Abstract Cutting Plane Method in Sparse Optimization

Part of the session
Theory and Algorithms in Sparse Optimization

Seta Rakotomandimby Jean-Philippe Chancelier

Michel De Lara Adrien Le Franc

École nationale des ponts et chaussées | Institut Polytechnique de Paris

Congrès de la ROADEF — February 26-28, 2025
École nationale des ponts et chaussées,
Champs-sur-Marne

Preprint on HAL

[Rakotomandimby, Chancelier, de Lara, and Le Franc, 2024]
Subgradient Selector in the Generalized Cutting Plane Method with an Application to Sparse Optimization

An abstract convex optimization method

[Kelley, 1960] Cutting plane method

Problem at hand

$$\min_{x \in X} f(x)$$

where

- $f : \mathbb{R}^n \to \mathbb{R}$ is finite continuous convex function
- $X \subset \mathbb{R}^n$ is a nonempty compact convex set

Illustration of the (Kelley's) cutting plane method

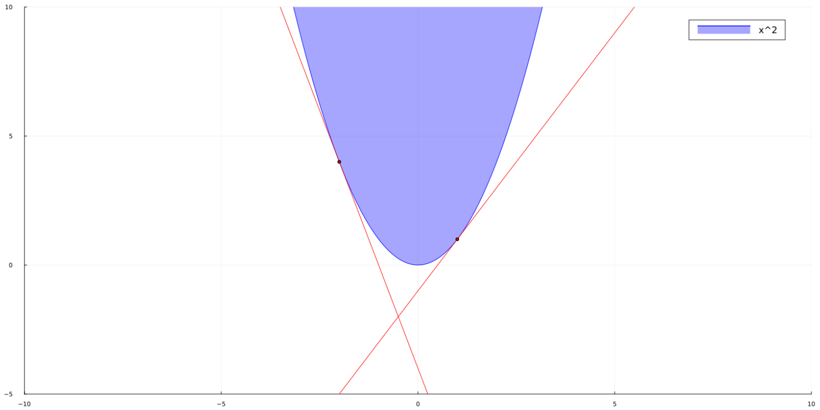


Illustration of the (Kelley) cutting plane method

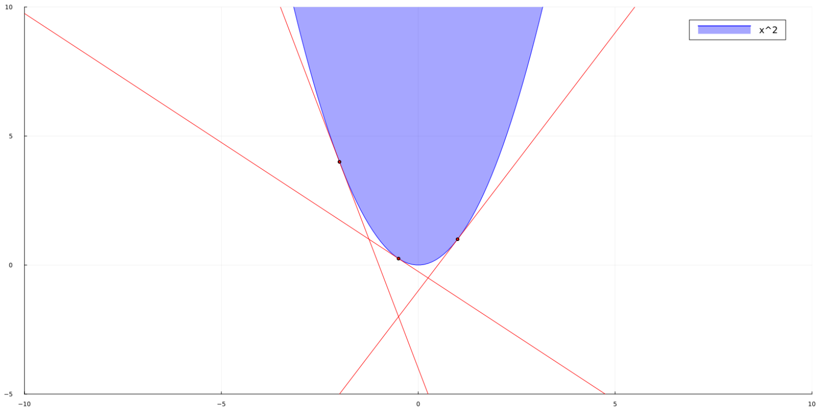


Illustration of the (Kelley) cutting plane method

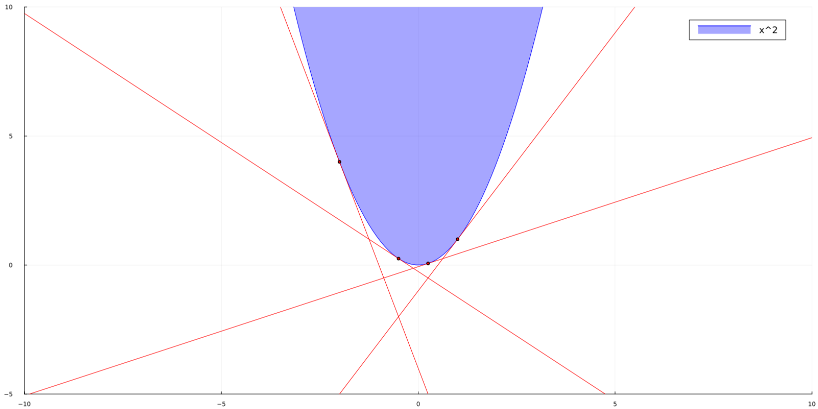


Illustration of the (Kelley) cutting plane method

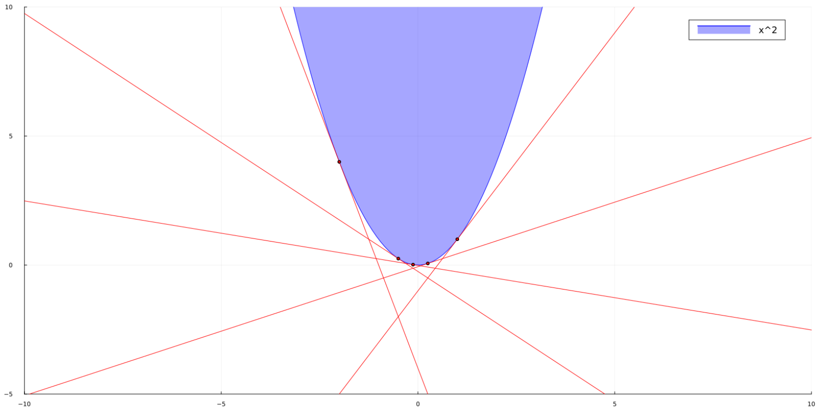
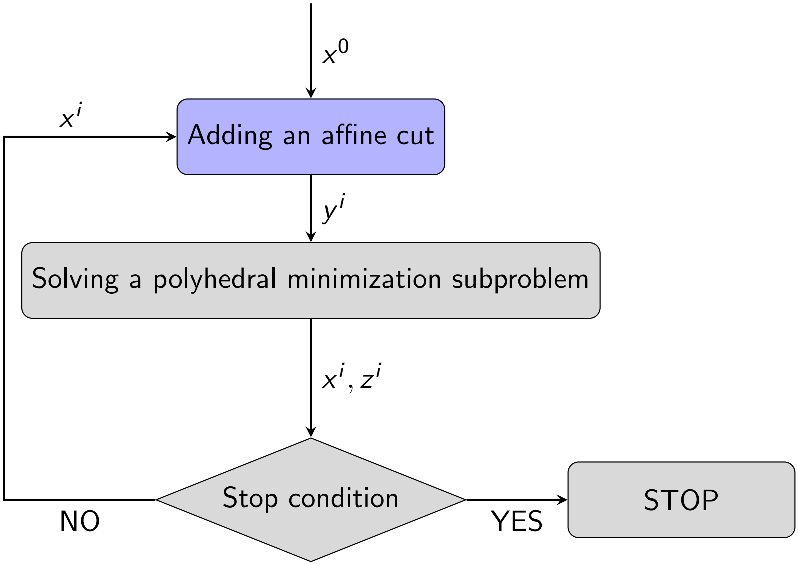


Diagram of the cutting plane method



Motivation: sparse optimization problem

The pseudonorm $\ell_0 : \mathbb{R}^n \to [\![0, n]\!]$

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

Archetypal sparse minimization problem

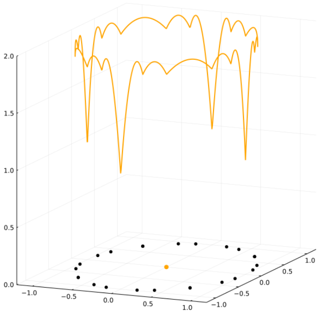
$$\min_{\substack{x \in \mathbb{R}^n \\ Ax = b}} \ell_0(x)$$

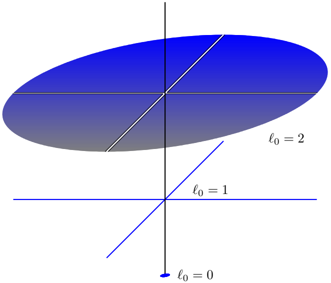
- $A \in \mathbb{R}^{m \times n}$
- $b \in \mathbb{R}^m$

Applications in

compressive sensing, spike deconvolution, model selection...

Capra "polyhedral" lower approximation of ℓ_0





[Le Franc, Chancelier, and De Lara, 2024]

Result of Capra-convexity

Theorem

[Chancelier and De Lara, 2022]

If both the source norm $\|\cdot\|$ and the dual norm $\|\cdot\|_{\star}$ are orthant-strictly monotonic (e.g. the Euclidean norm)

$$\partial_{\cent}\ell_0(x) \neq \emptyset \ , \quad \forall x \in \mathbb{R}^n$$

Outline of the presentation

Outline of the presentation

The abstract cutting plane method

- Definition of c-subgradient
- Definition of the abstract cutting plane method

Main results: convergence of the abstract cutting plane method

Application to sparse optimization

- Capra-subdifferential of ℓ_0
- Numerical results (spark computation)

Conclusion

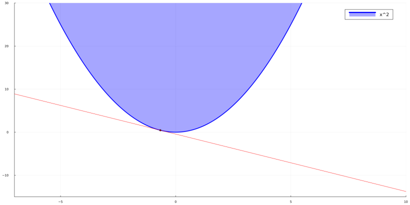
Usual subdifferential

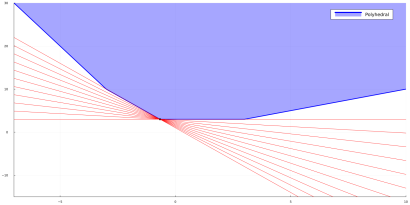
Definition

- Let $\langle \cdot \mid \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be the scalar product
- let $h : \mathbb{R}^n \to \overline{\mathbb{R}} \big(= \mathbb{R} \cup \{-\infty, +\infty\}\big)$ be a function

We define its subdifferential $\partial h : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ by

$$y \in \partial h(x) \iff \langle x' \mid y \rangle - h(x') \leq \langle x \mid y \rangle - h(x) \; , \; \forall x' \in \mathbb{R}^n$$





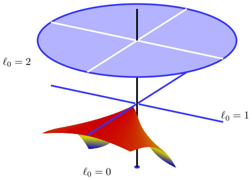
c-subdifferential

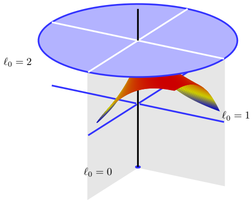
Definition

- let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a finite coupling
  where $\mathcal{X}$ and $\mathcal{Y}$ (nonempty primal set and nonempty dual set)
- let $h : \mathcal{X} \to \overline{\mathbb{R}}$ be a function

We define its c-subdifferential $\partial_c h : \mathcal{X} \rightrightarrows \mathcal{Y}$ by

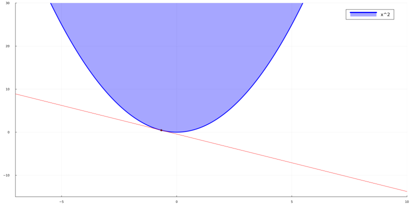
$$y \in \partial_c h(x) \iff c(x', y) - h(x') \leq c(x, y) - h(x) \ , \ \forall x' \in \mathcal{X}$$



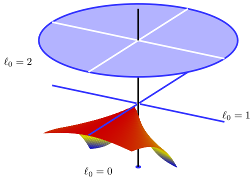


From usual convexity to generalized convexity

Usual convexity



Generalized convexity



Outline of the presentation

Problem at hand

$$\min_{x \in X} h(x)$$

where

- objective function: $h : \mathcal{X} \to \overline{\mathbb{R}}$
- c-subdifferentiability: $\partial_c h(x) \neq \emptyset\ , \ \forall x \in \underbrace{X}_{\text{constraint set}} \subset \mathcal{X}$
- finite coupling: $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$

Abstract cutting plane method

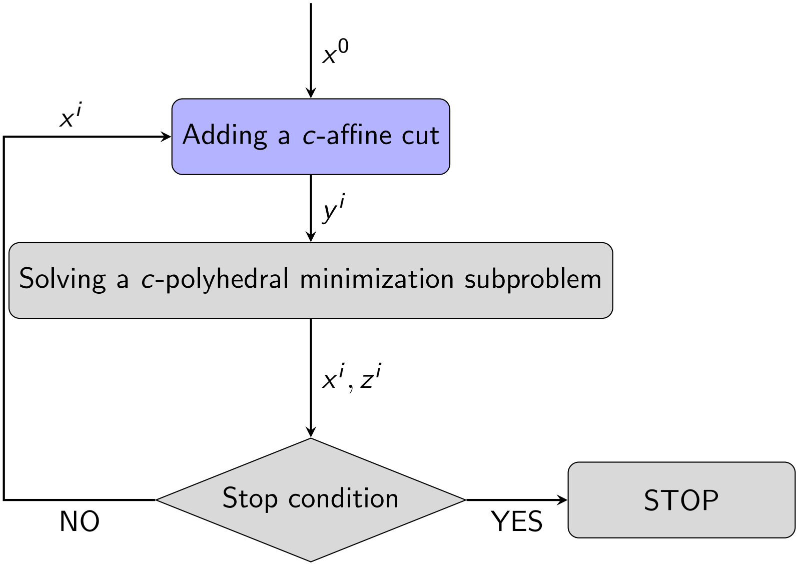
We say that $\underbrace{\{x^i\}_{i\geq 0}}_{\text{primal iterates}} \subset X$ and $\underbrace{\{z^i\}_{i\geq 1}}_{\text{lower bounds}} \subset \mathbb{R}$ are generated by $\mathrm{CP}(h, c; D, x^0)$, if

1. **Initialization.** $x^0 \in \underbrace{X}_{\text{constraint set}} \subset \mathcal{X}$
2. c**-subgradient selection.**
   $y^i = D(x^i)$, where $D : X \to \mathcal{Y}$ s.t. $\underbrace{D(x) \in \partial_c h(x)}_{c\text{-subgradient selector}}$
3. i**-th primal subproblem.**

$$(x^i, z^i) \in \operatorname*{arg\,min}_{(x,z)\in\mathcal{X}\times\mathbb{R}} z \text{ s.t. } \begin{cases} x \in X \,, \\ z \geq c(x, y^j) - c(x^j, y^j) + h(x^j) \\ \forall j \in [\![0, i-1]\!] \end{cases}$$

4. **Stop condition.** If not satisfied $i := i + 1$. Go to Step 2

Diagram of the abstract cutting plane method



Outline of the presentation

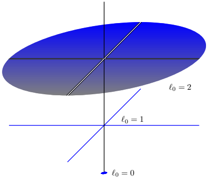
Lower semicontinuity

- We say that a function $h : \overbrace{(\mathcal{X}, d)}^{\text{metric space}} \to \overline{\mathbb{R}}$ is lower semicontinuous (l.s.c.) at $x \in \mathcal{X}$ if $h(x) \in \mathbb{R}$ and for all $\{x^i\}_{i \geq 0} \subset \mathcal{X}$ we have that

$$\lim_{i \to +\infty} x^i = x \implies \liminf_{i \to +\infty} h(x^i) \geq h(x)$$

- We say that h is l.s.c. if it is l.s.c. at x, for all $x \in \mathcal{X}$

The pseudonorm ℓ_0 is l.s.c.!



Key assumption

Lipschitz-like property: $\exists M > 0$ such that

$$|c(x, D(x)) - c(x', D(x))| \leq M d(x, x') \ , \ \ \forall x, x' \in \mathrm{dom}\, h$$

Key assumption

Lipschitz-like property: $\exists M > 0$ such that

$$|c(x, D(x)) - c(x', D(x))| \leq M d(x, x') \ , \ \ \forall x, x' \in \mathrm{dom} h$$

- Property on $c(\cdot, D(x)) : \mathcal{X} \to \mathbb{R} \ , \ \ \forall x \in \mathcal{X}$
- Let $h : \overbrace{(\mathcal{X}, d)}^{\text{metric space}} \to \overline{\mathbb{R}}$
  - The effective domain of h is defined by $\mathrm{dom} h = \left\{x \in \mathcal{X} : h(x) < +\infty\right\}$
  - We say h is proper if $\mathrm{dom} h \neq \emptyset$ and $h > -\infty$

Convergence result for c-cutting plane method

Theorem

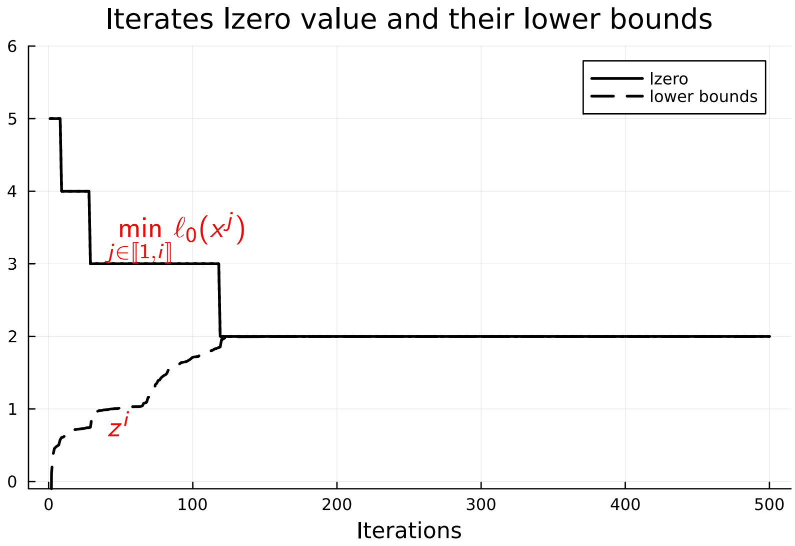
- $h : (\mathcal{X}, d) \to \overline{\mathbb{R}}$ is a proper l.s.c. function
- $\mathrm{dom} h$ is compact
- lower semicontinuity of the functions $c(\cdot, D(x))$, for all $x \in \mathrm{dom}\ h$
- Lipschitz-like property: $\exists M > 0$ such that

$$|c(x, D(x)) - c(x', D(x))| \leq Md(x, x')\ , \ \ \forall x, x' \in \mathrm{dom} h$$

Then, for all sequences $\{x^i\}_{i\geq 0}, \{z^i\}_{i\geq 1}$ generated by $\mathrm{CP}(h, c; D, x^0)$

- $\{z^i\}_{i\geq 1}$ increases to $h^* = \inf_{\mathcal{X}} h$
- $\{x^i\}_{i\geq 0}$ has a subsequence $\{x^{\nu(i)}\}_{i\geq 0}$ converging to some $x^* \in \arg\min_{\mathcal{X}} h$

Stop condition: lowerbound and upperbound sequences



Outline of the presentation

The abstract cutting plane method

Main results: convergence of the abstract cutting plane method

Application to sparse optimization

Conclusion

Minimizing ℓ_0 over a closed subset of the unit sphere

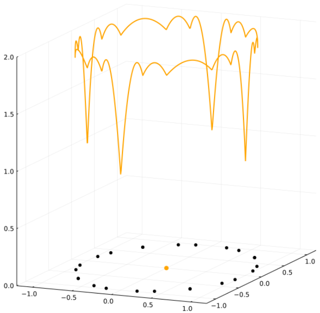
$$\min_{x \in \mathbb{R}^n} \ell_0(x) + \iota_X$$

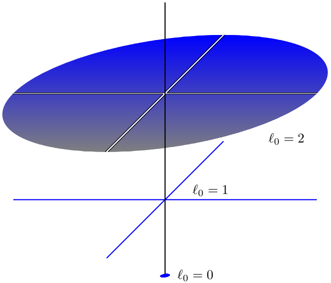
where

- X is a closed subset of the (Euclidean) unit sphere
- We will have to be cautious with the discontinuity points of ℓ_0 in X!

Outline of the presentation

Capra "polyhedral" lower approximation of ℓ_0





Generalized convexity of the ℓ_0 pseudonorm

Definition

- The pseudonorm $\ell_0 : \mathbb{R}^n \to [\![0, n]\!]$
  $\ell_0(x) = \sum_{i=1}^n \mathbf{1}_{\{x_i \neq 0\}}$, $\forall x \in \mathbb{R}^n$
- For a source norm $\|\cdot\|$, the Capra coupling $¢ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

$$¢(x, y) = \left\langle \frac{x}{\|x\|} \mid y \right\rangle \ , \ \text{where } \frac{0}{0} = 0$$

Theorem

[Chancelier and De Lara, 2022]
Let $\|\cdot\| = \sqrt{\langle \cdot \mid \cdot \rangle}$ be the source norm for the Capra coupling $¢$

$$\partial_{¢}\ell_0(x) \neq \emptyset \ , \ \forall x \in \mathbb{R}^n$$

Thus, $$\ell_0(x) = \max_{y \in \mathbb{R}^n} \underbrace{¢(x, y) - \ell_0^{¢}(y)}_{\text{Capra affine functions of } x}$$

¢-subdifferential

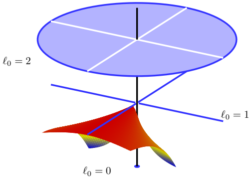
Definition

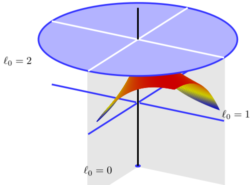
- Let $\cent : \mathbb{R}^n \times \mathcal{Y} \to \mathbb{R}$ be a Capra-coupling for a source norm $\|\cdot\|$
- let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function

We define its ¢-subdifferential $\partial_{\cent} h : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ by

$$y \in \partial_{\cent} h(x) \iff \frac{\langle x' \mid y \rangle}{\|x'\|} - h(x') \leq \frac{\langle x \mid y \rangle}{\|x\|} - h(x) \; , \quad \forall x' \in \mathbb{R}^n$$

with convention $0/0 = 0$

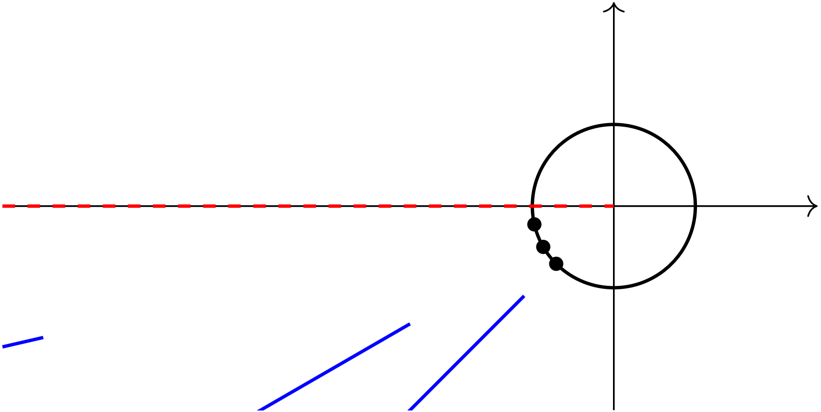




Problem with the norms of the Capra subgradients

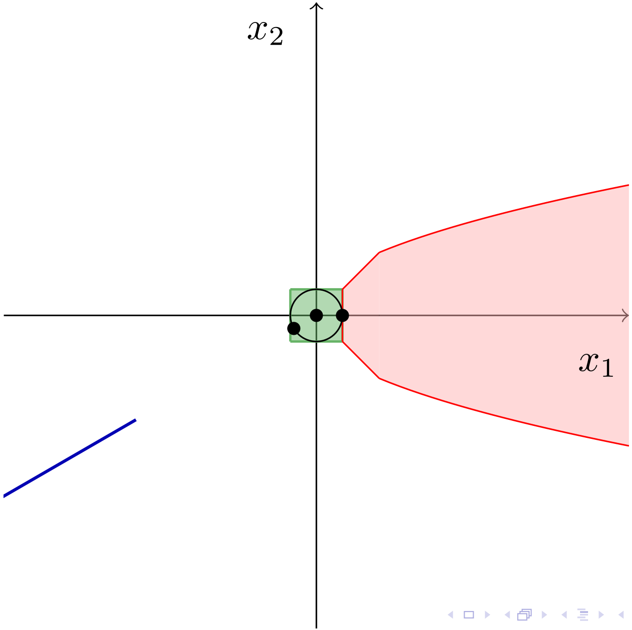
$$y \in \partial_{\cent} \ell_0(x) \; , \quad \text{for } x \in S$$

Problem with the norms of the Capra subgradients



Near sparse points, the norms of $\cent$-subgradients explode

Problem with the norms of the Capra subgradients



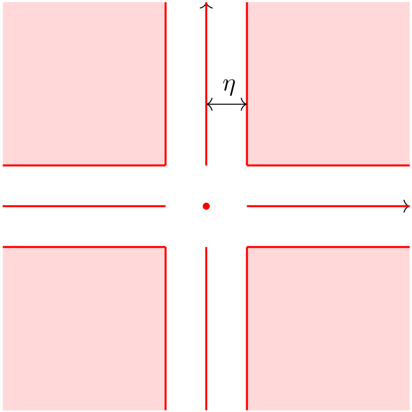
Proposed solutions: restricting the constraint set

$$\overbrace{\min_{x \in \mathbb{R}^n} \ell_0(x) + \iota_{X}}^{\text{original problem}} \leq \underbrace{\min_{x \in \mathbb{R}^n} \ell_0(x) + \iota_{X \cap R}}_{\text{surrogate problem}}$$

where $R \subset \mathbb{R}^n$ is a set such that
the Capra-subgradients of ℓ_0 are bounded

Proposed solutions: restricting the constraint set

$$R_\eta = \left\{ x \in \mathbb{R}^n : \min_{\substack{1 \le k \le n \\ x_k \neq 0}} |x_k| > \eta \right\}$$



Removing the primal points
whose minimal ¢-subgradients norm is exploding

A cutting plane method for the pseudonorm ℓ_0

$$\mathrm{CP}(h, ¢; D, x^0)$$

for the minimization problem $\min_{x \in \mathbb{R}^n} \ell_0(x) + \iota_{X \cap R_\eta}$, where $\eta > 0$

Example

- Objective function: $h = \ell_0 + \iota_{X \cap R_\eta}$
- Coupling:

$$¢(x, y) = \left\langle \frac{x}{\|x\|} \mid y \right\rangle \;, \quad \text{where } \frac{0}{0} = 0$$

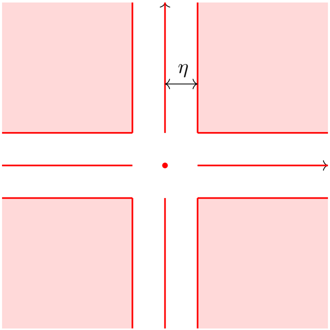
- ¢-subgradient selector: $D : X \to \mathcal{Y}$ is defined by

$$\{D(x)\} = \operatorname*{arg\,min}_{y \in \partial_{¢} \ell_0(x)} \|y\|^2 \;, \quad \forall x \in X$$

A cutting plane method for ℓ_0

Lipschitz property is satisfied for $(\cent, D)$ with

$$|\cent(x, D(x)) - \cent(x', D(x))| \leq \frac{\sqrt{1-\eta^2}+1}{\eta^2} \left\|x - x'\right\| \ , \ \forall x, x' \in X$$



Outline of the presentation

The computation of the spark of a matrix

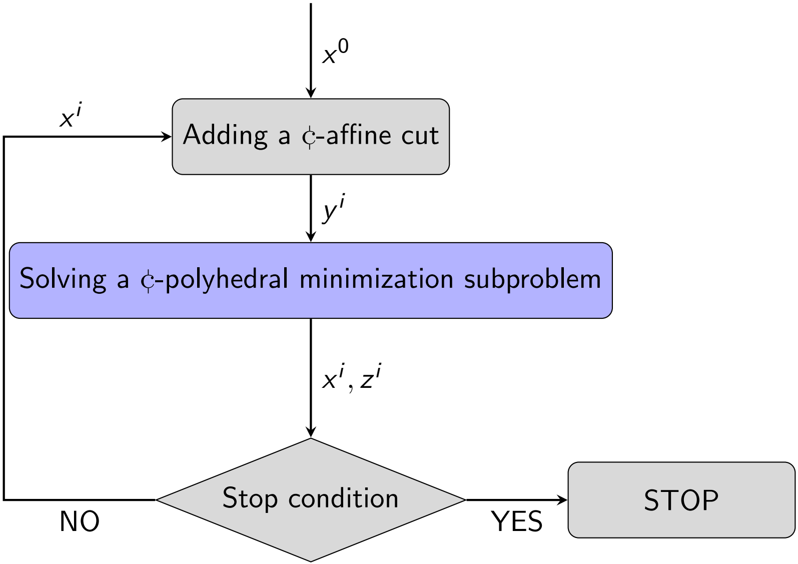
- Spark of the matrix $A \in \mathbb{R}^{m \times n}$

$$\mathrm{spark}(A) := \min_{\substack{x \in \mathbb{R}^n \setminus \{0\} \\ Ax = 0}} \ell_0(x)$$

- Interpretation of the spark of a matrix:

The spark of A is the smallest number of linearly dependent columns in A

Diagram of the Capra cutting plane method



Capra cutting plane (primal) subproblem

$$\min_{\substack{z \in \mathbb{R} \\ x \in \mathbb{R}^n}} z \text{ s.t. } \begin{cases} Ax = 0 \\ x \in R_\eta \\ z \geq \frac{\left\langle x \middle| y^j \right\rangle}{\|x\|} + \ell_0(x^j) - ¢(x^j, y^j) \\ \forall j \in [\![0, i-1]\!] \end{cases}$$

Capra cutting plane (primal) subproblem

Proposition

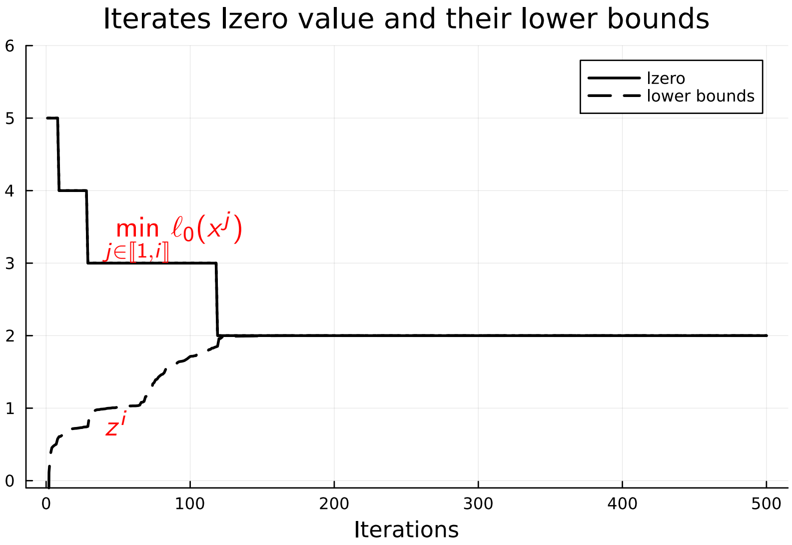
Given $\{x^j\}_{j \in [\![0, i-1]\!]}, \{y^j\}_{j \in [\![0, i-1]\!]} \subset \mathbb{R}^n$,
the i-th primal subproblem of a Capra cutting plane method is

$$\min_{\substack{z \in \mathbb{R} \\ \underbrace{s \in S}_{\text{sphere constraint}}}} z \text{ s.t. } \begin{cases} As = 0 \\ s \in R_\eta \\ \underbrace{z \geq \langle s \mid y^j \rangle + \ell_0(x^j) - ¢(x^j, y^j)}_{\text{linear constraint}} \\ \forall j \in [\![0, i-1]\!] \end{cases}$$

We use the General Norm Constraint of the solver `Gurobi`

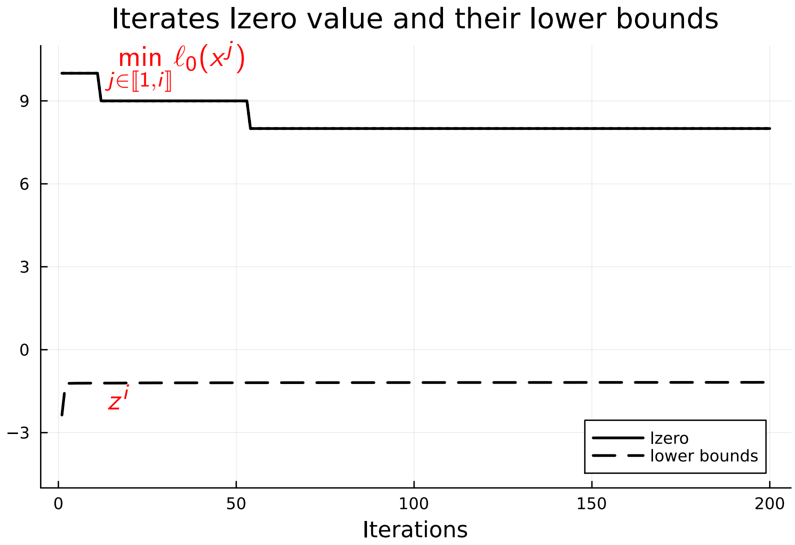
Numerical result in dimension 5

$\eta = 0.001$, gaussian $A \in \mathbb{R}^{2\times 5}$



Numerical result in dimension 10

$\eta = 0.001$, gaussian $A \in \mathbb{R}^{2\times 10}$



Numerical result in dimension 10

$\eta = 0.001$,   gaussian $A \in \mathbb{R}^{2\times 10}$

Outline of the presentation

Conclusion

- Abstract cutting plane for sparse optimization problems

Conclusion

- Abstract cutting plane for sparse optimization problems
- Linear programs on the sphere are crucial for the subproblems

Conclusion

- Abstract cutting plane for sparse optimization problems
- Linear programs on the sphere are crucial for the subproblems
- **Numerical challenge**:
  slow increase of the lowerbounds in higher dimension

Thank you for your attention!

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.

James E Kelley, Jr. The cutting-plane method for solving convex programs. *Journal of the society for Industrial and Applied Mathematics*, 8(4):703–712, 1960.

Adrien Le Franc, Jean-Philippe Chancelier, and Michel De Lara. The capra-subdifferential of the ℓ_0 pseudonorm. *Optimization*, 73(4):1229–1251, 2024. doi: 10.1080/02331934.2022.2145172.

Seta Rakotomandimby, Jean-Philippe Chancelier, Michel de Lara, and Adrien Le Franc. Subgradient Selector in the Generalized Cutting Plane Method with an Application to Sparse Optimization. working paper or preprint, December 2024. URL `https://hal.science/hal-04856783`.

Appendix

Formulas for the Capra subgradient of ℓ_0

Theorem

[Le Franc, Chancelier, and De Lara, 2024, Theorem 3.1]
Let $y \in \mathbb{R}^n$ and $\nu : [\![i, n]\!] \to [\![i, n]\!]$ such that $|y_{\nu(1)}| \geq \cdots \geq |y_{\nu(n)}|$

$$y \in \partial_{\text{¢}} \ell_0(x) \iff \begin{cases} y_{\mathrm{supp}(x)} = \lambda x_{\mathrm{supp}(x)} \ , \quad \lambda \geq 0 \\ |y_j| \leq \min_{i \in \mathrm{supp}(x)} |y_i| \ , \quad \forall j \notin \mathrm{supp}(x) \\ |y_{\nu(k+1)}|^2 \geq \left(\|y\|^{\mathrm{tn}}_{(k,2)} + 1\right)^2 - \left(\|y\|^{\mathrm{tn}}_{(k,2)}\right)^2 \\ \forall k \in [\![0, \ell_0(x) - 1]\!] \\ |y_{\nu(\ell_0(x)+1)}|^2 \leq \left(\|y\|^{\mathrm{tn}}_{(\ell_0(x),2)} + 1\right)^2 - \left(\|y\|^{\mathrm{tn}}_{(\ell_0(x),2)}\right)^2 \\ \left(\text{ when } \ell_0(x) \neq n\right) \end{cases}$$

Necessary continuity on the domain of the objective function

Proposition

Let $h : \mathcal{X} \to \overline{\mathbb{R}}$ be a l.s.c. proper function
If the couple $\left(c, D\right)$ satisfies the equicontinuous-like property, we have that

$$h_{|\mathrm{dom} h} \text{ is continuous}$$

R_η constraints with binary variables

- Let $\eta > 0$ and $E = \left\{x \in \mathbb{R}^n : \min_{\substack{1 \le k \le n \\ x_k \neq 0}} |x_i| > \eta\right\} \times \mathbb{R}$
- Linearization of R_η

$$x \in E \cap B_\infty(0, M) \iff \begin{array}{l} \text{Alternative} \\ \exists (b^-, b^0, b^+) \in \{0,1\}^n \\ b_j^- + b_j^0 + b_j^+ = 1 \\ \text{Sparse case} \\ -M(1 - b_j^0) \le x_j \le M(1 - b_i^0) \\ \text{Negative threshold case} \\ x_j \le M(1 - b_j^-) - \eta b_j^- \\ \text{Positive threshold case} \\ -M(1 - b_j^+) + \eta b_j^+ \le x_j \\ \forall j \in [\![1, n]\!] \end{array}$$

where $M > 0$