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 nationale des ponts et chaussées, France

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

GdR IASIS
2024, September 16
IHP Paris

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

The function ℓ_0 pseudonorm : $\mathbb{R}^n \to [\![0, n]\!]$ satisfies 3 out of 4 axioms of a norm

- $\ell_0(x) \geq 0$ ✓
- $\Big(\ell_0(x) = 0 \iff x = 0 \Big)$ ✓
- $\ell_0(x + x') \leq \ell_0(x) + \ell_0(x')$ ✓
- But... instead of absolute 1-homogeneity, it is absolute 0-homogeneity that holds true

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

$$\mathrm{supp}(\lambda x) = \mathrm{supp}(x) \; , \;\; \forall \lambda \neq 0$$

SNAPSHOTS OF OUR MAIN RESULTS

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

- Fenchel conjugacy ($\star$)

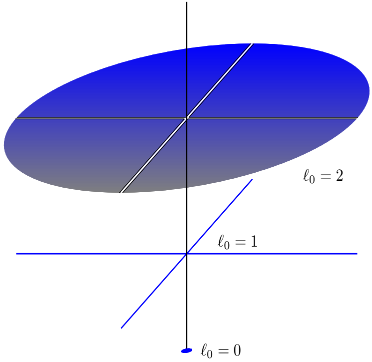
$$\ell_0^{\star\star'} = 0$$

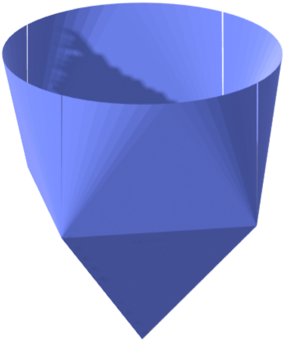
- E-Capra conjugacy ($\cent$)

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

[Chancelier and De Lara, 2021]

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





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,

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

Variational formulas for the ℓ_0 pseudonorm

Proposition

[Chancelier and De Lara, 2021]

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

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

END OF THE TEASER

Talk outline

Outline of the presentation

Couplings

Motivation: Legendre transform and Fenchel conjugacy in convex analysis

Definition

Two vector spaces $\mathcal{X}$ and $\mathcal{Y}$, paired by a bilinear form $\langle \ , \ \rangle$, (in the sense of convex analysis [Rockafellar, 1974, p. 13])) 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

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

Coupling function between sets

[Moreau, 1966-1967, 1970]

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

$$c : \mathcal{U} \times \mathcal{V} \to \overline{\mathbb{R}}$$

We also use the notation $\mathcal{U} \overset{c}{\leftrightarrow} \mathcal{V}$ for a coupling

Coupling c $c(u,v)$	c-convex functions $f^{cc'} = f$
$\langle u , v \rangle$	closed convex $f^{\star\star'} = f$
$u(v)$, u continuous	lower semicontinuous
$\log \langle u , v \rangle_+$	$\log \circ$ sublinear
$-N \lVert u - v \rVert^{\alpha}$, $0 < \alpha \leq 1$	α-Hölder continuous with constant N
$\min_{i, v_i > 0} u_i v_i$	increasing and convex-along-rays
Capra $¢(u,v) = \langle \frac{u}{\lVert u \rVert} , v \rangle$	$\ell_0^{¢¢'} = \ell_0$
$\mathcal{H}_0$	convex $\circ$ 0-homogeneous

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

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

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

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

E-Capra-conjugate of the ℓ_0 pseudonorm

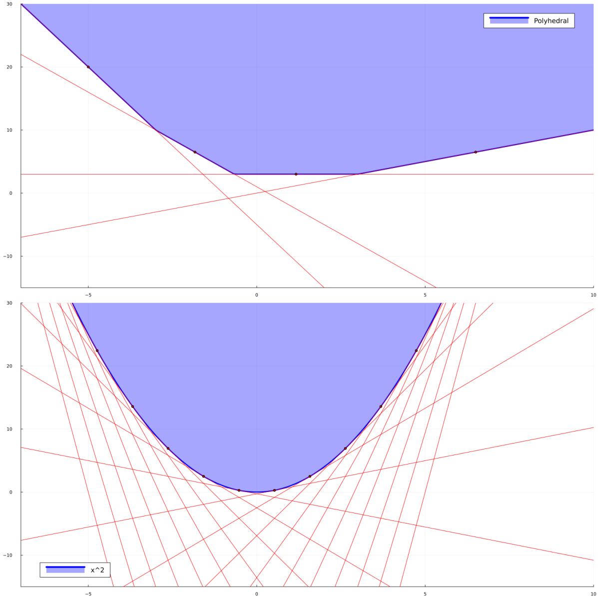
$$
\begin{aligned}
\ell_0^{\cent}(y) &= \sup_{x\in\mathbb{R}^n} \Big\{ \cent(x,y) \stackrel{\cdot}{+} \big(-\ell_0(x)\big) \Big\} \\
&= \sup \Big\{ 0, \sup_{x\neq 0} \Big\{ \frac{\langle x \mid y \rangle}{\|x\|_2} - \ell_0(x) \Big\} \Big\} \\
&= \sup \Big\{ 0, \sup_{s\in S_2} \big\{ \langle s \mid y \rangle - \ell_0(s) \big\} \Big\}
\end{aligned}
$$

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

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

Biconjugates and duality

Motivation: duality in convex analysis



Reverse coupling and Fenchel-Moreau biconjugate

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

$$c' : \mathcal{V} \times \mathcal{U} \to \overline{\mathbb{R}} \, , \;\; c'(v,u) = c(u,v) \, , \;\; \forall (v,u) \in \mathcal{V} \times \mathcal{U}$$

$$f \in \overline{\mathbb{R}}^{\mathcal{U}} \mapsto f^{c} \in \overline{\mathbb{R}}^{\mathcal{V}}$$

$$g \in \overline{\mathbb{R}}^{\mathcal{V}} \mapsto g^{c'} \in \overline{\mathbb{R}}^{\mathcal{U}}$$

Reverse coupling and Fenchel-Moreau biconjugate

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

$$c' : \mathcal{V} \times \mathcal{U} \to \overline{\mathbb{R}} \;, \;\; c'(v,u) = c(u,v) \;, \;\; \forall (v,u) \in \mathcal{V} \times \mathcal{U}$$

$$f \in \overline{\mathbb{R}}^{\mathcal{U}} \mapsto f^{c} \in \overline{\mathbb{R}}^{\mathcal{V}}$$

$$g \in \overline{\mathbb{R}}^{\mathcal{V}} \mapsto g^{c'} \in \overline{\mathbb{R}}^{\mathcal{U}}$$

$$g^{c'}(u) = \sup_{v \in \mathcal{V}} \Big(c(u,v) \dot{+} \big(-g(v)\big) \Big) \;, \;\; \forall u \in \mathcal{U}$$

$$f^{cc'}(u) = \big(f^{c}\big)^{c'}(u) = \sup_{v \in \mathcal{V}} \Big(c(u,v) \dot{+} \big(-f^{c}(v)\big) \Big) \;, \;\; \forall u \in \mathcal{U}$$

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)

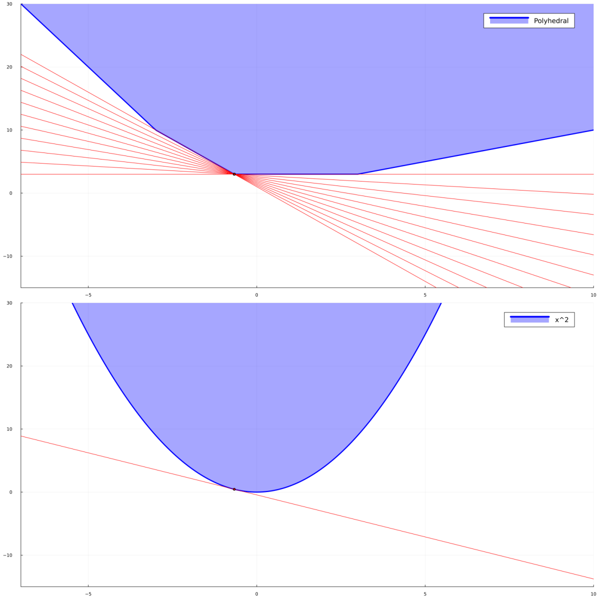
If the function $f : \mathcal{U} \to \overline{\mathbb{R}}$ is c-convex, we have that

$$f(u) = \sup_{v \in \mathcal{V}} \underbrace{\Big(c(u,v) \dotplus \big(-f^c(v)\big) \Big)}_{\text{elementary function of } u} , \quad \forall u \in \mathcal{U}$$

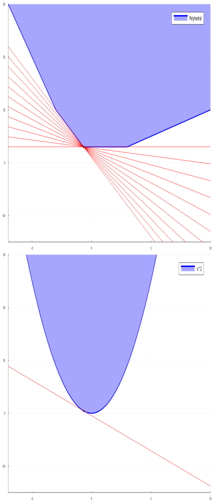
Example: $\star$-convex functions
$=$ closed convex functions
$=$ proper convex lsc or $\equiv -\infty$ or $\equiv +\infty$
$=$ suprema of affine functions

Subdifferential

Motivation: subgradients in convex analysis



Motivation: Rockafellar-Moreau subdifferential in convex analysis



$$
\begin{aligned}
& y \in \partial f(x) \\
\iff & f(x) + f^{\star}(y) = \langle x,\ y \rangle \\
\iff & f^{\star}(y) = \langle x,\ y \rangle - f(x) \\
\iff & x \in \arg\max_{u \in \mathcal{X}} \Big[\langle u,\ y \rangle - f(u) \Big] \\
\iff & \langle u,\ y \rangle - f(u) \leq \langle x,\ y \rangle - f(x) \\
& \forall u \in \mathcal{X}
\end{aligned}
$$

Subdifferentials of a conjugacy

For any function $f : \mathcal{U} \to \overline{\mathbb{R}}$ and $u \in \mathcal{U}$, $v \in \mathcal{V}$

Definition

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

$$v \in \partial^c f(u) \iff f(u) = c(u,v) \dot{+} \big(-f^c(v)\big)$$

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

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

Definition

Lower subdifferential

$$v \in \partial_c f(u) \iff f^c(v) = c(u,v) \dot{+} \big(-f(u)\big)$$

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

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

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

E-Capra = Fenchel coupling after primal normalization

- We introduce the Euclidean unit sphere S_2 and the pointed unit sphere $S_2^{(0)}$ by

$$S_2 = \{x \in \mathbb{R}^n \mid \|x\|_2 = 1\} , \quad S_2^{(0)} = S_2 \cup \{0\}$$

- and we define the primal radial projection ϱ as

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

- so that the coupling E-Capra

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

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

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

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

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*

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

- 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

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

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

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

The ℓ_0 pseudonorm displays hidden convexity

The ℓ_0 pseudonorm displays a convex factorization property

Theorem

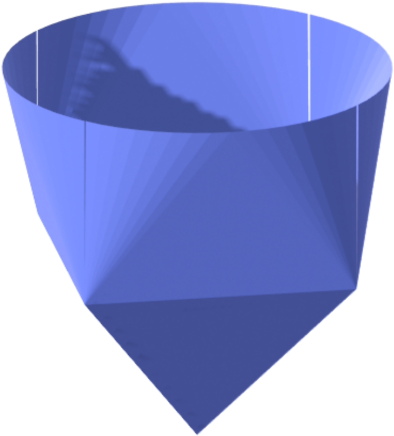
As the ℓ_0 pseudonorm is E-Capra-convex, we get that

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

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

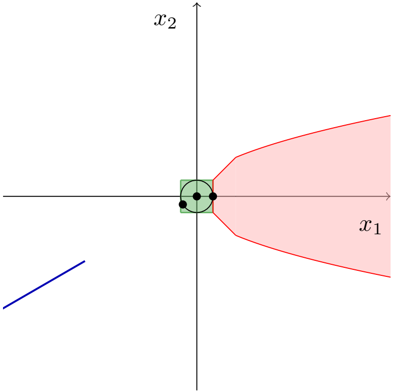
on the Euclidean unit sphere S_2

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

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

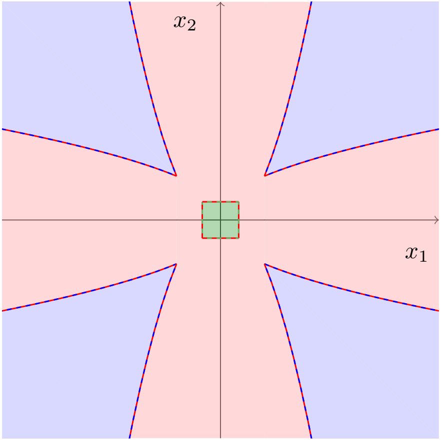
Capra-subdifferential of the ℓ_0 pseudonorm on $\mathbb{R}^2$

Illustration at three points (black dots)



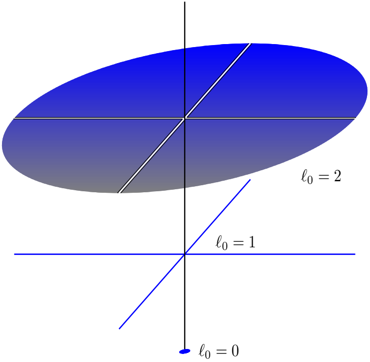
$\partial_{¢}\ell_0(0,0)$, $\partial_{¢}\ell_0(1,0)$, $\partial_{¢}\ell_0(-\frac{\sqrt{3}}{2},-\frac{1}{2})$

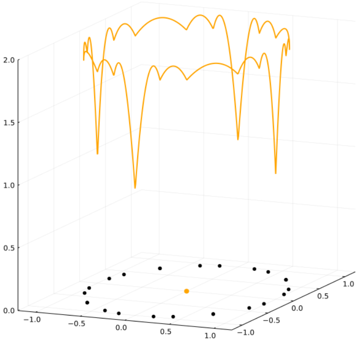
Capra-subdifferential of the ℓ_0 pseudonorm on $\mathbb{R}^2$



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

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

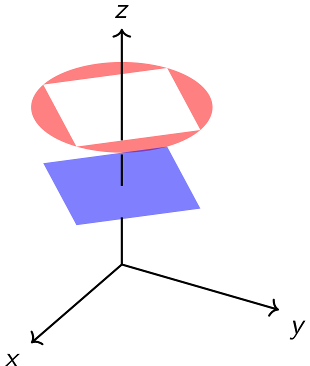
  $$B_{(1)}^{\top\star} \subset \cdots \subset B_{(\ell-1)}^{\top\star} \subset B_{(\ell)}^{\top\star} \subset \cdots \subset B_{(n)}^{\top\star} = B$$

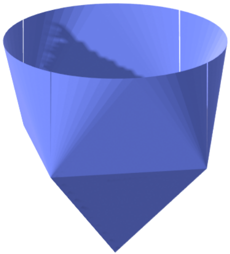
The ℓ_0-cup function as a convex envelope

Proposition

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

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





Variational formulas for the ℓ_0 pseudonorm

Proposition

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

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

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 ($\ell_0^{¢¢'} = \ell_0$) [Chancelier and De Lara, 2021] hidden convexity [Chancelier and De Lara, 2021] variational formula [Chancelier and De Lara, 2021] subdifferential [Le Franc et al., 2022]	difference of norms [Chancelier and De Lara, 2023]		
$\varphi \circ \ell_0$ $\varphi : \mathbb{N} \to \overline{\mathbb{R}}$ nondecreasing		¢-convex ($(\varphi \circ \ell_0)^{¢¢'} = \varphi \circ \ell_0$) [Chancelier and De Lara, 2022b] hidden convexity [Chancelier and De Lara, 2022b] variational formula [Chancelier and De Lara, 2022b] subdifferential [Chancelier and De Lara, 2022b]		
$\varphi \circ \ell_0$ $\varphi : \mathbb{N} \to \overline{\mathbb{R}}$ any			$(\varphi \circ \ell_0)^{¢¢'}$ [Chancelier and De Lara, 2022a] variational inequality [Chancelier and De Lara, 2022a] subdifferential [Chancelier and De Lara, 2022a]	
$F \circ$ support $F : 2^{[\![1,d]\!]} \to \overline{\mathbb{R}}$ any			$(F \circ \mathrm{support})^{¢¢'}$ [preprint] variational inequality [preprint] 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-strictly monotonic norms

Orthant-strictly monotonic norms

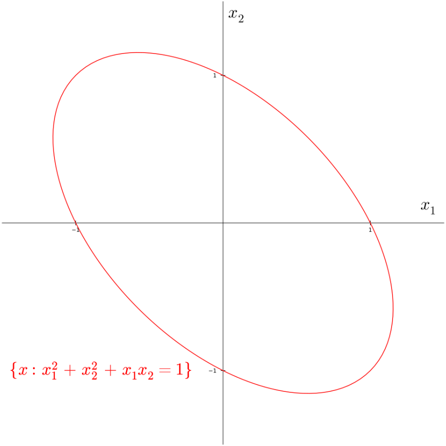
For any $x \in \mathbb{R}^n$, we denote by $|x|$
the vector of $\mathbb{R}^n$ with components $|x_i|$, $i \in [\![1, n]\!]$

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
  $\Big(|x| \leq |x'| \text{ and } x \circ x' \geq 0 \Rightarrow \|x\| \leq \|x'\| \Big)$,
  where $x \circ x' = (x_1 x'_1, \ldots, x_d x'_d)$
  is the Hadamard (entrywise) product
- orthant-strictly monotonic [Chancelier and De Lara, 2023]
  if, for all x, x' in $\mathbb{R}^n$, we have
  $\Big(|x| < |x'| \text{ and } x \circ x' \geq 0 \Rightarrow \|x\| < \|x'\| \Big)$,
  where $|x| < |x'|$ means that there exists $j \in [\![1, n]\!]$
  such that $|x_j| < |x'_j|$

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

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

Examples of orthant-strictly monotonic norms

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

- 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

$$\ell_1, \ell_2$$

- The ℓ_1-norm $\|\cdot\|_1$ is orthant-strictly monotonic, whereas its dual norm, the ℓ_∞-norm $\|\cdot\|_\infty$, is orthant-monotonic, but is not orthant-strictly monotonic

Orthant-strictly monotonic norms and Capra-convexity

Capra-subdifferentiability properties of the ℓ_0 pseudonorm

[Chancelier and De Lara, 2022b]

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 function $\mathcal{L}_0$ is the best convex lsc lower approximation of ℓ_0

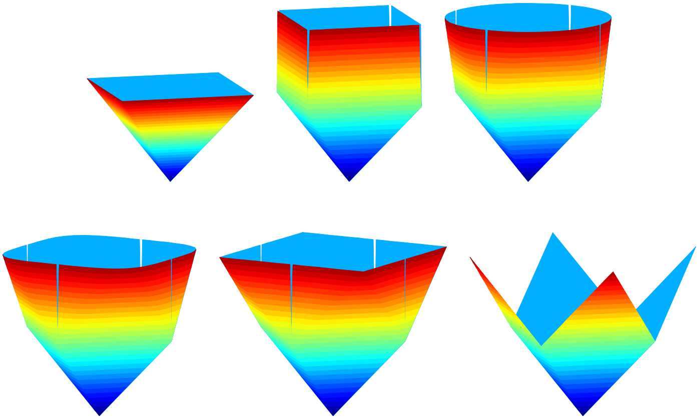
$$\text{best convex lsc function} \quad \mathcal{L}_0(x) \leq \ell_0(x) \ , \ \forall x \in B_p$$

on the unit ball B_p, 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 $p \in \{1, 1.1, 2, 4, 300, \infty\}$



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 $X \subset \mathbb{R}^d$ a nonempty set,

$$\text{minimal } \ell_0 \text{ pseudonorm} \qquad \min_{x \in X} \ell_0(x)$$

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 $f : \mathbb{R}^d \to \overline{\mathbb{R}}$,

$$\text{optimal } k\text{-sparse vector} \qquad \min_{\underbrace{\ell_0(x) \leq k}_{k\text{-sparse vectors}}} f(x)$$

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

$$\min_{x \in \mathbb{R}^d} \Big(f(x) + \underbrace{\gamma \ell_0(x)}_{\text{sparse penalty}} \Big)$$

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

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

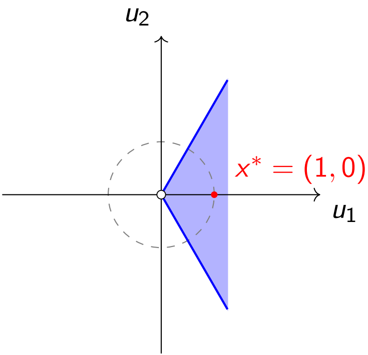
As an application, we get that

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

But. . .

Bad news :-(

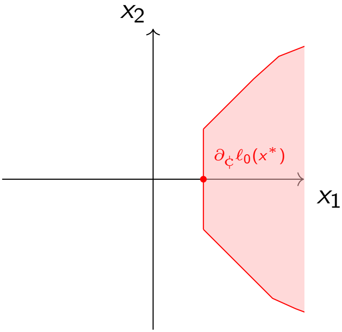
when zero is in the subdifferential of the sum...

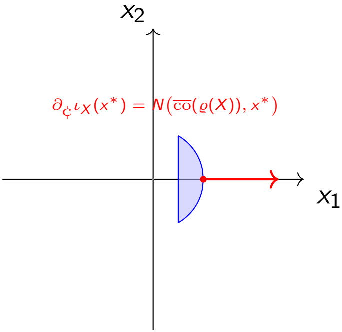


$$x^* \in \arg\min_{X} \ell_0 \iff 0 \in \underbrace{\partial_{\text{ç}}(\ell_0 + \iota_X)(x^*)}_{\text{subdifferential of the sum}}$$

… but zero is not in the sum of the subdifferentials

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





Where do we stand?

- We had good hope to handle sparse optimization problems with the E-Capra coupling that makes the pseudonorm ℓ_0 E-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

Fenchel-Moreau theorem

$$f(x) = x^2 \;, \quad \forall x \in \mathbb{R}$$

$$\Longrightarrow f(x) = \max_{y \in \mathbb{R}} \Big(\underbrace{xy - \frac{1}{4}y^2}_{\text{affine function of } x} \Big) \;, \quad \forall x \in \mathbb{R}$$

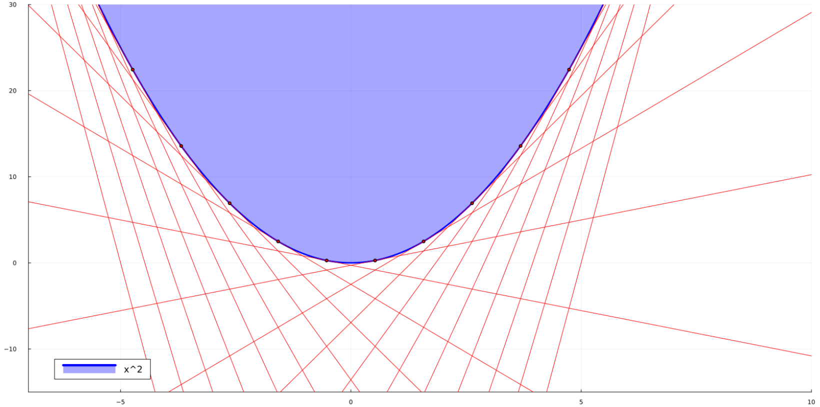


Illustration of the (Kelley) cutting plane method

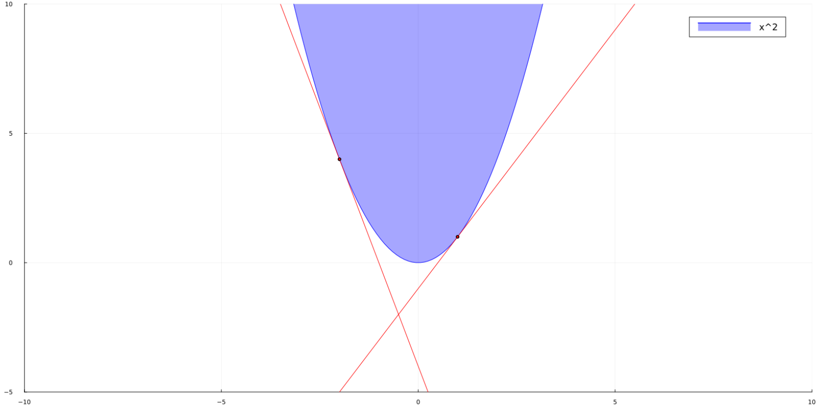


Illustration of the (Kelley) cutting plane method

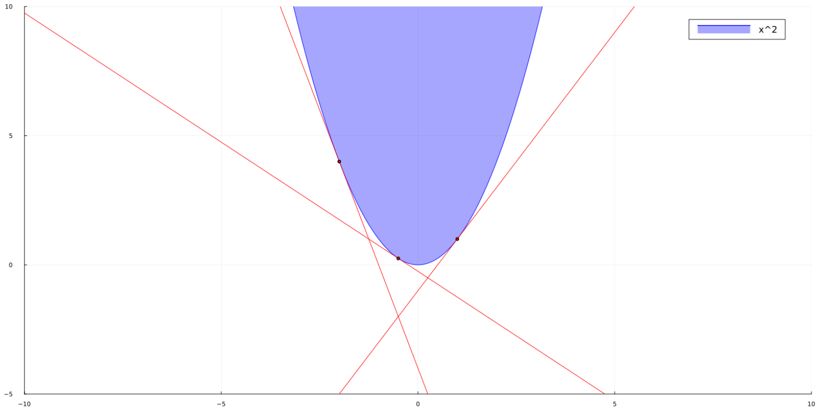


Illustration of the (Kelley) cutting plane method

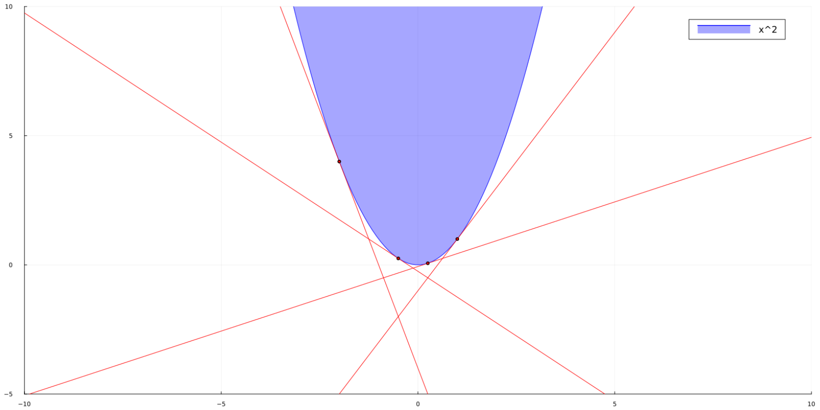
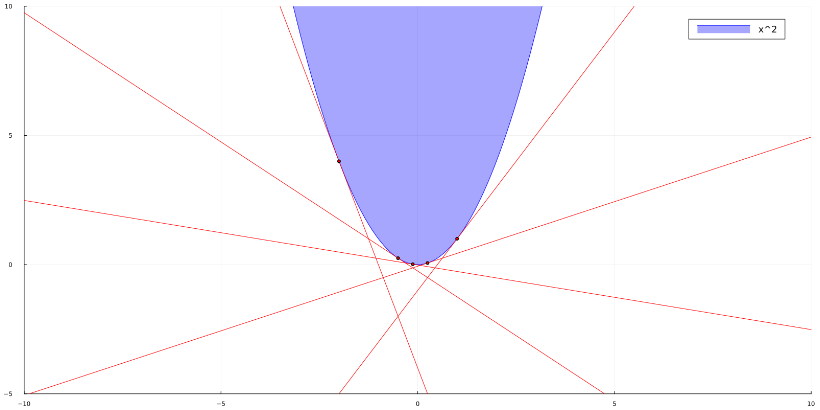


Illustration of the (Kelley) cutting plane method



Generalized convexity of the ℓ_0 pseudonorm

[Chancelier and De Lara, 2022b]

Theorem

Let $\|\cdot\| = \sqrt{\langle \cdot \mid \cdot \rangle}$ be the source norm for the Capra coupling ¢

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

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

Minimization of ℓ_0 under constraints

- Let $X \subset \mathbb{R}^n \setminus \{0\}$ be a compact set the problem

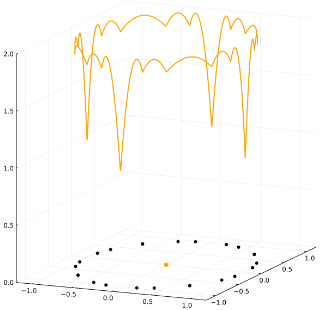
$$\min_{x \in X} \ell_0(x)$$

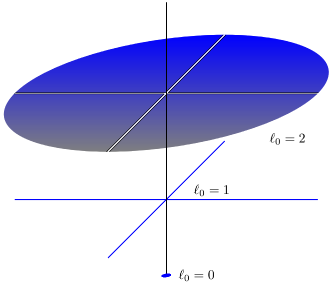
- Idea: using a finite number of Capra cuts $\underbrace{Y}_{\text{finite}} \subset \mathbb{R}^n$

$$\begin{aligned} \min_{x \in X} \ell_0(x) &= \min_{x \in X} \max_{y \in \mathbb{R}^n} ¢(x, y) - \ell_0^{¢}(y) \\ &\geq \min_{x \in X} \max_{y \in Y} ¢(x, y) - \ell_0^{¢}(y) \end{aligned}$$

- Cutting plane method:
alternatively minimizing and improving lower approximations

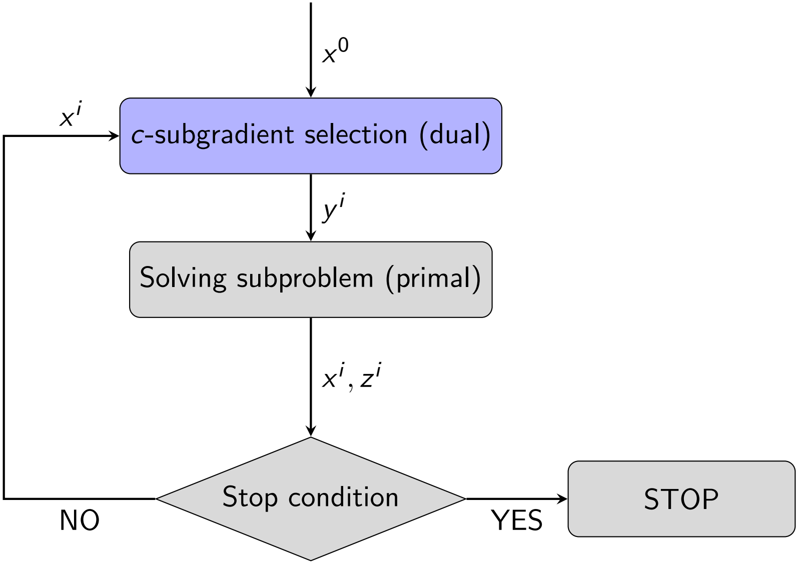
Capra "polyhedral" lower approximation of ℓ_0





Presentation of the abstract cutting plane method

Diagram of the abstract cutting plane method



Abstract cutting plane method

We say $\underbrace{\{x^i\}_{i\geq 0}}_{\text{primal iterates}} \subset X$, $\underbrace{\{y^i\}_{i\geq 0}}_{\text{dual iterates}} \subset \mathcal{Y}$ and $\underbrace{\{z^i\}_{i\geq 1}}_{\text{lower bounds}} \subset \mathbb{R}$

are generated by $\mathrm{CP}(X, x^0, f, c, Y, E)$, if

1. **Initialization**

   $x^0 \in \underbrace{X}_{\text{optimization set}} \subset \mathcal{X}$

2. ***c*-subgradient selection**

   $y^i = Y(x^i)$, where $Y : X \to \mathcal{Y}$ s.t. $\underbrace{Y(x) \in \partial_c f(x)}_{c\text{-subgradient selector}}$

3. ***i*-th primal subproblem**

$$(x^i, z^i) \in \underset{(x,z)\in\mathcal{X}\times\mathbb{R}}{\arg\min}\; z \text{ s.t. } \begin{cases} x \in X\,, \quad (x,z) \in \overbrace{E \subset \mathcal{X}\times\mathbb{R}}^{\text{additional constraints}} \\ z \geq f(x^j) + c(x, y^j) - c(x^j, y^j) \\ \forall j \in [\![0, i-1]\!] \end{cases}$$

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

Convergence result

Convergence result for c-cutting plane method

Theorem

Let $\mathrm{CP}(X, x^0, f, c, Y, E)$ be a cutting plane method generating $\{x^i\}_{i\geq 0} \subset X$, $\{y^i\}_{i\geq 0} \subset \mathcal{Y}$ and $\{z^i\}_{i\geq 1} \subset \mathbb{R}$

If

- $(\mathcal{X}, d)$ metric space, $X \subset \mathcal{X}$ compact, $f : \mathcal{X} \to \overline{\mathbb{R}}$ l.s.c. on X
- $\boxed{\partial_c f(x) \neq \emptyset \ , \quad \forall x \in X}$ (so $f(x) = f^{cc'}(x)\ , \quad \forall x \in X$)
- $\left(\arg\min_X f\right) \times \{\min_X f\} \subset E \subset \mathcal{X} \times \mathbb{R}$
- there exists $M > 0$ such that

$$\boxed{\begin{aligned} |c(x,y) - c(x',y)| \leq Md(x,x')\ , \quad \forall x, x' \in X \\ \forall y \in \textstyle\bigcup_{i\in\mathbb{N}} Y(X \cap P_{\mathcal{X}}(E)) \end{aligned}}$$

Then

- $z^i \nearrow \min_X f$
- $\{x^i\}_{i\geq 0}$ has a subsequence $\{x^{\nu(i)}\}_{i\geq 0} \xrightarrow[i\to+\infty]{} x^* \in \arg\min_X f$

Capra cutting plane (primal) subproblem

Capra cutting plane (primal) subproblem

$$\min_{\substack{z\in\mathbb{R}\\ x\in\mathbb{R}^n}} z \qquad \text{s.t.} \quad \begin{cases} x \in X \\ (x, z) \in E \\ z \geq \dfrac{\left\langle x \middle| \, y^j \right\rangle}{\|x\|} + f(x^j) - ¢(x^j, y^j) \\ \forall j \in [\![0, i-1]\!] \end{cases}$$

Capra cutting plane (primal) subproblem

Proposition

- Let $X \subset \mathbb{R}^n \setminus \{0\}$ be a set and S be the Euclidean unit sphere
- let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function
- let $E \subset \mathbb{R}^n \times \mathbb{R}$ be such that $\left(\arg\min_X f\right) \times \{\min_X f\} \subset E$

Then, given $\{x^j\}_{1 \leq j \leq i-1}, \{y^j\}_{1 \leq j \leq i-1} \subset \mathbb{R}^n$ the i-th primal subproblem of a Capra cutting plane method is

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

Capra subproblem is a linear program on the sphere

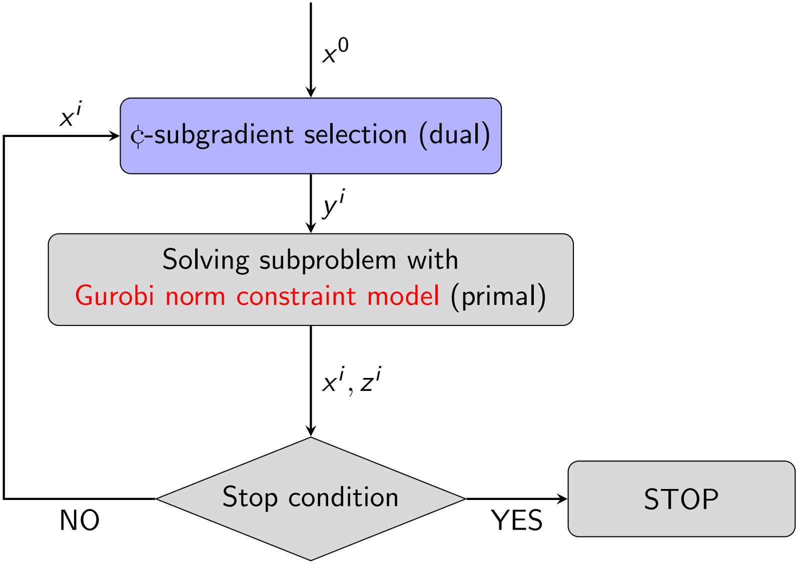
Capra cutting plane subproblem: LP on sphere

When X and E are polyhedral (e.g. $X = \{x : Ax = 0\}$)

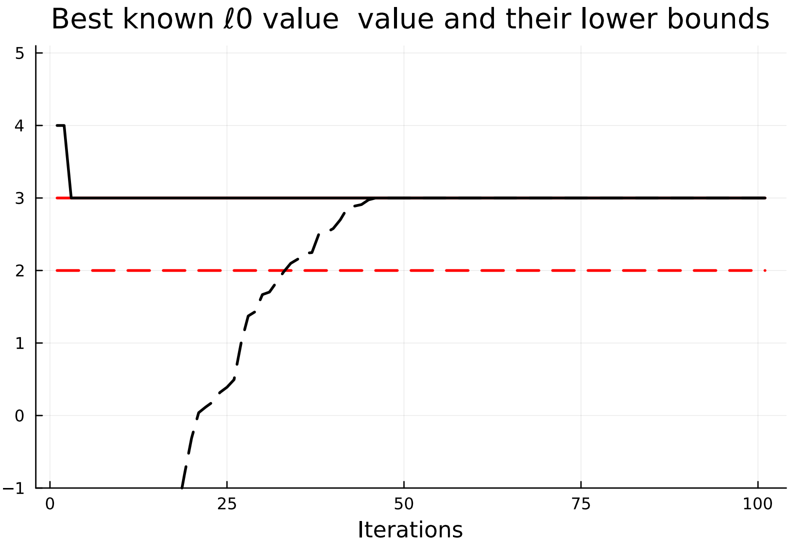
$$
\underbrace{\min_{\substack{z\in\mathbb{R}\\ s \in S}}}_{\text{sphere constraint}} \quad z \quad \text{s.t.} \quad
\begin{cases}
\underbrace{s \in \mathrm{cone}(X)}_{\text{linear constraint}} \\
\underbrace{(s, z) \in E}_{\text{linear constraint}} \\
\underbrace{z \geq \langle s \mid y^j \rangle +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

How to solve a LP on the unit sphere?

Diagram of the abstract cutting plane method



An abstract cutting “plane” method for ℓ_0

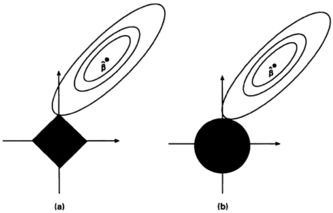


Outline of the presentation

What are sparsity-inducing norms/balls?

The intuition behind lasso

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



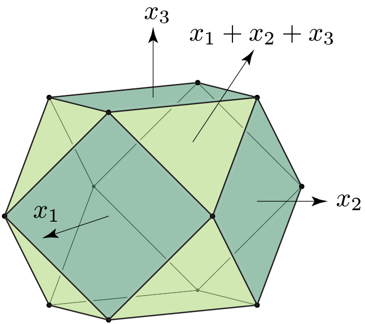
$$\min_{x \in \mathbb{R}^d} \left(f(x) + \gamma \|x\|_2 \right)$$

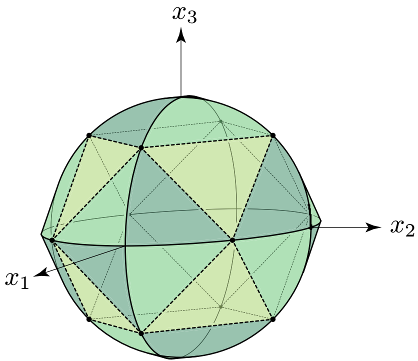
Comments of [Tibshirani, 1996, Figure 2]

> *"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."*

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996

Kinks stand at sparse points





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 $\mathcal{N}(C)$

Geometric (alignment) expression of optimality condition

- We consider an optimal solution $x^* \neq 0$ of

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

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

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

- We expect that the support of x^*
can be recovered from dual information $-\nabla f(x^*)$

Exposed faces of unit balls with k-sparse extreme points

We reformulate sparsity in terms of coordinate subspaces

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

- For any subset $K \subset [\![1, n]\!]$ of indices, we set

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

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

$$\ell_0^{\leq k} = \underbrace{\left\{ x \in \mathbb{R}^n \,\middle|\, \ell_0(x) \leq k \right\}}_{k\text{-sparse vectors}} = \bigcup_{|K| \leq k} \mathcal{R}_K \; , \quad \forall k \in [\![0, n]\!]$$

- We denote by $\pi_K : \mathbb{R}^n \to \mathcal{R}_K$ the orthogonal projection

  For any vector $y \in \mathbb{R}^n$, $\pi_K(y) = y_K \in \mathcal{R}_K \subset \mathbb{R}^n$ is the vector whose entries coincide with those of y,
  except for those outside of K that vanish

Design of unit ball
with k-sparse extreme points

(for example, 2-sparse points in $\mathbb{R}^3$)

Design of unit ball with k-sparse extreme points

For given sparsity threshold $k \in [\![1, d]\!]$,
we consider a source norm $\|\cdot\|$, with unit ball B, and we

- project B onto $\ell_0^{\leq k}$,
  form the convex hull and get

$$B^{\top\star}_{\star,(k)} = \mathrm{co}\Big(\bigcup_{|K| \leq k} \pi_K(B) \Big)$$

  unit ball of the generalized k-support dual norm $\|\cdot\|^{\top\star}_{\star,(k)}$ [Chancelier and De Lara, 2022b]
- the extreme points belong to $\bigcup_{|K| \leq k} \mathcal{R}_K = \ell_0^{\leq k}$,
  hence are k-sparse vectors

Generalized top-k and k-support dual norms

Chancelier and De Lara [2022b].

Definition

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

- the generalized k-support dual norm $\|\cdot\|_{\star,(k)}^{\top\star}$

$$\text{is the dual norm} \qquad \|\cdot\|_{\star,(k)}^{\top\star} = \left(\|\cdot\|_{\star,(k)}^{\top}\right)_{\star}$$

- of the generalized top-k dual norm $\|\cdot\|_{\star,(k)}^{\top}$ defined by

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

Exposed faces characterization

Exposed faces characterization

Theorem

Let $k \in [\![1, n]\!]$
Then, for any nonzero dual vector $y \in \mathbb{R}^d \setminus \{0\}$,
the exposed face of the unit ball $B^{\top\star}_{\star,(k)}$ is given by

$$F_{\perp}(B^{\top\star}_{\star,(k)}, y) = \overline{\mathrm{co}}\Big\{ \overbrace{\pi_{K^*}\big(\underbrace{F_{\perp}(B, \pi_{K^*} y)}_{\substack{\text{exposed face} \\ \text{of the original} \\ \text{unit ball}}}\big)}^{\text{projection on } \mathcal{R}_{K^*}} : K^* \in \arg\max_{|K| \leq k} |\!|\!| \pi_K y |\!|\!|_{\star} \Big\}$$

Exposed faces characterization

Theorem

Let $k \in [\![1, n]\!]$
Suppose that the source norm $|\!|\!|\cdot|\!|\!|$ is orthant-strictly monotonic

Then, for any nonzero dual vector $y \in \mathbb{R}^d \setminus \{0\}$,
the exposed face of the unit ball $B^{\top\star}_{\star,(k)}$ is given by

$$F_{\perp}(B^{\top\star}_{\star,(k)}, y) = \overline{\mathrm{co}}\Big\{\underbrace{F_{\perp}(B, \pi_{K^*} y)}_{\substack{\text{exposed face}\\ \text{of the original}\\ \text{unit ball}}} : \; K^* \in \mathop{\arg\max}_{|K| \leq k} |\!|\!|\pi_K y|\!|\!|_{\star}\Big\}$$

Support identification using k-sparsity inducing norms

Support identification: main result

Theorem

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

For given sparsity threshold $k \in [\![1, d]\!]$,
an optimal solution x^* of

$$\min_{x \in \mathbb{R}^d} \left(f(x) + \gamma \overbrace{\|x\|_{\star,(k)}^{\top\star}}^{\substack{\text{generalized} \\ k\text{-support} \\ \text{dual norm}}} \right)$$

has support

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

Sparse support identification: corollary

Corollary

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a smooth convex function and $\gamma > 0$

For given sparsity threshold $k \in [\![1, d]\!]$, if an optimal solution x^* of

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

satisfies

$$\arg\max_{|K| \leq k} \|\pi_K(-\nabla f(x^*))\|_\star = K^* \quad \text{is unique}$$

then it has support

$$\mathrm{supp}(x^*) \subset K^* \text{ with } |K^*| \leq k$$

so that the optimal solution x^* is k-sparse

Support identification: Lasso

Corollary

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a smooth convex function,
$\gamma > 0$ and $\|\cdot\|_1$ be the ℓ_1 norm

An optimal solution x^* of

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

has support

$$\mathrm{supp}(x^*) \subset \arg\max_{j \in [\![1, d]\!]} |\nabla_j f(x^*)|$$

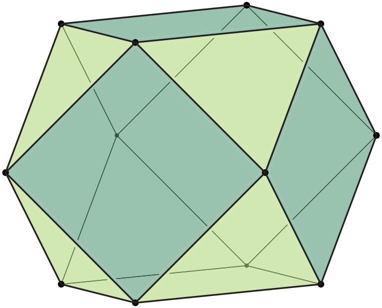
Geometry of sparsity-inducing balls

source norm $\vert\!\vert\!\vert\cdot\vert\!\vert\!\vert$	$\vert\!\vert\!\vert\cdot\vert\!\vert\!\vert^{\top}_{\star,(k)}$, $k \in [\![1, d]\!]$	$\vert\!\vert\!\vert\cdot\vert\!\vert\!\vert^{\top\star}_{\star,(k)}$, $k \in [\![1, d]\!]$
$\Vert\cdot\Vert_p$	top-(q,k) norm $\Vert y\Vert^{\top}_{q,k}$ $\Vert y\Vert^{\top}_{q,k} = \left(\sum_{l=1}^{k} \vert y_{\nu(l)}\vert^q\right)^{\frac{1}{q}}$	(p,k)-support norm $\Vert x\Vert^{\top\star}_{p,k}$ no analytic expression
$\Vert\cdot\Vert_1$	top-(∞,k) norm ℓ_∞-norm $\Vert y\Vert^{\top}_{\infty,k} = \Vert y\Vert_\infty$, $\forall k \in [\![1, d]\!]$	$(1,k)$-support norm ℓ_1-norm $\Vert x\Vert^{\top\star}_{1,k} = \Vert x\Vert_1$, $\forall k \in [\![1, d]\!]$
$\Vert\cdot\Vert_2$	top-$(2,k)$ norm $\Vert y\Vert^{\top}_{2,k} = \sqrt{\sum_{l=1}^{k} \vert y_{\nu(l)}\vert^2}$ $\Vert y\Vert^{\top}_{2,1} = \Vert y\Vert_\infty$	$(2,k)$-support norm $\Vert x\Vert^{\top\star}_{2,k}$ no analytic expression $\Vert x\Vert^{\top\star}_{2,1} = \Vert x\Vert_1$
$\Vert\cdot\Vert_\infty$	top-$(1,k)$ norm $\Vert y\Vert^{\top}_{1,k} = \sum_{l=1}^{k} \vert y_{\nu(l)}\vert$ $\Vert y\Vert^{\top}_{1,1} = \Vert y\Vert_\infty$	(∞,k)-support norm $\Vert x\Vert^{\top\star}_{\infty,k} = \max\{\frac{\Vert x\Vert_1}{k}, \Vert x\Vert_\infty\}$ $\Vert x\Vert^{\top\star}_{1,1} = \Vert x\Vert_1$

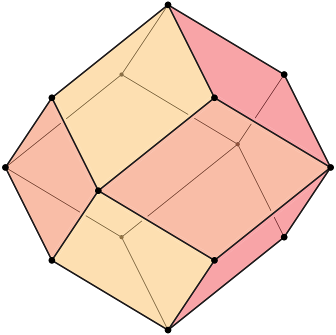
Table: Examples of generalized top-k and k-support dual norms generated by the ℓ_p source norms $\vert\!\vert\!\vert\cdot\vert\!\vert\!\vert = \Vert\cdot\Vert_p$ for $p \in [1, \infty]$, where $1/p + 1/q = 1$. For $y \in \mathbb{R}^n$, ν denotes a permutation of $\{1, \ldots, d\}$ such that $\vert y_{\nu(1)}\vert \geq \vert y_{\nu(2)}\vert \geq \cdots \geq \vert y_{\nu(d)}\vert$.

When the source norm is the ℓ_∞-norm

Case $k = 2$ in $\mathbb{R}^3$ with source norm the ℓ_∞-norm



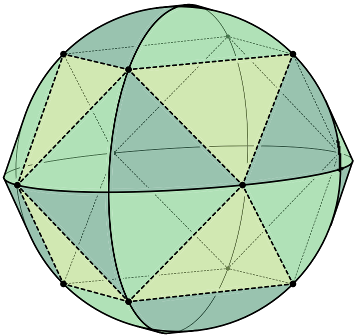
(a) Unit ball $B^{\top\star}_{\infty,2}$ (support norm)



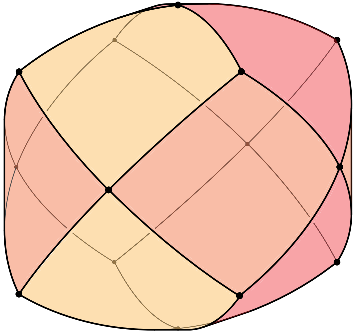
(b) Unit ball $B^{\top}_{1,2}$ (top norm)

When the source norm is the ℓ_2-norm

Case $k = 2$ in $\mathbb{R}^3$ with source norm the ℓ_2-norm



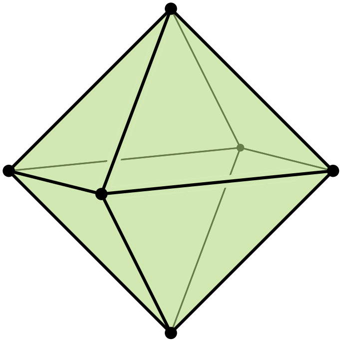
(a) Unit ball $B_{2,2}^{\top\star}$ (support norm)



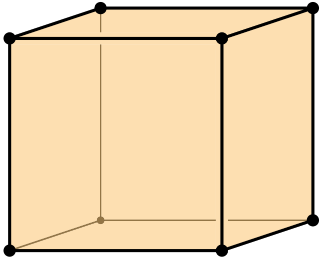
(b) Unit ball $B_{2,2}^{\top}$ (top norm)

When the source norm is the ℓ_1-norm

Case $k = 2$ in $\mathbb{R}^3$ with source norm the ℓ_1-norm



(a) Unit ball $B^{\top\star}_{1,2}$
(support norm)



(b) Unit ball $B^{\top}_{\infty,2}$
(top norm)

Outline of the presentation

- 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 $\partial_{¢}\ell_0 \neq \emptyset$, 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 $\mathcal{L}_0 = \ell_0^{¢\star'}$
  - Capra-cuts based algorithms
  - lower bound convex programs
  - $\mathcal{H}_0$-couplings to go beyond Capra-couplings

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. Capra-convexity, convex factorization and variational formulations for the l_0 pseudonorm. *Set-Valued and Variational Analysis*, 30:597–619, 2022b.

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*, 30(3):743–769, 2023.

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.

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.

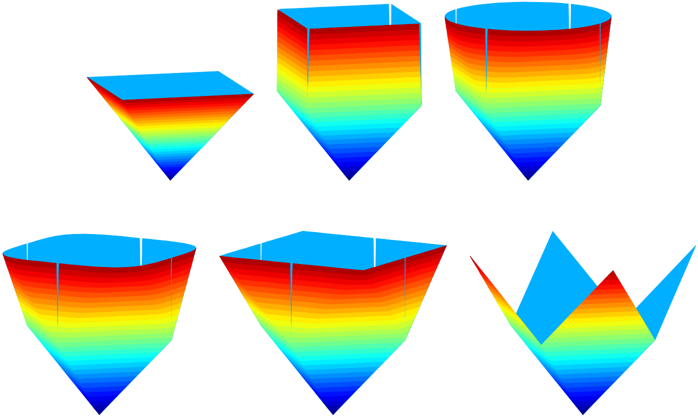
J. J. Moreau. Fonctionnelles convexes. *Séminaire Jean Leray*, 2:1–108, 1966-1967.

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

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

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

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, 2022b]
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 bound convex programs for exact sparse optimization

Concave dual problem for exact sparse optimization

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

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

Proposition

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

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

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 $I : \mathbb{R}^n \to \overline{\mathbb{R}}$, $\varphi : [\![0, n]\!] \to \overline{\mathbb{R}}$ nondecreasing (ex: identity, $\iota_{\{0,1,\dots,k\}}$) and original minimization problem

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

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

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

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

- Value function

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

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

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

$$= I(w) \dotplus \big(\langle \varrho(w),\ y \rangle - (\varphi \circ \ell_0)^{¢}(y)\big)$$

- Dual maximization problem

$$\varphi^{\star\star'}(0) = \sup_{y \in \mathbb{R}^n} \inf_{w \in \mathbb{R}^n} \mathcal{L}(w, y) = \sup_{y \in \mathbb{R}^n} \left\{ \big(-I^{-¢}(y)\big) \dotplus \big(-(\varphi \circ \ell_0)^{¢}(y)\big) \right\}$$

- Original minimization problem (case "$\dotplus = +$" when I proper)

$$\varphi(0) = \inf_{w \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^n} \mathcal{L}(w, y) = \inf_{w \in \mathbb{R}^n} \left\{ I(w) \dotplus \varphi\big(\ell_0(w)\big) \right\}$$

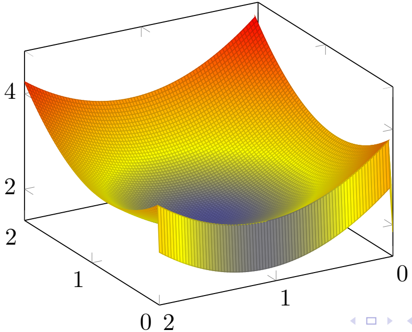
Numerics

A toy example

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

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

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



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

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

with

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

Generalized primal-dual proximal splitting

GPDPS Algorithm [Clason, Mazurenko, and Valkonen, 2020]

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

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

The prox of I 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