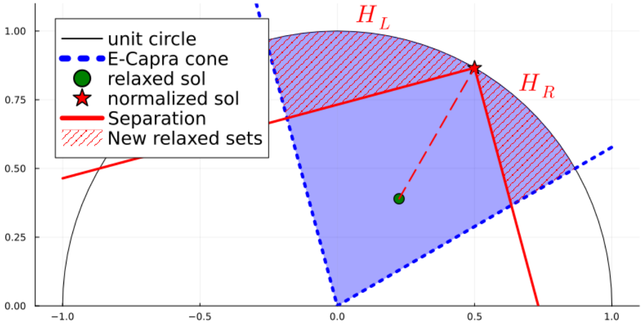
Generalized convexity for PILP

- Clarification on the notion of dual problems
- A new dual problem for PILP
- Sensitivity analysis in PILP [Wolsey, 1981]

Abstract convex methods for generalized convexity

- Cutting plane method $\leftrightarrow$ Gomory cutting plane method
- Other : Branch-and-Bound, Tabu search, variants with local search [Rubinov, 2000]

Spatial branch-and-bound



Thank you for your attention!

Bibliographie

E. J. Balder. An extension of duality-stability relations to nonconvex optimization problems. *SIAM Journal on Control and Optimization*, 15(2):329–343, 1977.

Charles E Blair and Robert G Jeroslow. The value function of an integer program. *Mathematical programming*, 23(1):237–273, 1982.

Jean-Philippe Chancelier and Michel De Lara. Variational formulations for the l_0 pseudonorm and applications to sparse optimization. Preprint hal-02459688, arXiv:2002.01314, 2020.

Jean-Philippe Chancelier and Michel De Lara. Hidden convexity in the l_0 pseudonorm. *Journal of Convex Analysis*, 28(1):203–236, 2021.

Jean-Philippe Chancelier and Michel De Lara. Capra-convexity, convex factorization and variational formulations for the l_0 pseudonorm. *Set-Valued and Variational Analysis*, 30:597–619, 2022.

Robert G Jeroslow. Minimal inequalities. *Mathematical programming*, 17(1):1–15, 1979.

Diego Klabjan. Subadditive approaches in integer programming. *European Journal of Operational Research*, 183(2):525–545, 2007. ISSN 0377-2217. doi: https://doi.org/10.1016/j.ejor.2006.10.009. URL https://www.sciencedirect.com/science/article/pii/S0377221706010423.

Adrien Le Franc. *Subdifferentiability in convex and stochastic optimization applied to renewable power systems*. Theses, École des Ponts ParisTech, December 2021. URL https://pastel.archives-ouvertes.fr/tel-03657075.

Adrien Le Franc, Jean-Philippe Chancelier, and Michel De Lara. The capra-subdifferential of the l_0 pseudonorm. *Optimization*, pages 1–23, 2022. doi: 10.1080/02331934.2022.2145172. accepted for publication.

J. E. Martínez-Legaz. Generalized convex duality and its economic applications. In Schaible S. Hadjisavvas N., Komlósi S., editor, *Handbook of Generalized Convexity and Generalized Monotonicity. Nonconvex Optimization and Its Applications*, volume 76, pages 237–292. Springer-Verlag, 2005.

Juan-Enrique Martinez-Legaz and Ivan Singer. Subdifferentials with respect to dualities. *Mathematical Methods of Operations Research*, 42(1):109–125, February 1995.

Diethard Pallaschke and Stefan Rolewicz. *Foundations of mathematical optimization*, volume 388 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1997. ISBN 0-7923-4424-3.

R. Tyrrell Rockafellar. *Conjugate Duality and Optimization*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1974.

Alexander Rubinov. *Abstract convexity and global optimization*, volume 44 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 2000. ISBN 0-7923-6323-X.

Andreas M. Tillmann and Marc E. Pfetsch. The computational complexity of the restricted isometry

Outline

Outline

Moreau lower and upper additions

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\} = [-\infty, +\infty]$$

Moreau lower and upper additions extend the usual addition with

$$(+\infty) \,\underset{\cdot}{+}\, (-\infty) = (-\infty) \,\underset{\cdot}{+}\, (+\infty) = -\infty$$
$$(+\infty) \,\dot{+}\, (-\infty) = (-\infty) \,\dot{+}\, (+\infty) = +\infty$$

Coupling

Background on couplings and Fenchel-Moreau conjugacies

Definition

Two vector spaces $\mathbb{X}$ and $\mathbb{Y}$, paired by a bilinear form $\langle \, , \rangle$ give rise to the classic Fenchel conjugacy

$$f \in \overline{\mathbb{R}}^{\mathbb{X}} \mapsto f^{\star} \in \overline{\mathbb{R}}^{\mathbb{Y}}$$

$$f^{\star}(y) = \sup_{x \in \mathbb{X}} \Big(\langle x , y \rangle + \big(-f(x) \big) \Big) \; , \quad \forall y \in \mathbb{Y}$$

- Let be given two sets $\mathbb{X}$ ("primal") and $\mathbb{Y}$ ("dual") not necessarily paired vector spaces (nodes and arcs, etc.)
- We consider a coupling function

$$c : \mathbb{X} \times \mathbb{Y} \to \overline{\mathbb{R}}$$

We also use the notation $\mathbb{X} \overset{c}{\leftrightarrow} \mathbb{Y}$ for a coupling

[Martínez-Legaz, 2005]

Conjugacy

Fenchel-Moreau conjugate and biconjugate

$$f \in \overline{\mathbb{R}}^{\mathbb{X}} \mapsto f^c \in \overline{\mathbb{R}}^{\mathbb{Y}}$$

Definition

The c-Fenchel-Moreau conjugate of a function $f : \mathbb{X} \to \overline{\mathbb{R}}$, with respect to the coupling c, is the function $f^c : \mathbb{Y} \to \overline{\mathbb{R}}$ defined by

$$f^c(y) = \sup_{x \in \mathbb{X}} \Big(c(x,y) \dotplus \big(-f(x)\big) \Big) \ , \quad \forall y \in \mathbb{Y}$$

The c-Fenchel-Moreau biconjugate $f^{cc'} : \mathbb{X} \to \overline{\mathbb{R}}$ is given by

$$f^{cc'}(x) = \big(f^c\big)^{c'}(x) = \sup_{y \in \mathbb{Y}} \Big(c(x,y) \dotplus \big(-f^c(y)\big) \Big) \ , \quad \forall x \in \mathbb{X}$$

Fenchel-Moreau biconjugate

With the coupling c, we associate the reverse coupling c'

$$c' : \mathbb{Y} \times \mathbb{X} \to \overline{\mathbb{R}} \,, \quad c'(y,x) = c(x,y) \,, \quad \forall (y,x) \in \mathbb{Y} \times \mathbb{X}$$

- The c'-Fenchel-Moreau conjugate of a function $g : \mathbb{Y} \to \overline{\mathbb{R}}$, with respect to the coupling c', is the function $g^{c'} : \mathbb{X} \to \overline{\mathbb{R}}$

$$g^{c'}(x) = \sup_{y \in \mathbb{Y}} \Big(c(x,y) \dot{+} \big(-g(y)\big) \Big) \,, \quad \forall x \in \mathbb{X}$$

- The c-Fenchel-Moreau biconjugate $f^{cc'} : \mathbb{X} \to \overline{\mathbb{R}}$ of a function $f : \mathbb{X} \to \overline{\mathbb{R}}$ is given by

$$f^{cc'}(x) = \big(f^c\big)^{c'}(x) = \sup_{y \in \mathbb{Y}} \Big(c(x,y) \dot{+} \big(-f^c(y)\big) \Big) \,, \quad \forall x \in \mathbb{X}$$

So-called c-convex functions have dual representations

For any function $f : \mathbb{X} \to \overline{\mathbb{R}}$, one has that

$$\boxed{f^{cc'} \leq f}$$

Definition

The function $f : \mathbb{X} \to \overline{\mathbb{R}}$ is c-convex if $f^{cc'} = f$

If the function $f : \mathbb{X} \to \overline{\mathbb{R}}$ is c-convex, we have

$$f(x) = \sup_{y \in \mathbb{Y}} \underbrace{\Big(c(x, y) \dotplus \big(-f^c(y)\big) \Big)}_{\text{elementary function of } x} , \quad \forall x \in \mathbb{X}$$

Example: $\star$-convex functions
= closed convex functions [Rockafellar, 1974, p. 15]
= proper convex lsc or $\equiv -\infty$ or $\equiv +\infty$
= suprema of affine functions

Subdifferential

Subdifferential(s) $\partial^c f, \partial_c f, \partial_c^c f : \mathbb{X} \rightrightarrows \mathbb{Y}$ of a conjugacy

For any function $f : \mathbb{X} \to \overline{\mathbb{R}}$ and $x \in \mathbb{X}$, $y \in \mathbb{Y}$,

Definition

Upper subdifferential (following Martinez-Legaz and Singer [1995])

$$y \in \partial^c f(x) \iff f(x) = c(x,y) \overset{\cdot}{+} \left(-f^c(y)\right)$$

Middle subdifferential ("à la Fenchel-Young")

$$y \in \partial_c^c f(x) \iff f(x) \dot{+} f^c(y) = c(x,y)$$

Lower subdifferential ("à la Rockafellar-Moreau")

$$y \in \partial_c f(x) \iff f^c(y) = c(x,y) \overset{\cdot}{+} \left(-f(x)\right)$$

Properties of subdifferentials

- The upper subdifferential $\partial^c f$ has the property that

$$\partial^c f(x) \neq \emptyset \Rightarrow \underbrace{f^{cc'}(x) = f(x)}_{\text{the function } f \text{ is } c\text{-convex at } x}$$

- The lower subdifferential $\partial_c f$ is characterized by

$$\begin{aligned} y \in \partial_c f(x) &\iff x \in \mathop{\arg\max}_{x' \in \mathbb{X}} \left[c(x', y) \dotplus (-f(x')) \right] \\ &\iff c(x', y) \dotplus (-f(x')) \\ &\qquad \leq c(x, y) \dotplus (-f(x)) \,, \quad \forall x' \in \mathbb{X} \end{aligned}$$

- All definitions coincide when
$-\infty < c < +\infty$ and $-\infty < f(x) < +\infty$

Outline

Dual problems: perturbation scheme [Rockafellar, 1974]

- Set $\mathbb{W}$, function $h : \mathbb{W} \to \overline{\mathbb{R}}$
  and original minimization problem

$$\inf_{w \in \mathbb{W}} h(w)$$

- Embedding/perturbation scheme given by
  a nonempty set $\mathbb{X}$ (perturbations), an element $\overline{x} \in \mathbb{X}$ (anchor)
  and a function (Rockafellian) $\mathcal{R} : \mathbb{W} \times \mathbb{X} \to \overline{\mathbb{R}}$ such that

$$h(w) = \mathcal{R}(w, \overline{x})$$

- Perturbation function

$$\phi(x) = \inf_{w \in \mathbb{W}} \mathcal{R}(w, x)$$

- Original minimization problem

$$\phi(\overline{x}) = \inf_{w \in \mathbb{W}} \mathcal{R}(w, \overline{x}) = \inf_{w \in \mathbb{W}} h(w)$$

Dual problems: conjugacy, weak and strong duality

- Coupling $\mathbb{X} \overset{c}{\leftrightarrow} \mathbb{Y}$, and Lagrangian $\mathcal{L} : \mathbb{W} \times \mathbb{Y} \to \overline{\mathbb{R}}$ given by

$$\mathcal{L}(w, y) = \inf_{x \in \mathbb{X}} \Big\{ \mathcal{R}(w, x) \dotplus \big(-c(x, y)\big) \Big\}$$

- Dual function

$$\psi(y) = -\phi^{c}(y) = \inf_{w \in \mathbb{W}} \mathcal{L}(w, y)$$

- Dual maximization problem (weak duality)

$$\phi^{cc'}(\overline{x}) = \sup_{y \in \mathbb{Y}} \big\{ c(\overline{x}, y) \dotplus \psi(y) \big\} \leq \inf_{w \in \mathbb{W}} h(w) = \phi(\overline{x})$$

- Strong duality holds true when ϕ is c-convex at $\overline{x}$, that is,

$$\phi^{cc'}(\overline{x}) = \sup_{y \in \mathbb{Y}} \big\{ c(\overline{x}, y) \dotplus \psi(y) \big\} = \inf_{w \in \mathbb{W}} h(w) = \phi(\overline{x})$$

Dual problems: perturbation scheme [Rockafellar, 1974]

sets	optimization set $\mathbb{W}$	primal set $\mathbb{X}$	coupling $\mathbb{X} \overset{c}{\leftrightarrow} \mathbb{Y}$	dual set $\mathbb{Y}$
variables	decision $w \in \mathbb{W}$	perturbation $x \in \mathbb{X}$	$c(x, y) \in \overline{\mathbb{R}}$	sensitivity $y \in \mathbb{Y}$
bivariate functions		Rockafellian $\mathcal{R} : \mathbb{W} \times \mathbb{X} \to \overline{\mathbb{R}}$		Lagrangian $\mathcal{L} : \mathbb{W} \times \mathbb{Y} \to \overline{\mathbb{R}}$
definition				$\mathcal{L}(w, y) = \inf_{x\in\mathbb{X}} \left\{\mathcal{R}(w, x) \dotplus (-c(x, y))\right\}$
property				$-\mathcal{L}(w, \cdot) = \big(\mathcal{R}(w, \cdot)\big)^{c}$
property				$-\mathcal{L}(w, \cdot)$ is c'-convex
univariate functions		perturbation function $\phi : \mathbb{X} \to \overline{\mathbb{R}}$		dual function $\psi : \mathbb{Y} \to \overline{\mathbb{R}}$
definition		$\phi(x) = \inf_{w\in\mathbb{W}} \mathcal{R}(w, x)$		$\psi(y) = \inf_{w\in\mathbb{W}} \mathcal{L}(w, y)$
property				$-\psi = \phi^{c}$

Anchor $\overline{x} \in \mathbb{X}$ and dual maximization problem (weak duality)
$\phi^{cc'}(\overline{x}) = \sup_{y\in\mathbb{Y}} \left\{c(\overline{x}, y) \dotplus \psi(y)\right\} \leq \inf_{w\in\mathbb{W}} h(w) = \phi(\overline{x})$
Strong duality iff ϕ is c-convex at $\overline{x}$ iff $\phi^{cc'}(\overline{x}) = \phi(\overline{x})$

Outline

Abstract cutting plane method

[Rubinov, 2000, §9.2.3]

Definition

Let $\mathbb{W}$ be a set, $H \subset \overline{\mathbb{R}}^{\mathbb{W}}$ be a set of elementary functions, and $f : \mathbb{W} \to \overline{\mathbb{R}}$ be a H-convex function

1. Set $k := 0$. Choose an arbitrary initial point $w_0 \in \mathbb{W}$
2. Calculate an abstract subgradient $h_k \in \partial^H f(w_k)$
   Let $f_{-1} = -\infty$ and
   $$f_k = \max\{f_{k-1}, \underbrace{h_k}_{\text{new cut}}\}$$
3. Calculate an optimal solution $\widehat{w} \in \arg\min_{w \in \mathbb{W}} f_k(w)$
4. Set $k := k + 1$, $w_k = \widehat{w}$
   Repeat from Step 2 until a stop condition is satisfied

Abstract cutting plane method: convergence result

[Pallaschke and Rolewicz, 1997, Theorem 9.1.1]

Theorem

Let

- $(\mathbb{W}, d)$ be a metric space
- H be a family of real-valued locally uniform continuous functions $h : \mathbb{W} \to \mathbb{R}$,
- $f : \mathbb{W} \to \overline{\mathbb{R}}$ be a continuous H-convex function

Then, all accumulation points of the sequence $\{w_k\}_{k \in \mathbb{N}}$ generated by the abstract cutting plane method are minimizers of the function f

Outline

E-Capra conjugacy

The $\ell_2 : \mathbb{R}^d \to \mathbb{R}_+$ norm is defined by $\ell_2(x) = \sqrt{\sum_{i=1}^d x_i^2}$

Definition

Let $n : \mathbb{R}^d \to \mathbb{R}^d$ be the normalization mapping given by

$$\forall x \in \mathbb{R}^d \ , \quad n(x) = \begin{cases} x/\ell_2(x), & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

We define the Euclidean Constant Along PRimal RAy (E-CAPRA) coupling $¢ : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by

$$¢(x, y) = \langle n(x), y \rangle \ , \quad \forall x, y \in \mathbb{R}^d$$

Definition and characterization of E-Capra convex sets

Definition

We say that the set $D \subset \mathbb{R}^d$ is E-*Capra* convex if $\iota_D = \iota_D^{¢¢'}$ meaning the indicator function ι_D is a E-*Capra* convex function

[Le Franc, 2021, Proposition 6.2.6]

Proposition

Let $D \subseteq \mathbb{R}^d$ be a nonempty set

$$D \text{ is E-}\textit{Capra}\text{ convex} \iff \begin{cases} D \text{ is a cone,} \\ D \cup \{0\} \text{ is closed,} \\ D \cap \{0\} = \overline{\mathrm{co}}\big(n(D)\big) \cap \{0\} \end{cases}$$

where $\overline{\mathrm{co}}$ is the closed convex hull

The ℓ_0 pseudonorm is ~~not~~ a ~~norm~~

Let $d \in \mathbb{N}^*$

$$\boxed{\ell_0(x) = \sum_{i=1}^{d} 1_{\{x_i \neq 0\}}} \;, \quad \forall x \in \mathbb{R}^d$$

- The pseudonorm $\ell_0 : \mathbb{R}^d \to [\![0, d]\!] = \{0, 1, \ldots, d\}$ satisfies 3 out of 4 axioms of a norm
  - we have $\ell_0(x) \geq 0$ ✓
  - we have $\Big(\ell_0(x) = 0 \iff x = 0 \Big)$ ✓
  - we have $\ell_0(x + x') \leq \ell_0(x) + \ell_0(x')$ ✓
  - But... <u>0-homogeneity</u> holds true
$$\ell_0(\rho x) = \ell_0(x) \;, \quad \forall \rho \neq 0$$
- We denote the level sets of the ℓ_0 pseudonorm by
$$\ell_0^{\leq k} = \left\{ x \in \mathbb{R}^d \,\middle|\, \ell_0(x) \leq k \right\} \;, \quad \forall k \in [\![0, d]\!]$$

E-*Capra* subdifferential of pseudonorm ℓ_0

[Chancelier and De Lara, 2022]

Proposition

The pseudonorm ℓ_0 is E-*Capra* convex, meaning $\ell_0 = \ell_0^{¢¢'}$

[Le Franc, Chancelier, and De Lara, 2022]

Proposition

Let $x \in \mathbb{R}^d \setminus \{0\}$ and $\mathrm{supp}(x) = \Big\{ i \in \{1, \ldots, d\} \Big| x_i \neq 0 \Big\}$
For $y \in \mathbb{R}^d$, let the permutation $\nu : \{1, \ldots, d\} \to \{1, \ldots, d\}$be such that $|y_{\nu(1)}| \geq \cdots \geq |y_{\nu(n)}|$

$$
y \in \partial_{¢} \ell_0(x) \iff \begin{cases} \exists \lambda \in \mathbb{R}_+ \ , \ y_i = & \lambda x_i \ , \ \forall i \in \mathrm{supp}(x) \ , \\ |y_j| \leq & \min_{i \in \mathrm{supp}(x)} |y_i| \ , \ \forall j \notin \mathrm{supp}(x) \ , \\ |y_{\nu(k+1)}|^2 \geq & (\|y\|^{\mathrm{tn}}_{k,2} + 1)^2 - (\|y\|^{\mathrm{tn}}_{k,2})^2 \ , \\ & \forall k \in \{0, \ldots, \ell_0(x) - 1\} \ , \\ |y_{\nu(\ell_0(x)+1)}|^2 \leq & (\|y\|^{\mathrm{tn}}_{\ell_0(x),2} + 1)^2 - (\|y\|^{\mathrm{tn}}_{\ell_0(x),2})^2 \ , \\ & \forall k \in \{0, \ldots, \ell_0(x) - 1\} \ . \end{cases}
$$

E-*Capra* subdifferential ratio ℓ_1/ℓ_2

Proposition

We define $\frac{\ell_1}{\ell_2}(0) = 0$

Then, the function $\frac{\ell_1}{\ell_2}$ is E-*Capra* convex, meaning $\frac{\ell_1}{\ell_2} = \left(\frac{\ell_1}{\ell_2}\right)^{¢¢'}$

Proposition

For any $x \in \mathbb{R}^d$, we have that

$$y \in \partial_{¢}(\frac{\ell_1}{\ell_2})(x) \iff y = \mathrm{sign}(x)$$

where the sign function $\mathrm{sign} : \mathbb{R}^d \to \{-1, 0, 1\}^d$ is defined by

$$\forall x \in \mathbb{R}^d \ , \ \ \mathrm{sign}(x) = \begin{cases} -1, & \text{if } x_i < 0 \\ 0, & \text{if } x_i = 0 \\ 1, & \text{if } x_i > 0 \end{cases}$$

Three problems with E-Capra convex objective function

- $\overline{\text{cone}}$ is the closed convex conical hull

Problems	Min of the ratio of two norms	Min of ℓ_0	Spark of a matrix
Objective function	ℓ_1/ℓ_2	ℓ_0	ℓ_0
Objective E-Capra convex	✓	✓	✓
Feasible set	$\overline{\text{cone}}(g_1, \ldots g_r) \setminus \{0\}$	$\overline{\text{cone}}(g_1, \ldots g_r) \setminus \{0\}$	$\{x \in \mathbb{R}^d \setminus \{0\} : Ax = 0\}$
Feasible set E-Capra convex	✓	✓	

- The cone generators $\{g_1, \ldots g_r\} \subset \mathbb{R}^d$ are such that

$$0 \notin \overline{\text{co}}\Big(n\big(\overline{\text{cone}}(g_1, \ldots g_r) \setminus \{0\}\big)\Big)$$

So $\overline{\text{cone}}(g_1, \ldots g_r) \setminus \{0\}$ is a E-Capra convex set

- The set $\{x \in \mathbb{R}^d \setminus \{0\} : Ax = 0\}$ is not E-Capra convex when the matrix A is singular

Three problems with E-Capra convex objective function

Problems	Min of the ratio of two norms	Min of ℓ_0	Spark of a matrix
Objective function	ℓ_1/ℓ_2	ℓ_0	ℓ_0
Objective E-Capra convex	✓	✓	✓
Feasible set	$\overline{\mathrm{cone}}(g_1, \ldots g_r) \setminus \{0\}$	$\overline{\mathrm{cone}}(g_1, \ldots g_r) \setminus \{0\}$	$\{x \in \mathbb{R}^d \setminus \{0\} : Ax = 0\}$
Feasible set E-Capra convex	✓	✓	

- Minimization of ℓ_1/ℓ_2:
  toy example which satisfies the convergence assumptions[Pallaschke and Rolewicz, 1997, Theorem 9.1.1]
- Minimization of the pseudonorm ℓ_0 on a cone without 0:
  more realistic, does not satisfy the convergence assumptions (ℓ_0 not continuous)
- Computation of the spark of a matrix:
  'semi' E-Capra convex problem
  useful in compress sensing [Tillmann and Pfetsch, 2014]

Reformulation as a minimization program on the sphere

Proposition

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a E-Capra convex function,
and let $K \subset \mathbb{R}^d$ be a E-Capra convex set

Then, the problem

$$\inf_{x \in K \setminus \{0\}} f(x)$$

has the same value than

$$\inf_{\substack{x \in K \\ \ell_2(x) = 1}} f(x)$$

and their solutions are the same up to normalization by the norm ℓ_2

E-Capra cutting plane method

Definition

Let $K = \mathrm{cone}(g_1, \ldots, g_r) \subset \mathbb{R}^n$ be an E-Capra convex cone
Let $f : \mathbb{R}^n \to \mathbb{R}$ be an E-Capra convex function
We call the following algorithm the E-Capra cutting plane method

1. Set $k := 0$. Find $x_0 \in K$ such that $\ell_2(x_0) = 1$
2. Calculate an E-Capra subgradient $y^k \in \partial^{¢} f(x^k)$
   Let $f_{-1} = -\infty$ and
   $$f_k = \max\{f_{k-1}, \underbrace{\langle \cdot, y_k \rangle - f^{¢}(y_k)\langle \cdot, y^k \rangle - f^{¢}(y^k)}_{\text{new cut}}\}$$
3. Calculate an optimal solution $\widehat{x} \in \arg\min_{\substack{x \in K \\ \ell_2(x)=1}} f_k(x)$
4. Set $k := k + 1$, $x_k = \widehat{x}$
   Repeat from Step 2 until a stop condition is satisfied

Difficulties with the E-Capra cutting plane method

- The norms of subgradients explode

$$\ell_2(y^k) \xrightarrow[k\to\infty]{} \infty$$

$\to$ Solution: project x^k on the i-th axis when $|x_i^k| \approx 0$ before computing $y^k \in \partial^{¢} f(x^k)$

- The sphere constraint $\ell_2(x) = 1$ is not a convex constraint

$\to$ Solution: use a nonlinear solver (here `IpOpt`) and add the constraint $\ell_1(x) \leq \underbrace{\bar{\ell}_0^k}_{\text{minimal known value of } \ell_0 \text{ at step } k}$ to the subproblem

E-Capra cutting plane method with local search

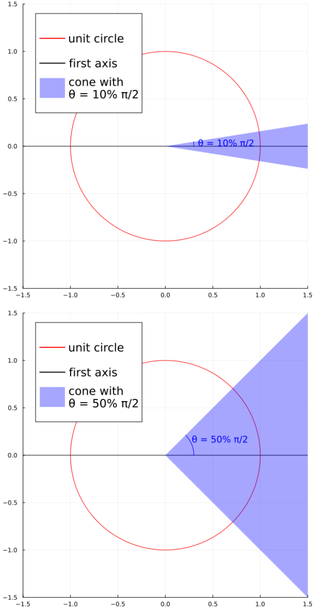
Definition

We call the following algorithm the E-Capra cutting plane method with local search for the pseudonorm ℓ_0

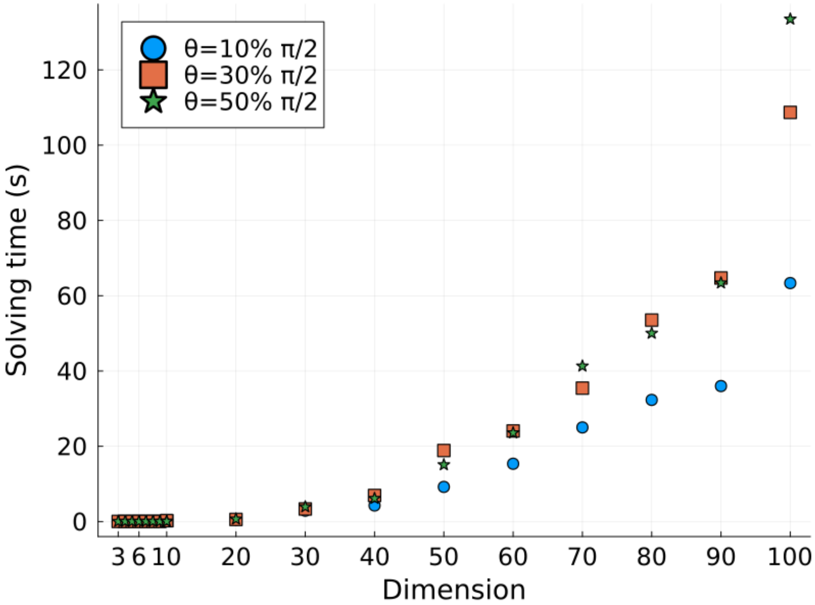
1. Set a threshold $\varepsilon > 0$ Set $k := 0$ Set the upper bound $\overline{\ell}_0^k = n$
   Find $x_0 \in K$ such that $\ell_2(x_0) = 1$
2. For each $i \in \{1, \ldots, n\}$, if $|x_i^k| < \varepsilon$, set $x_i^k := 0$
3. Calculate an E-Capra subgradient $y^k \in \partial^{\cent} f(x^k)$
   Let $f_{-1} = -\infty$ and $f_k = \max\{f_{k-1}, \langle \cdot, y_k \rangle - f^{\cent}(y_k)\}$
4. Calculate an optimal solution $\widehat{x} \in \operatorname*{arg\,min}_{\substack{x \in K \\ \ell_2(x)=1 \;,\; \ell_1(x) \leq \overline{\ell}_0^k}} f_k(x)$
5. (Local search) Set $x^{k+1} := \widehat{x}$.
   Set the $1 + \overline{\ell}_0^k$ smallest components of $\widehat{x}$ to 0.
   If $\widehat{x} \in K \setminus \{0\}$, set $\overline{\ell}_0^{k+1} := \overline{\ell}_0^k - 1$ and $x^{k+1} := \widehat{x}$.
   Otherwise, set $\overline{\ell}_0^{k+1} := \overline{\ell}_0^k$
6. Set $k := k + 1$ Repeat from Step 2 until a stop condition is satisfied

Minimization of the ratio of two norms

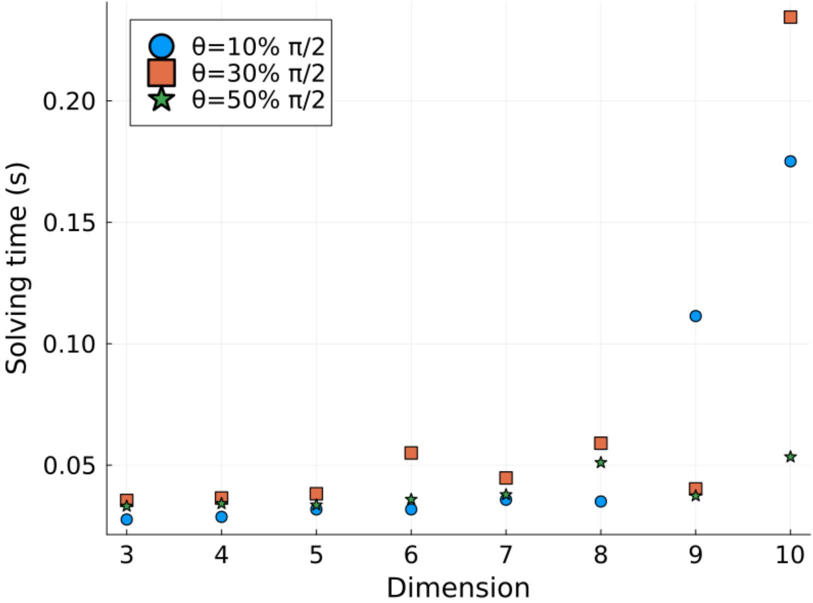
Instances: visualization in the 2D case



Solving time for the ratio of two norms



Zoom on the low dimensions



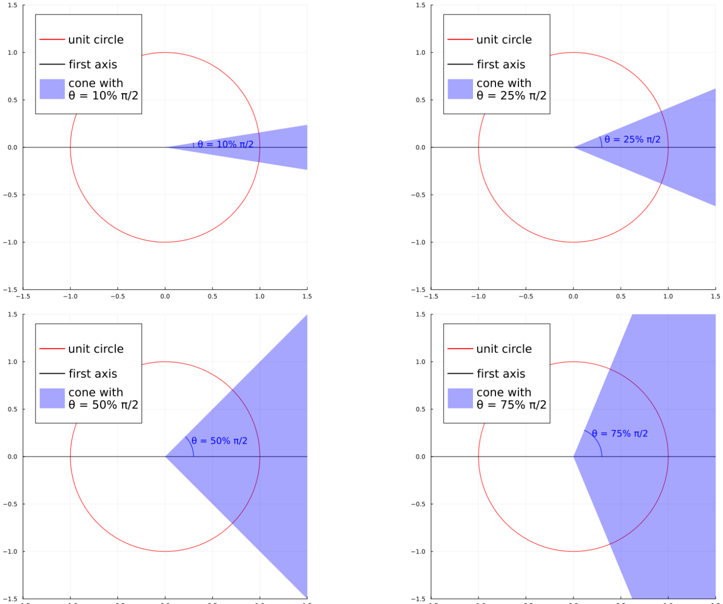
Conclusion for the ratio of two norms

- The toy example converges (no surprise, convergence theorem assumptions are satisfied)
- Tighter cones lead to faster convergence
- Experimental observation: when the method finds the optimal solution it sticks to it for the following iterations
- Future tests: $\inf_{x \neq 0} \frac{\|Ax\|}{\|x\|}$

Minimization of the pseudonorm ℓ_0 over a finitely generated cone

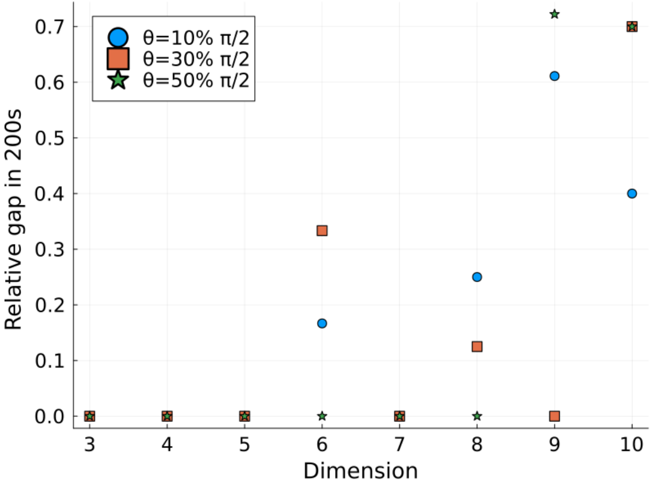
Instances

Same instances as the minimization of the ratio of two norms



Solving time for pseudonorm ℓ_0

$$\text{Relative gap} = \frac{\text{best value found} - \text{optimal value}}{\text{dimension}}$$



Conclusion for the minimization of pseudonorm ℓ_0 in a cone

- E-Capra cutting plane method does not converge for ℓ_0
- E-Capra cutting plane method with local search does not converge for ℓ_0 beyond dimension 4
- Maybe the noncontinuity of ℓ_0 is in cause

Computation spark of matrix

Definition of the spark of a matrix

Definition

Let $A \in \mathbb{R}^{m\times d}$ be a real matrix
Then, we call $\mathrm{spark}(A) \in = 1, d, \ldots, \cup\{+\infty\}$ the spark of A which is given by

$$\mathrm{spark}(A) = \min\left\{\ell_0(x) \,\middle|\, Ax = 0\ ,\ \ x \neq 0\right\}$$

Proposition

Let $A \in \mathbb{R}^{m\times d}$ be a real matrix
Then, $\mathrm{spark}(A)$ is the smallest number of dependent columns of the matrix A

Examples for the spark of a matrix

- $\operatorname{spark}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{pmatrix} = 1$
- $\operatorname{spark}\begin{pmatrix}-1 & 1 & 0\\ 2 & -2 & 0\\ 3 & -3 & 1\end{pmatrix} = 2$
- $\operatorname{spark}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} = +\infty$
- $\operatorname{spark}\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 1 & 1 & 1\end{pmatrix} = 3$

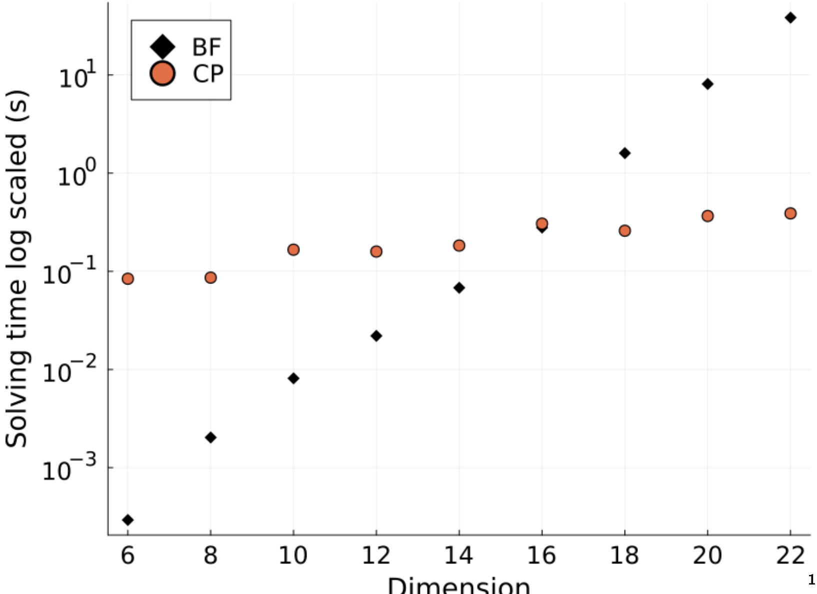
Brute force computing of the spark of a matrix

1. Set $s := 1$
2. For every family $\{A_{i_1}, \ldots, A_{i_s}\}$ of s columns of A
   if the family $\{A_{i_1}, \ldots, A_{i_s}\}$ is not free, stop and return s
3. Set $s := s + 1$
   If $s \leq d$, repeat from Step
   Otherwise, return $+\infty$

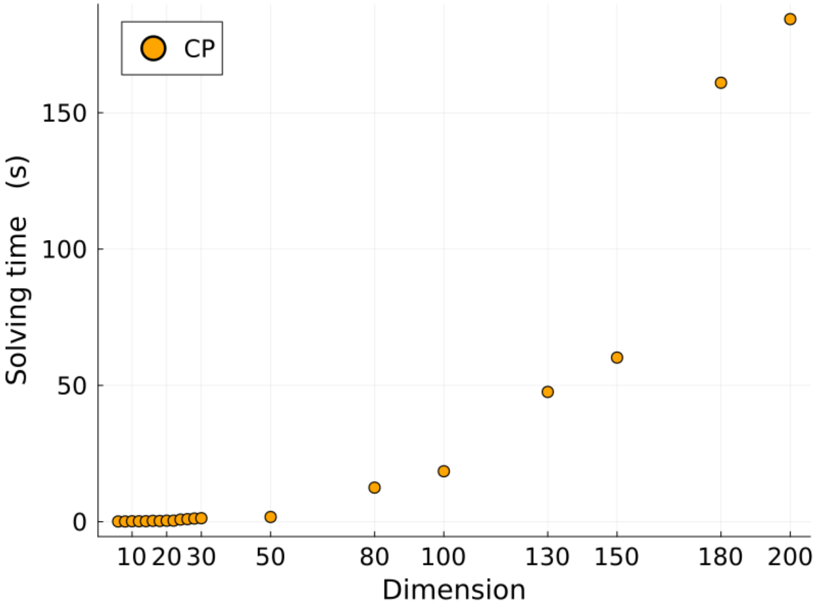
Generation of instances

- Instances:
  square matrices $A \in \mathbb{R}^{d \times d}$ such that $\mathrm{spark}(A) = d/2$
- We have used the following algorithm
  1. Randomly choose $s-1$ vectors $A_i \in \mathbb{R}^d$.
  2. Randomly choose $s-1$ real numbers $\mu_i \in \mathbb{R}$.
  3. Compute the vector $A_s = \sum_{i=1}^{s-1} \mu_i A_i$.
  4. Randomly choose $n-s$ vectors $A_h \in \mathbb{R}^d$.
  5. Set the matrix $A = \big(A_1, \ldots, A_n\big)$.
  6. Shuffle the columns of the matrix A.

Solving time comparison between brute force and E-CAPRA cutting plane for spark



Solving time for the E-Capra cutting plane method for spark



Conclusion for the computing of the spark of a Matrix

- Convergence even though the feasible set $\{x \neq 0 | Ax = 0\}$ is not E-CAPRA convex
- Converges faster than bruteforce
- No convergence for ℓ_0 and convergence for spark maybe because
  - $\overline{\mathrm{cone}}(g_1, \ldots, g_r)$ is a cone
  - $\{x \neq 0 | Ax = 0\}$ is a vector space

Discussion

Conclusion of the numerical tests (1)

- Ranking of the problems by difficulty
  1. (Toy example) the minimization of the ratio of the ℓ_1 norm over the ℓ_2 norm;
  2. the computation of the spark of a square matrix;
  3. the minimization of the ℓ_0 pseudonorm in a blunt convex cone.
- Computing Spark is not E-Capra convex but converges
  Minimization of ℓ_0 in a blunt cone is E-Capra but does not converge
- Future tests: see if the minimization of ℓ_0 converges when the constraints are in 'dual' form $\{x|Ax \leq 0\}$