Hidden Convexity in the ℓ_0 Pseudonorm

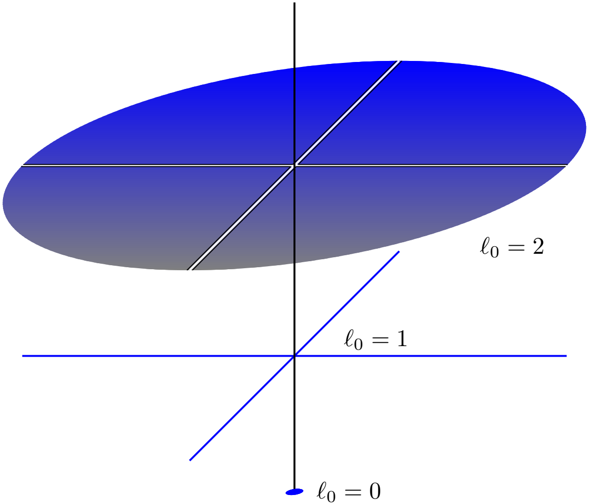
Algorithms in Generalized Convexity and Application to Sparse Optimization

Jean-Philippe Chancelier and *Michel De Lara*
Cermics, École des Ponts ParisTech, France

with the contributions of
Adrien Le Franc, Seta Rakotomandimby,
Antoine Deza, Lionel Pournin

McTAO Seminar
Centre Inria d'Université Côte d'Azur
22 April 2024

Here are the level sets
of the (highly nonconvex) ℓ_0 pseudonorm in $\mathbb{R}^2$



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

Let $n \in \mathbb{N}^*$ be a fixed natural number

- For any vector $x \in \mathbb{R}^n$, we define its ℓ_0 pseudonorm(x) by

$$\ell_0(x) = \text{number of nonzero components of } x = \sum_{i=1}^{n} \mathbf{1}_{\{x_i \neq 0\}}$$

- The function ℓ_0 pseudonorm : $\mathbb{R}^n \to [\![0, n]\!] = \{0, 1, \ldots, n\}$ 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... instead of 1-homogeneity, it is 0-homogeneity that holds true

$$\ell_0(\rho x) = \ell_0(x) \; , \;\; \forall \rho \neq 0$$

WHY STUDY A FUNCTION
THAT IS
ALMOST SURELY CONSTANT?

The ℓ_0 pseudonorm is used in typical sparse optimization problems

- Spark of a matrix A

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

- Compressed sensing: recovery of a sparse signal $x \in \mathbb{R}^n$ from a measurement $b = Ax$

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

- Least squares sparse regression (best subset selection):

$$\text{for } k \in [\![1, n]\!] \qquad \min_{\substack{x \in \mathbb{R}^n \\ \ell_0(x) \leq k}} \|Ax - b\|^2$$

"explaining" the output b by at most k components of x

Fenchel conjugacy ($\star$) *versus* E-Capra conjugacy ($\cent$) for the ℓ_0 pseudonorm

- Fenchel conjugacy ($\star$)

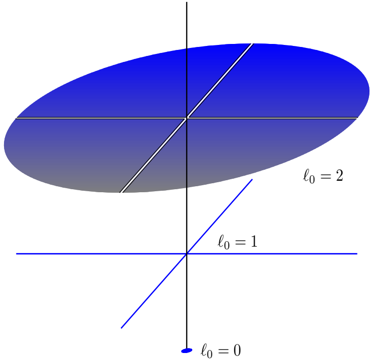
$$\ell_0^{\star\star'} = 0$$

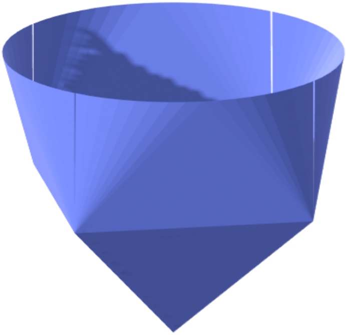
- E-Capra conjugacy ($\cent$)

$$\boxed{\ell_0^{\cent\cent'} = \ell_0}$$

[Chancelier and De Lara, 2021]

The ℓ_0 pseudonorm coincides, on the unit sphere, with the proper convex lower semicontinuous ℓ_0-cup function $\mathcal{L}_0 = \ell_0^{¢\star\prime}$





Towards algorithms?

- As motivation, we consider the sparse optimization problem, where C is a nonempty closed convex subset of $\mathbb{R}^n$,

$$\min_{x \in C} \ell_0(x) = \min_{x \in \mathbb{R}^n} \Big\{ \underbrace{\ell_0(x)}_{\substack{\text{E-Capra convex} \\ \ell_0^{\cent\cent'} = \ell_0}} + \overbrace{\underbrace{\iota_C(x)}_{\substack{\text{proper convex lsc} \\ \iota_C^{\star\star'} = \iota_C}}}^{\substack{\text{indicator} \\ \text{function}}} \Big\}$$

  where $\cent$ is the so-called E-Capra coupling
- Can we design algorithms using the above property that the pseudonorm ℓ_0 is E-Capra convex?

END OF THE TEASER

Talk outline

Outline of the presentation

Crash course on generalized convexity [5 min]

Capra conjugacies [20 min]

Towards Capra-algorithms in sparse optimization? [15 min]

Conclusion [1 min]

Additional material

Motivation: Legendre transform and Fenchel conjugacy in convex analysis

Definition

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

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

given by the Legendre transform

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

Coupling functions

Coupling function between sets

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

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

We also use the notation $\mathcal{X} \overset{c}{\leftrightarrow} \mathcal{Y}$ for a coupling [Moreau, 1966-1967, 1970]

In duality in convex analysis, one uses the bilinear coupling

$$c(x, y) = \langle x \, , \, y \rangle$$

and, on a Hilbert space, the scalar product

$$c(x, y) = \langle x \mid y \rangle$$

Euclidean Constant Along Primal RAys (Capra) coupling

- On the Euclidean space $\mathbb{R}^n$, the Euclidean-Capra coupling (E-Capra) $\mathbb{R}^n \overset{¢}{\longleftrightarrow} \mathbb{R}^n$ is given by

$$\forall y \in \mathbb{R}^n \; , \quad \begin{cases} ¢(x,y) & = \dfrac{\langle x \mid y \rangle}{\|x\|_2} = \dfrac{\langle x \mid y \rangle}{\sqrt{\langle x \mid x \rangle}} \; , \quad \forall x \in \mathbb{R}^n \backslash \{0\} \\ ¢(0,y) & = 0 \end{cases}$$

- The coupling E-Capra has the property of being Constant Along Primal RAys (Capra)

Fenchel-Moreau conjugacies

Fenchel-Moreau conjugate of a function

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

Definition

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

$$f^c(y) = \sup_{x \in \mathcal{X}} \Big(c(x,y) \stackrel{\cdot}{+} \big(-f(x)\big) \Big) , \quad \forall y \in \mathcal{Y}$$

We use the Moreau lower and upper additions on $\overline{\mathbb{R}}$
that extend the usual addition with

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

E-Capra-conjugate of the ℓ_0 pseudonorm

$$\ell_0^{\cent}(y) = \sup_{x\in\mathbb{R}^n} \Big\{ \cent(x,y) \dot{+} \big(-\ell_0(x)\big) \Big\}$$

$$= \sup \Big\{ 0, \sup_{x\neq 0} \Big\{ \frac{\langle x \mid y \rangle}{\|x\|_2} - \ell_0(x) \Big\} \Big\}$$

$$= \sup \Big\{ 0, \sup_{s\in S_2} \big\{ \langle s \mid y \rangle - \ell_0(s) \big\} \Big\}$$

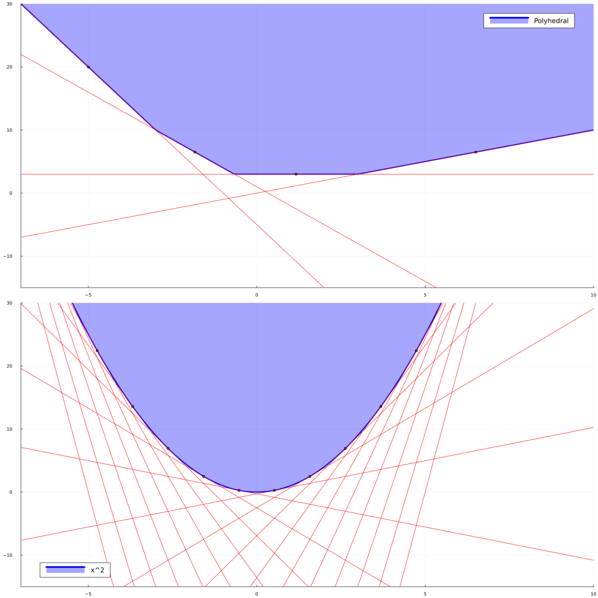
where $S_2 \subset \mathbb{R}^n$ is the Euclidean unit sphere

$$= \sup \Big\{ 0, \sup_{j\in[\![1,d]\!]} \Big\{ \underbrace{\sup_{\substack{s\in S_2 \\ \ell_0(s)=j}} \langle s \mid y \rangle}_{\substack{\text{top-}(2,j)\text{ norm} \\ \|y\|^{\top}_{2,j} = \sqrt{\sum_{l=1}^{k} |y_{\nu(l)}|^2}}} - j \Big\} \Big\}$$

$$= \sup_{j\in[\![1,d]\!]} \Big[\|y\|^{\top}_{2,j} - j \Big]_{+}$$

Biconjugates and duality

Motivation: duality in convex analysis



Reverse coupling and Fenchel-Moreau biconjugate

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

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

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

$$g \in \overline{\mathbb{R}}^{\mathcal{Y}} \mapsto g^{c'} \in \overline{\mathbb{R}}^{\mathcal{X}}$$

Reverse coupling and Fenchel-Moreau biconjugate

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

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

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

$$g \in \overline{\mathbb{R}}^{\mathcal{Y}} \mapsto g^{c'} \in \overline{\mathbb{R}}^{\mathcal{X}}$$

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

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

In generalized convexity,
one defines so-called c-convex functions

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

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

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

Definition

The function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is said to be c-convex if

$$f^{cc'} = f$$

c-convex functions have dual representations as suprema of elementary functions (abstract convexity)

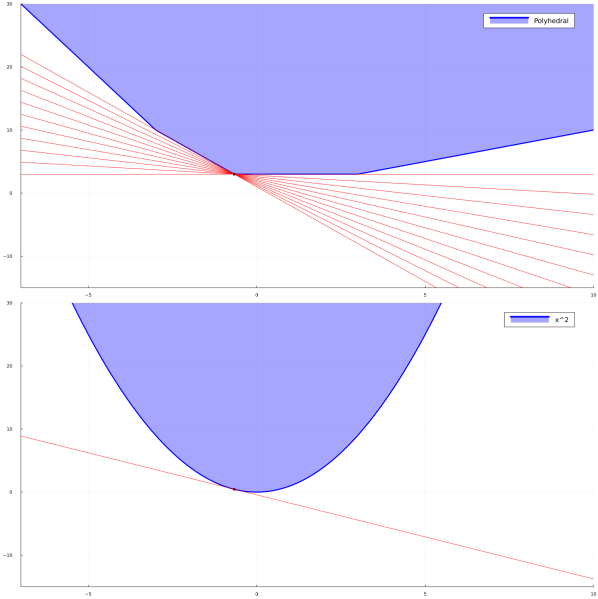
If the function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is c-convex, we have that

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

Example: $\star$-convex functions
$=$ closed convex functions
$=$ proper convex lsc or $\equiv -\infty$ or $\equiv +\infty$
$=$ suprema of affine functions

Subdifferential

Motivation: subgradients in convex analysis



(Upper) subdifferential $\partial^c f : \mathcal{X} \rightrightarrows \mathcal{Y}$ of a conjugacy

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

Definition

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

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

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

$$\partial^c f(x) \neq \emptyset \implies f(x) = \max_{y \in \partial^c f(x)} \Big(c(x,y) \dot{+} \left(-f^c(y)\right) \Big)$$

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

Wrap-up on generalized/abstract convexity

- Generalized convexity
  - coupling function between two sets
    $c : \mathcal{X} \times \mathcal{Y} \to \overline{\mathbb{R}}$
  - conjugacy and biconjugacy
    $f \in \overline{\mathbb{R}}^{\mathcal{X}} \mapsto f^{c} \in \overline{\mathbb{R}}^{\mathcal{Y}} \mapsto f^{cc'} \in \overline{\mathbb{R}}^{\mathcal{X}}$
  - generalized convex functions
    $f = f^{cc'}$
  - subdifferential
    $\partial^{c} f(x) \subset \mathcal{Y}$
- Abstract convexity
  - set of elementary functions
  - abstract convex envelope:
    supremum of lower elementary functions
  - abstract convex function:
    equal to its abstract convex envelope
  - subdifferential:
    tight lower elementary functions

Outline of the presentation

Outline of the presentation

We introduce the coupling E-Capra between $\mathbb{R}^n$ and itself

Definition

The Euclidean-Capra coupling (E-Capra) $\mathbb{R}^n \overset{¢}{\longleftrightarrow} \mathbb{R}^n$ is given by

$$\forall y \in \mathbb{R}^n , \quad \begin{cases} ¢(x,y) & = \dfrac{\langle x \mid y \rangle}{\|x\|_2} = \dfrac{\langle x \mid y \rangle}{\sqrt{\langle x \mid x \rangle}} , \quad \forall x \in \mathbb{R}^n \backslash \{0\} \\ ¢(0,y) & = 0 = \dfrac{0}{0} \end{cases}$$

The coupling E-Capra has the property of being
Constant Along Primal RAys (Capra)

E-Capra = Fenchel coupling after primal normalization

- We define the primal radial projection ϱ as

$$\varrho : \mathbb{R}^n \to S_2 \cup \{0\} \; , \;\; \varrho(x) = \begin{cases} \dfrac{x}{\|x\|_2} & \text{if } \; x \neq 0 \\ \\ \dfrac{0}{0} = 0 & \text{if } \; x = 0 \end{cases}$$

- so that the coupling E-Capra

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

appears as the Fenchel coupling after primal normalization (and the coupling E-Capra is one-sided linear)

The E-Capra conjugacy shares properties with the Fenchel conjugacy

Proposition

- For any function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$,
  the $¢$-Fenchel-Moreau conjugate is given by

$$f^{¢} = \big(\inf \left[f \mid \varrho \right]\big)^{\star} \qquad \text{where}$$

$$\inf \left[f \mid \varrho \right](x) = \begin{cases} \inf_{\rho > 0} f(\rho x) & \text{if } x \in S_2 \cup \{0\} \\ +\infty & \text{if } x \notin S_2 \cup \{0\} \end{cases}$$

- For any function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$,
  the $¢'$-Fenchel-Moreau conjugate is given by

$$g^{¢'} = g^{\star'} \circ \varrho$$

The E-Capra-convex functions are 0-homogeneous and coincide, on the unit sphere, with a closed convex function

Proposition

¢-convexity of the function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$

$$\iff h = h^{¢¢'}$$

$$\iff h = \underbrace{\left(h^{¢}\right)^{\star'}}_{\text{convex lsc function}} \circ \varrho$$

$\iff$ hidden convexity in the function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$

there exists a closed convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$

such that $h = f \circ \varrho$, that is, $h(x) = f\left(\frac{x}{\|x\|_2}\right)$

The ℓ_0 pseudonorm is E-Capra-convex

Notation

- The Euclidean top-$(2,k)$ norm is also known as the $(2,k)$-*symmetric gauge norm*, or *Ky Fan vector norm*

$$\|y\|^{\top}_{2,k} = \sqrt{\sum_{l=1}^{k} |y_{\nu(l)}|^2} \ , \quad |y_{\nu(1)}| \geq |y_{\nu(2)}| \geq \cdots \geq |y_{\nu(n)}|$$

- We denote the level sets of the ℓ_0 pseudonorm by

$$\ell_0^{\leq k} = \left\{ x \in \mathbb{R}^n \,\middle|\, \ell_0(x) \leq k \right\} \ , \quad \forall k \in [\![0, n]\!]$$

  and its elements are call k-sparse vectors
- For any subset $W \subset \mathbb{R}^n$, its indicator function ι_W is

$$\iota_W(w) = \begin{cases} 0 & \text{if } w \in W \\ +\infty & \text{if } w \notin W \end{cases}$$

The ℓ_0 pseudonorm and the E-Capra-coupling

Theorem

The ℓ_0 pseudonorm,
the indicator functions $\iota_{\ell_0^{\leq k}}$ of its level sets
and the Euclidean top-$(2,k)$ norms $\|\cdot\|_{2,k}^{\top}$ are related by

$$\iota_{\ell_0^{\leq k}}^{\cent} = \|\cdot\|_{2,k}^{\top} \ , \quad k \in [\![0, n]\!]$$

$$\ell_0^{\cent} = \sup_{j \in [\![0,n]\!]} \Big[\|\cdot\|_{2,j}^{\top} - j \Big]$$

$$\ell_0^{\cent\cent'} = \ell_0$$

The ℓ_0 pseudonorm displays hidden convexity

The ℓ_0 pseudonorm displays a convex factorization property

Theorem

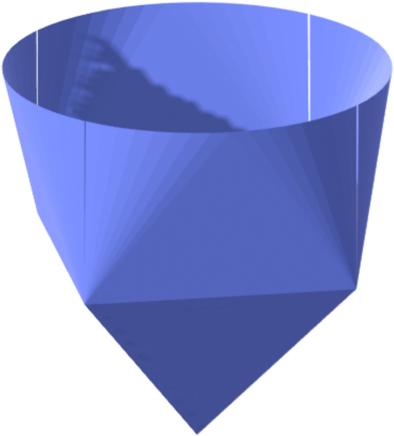
As the ℓ_0 pseudonorm is E-Capra-convex, we get that

$$\ell_0 = \ell_0^{¢¢'} = \ell_0^{¢\star'} \circ \varrho = \underbrace{\left(\ell_0^{¢}\right)^{\star'}}_{\text{convex lsc function } \mathcal{L}_0} \circ \overbrace{\varrho}^{\text{radial projection}}$$

As a consequence, the ℓ_0 pseudonorm coincides,
on the Euclidean unit sphere S_2,
with a proper convex lsc function,
the Euclidean ℓ_0-cup function $\mathcal{L}_0 = \ell_0^{¢\star'}$

$$\ell_0(x) = \mathcal{L}_0(x) \ , \ \ \forall x \in S_2$$

Graph of the Euclidean ℓ_0-cup function $\mathcal{L}_0 = \ell_0^{¢\star\prime}$



Best proper convex lsc lower approximation of the ℓ_0 pseudonorm on the Euclidean unit ball

Theorem

The Euclidean ℓ_0-cup function $\mathcal{L}_0 = \ell_0^{¢\star'}$ is
the best convex lsc lower approximation of the ℓ_0 pseudonorm
on the Euclidean unit ball B_2

$$\text{best convex lsc function} \quad \mathcal{L}_0(x) \leq \ell_0(x) \;, \;\; \forall x \in B_2$$

and, as seen above, coincides with the ℓ_0 pseudonorm

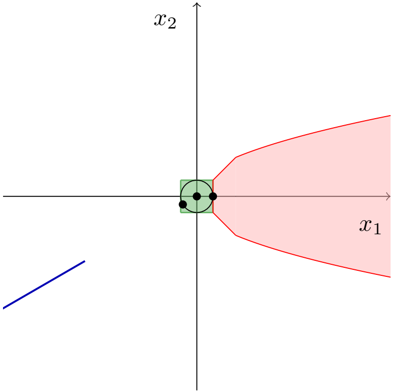
on the Euclidean unit sphere S_2

$$\ell_0(x) = \mathcal{L}_0(x) \;, \;\; \forall x \in S_2$$

E-Capra subdifferential of the ℓ_0 pseudonorm
(with Adrien Le Franc)

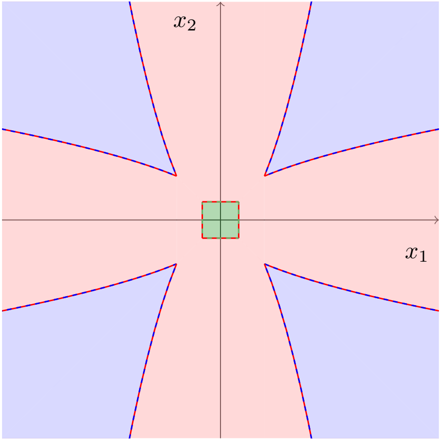
Capra-subdifferential of the ℓ_0 pseudonorm on $\mathbb{R}^2$

Illustration at three points (black dots)



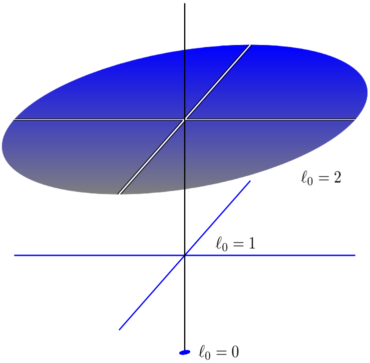
$$\partial_{¢}\ell_0(0,0)\ ,\ \ \partial_{¢}\ell_0(1,0)\ ,\ \ \partial_{¢}\ell_0(-\tfrac{\sqrt{3}}{2},-\tfrac{1}{2})$$

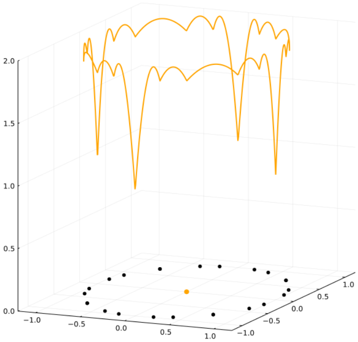
Capra-subdifferential of the ℓ_0 pseudonorm on $\mathbb{R}^2$



$$\partial_{¢}\ell_0(0) \bigcup \left\{ \bigcup_{\ell_0(x)=1} \partial_{¢}\ell_0(x) \right\} \bigcup \left\{ \bigcup_{\ell_0(x)=2} \partial_{¢}\ell_0(x) \right\}$$

Lower approximation of the ℓ_0 pseudonorm by a finite number of elementary E-Capra-functions





Variational formulas

We recall the Euclidean $(2,k)$-support norms $\|\cdot\|_{2,k}^{\top\star}$

- The dual norm of the top-$(2,k)$ norm $\|\cdot\|_{2,k}^{\top}$

$$\|\cdot\|_{2,k}^{\top\star} = \left(\|\cdot\|_{2,k}^{\top}\right)_{\star}$$

  is called the (Euclidean) $(2,k)$-support norm [Argyriou, Foygel, and Srebro, 2012]

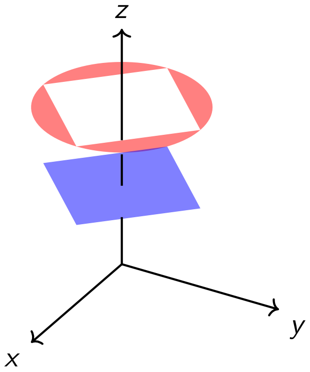
- We have the following inclusions between unit balls

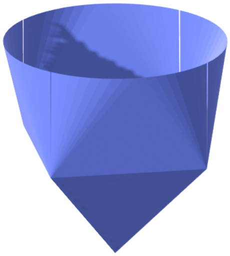
$$B_{(1)}^{\top\star} \subset \cdots \subset B_{(\ell-1)}^{\top\star} \subset B_{(\ell)}^{\top\star} \subset \cdots \subset B_{(n)}^{\top\star} = B$$

Proposition

The proper convex lsc function $\mathcal{L}_0$ is the convex envelope of the following piecewise constant function

$$L_0(x) = \begin{cases} 0 & \text{if } x = 0, \\ \ell & \text{if } x \in B^{\top\star}_{(\ell)} \setminus B^{\top\star}_{(\ell-1)}\ , \quad \ell \in [\![1, n]\!] \\ +\infty & \text{if } x \notin B^{\top\star}_{(n)} = B \end{cases}$$





Variational formulas for the ℓ_0 pseudonorm

Proposition

$$\ell_0(x) = \frac{1}{\|x\|_2} \min_{\substack{x^{(1)}\in\mathbb{R}^n,\ldots,x^{(n)}\in\mathbb{R}^n \\ \sum_{\ell=1}^n \left\|x^{(\ell)}\right\|_{2,\ell}^{\top\star} \leq \|x\|_2 \\ \sum_{\ell=1}^n x^{(\ell)} = x}} \sum_{\ell=1}^n \ell \left\|x^{(\ell)}\right\|_{2,\ell}^{\top\star}, \ \forall x \in \mathbb{R}^n$$

$$\ell_0(x) = \sup_{y\in\mathbb{R}^n} \inf_{\ell\in[\![1,n]\!]} \left(\frac{\langle x \mid y \rangle}{\|x\|_2} - \left[\|y\|_{2,\ell}^{\top} - \ell \right]_+ \right), \ \forall x \in \mathbb{R}^n \setminus \{0\}$$

Outline of the presentation

Work has gone on along two paths

	Norm Euclidean	Norm orthant-strictly monotonic	Norm any	1-homogeneous nonnegative function
ℓ_0 pseudonorm	¢-convex ($\ell_0^{¢¢'} = \ell_0$) hidden convexity variational formula [Chancelier and De Lara, 2021] subdifferential [Le Franc et al., 2022]	difference of norms [Chancelier and De Lara, 2022b]		
$\varphi \circ \ell_0$ $\varphi : \mathbb{N} \to \overline{\mathbb{R}}$ nondecreasing		¢-convex ($(\varphi \circ \ell_0)^{¢¢'} = \varphi \circ \ell_0$) hidden convexity variational formula subdifferential [Chancelier and De Lara, 2022c]		
$\varphi \circ \ell_0$ $\varphi : \mathbb{N} \to \overline{\mathbb{R}}$ any			$(\varphi \circ \ell_0)^{¢¢'}$ variational inequality subdifferential [Chancelier and De Lara, 2022a]	
$F \circ$ support $F : 2^{[\![1,d]\!]} \to \overline{\mathbb{R}}$ any			$(F \circ \mathrm{support})^{¢¢'}$ variational inequality subdifferential [preprint]	
0-homogeneous function				best lower approximation [preprint]

We introduce the coupling Capra

- Let be given $\mathcal{X}$ and $\mathcal{Y}$, two vector spaces paired by a bilinear form $\langle \cdot , \cdot \rangle$
- Suppose that $\mathcal{X}$ is equipped with a (source) norm $\|\cdot\|$

Definition

[Chancelier and De Lara, 2022a]

The coupling Capra $\mathcal{X} \overset{¢}{\longleftrightarrow} \mathcal{Y}$ is given by

$$\forall y \in \mathcal{Y} \, , \quad \begin{cases} ¢(x,y) & = \dfrac{\langle x , y \rangle}{\|x\|} \, , \quad \forall x \in \mathcal{X} \backslash \{0\} \\ ¢(0,y) & = 0 \end{cases}$$

In what follows, $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$
with norm $\|\cdot\|$ having unit ball B and unit sphere S

Orthant-monotonic and orthant-strictly monotonic norms

Orthant-monotonic norms

For any $x \in \mathbb{R}^n$, we denote by $|x|$
the vector of $\mathbb{R}^n$ with components $|x_i|$, $i \in [\![1, n]\!]$

Definition

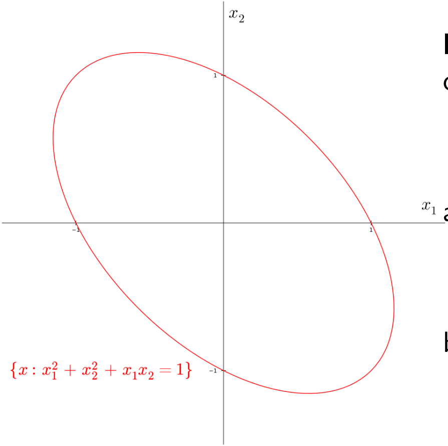
A norm $\|\cdot\|$ on the space $\mathbb{R}^n$ is called orthant-monotonic [Gries, 1967]
if, for all x, x' in $\mathbb{R}^n$, we have

$$|x| \leq |x'| \text{ and } x \circ x' \geq 0 \implies \|x\| \leq \|x'\|$$

where $x \circ x' = (x_1 x'_1, \ldots, x_n x'_n)$
is the Hadamard (entrywise) product

$$\left.\begin{array}{c} |x_1| \leq |x'_1| \,, \quad \ldots \,, \quad |x_n| \leq |x'_n| \\ \text{and} \\ x_1 x'_1 \geq 0 \,, \quad \ldots \,, \quad x_n x'_n \geq 0 \end{array}\right\} \implies \|x\| \leq \|x'\|$$

Example of unit sphere of a non orthant-monotonic norm



In the bottom right orthant, consider

$$|(0, -1)| \leq |(0.5, -1)|$$

and

$$(0, -1) \circ (0.5, -1) \geq (0, 0)$$

but

$$1 = \|(0, -1)\| > \|(0.5, -1)\|$$

Orthant-strictly monotonic norms

[Chancelier and De Lara, 2022b]

Definition

A norm $\|\cdot\|$ on the space $\mathbb{R}^n$ is called orthant-strictly monotonic if, for all x, x' in $\mathbb{R}^n$, we have

$$|x| < |x'| \text{ and } x \circ x' \geq 0 \implies \|x\| < \|x'\|$$

where $|x| < |x'|$ means that
there exists $j \in [\![1, n]\!]$ such that $|x_j| < |x'_j|$

Intuition: $\epsilon \neq 0 \implies \|(0, *, 0, *, *, 0)\| < \|(0, *, \epsilon, *, *, 0)\|$

Examples of orthant-strictly monotonic norms

$$\|x\|_\infty = \sup_{i \in [\![1,n]\!]} |x_i| \text{ and } \|x\|_p = \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p} \text{ for } p \in [1, \infty[$$

with unit ball B_p and unit sphere S_p

- All the ℓ_p-norms $\|\cdot\|_p$ on the space $\mathbb{R}^n$, for $p \in [1, \infty]$, are monotonic, hence orthant-monotonic

$$\ell_1, \ell_2, \ell_\infty$$

- All the ℓ_p-norms $\|\cdot\|_p$ on the space $\mathbb{R}^n$, for $p \in [1, \infty[$, are orthant-strictly monotonic

$$\ell_1, \ell_2, \cancel{\ell_\infty}$$

$$|\epsilon| < 1 \implies \|(1,0)\|_\infty = 1 = \|(1,\epsilon)\|_\infty$$

Orthant-strictly monotonic norms and Capra-convexity

Capra-subdifferentiability properties of the ℓ_0 pseudonorm

[Chancelier and De Lara, 2022c]

Proposition

If both the norm $\|\cdot\|$ and the dual norm $\|\cdot\|_\star$ are orthant-strictly monotonic, we have that

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

that is, the pseudonorm ℓ_0 is Capra-subdifferentiable on $\mathbb{R}^n$ and, as a consequence,

$$\ell_0^{¢¢'} = \ell_0$$

Best convex lower approximation of the ℓ_0 pseudonorm on the ℓ_p-unit balls, $p \in [1, \infty]$

Theorem

The best convex lsc lower approximation $\mathcal{L}_0$ of ℓ_0

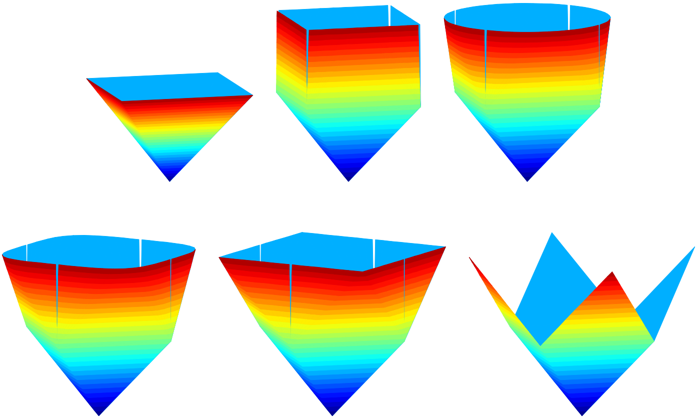
$$\text{best convex lsc function} \quad \mathcal{L}_0(x) \leq \ell_0(x) \ , \ \ \forall x \in B_p$$

on the unit ball B_p is $\ell_0^{¢\star\prime}$, and coincides with the ℓ_0 pseudonorm

$$\ell_0(x) = \mathcal{L}_0(x) \ , \ \ \forall x \in S_p$$

on the unit sphere S_p

Tightest closed convex function below the ℓ_0 pseudonorm on the ℓ_p-unit balls on $\mathbb{R}^2$ for $p \in \{1, 1.1, 2, 4, 300, \infty\}$



Capra-subdifferential of the ℓ_0 pseudonorm

Capra-subdifferential of the ℓ_0 pseudonorm

- $\{\|\cdot\|^{\mathcal{R}}_{(j)}\}_{j\in[\![1,n]\!]}$ and $\{\|\cdot\|^{\mathcal{R}}_{(j),\star}\}_{j\in[\![1,n]\!]}$,
associated coordinate-k and dual coordinate-k norms
- $\{B^{\mathcal{R}}_{(j)}\}_{j\in[\![1,n]\!]}$ and $\{B^{\mathcal{R}}_{(j),\star}\}_{j\in[\![1,n]\!]}$, corresponding unit balls

Proposition

[Chancelier and De Lara, 2022a]
The Capra-subdifferential of the ℓ_0 pseudonorm is given by

$$\text{if } x = 0, \qquad \partial_{\text{¢}}\ell_0(0) = \bigcap_{j\in[\![1,n]\!]} jB^{\mathcal{R}}_{(j),\star}$$

$$\text{if } x \neq 0 \text{ and } \ell_0(x) = \ell, \qquad \partial_{\text{¢}}\ell_0(x) = N\Big(B^{\mathcal{R}}_{(\ell)}, \frac{x}{\|x\|^{\mathcal{R}}_{(\ell)}}\Big) \cap Y_\ell$$

$$\text{where} \qquad Y_\ell = \Big\{ y \in \mathcal{Y} \,\Big|\, \ell \in \operatorname*{arg\,max}_{j\in[\![0,n]\!]} \big(\|y\|^{\mathcal{R}}_{(j),\star} - j\big) \Big\}\,, \;\; \forall \ell \in [\![0,n]\!]$$

Exposed faces and normal cones

For any nonempty closed convex subset $C \subset \mathcal{X}$,
where $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$,

- the exposed face $F_\perp(C, y)$ of C by any dual vector $y \in \mathcal{Y}$ is

$$F_\perp(C, y) = \arg\max_{x \in C} \langle x \mid y \rangle$$

- the normal cone $N(C, x)$ of C at any primal vector $x \in C$ is defined by the conjugacy relation

$$x \in C \text{ and } y \in N(C, x) \iff x \in F_\perp(C, y)$$

The family of all normal cones is the normal fan $\mathcal{N}(C)$

Coordinate-k norms and their dual norms

Courtesy of Basile and Lionel Pournin

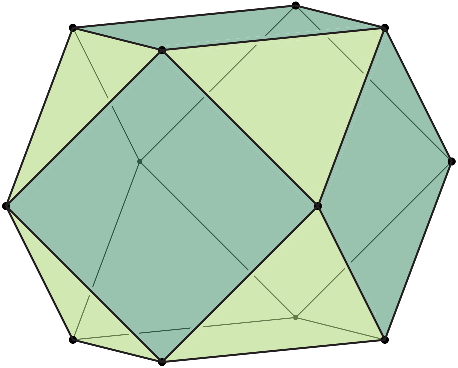


Figure: Unit ball $\overline{\mathrm{co}}\left(\ell_0^{\leq 2} \cap S_1\right)$ when $n = 3$

Extreme points of the coordinate-k norm unit ball are k-sparse

For any source norm $\|\cdot\|$ on $\mathbb{R}^n$, and for $k \in [\![1, d]\!]$,

- the coordinate-k norm $\|\cdot\|^{\mathcal{R}}_{(k)}$ has unit ball

$$B^{\mathcal{R}}_{(k)} = \underbrace{\overline{\mathrm{co}}\left(\ell_0^{\leq k} \cap S\right)}_{\text{closed convex hull}} = \underbrace{\mathrm{co}\left(\ell_0^{\leq k} \cap S\right)}_{\text{convex hull}}$$

- hence the extreme points of $B^{\mathcal{R}}_{(k)}$ belong to $\ell_0^{\leq k} \cap S \subset \ell_0^{\leq k}$, hence are k-sparse vectors

Extreme points of the coordinate-k norm unit ball are k-sparse

For any source norm $\|\cdot\|$ on $\mathbb{R}^n$, and for $k \in [\![1, d]\!]$,

- the coordinate-k norm $\|\cdot\|^{\mathcal{R}}_{(k)}$ has unit ball

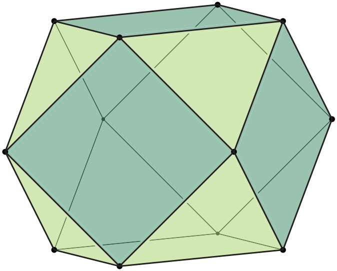
$$B^{\mathcal{R}}_{(k)} = \underbrace{\overline{\mathrm{co}}\left(\ell_0^{\leq k} \cap S\right)}_{\text{closed convex hull}} = \underbrace{\mathrm{co}\left(\ell_0^{\leq k} \cap S\right)}_{\text{convex hull}}$$

- hence the extreme points of $B^{\mathcal{R}}_{(k)}$ belong to $\ell_0^{\leq k} \cap S \subset \ell_0^{\leq k}$, hence are k-sparse vectors

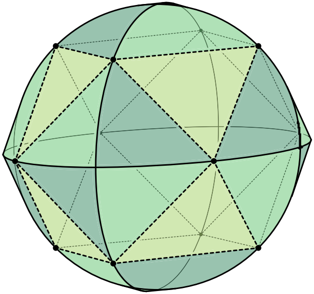
This is how we define

- a sequence $\left\{\|\cdot\|^{\mathcal{R}}_{(k)}\right\}_{k \in [\![1,n]\!]}$ of coordinate-k norms
- a sequence $\left\{\|\cdot\|^{\mathcal{R}}_{(k),\star}\right\}_{k \in [\![1,n]\!]}$ of dual coordinate-k norms

Courtesy of Basile and Lionel Pournin



(a) Unit ball $\overline{\mathrm{co}}\left(\ell_0^{\leq 2} \cap S_1\right)$ when $n = 3$



(b) Unit ball $\overline{\mathrm{co}}\left(\ell_0^{\leq 2} \cap S_2\right)$ when $n = 3$

Coordinate and dual coordinate norms induced by the ℓ_p-norms $\|\cdot\|_p$

For $y \in \mathbb{R}^n$, ν is a permutation of $[\![1, n]\!]$ such that $|y_{\nu(1)}| \geq |y_{\nu(2)}| \geq \cdots \geq |y_{\nu(n)}|$

source norm $\|\cdot\|$	$\|\cdot\|^{\mathcal{R}}_{(k),\star}$	$\|\cdot\|^{\mathcal{R}}_{(k)}$
$\|\cdot\|_p$	top-(q,k) norm $\|y\|^{\top}_{q,k}$ $= \left(\sum_{l=1}^{k} |y_{\nu(l)}|^q\right)^{\frac{1}{q}}$	(p,k)-support norm $\|x\|^{\top\star}_{p,k}$ no analytic expression
$\|\cdot\|_1$	top-(∞,k) norm ℓ_∞-norm $\|y\|^{\top}_{\infty,k} = \|y\|_\infty$	$(1,k)$-support norm ℓ_1-norm $\|x\|^{\top\star}_{1,k} = \|x\|_1$
$\|\cdot\|_2$	top-$(2,k)$ norm $\|y\|^{\top}_{2,k} = \sqrt{\sum_{l=1}^{k} |y_{\nu(l)}|^2}$ $\|y\|^{\top}_{2,1} = \|y\|_\infty$	$(2,k)$-support norm $\|x\|^{\top\star}_{2,k}$ no analytic expression $\|x\|^{\top\star}_{2,1} = \|x\|_1$
$\|\cdot\|_\infty$	top-$(1,k)$ norm $\|y\|^{\top}_{1,k} = \sum_{l=1}^{k} |y_{\nu(l)}|$ $\|y\|^{\top}_{1,1} = \|y\|_\infty$	(∞,k)-support norm $\|x\|^{\top\star}_{\infty,k} = \max\{\frac{\|x\|_1}{k}, \|x\|_\infty\}$ $\|x\|^{\top\star}_{1,1} = \|x\|_1$

Why do top-k and k-support norms pop up?

Generalized top and support norms

We reformulate sparsity in terms of coordinate subspaces

- For any $K \subset [\![1, n]\!]$, we introduce the (coordinate) subspace

$$\mathcal{R}_K = \left\{ y \in \mathbb{R}^n \,\middle|\, y_j = 0 \,, \ \forall j \notin K \right\} \subset \mathbb{R}^n$$

- The connection with the level sets of the ℓ_0 pseudonorm is

$$\ell_0^{\leq k} = \bigcup_{|K| \leq k} \mathcal{R}_K \,, \ \forall k \in [\![0, n]\!]$$

- We denote by $\pi_K : \mathbb{R}^n \to \mathcal{R}_K$ the orthogonal projection
- For any vector $y \in \mathbb{R}^n$, $\pi_K(y) \in \mathbb{R}^n$ is the vector whose components coincide with those of y, except for those outside of K that vanish

$$y = (*, *, *, *, *, *) \to \pi_{\{2,4,5\}}(y) = (0, *, 0, *, *, 0)$$

We define generalized top-k and k-support dual norms

Definition

For any source norm $\|\cdot\|$ on $\mathbb{R}^n$, for any $k \in [\![1, n]\!]$, we call

- generalized top-k dual norm the norm

$$\|y\|^{\top}_{\star,(k)} = \underbrace{\sup_{|K| \leq k} \| \overbrace{\pi_K(y)}^{\substack{k\text{-sparse}\\ \text{projection}\\ \text{on } \mathcal{R}_K}} \|_\star}_{\text{exploring all } k\text{-sparse projections}} , \quad \forall y \in \mathbb{R}^n$$

- generalized k-support dual norm the dual norm

$$\|\cdot\|^{\top\star}_{\star,(k)} = \left(\|\cdot\|^{\top}_{\star,(k)}\right)_\star$$

Coordinate norms and dual norms *versus* generalized top-k and k-support dual norms

Proposition

If the source norm $\|\cdot\|$ is orthant monotonic,
for all $k \in [\![1, n]\!]$,

k-coordinate norm		k-support dual norm
$\|\cdot\|^{\mathcal{R}}_{(k)}$	$=$	$\|\cdot\|^{\top\star}_{\star,(k)}$

dual k-coordinate norm		top-k dual norm
$\|\cdot\|^{\mathcal{R}}_{(k),\star}$	$=$	$\|\cdot\|^{\top}_{\star,(k)}$

so that, if S is the unit sphere of the source norm $\|\cdot\|$,

$$B^{\mathcal{R}}_{(k)} = \mathrm{co}\left(\ell_0^{\leq k} \cap S\right) = B^{\top\star}_{\star,(k)}$$

Where do we stand?

- We have Capra couplings $¢$ for which the pseudonorm ℓ_0
  - has nonempty Capra-subdifferential

$$\partial_{¢}\ell_0 \neq \emptyset$$

  - hence is Capra-convex (equal to its Capra-biconjugate)

$$\ell_0^{¢¢'} = \ell_0$$

- This looks promising to study sparse optimization problems

But...

Outline of the presentation

Archetypal sparse optimization problems

- For $X \subset \mathbb{R}^n$ a nonempty set,

$$\min_{x \in X} \ell_0(x)$$

is an optimization problem for which any point in X is a local minimizer! Jean-Baptiste Hiriart-Urruty and Hai Le. A variational approach of the rank function. *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research*, 21 (2):207–240, 2013.

- For $k \in [\![1, n]\!]$ and a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$,

$$\min_{\ell_0(x) \leq k} f(x)$$

- For $\gamma > 0$ and a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$,

$$\min_{x \in \mathbb{R}^n} \left(f(x) + \gamma \ell_0(x) \right)$$

Outline of the presentation

Good news :-)
the Fermat rule holds true for the Capra coupling

$$x^* \in \arg\min f \iff 0 \in \partial_{\text{¢}} f(x^*)$$

Good news :-)
the Fermat rule holds true for the Capra coupling

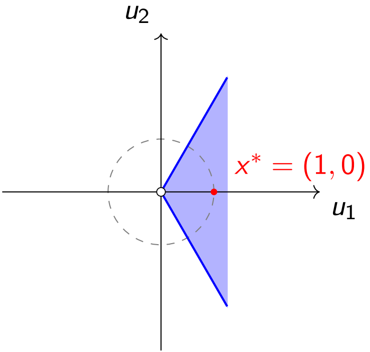
$$x^* \in \arg\min f \iff 0 \in \partial_{¢} f(x^*)$$

As an application, we get that

$$x^* \in \arg\min_{x \in X} \ell_0(x) \iff 0 \in \partial_{¢}(\ell_0 + \iota_X)(x^*)$$

But. . .

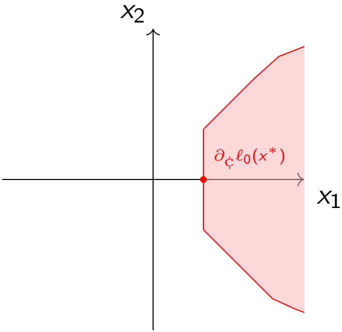
Bad news :-(
when zero is in the subdifferential of the sum...

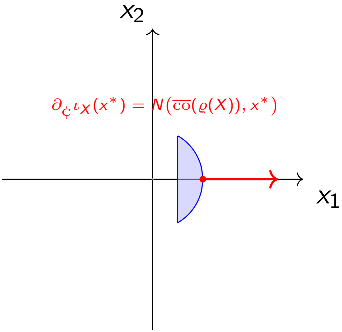


$$x^* \in \arg\min_{X} \ell_0 \iff 0 \in \underbrace{\partial_{\text{Ç}}(\ell_0 + \iota_X)(x^*)}_{\text{subdifferential of the sum}}$$

… but zero is not in the sum of the subdifferentials

$$\underbrace{\partial_{\text{ç}}\ell_0(x^*) + \partial_{\text{ç}}\iota_X(x^*)}_{0\notin} \subsetneq \underbrace{\partial_{\text{ç}}\big(\ell_0 + \iota_X\big)(x^*)}_{0\in}$$





Who is to blame? Capra or ℓ_0?
(with Seta Rakotomandimby)

Primal-dual pair in the Capra-subdifferential of an absolute function

Proposition

Let $f : \mathbb{R}^n \to \mathbb{R}$ be an absolute function and $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ be an absolute norm, meaning that

$$f(x) = f(|x|) \; , \;\; \forall x \in \mathbb{R}^n$$
$$\|x\| = \||x|\| \; , \;\; \forall x \in \mathbb{R}^n$$

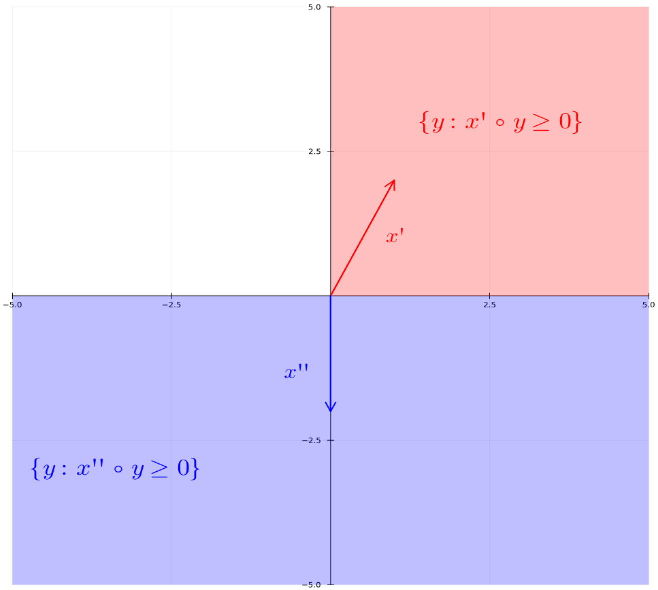
Then, we have that

$$y \in \partial_{\text{¢}} f(x) \implies x \circ y \geq 0$$

where $x \circ y = (x_1 y_1, \ldots, x_n y_n)$

NB: this property also holds true with the classic Rockafellar-Moreau subdifferential in convex analysis

Illustration of $x \circ y \geq 0$

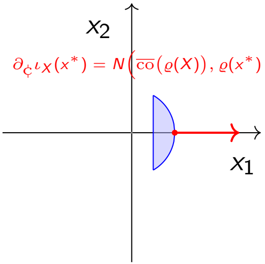


Capra-subdifferential of an indicator function

Proposition

Let $X \subset \mathbb{R}^n$ be a nonempty set. Then, for any $x \in \mathbb{R}^n$

$$\partial_{¢}\iota_X(x) = \begin{cases} \overbrace{N\Big(\overline{\mathrm{co}}\big(\varrho(X)\big), \varrho(x)\Big)}^{\text{normal cone}} & \text{if } x \in X \\ \emptyset & \text{if } x \notin X \end{cases}$$



- The Capra-subdifferential of ι_X at x^* is the normal cone of the convex subset $\overline{\mathrm{co}}\big(\varrho(X)\big) \subset B$ at $\varrho(x^*) \in S$, hence points outward
- The Rockafellar-Moreau subdifferential of ι_X at x^* is the normal cone of X at x^*

$0 \in \partial_{\text{¢}} f(x) + \partial_{\text{¢}} \iota_X(x)$ is much too strong a condition

Under the previous assumptions, we get that

$$0 \in \partial_{\text{¢}} f(x) + \partial_{\text{¢}} \iota_X(x) \implies 0 = \underbrace{\overbrace{y'}^{\partial_{\text{¢}} f(x)}}_{x \circ y' \geq 0} + \overbrace{y''}^{\partial_{\text{¢}} \iota_X(x)}$$

$$\implies \underbrace{y'' \in N\Big(\overline{\mathrm{co}}\big(\varrho(X)\big), \varrho(x)\Big)}_{y'' \text{ is outward}} \text{ and } \underbrace{x \circ y'' \leq 0}_{y'' \text{ is inward}}$$

- In general, this will give $y'' = 0$, that is, $0 \in \partial_{\text{¢}} f(x)$
- Thus, necessarily, $x \in X$ must be a global minimum of f over all $\mathbb{R}^n$, which is much too strong...

Where do we stand?

- We had good hope to handle sparse optimization problems with the Capra coupling that makes the pseudonorm ℓ_0 Capra convex
- But, in a simple sparse optimization problem, it is not true that the subdifferential of the sum is equal to the sum of the subdifferentials
- And not having practical qualification conditions is an obstacle to many numerical methods

Outline of the presentation

Crash course on generalized convexity [5 min]

Capra conjugacies [20 min]
- Euclidean Capra conjugacy
- Capra conjugacies

Towards Capra-algorithms in sparse optimization? [15 min]
- Good and bad news about the Fermat rule (with Adrien Le Franc and Seta Rakotomandimby)
- Capra-cuts method (with Seta Rakotomandimby)
- The geometry of sparsity-inducing unit balls (with Antoine Deza and Lionel Pournin)

Conclusion [1 min]

Additional material

Minimization problems from compressed sensing

- Goal: recovery of a sparse signal $x \in \mathbb{R}^n$ from a measurement $b \in \mathbb{R}^m \setminus \{0\}$, where $m < n$
- Measurements are modeled by $A \in \mathbb{R}^{m \times n}$ such that

$$Ax = b$$

- Minimization approach for the recovery

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

Using a Capra-polyhedral approximation for ℓ_0

- For a suitable (infinite) subset $Y \subset \bigcup_{x'} \partial_{\cent} \ell_0(x')$ of Capra-subgradients of ℓ_0, we have that

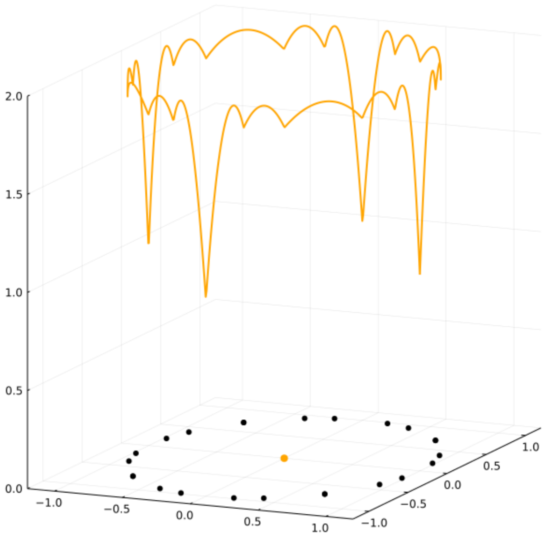
$$\ell_0(x) = \sup_{y \in Y} \langle \varrho(x), y \rangle - \ell_0^{\cent}(y) \; , \;\; \forall x \in \mathbb{R}^n$$

- Idea: using a Capra-"polyhedral" approximation f of ℓ_0 in the minimization problem

$$f(x) = \max_{y \in \tilde{Y}} \langle \varrho(x), y \rangle - \ell_0^{\cent}(y)$$

where $\tilde{Y} \subset Y$ and $\tilde{Y}$ finite $\rightsquigarrow$ cutting plane-like method

Illustration of a Capra-polyhedral approximation for ℓ_0



Abstract cutting plane method

[Rubinov, 2000, §9.2.3]

Definition

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

1. Set $k := 0$. Choose an arbitrary initial point $w_0 \in \mathcal{W}$
2. Find an abstract subgradient $h_k \in \partial^H f(w_k)$
   Let $f_{-1} = -\infty$ and set

$$f_k = \max\{f_{k-1}, \underbrace{h_k}_{\substack{\text{new cut} \\ \text{in } \partial^H f(w_k)}} \}$$

3. Find an optimal solution $\widehat{w} \in \arg\min_{w \in \mathcal{W}} f_k(w)$
4. Set $k := k + 1$, $w_k = \widehat{w}$
   Repeat from Step 2 until a stop condition is satisfied

Still problems with ℓ_0

- The pseudonorm ℓ_0 is abstract Capra-convex
- ... but ℓ_0 is not continuous and its abstract Capra-subgradients

$$\Big\{x \mapsto \langle \varrho(x), y\rangle - \ell_0^{\cent}(y)\Big\}_{y \in \cup_{x'} \partial_{\cent} \ell_0(x')}$$

  are not uniformly continuous
- So the pseudonorm ℓ_0 does not satisfy any assumptions of established theoretical convergence results [Pallaschke and Rolewicz, 1997, Theorem 9.1.1]
- Also, numerically, we observe no convergence for simple examples in dimension $n = 3$

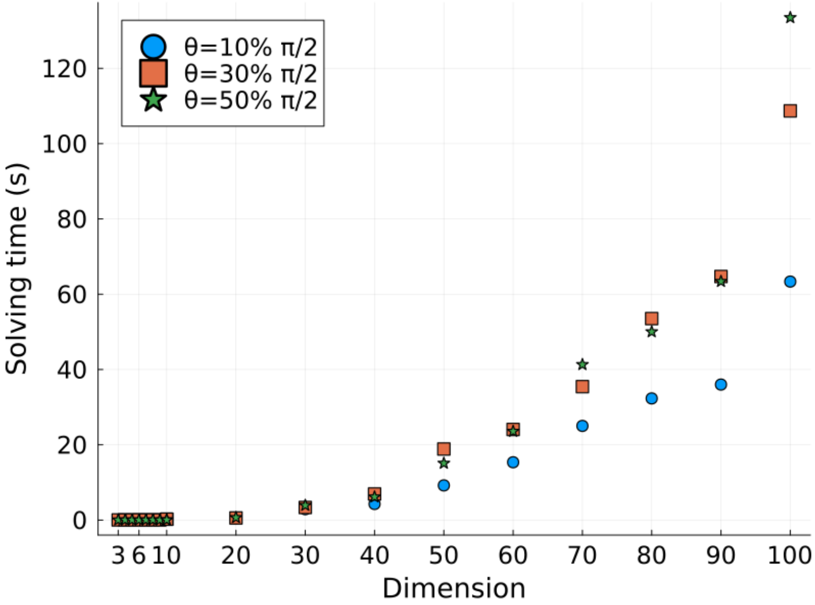
However for ℓ_1/ℓ_2 !

- ℓ_1/ℓ_2 is a surrogate function for ℓ_0 in compressed sensing
- ℓ_1/ℓ_2 is Capra-convex
  (and an absolute function so Fermat rule is no help)
- and ℓ_1/ℓ_2 is continuous
  and the following Capra-abstract subgradients

$$\left\{x \mapsto \langle \varrho(x), y \rangle - \ell_0^{¢}(y)\right\}_{y \in \{-1,0,1\}^n}$$

  are uniformly continuous
- Most assumptions of theoretical convergence results
  [Pallaschke and Rolewicz, 1997, Theorem 9.1.1] are satisfied

Solving time for the ratio of two norms



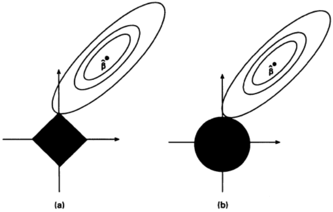
Work needs to be done for theoretical guarantees

- Convergence results
  [Pallaschke and Rolewicz, 1997, Theorem 9.1.1]
  [Rubinov, 2000, Proposition 9.2]
- But the assumptions do not fit our case:
  need to be adapted

Outline of the presentation

The intuition behind lasso

$$\min_{x \in \mathbb{R}^n} \left(f(x) + \gamma \|x\|_1\right)$$



$$\min_{x \in \mathbb{R}^n} \left(f(x) + \gamma \|x\|_2\right)$$

Comments of [Tibshirani, 1996, Figure 2]

> *"The lasso solution is the first place that the contours touch the square, and this will sometimes occur at a corner, corresponding to a zero coefficient. The picture for ridge regression is shown in Fig. 2(b): there are no corners for the contours to hit and hence zero solutions will rarely result."*

Geometric (alignment) expression of optimality condition

- We consider an optimal solution x^* of

$$\min_{x \in \mathbb{R}^n} \left(f(x) + \gamma \|x\| \right)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth convex function, $\gamma > 0$ and $\|\cdot\|$ is a norm with unit ball B

- By the Fermat rule, when $x^* \neq 0$,

$$0 \in \nabla f(x^*) + \gamma \partial \|\cdot\|(x^*) \iff \frac{x^*}{\|x^*\|} \in \underbrace{F_\perp(B, -\nabla f(x^*))}_{\substack{\text{face of the unit ball } B \\ \text{exposed by } -\nabla f(x^*)}}$$

- The norm $\|\cdot\|$ may be qualified as sparsity-inducing if information about the support of x^* and the exposed faces of the unit ball B can be recovered from one another [Fan, Jeong, Sun, and Friedlander, 2020]

Design of sparsity inducing norms/balls
for k-sparse vectors with given k

Courtesy of Basile and Lionel Pournin

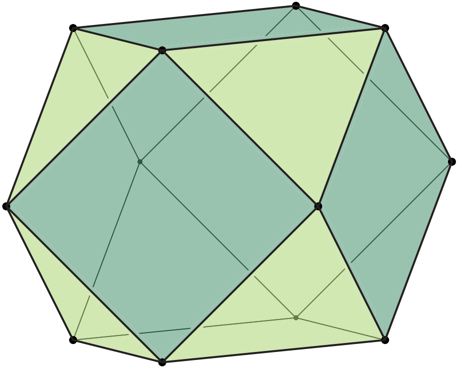


Figure: Unit ball $\overline{\mathrm{co}}\left(\ell_0^{\leq 2} \cap S_1\right)$ when $n = 3$

How to design a sparsity inducing unit ball?

For $k \in [\![1, d]\!]$

- consider the k-sparse vectors in $\ell_0^{\leq k}$
- as they do not form a compact set,
  intersect $\ell_0^{\leq k}$ with a unit sphere S (or a unit ball B)
- form the convex hull and obtain a new

$$\text{unit ball} \quad B_{(k)}^{\mathcal{R}} = \mathrm{co}\big(\ell_0^{\leq k} \cap S\big)$$

whose extreme points belong to $\ell_0^{\leq k} \cap S \subset \ell_0^{\leq k}$,
hence are k-sparse vectors

Does this procedure induces sparsity? If yes, in what sense?

Support identification of a k-sparse vector in the exposed face of a generalized k-support dual norm (1/2)

Theorem

Let $k \in [\![1, n]\!]$. If the source norm $\|\cdot\|$ is orthant-monotonic, then

$$B^{\mathcal{R}}_{(k)} = \mathrm{co}\big(\ell_0^{\leq k} \cap S\big) = B^{\top\star}_{\star,(k)}$$

and, for any nonzero dual vector $y \in \mathcal{Y} \setminus \{0\}$,
the two following statements are equivalent

(i) $x \in \ell_0^{\leq k} \cap F_{\perp}(B^{\top\star}_{\star,(k)}, y)$

(ii) there exists $K^* \in \arg\max_{|K| \leq k} \|\pi_K(y)\|_\star$
such that $x \in \pi_{K^*}\big(B \cap F_{\perp}(B, \pi_{K^*}(y))\big) \subset \mathcal{R}_{K^*}$

As a consequence, we get that

$$\mathrm{supp}(x) \subset K^*$$

Support identification of a k-sparse vector in the exposed face of a generalized k-support dual norm (2/2)

$$\text{Consider a vector } x \in \overbrace{\ell_0^{\leq k}}^{k\text{-sparse}} \cap \overbrace{F_{\perp}(B_{\star,(k)}^{\top\star}, y)}^{\text{exposed face}}$$

1. From $x \in \ell_0^{\leq k}$, we only know that

$$\text{there exists } K \subset [\![1, n]\!] \text{ with } |K| \leq k \text{ such that}$$

$$\mathrm{supp}(x) \subset K$$

2. From $x \in F_{\perp}(B_{\star,(k)}^{\top\star}, y)$, we add information and obtain that

$$\text{there exists } K^* \in \arg\max_{|K| \leq k} \|\pi_K(y)\|_{\star} \text{ such that}$$

$$\mathrm{supp}(x) \subset K^*$$

Support identification

Corollary

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth convex function, $\gamma > 0$ and $\|\cdot\|$ be a norm

Then, an optimal solution x^* of

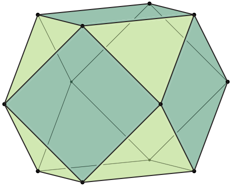
$$\min_{x \in \mathbb{R}^n} \left(f(x) + \gamma \|x\|_{\star,(k)}^{\top\star} \right)$$

has support

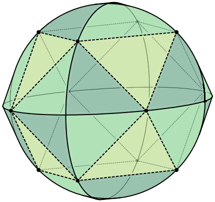
$$\mathrm{supp}(x^*) \subset \bigcup_{\substack{K^* \in \arg\max_{|K| \leq k} \\ \|\pi_K(-\nabla f(x^*))\|_\star}} K^*$$

Especially interesting when the $\arg\max_{|K|\leq k}$ is unique, because then the optimal solution x^* is k-sparse

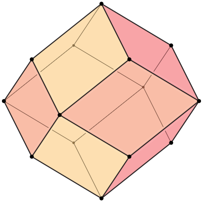
Geometry of sparsity inducing balls



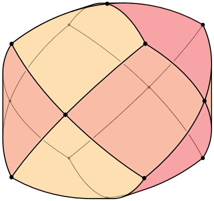
(a) Unit ball $B^{\top\star}_{\infty,2}$ when $n = 3$



(b) Unit ball $B^{\top\star}_{2,2}$ when $n = 3$



(c) Unit ball $B^{\top}_{1,2}$ when $n = 3$



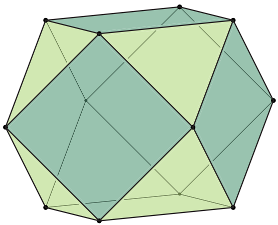
(d) Unit ball $B^{\top}_{2,2}$ when $n = 3$

Figure: Four top (6c and 6d) and support (7a and 7b) unit balls, either obtained from the ℓ_1 source norm (7a and 6c) or from the ℓ_2 source norm (7b and 6d)

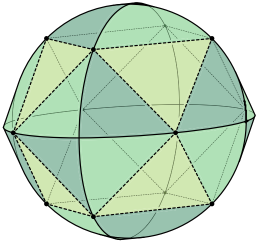
Additional geometric properties

Proposition

For any $k \in [\![1, n]\!]$, all the proper faces of $B_{2,k}^{\top\star}$ are hypersimplices, and the normal fan of $B_{2,k}^{\top\star}$ refines the normal fan of $B_{\infty,k}^{\top\star}$



(a) Unit ball $B_{\infty,2}^{\top\star}$ when $n = 3$



(b) Unit ball $B_{2,2}^{\top\star}$ when $n = 3$

Figure: Two support norm unit balls, either obtained from the ℓ_1 source norm (7a) or from the ℓ_2 source norm (7b)

Outline of the presentation

Crash course on generalized convexity [5 min]

Capra conjugacies [20 min]

Towards Capra-algorithms in sparse optimization? [15 min]

Conclusion [1 min]

Additional material

- So-called generalized convexity and Fenchel-Moreau conjugacy are extensions of duality beyond convex analysis
- The Capra-coupling ¢ and induced Capra-conjugacy seem promising to handle sparsity in optimization as the pseudonorm ℓ_0 satisfies

$$\partial_{¢}\ell_0 \neq \emptyset \text{ hence } \ell_0^{¢¢'} = \ell_0$$

  but we have problems handling sums like $\ell_0 + \iota_X$:-(
- So, our working program is now to study
  - the ℓ_0-cup function $\mathcal{L}_0 = \ell_0^{¢\star'}$
  - the geometry of unit balls of norms related to the Capra-coupling ¢ and to the pseudonorm ℓ_0
  - lower bound convex programs

Andreas Argyriou, Rina Foygel, and Nathan Srebro. Sparse prediction with the k-support norm. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pages 1457–1465, USA, 2012. Curran Associates Inc.

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. Constant along primal rays conjugacies and the l_0 pseudonorm. *Optimization*, 71(2):355–386, 2022a. doi: 10.1080/02331934.2020.1822836.

Jean-Philippe Chancelier and Michel De Lara. Orthant-strictly monotonic norms, generalized top-k and k-support norms and the l0 pseudonorm. *Journal of Convex Analysis (to appear)*, 2022b.

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, 2022c.

Zhenan Fan, Halyun Jeong, Yifan Sun, and Michael P. Friedlander. Atomic decomposition via polar alignment. *Foundations and Trends® in Optimization*, 3(4):280–366, 2020.

D. Gries. Characterization of certain classes of norms. *Numerische Mathematik*, 10:30–41, 1967.

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.

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

J. J. Moreau. Fonctionnelles convexes. *Séminaire Jean Leray*, 2:1–108, 1966-1967.

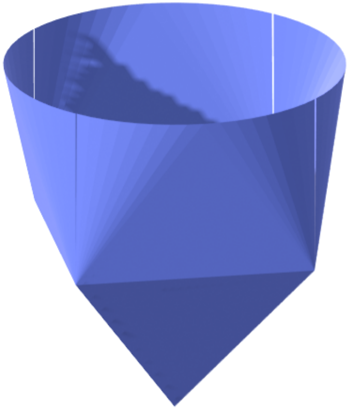
Jean Jacques Moreau. Inf-convolution, sous-additivité, convexité des fonctions numériques. *J. Math. Pures Appl. (9)*, 49:109–154, 1970.

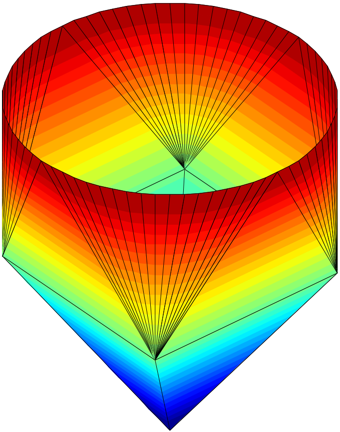
Diethard Pallaschke and Stefan Rolewicz. *Foundations of mathematical optimization*, volume 388 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1997. ISBN 0-7923-4424-3.

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.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996. ISSN 00359246. URL http://www.jstor.org/stable/2346178.

Thank you :-)





Outline of the presentation

The ℓ_0 pseudonorm is (almost) a convex-composite function

- [Chancelier and De Lara, 2021]

$$\boxed{\ell_0(x) = \underbrace{\mathcal{L}_0}_{\text{proper convex lsc}}\Big(\frac{x}{\|x\|}\Big)} \,, \quad \forall x \in \mathbb{R}^n \setminus \{0\}$$

- As a consequence,
  if $C \subset \mathbb{R}^n$ is a closed convex set with $0 \notin C$,

$$\min_{x \in C} \ell_0(x) = \min_{x \in \mathbb{R}^n} \Big\{ \mathcal{L}_0\big(\frac{x}{\|x\|}\big) + \iota_C(x) \Big\}$$

  or if $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a proper convex lsc function,

$$\min_{x \in \mathbb{R}^n, \ell_0(x) \leq k} f(x) = \min_{x \in \mathbb{R}^n} \Big\{ f(x) + \underbrace{\iota_{B^{\top\star}_{(k)}}}_{\substack{(2,k)\text{-support norm} \\ \text{unit ball}}}\big(\frac{x}{\|x\|}\big) \Big\}$$

Graded sequence of norms

We define graded sequence of norms

A graded sequence of norms detects the number of nonzero components of a vector in $\mathbb{R}^n$

when the sequence becomes stationary

Definition

We say that a sequence $\{\|\cdot\|_k\}_{k \in [\![1,n]\!]}$ of norms is (increasingly) graded with respect to the ℓ_0 pseudonorm if, for any $y \in \mathbb{R}^n$ and $l \in [\![1,n]\!]$, we have

$$\ell_0(y) = \ell \iff \|y\|_1 \leq \cdots \leq \|y\|_{\ell-1} < \|y\|_\ell = \cdots = \|y\|_n$$

or, equivalently, $k \in [\![1,n]\!] \mapsto \|y\|_k$ is nondecreasing and

$$\ell_0(y) \leq \ell \iff \|y\|_\ell = \|y\|_n$$

Graded sequences are suitable for so-called "difference of convex" (DC) optimization methods to tackle sparse $\ell_0(y) \leq l$ constraints

Orthant-strictly monotonic dual norms produce graded sequences of norms

Proposition

If the dual norm $\|\cdot\|_\star$ of the source norm $\|\cdot\|$ is orthant-strictly monotonic, then the sequence

$$\underbrace{\left\{\|\cdot\|^{\top}_{\star,(k)}\right\}_{k\in[\![1,n]\!]}}_{\text{generalized top-}k\text{ dual norm}} = \underbrace{\left\{\|\cdot\|^{\mathcal{R}}_{(k),\star}\right\}_{k\in[\![1,n]\!]}}_{\text{dual-}k\text{ coordinate norm}}$$

is graded with respect to the ℓ_0 pseudonorm

Thus, we can produce families of graded sequences of norms suitable for "difference of convex" (DC) optimization methods to tackle sparse constraints

Fenchel *versus* Capra conjugacies for ℓ_0

[Chancelier and De Lara, 2022a], [Chancelier and De Lara, 2022c]
If both the source norm and its dual are orthant-strictly monotonic

Fenchel conjugacy	Capra conjugacy
$\iota^{\star}_{\ell_0^{\leq k}} = \iota_{\{0\}}, \; k \neq 0$	$\iota^{\cent}_{\ell_0^{\leq k}} = \|\cdot\|^{\mathcal{R}}_{(k),\star} = \|\cdot\|^{\top}_{\star,(k)}$
$\ell_0^{\star} = \iota_{\{0\}}$	$\ell_0^{\cent} = \sup_{\ell \in [\![0,n]\!]} \Big[\|\cdot\|^{\mathcal{R}}_{(\ell),\star} - \ell \Big]$ $= \sup_{\ell \in [\![0,n]\!]} \Big[\|\cdot\|^{\top\star}_{\star,(\ell)} - \ell \Big]$
$\iota^{\star\star'}_{\ell_0^{\leq k}} = 0$	$\iota^{\cent\cent'}_{\ell_0^{\leq k}} = \iota_{\ell_0^{\leq k}}$
$\ell_0^{\star\star'} = 0$	$\ell_0^{\cent\cent'} = \ell_0$

Lower bounds for the pseudonorm ℓ_0

Best ratio of norms [Chancelier and De Lara, 2022a]

- For any $\varphi : [\![0, d]\!] \to [0, +\infty[$, such that $\varphi(j) > \varphi(0) = 0$ for all $j \in [\![1, d]\!]$, there exists a norm $\|\cdot\|^{\mathcal{R}}_{(\varphi)}$ such that

$$\frac{\|x\|^{\mathcal{R}}_{(\varphi)}}{\|x\|} \leq \varphi\big(\ell_0(x)\big) \;, \;\; \forall x \in \mathbb{R}^n \backslash \{0\}$$

where $\|\cdot\|^{\mathcal{R}}_{(\varphi)}$ is characterized by its dual norm

$$\|y\|^{\mathcal{R}}_{(\varphi),\star} = \sup_{j \in [\![1,d]\!]} \frac{\|y\|^{\mathcal{R}}_{(j),\star}}{\varphi(j)} \;, \;\; \forall y \in \mathbb{R}^n$$

- For $\|\cdot\| = \|\cdot\|_p$ with $p > 1$, and $\varphi_\alpha(j) = j^{1/\alpha}$ for $\alpha > 0$,

$$\left(\frac{\big(\|x\|_p\big)^{\mathcal{R}}_{(\varphi_\alpha)}}{\|x\|_p}\right)^{\alpha} \leq \ell_0(x) \;, \;\; \forall x \in \mathbb{R}^n \backslash \{0\}$$

$$\left(\frac{\|x\|_1}{\|x\|_p}\right)^{p} \leq \ell_0(x) \;, \;\; \forall x \in \mathbb{R}^n \backslash \{0\}$$

Lower bound convex programs for exact sparse optimization

Concave dual problem for exact sparse optimization

From
$$\sup_{y \in \mathcal{Y}} \Big(\big(-f^{¢}(y)\big) \dotplus \big(-\iota_X^{-¢}(y)\big) \Big) \leq \inf_{x \in \mathcal{X}} \Big(f(x) \dotplus \iota_X(x) \Big)$$

we deduce that

$$\sup_{y \in \mathbb{R}^n} \Big(-\big(\inf \left[f \mid \varrho \right]\big)^{\star}(y) \dotplus \big(-\underbrace{\iota_{\ell_0^{\leq k}}^{-¢}(y)}_{\|y\|_{2,k}^{\top}}\big) \Big) \leq \inf_{\ell_0(x) \leq k} f(x)$$

Proposition

For any function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, we have the following lower bound

$$\sup_{y \in \mathbb{R}^n} \overbrace{\Big(-\big(\inf \left[f \mid \varrho \right]\big)^{\star}(y) - \|y\|_{2,k}^{\top} \Big)}^{\text{concave usc function}} \leq \inf_{\ell_0(x) \leq k} f(x)$$
$$= \inf_{\ell_0(x) \leq k} \inf \left[f \mid \varrho \right](x)$$

Convex primal problem for exact sparse optimization

Proposition

Under a mild technical assumption (“à la” Fenchel-Rockafellar), namely if $\big(\inf\big[f \mid \varrho\big]\big)^{\star}$ is a proper function, we have the following lower bound

$$\min_{\|x\|_{2,k}^{\top\star}\leq 1} \big(\inf\big[f \mid \varrho\big]\big)^{\star\star'}(x) \leq \inf_{\ell_0(x)\leq k} f(x) = \inf_{\ell_0(x)\leq k} \inf\big[f \mid \varrho\big](x)$$

The primal problem is the minimization of a closed convex function on the unit ball of the $(2,k)$-support norm $\|\cdot\|_{2,k}^{\top\star}$ (introduced in [Argyriou, Foygel, and Srebro, 2012])

Duality

Perturbation scheme

- Functions $k : \mathbb{R}^n \to \overline{\mathbb{R}}$, $\varphi : [\![0,n]\!] \to \overline{\mathbb{R}}$ nondecreasing (ex: identity, $\iota_{[\![0,k]\!]}$) and original minimization problem

$$\inf_{w\in\mathbb{R}^n} \Big\{ k(w) \dotplus \varphi\big(\ell_0(w)\big) \Big\} = \inf_{w\in\mathbb{R}^n} \Big\{ k(w) \dotplus \big(\varphi \circ \ell_0\big)^{¢\star'}\big(\varrho(w)\big) \Big\}$$

because $\varphi \circ \ell_0 = \big(\varphi \circ \ell_0\big)^{¢¢'} = \big(\varphi \circ \ell_0\big)^{¢\star'} \circ \varrho$ [Chancelier and De Lara, 2022c]

- Rockafellian (perturbation scheme) $R : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$

$$R(w,x) = k(w) \dotplus \big(\varphi \circ \ell_0\big)^{¢\star'}\big(\varrho(w) + x\big) \; , \quad \forall (w,x) \in \mathbb{R}^n \times \mathbb{R}^n$$

- Value function

$$\varphi(x) = \inf_{w\in\mathbb{R}^n} \Big\{ k(w) \dotplus \big(\varphi \circ \ell_0\big)^{¢\star'}\big(\varrho(w) + x\big) \Big\} \; , \quad \forall x \in \mathbb{R}^n$$

Lagrangian and dual problem

- Fenchel coupling $\mathbb{R}^n \overset{\langle\cdot|\cdot\rangle}{\leftrightarrow} \mathbb{R}^n$, and Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ given, for any $(w, y) \in \mathbb{R}^n \times \mathbb{R}^n$, by

$$\mathcal{L}(w,y) = \inf_{x\in\mathbb{R}^n} \Big\{ k(w) \dot{+} (\varphi \circ \ell_0)^{¢\star'}(\varrho(w) + x) - \langle x,\ y\rangle \Big\}$$
$$= k(w) \dot{+} \big(\langle \varrho(w),\ y\rangle - (\varphi \circ \ell_0)^{¢}(y)\big)$$

- Dual maximization problem

$$\varphi^{\star\star'}(0) = \sup_{y\in\mathbb{R}^n} \inf_{w\in\mathbb{R}^n} \mathcal{L}(w,y) = \sup_{y\in\mathbb{R}^n} \Big\{ \big(-k^{-¢}(y)\big) \underset{\cdot}{+} \big(-(\varphi \circ \ell_0)^{¢}(y)\big) \Big\}$$

- Original minimization problem (case "$\dot{+} = +$" when k proper)

$$\varphi(0) = \inf_{w\in\mathbb{R}^n} \sup_{y\in\mathbb{R}^n} \mathcal{L}(w,y) = \inf_{w\in\mathbb{R}^n} \Big\{ k(w) \dot{+} \varphi(\ell_0(w)) \Big\}$$

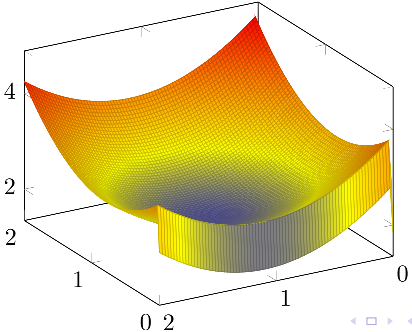
Numerics

A toy example

$$\min_{w \in \mathbb{R}^2} \overbrace{\left((w_1 - b_1)^2 + (w_2 - b_2)^2\right)}^{k(w)} + \ell_0(w)$$

$$\text{with} \quad b = (0.8, 1.1)$$

We have that $\{(0, b_2)\} = \{(0, 1.1)\} = \arg\min_{w \in \mathbb{R}^2} \left\{ k(w) + \ell_0(w) \right\}$



The toy example as a min-max problem

As $\ell_0(w) = \max_{y \in \mathbb{R}^2} \left\{ ¢(w, y) - \ell_0^{¢}(y) \right\}$, we obtain that

$$\min_{w \in \mathbb{R}^2} \left\{ k(w) + \ell_0(w) \right\} = \min_{w \in \mathbb{R}^2} \max_{y \in \mathbb{R}^2} \left\{ k(w) + ¢(w, y) - \ell_0^{¢}(y) \right\}$$

with

$$\ell_0^{¢}(y) = \sup_{k \in [\![1, n]\!]} \left[\|y\|_{2,k}^{\top} - k \right]_+$$

Generalized primal-dual proximal splitting

GPDPS Algorithm Christian Clason, Stanislav Mazurenko, and Tuomo Valkonen. Primal–dual proximal splitting and generalized conjugation in non-smooth non-convex optimization. *Applied Mathematics and Optimization*, 84(2):1239–1284, apr 2020.

Given a starting point (w_0, y_0) and step lengths $\tau_i, \omega_i, \sigma_i > 0$, iterate

$$
\begin{aligned}
w^{(i+1)} &:= \operatorname{prox}_{\tau_i k}\big(w^{(i)} - ¢_w(w^{(i)}, y^{(i)})\big)\\
\overline{w}^{(i+1)} &:= w^{(i+1)} + \omega_i(w^{(i+1)} - w^{(i)})\\
y^{(i+1)} &:= \operatorname{prox}_{\sigma_i \ell_0^{¢}}\big(y^{(i)} + \sigma_i ¢_y(\overline{w}^{(i+1)}, y^{(i)})\big)
\end{aligned}
$$

The prox of k is analytically computed (quadratic function), whereas the prox of $\ell_0^{¢}$ is numerically computed with the optimization algorithm `newuoa` by M.J.D. Powell

GPDPS convergence, varying the starting point