The Geometry of Sparsity-Inducing Balls

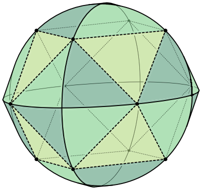
Jean-Philippe Chancelier, *Michel De Lara*

Cermics, École des Ponts ParisTech, France

Antoine Deza McMaster University, Hamilton, Ontario, Canada

Lionel Pournin Université Paris 13, Villetaneuse, France

EURO 2024, Copenhaguen, Denmark
30 June – 3 July 2024



Support and the ℓ_0 pseudonorm

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

$$[\![0,d]\!] = \{0,1,\ldots,d\} \;,\;\; [\![1,d]\!] = \{1,\ldots,d\}$$

For any vector $x \in \mathbb{R}^d$, we define

- its support by

$$\mathrm{supp}(x) = \{ j \in [\![1,d]\!] \mid x_j \neq 0 \}$$

$$\mathrm{supp}((0,*,0,*,*,0)) = \{2,4,5\} \subset [\![1,6]\!]$$

- its ℓ_0 pseudonorm(x) by

$$\ell_0(x) = \overbrace{|\mathrm{supp}(x)|}^{\text{cardinality}} = \overbrace{\sum_{i=1}^{d} \mathbf{1}_{\{x_i \neq 0\}}}^{\substack{\text{number of}\\ \text{nonzero entries}}}$$

$$\ell_0\big((0,*,0,*,*,0)\big) = |\{2,4,5\}| = 3 \in [\![0,6]\!]$$

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

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

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

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

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

Talk outline

Outline of the presentation

Design of sparsity-inducing unit balls [10 min]

Geometry of sparsity-inducing balls [5 min]

Orthant-strictly monotonicity and Capra-convexity of ℓ_0 [5 min]

Conclusion

Outline of the presentation

- Design of sparsity-inducing unit balls [10 min]
  - What are sparsity-inducing norms/balls?
  - Exposed faces of unit balls with k-sparse extreme points
  - Support identification using k-sparsity inducing norms
- Geometry of sparsity-inducing balls [5 min]
- Orthant-strictly monotonicity and Capra-convexity of ℓ_0 [5 min]
  - Orthant-strictly monotonic (OSM) norms
  - OSM norms and hidden convexity in the ℓ_0 pseudonorm
  - Crash course on generalized convexity
  - OSM norms, Capra conjugacies and the ℓ_0 pseudonorm
- Conclusion

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

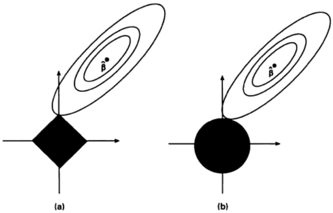
$$\text{optimal } k\text{-sparse vector} \qquad \min_{\underbrace{\ell_0(x) \leq k}_{k\text{-sparse vectors}}} f(x)$$

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

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

The intuition behind lasso

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



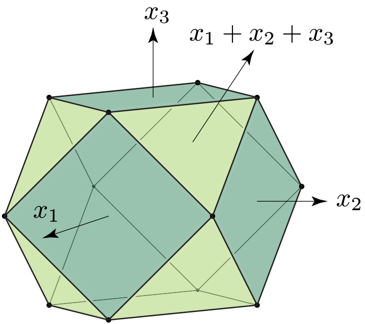
$$\min_{x\in\mathbb{R}^d} \left(f(x) + \gamma \|x\|_2\right)$$

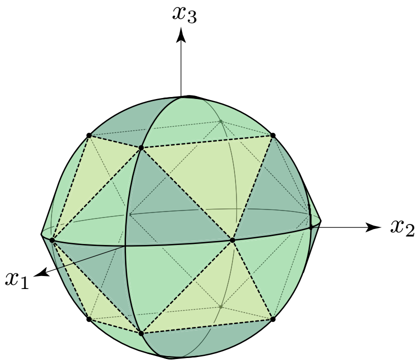
Comments of [Tibshirani, 1996, Figure 2]

> *"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





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 $-\nabla f(x^*)$

Outline of the presentation

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 $K \subset [\![1, d]\!]$, we introduce the (coordinate) subspace

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

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

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

- We denote by $\pi_K : \mathbb{R}^d \to \mathcal{R}_K$ the orthogonal projection
For any vector $y \in \mathbb{R}^d$, $\pi_K(y) \in \mathcal{R}_K \subset \mathbb{R}^d$ is the vector whose components 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

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

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

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

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

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

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

Exposed faces characterization

Exposed faces characterization

Theorem

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

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

Exposed faces characterization

Theorem

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

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

Outline of the presentation

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

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

has support

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

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

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

Outline of the presentation

The case of ℓ_p-norms $\|\cdot\|_p$

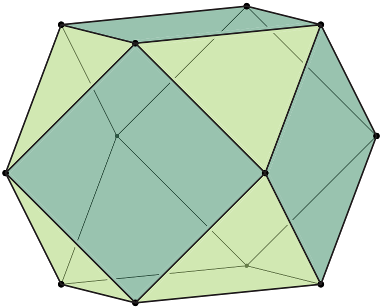
$$\|x\|_\infty = \sup_{i \in [\![1,d]\!]} |x_i| \text{ and } \|x\|_p = \left(\sum_{i=1}^{d} |x_i|^p\right)^{1/p} \text{ for } p \in [1, \infty[$$

source norm $\|\cdot\|$	$\|\cdot\|^{\top}_{\star,(k)}$, $k \in [\![1, d]\!]$	$\|\cdot\|^{\top\star}_{\star,(k)}$, $k \in [\![1, d]\!]$
$\|\cdot\|_p$	top-(q,k) norm $\|y\|^{\top}_{q,k}$ $\|y\|^{\top}_{q,k} = \left(\sum_{l=1}^{k} \|y_{\nu(l)}\|^q\right)^{\frac{1}{q}}$	(p,k)-support norm $\|x\|^{\top\star}_{p,k}$ no analytic expression
$\|\cdot\|_1$	top-(∞,k) norm ℓ_∞-norm $\|y\|^{\top}_{\infty,k} = \|y\|_\infty$, $\forall k \in [\![1, d]\!]$	$(1,k)$-support norm ℓ_1-norm $\|x\|^{\top\star}_{1,k} = \|x\|_1$, $\forall k \in [\![1, d]\!]$
$\|\cdot\|_2$	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 (computation [Argyriou et al., 2012, Prop. 2.1]) $\|x\|^{\top\star}_{2,1} = \|x\|_1$
$\|\cdot\|_\infty$	top-$(1,k)$ norm $\|y\|^{\top}_{1,k} = \sum_{l=1}^{k} \|y_{\nu(l)}\|$ $\|y\|^{\top}_{1,1} = \|y\|_\infty$	(∞,k)-support norm $\|x\|^{\top\star}_{\infty,k} = \max\{\frac{\|x\|_1}{k}, \|x\|_\infty\}$ $\|x\|^{\top\star}_{1,1} = \|x\|_1$

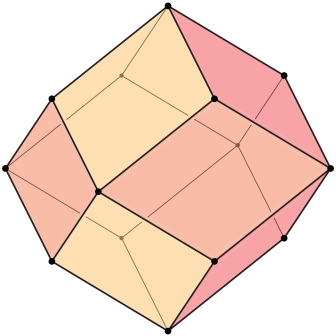
Table: Examples of generalized top-k and k-support dual norms generated by the ℓ_p source norms $\|\cdot\| = \|\cdot\|_p$ for $p \in [1, \infty]$, where $1/p + 1/q = 1$. For $y \in \mathbb{R}^d$, ν denotes a permutation of $\{1, \ldots, d\}$ such that $|y_{\nu(1)}| \geq |y_{\nu(2)}| \geq \cdots \geq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

When the source norm is the ℓ_∞-norm

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



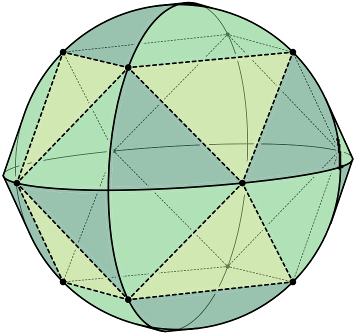
(a) Unit ball $B^{\top\star}_{\infty,2}$
(support norm)



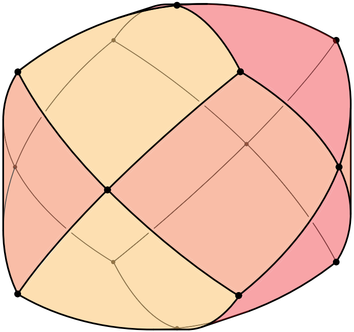
(b) Unit ball $B^{\top}_{1,2}$
(top norm)

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 $B^{\top\star}_{2,2}$ (support norm)

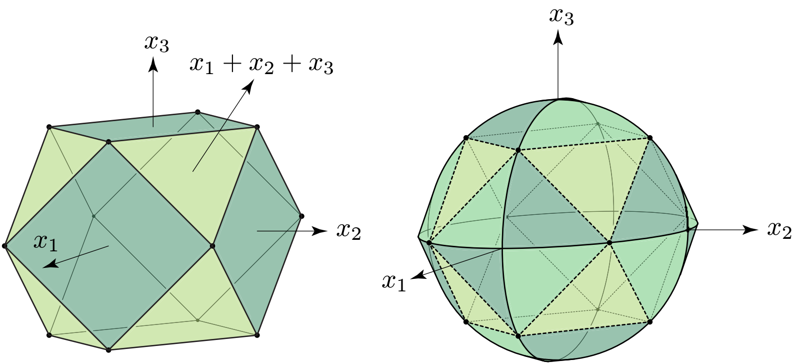


(b) Unit ball $B^{\top}_{2,2}$ (top norm)

Geometric description

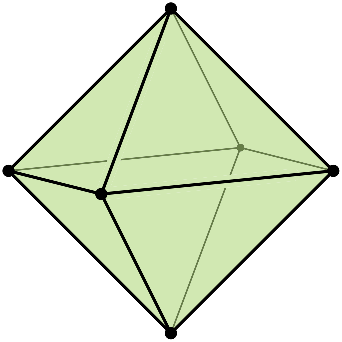
Proposition

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

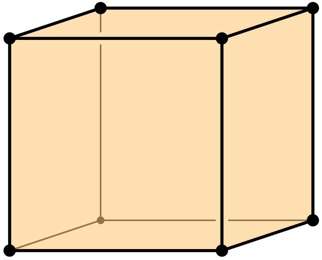


When the source norm is the ℓ_1-norm

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)

What comes next?

- What are orthant-strictly monotonic norms?
- In what are they related to the ℓ_0 pseudonorm?

Background on the original motivation
Jean-Philippe Chancelier, Michel De Lara

Outline of the presentation

Outline of the presentation

Orthant-monotonic norms

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

Definition

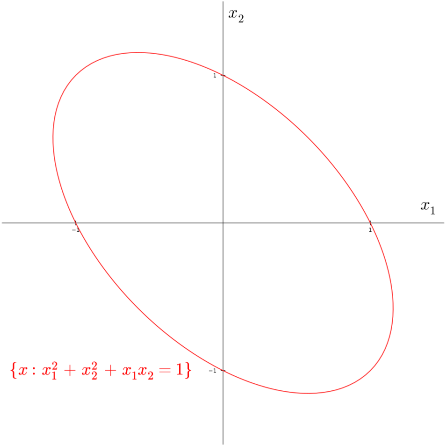
A norm $\|\cdot\|$ on the space $\mathbb{R}^d$ is called orthant-monotonic [Gries, 1967]
if, for all x, x' in $\mathbb{R}^d$, we have

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

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

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

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, 2023]

Definition

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

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

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

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

Examples of orthant-strictly monotonic norms

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

with unit ball B_p and unit sphere S_p

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

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

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

Outline of the presentation

Design of sparsity-inducing unit balls [10 min]

What are sparsity-inducing norms/balls?

Exposed faces of unit balls with k-sparse extreme points

Support identification using k-sparsity inducing norms

Geometry of sparsity-inducing balls [5 min]

Orthant-strictly monotonicity and Capra-convexity of ℓ_0 [5 min]

Orthant-strictly monotonic (OSM) norms

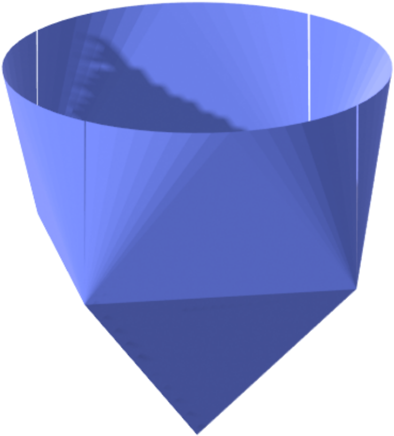
OSM norms and hidden convexity in the ℓ_0 pseudonorm

Crash course on generalized convexity

OSM norms, Capra conjugacies and the ℓ_0 pseudonorm

Conclusion

Graph of the Euclidean ℓ_0-cup function $\mathcal{L}_0$



Orthant-strictly monotonic norms and hidden convexity in the ℓ_0 pseudonorm

[Chancelier and De Lara, 2022b]

Theorem

If both the norm $\|\cdot\|$ and the dual norm $\|\cdot\|_\star$ are orthant-strictly monotonic, there exists a proper convex lsc function $\mathcal{L}_0$, the ℓ_0-cup function, with domain the unit ball B, such that

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

and, as a consequence, the ℓ_0 pseudonorm coincides, on the unit sphere S, with the proper convex lsc function $\mathcal{L}_0$

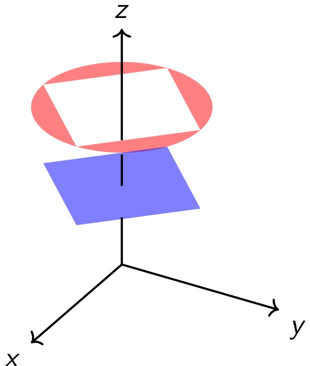
$$\ell_0(x) = \mathcal{L}_0(x) \; , \; \forall x \in S$$

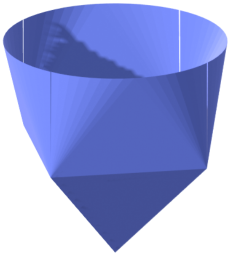
The ℓ_0-cup function as a convex envelope

Proposition

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

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





The ℓ_0-cup function as best proper convex lsc lower approximation of the ℓ_0 pseudonorm on the unit ball

Theorem

The ℓ_0-cup function $\mathcal{L}_0$ is
the best convex lsc lower approximation of the ℓ_0 pseudonorm
on the unit ball B

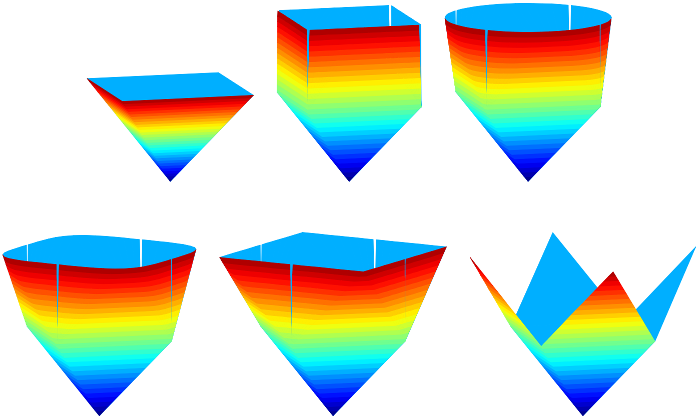
$$\text{best convex lsc function} \quad \mathcal{L}_0(x) \leq \ell_0(x) \ , \ \ \forall x \in B$$

and, as seen above, coincides with the ℓ_0 pseudonorm

on the unit sphere S

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

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



Outline of the presentation

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 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

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

Constant Along Primal RAys (Capra) coupling

[Chancelier and De Lara, 2021, 2022a]

Definition

On the vector space $\mathbb{R}^d$, equipped with a (source) norm $\|\cdot\|$, the Capra coupling (Capra) $\mathbb{R}^d \overset{¢}{\longleftrightarrow} \mathbb{R}^d$ is given by

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

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

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

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

Capra-conjugate of the ℓ_0 pseudonorm

[Chancelier and De Lara, 2021, 2022a]

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

where $S \subset \mathbb{R}^d$ is the unit sphere

$$
\begin{aligned}
&= \sup \Big\{ 0, \sup_{j \in [\![1,d]\!]} \Big\{ \underbrace{\sup_{\substack{s \in S \\ \ell_0(s) = j}} \langle s \mid y \rangle}_{\text{coordinate-}j \text{ norm } \|y\|_{(j)}^{\mathcal{R}}} - j \Big\} \Big\} \\
&= \sup_{j \in [\![1,d]\!]} \Big[\|y\|_{(j)}^{\mathcal{R}} - j \Big]_+
\end{aligned}
$$

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

Capra $=$ Fenchel coupling after primal normalization

- We define the primal radial projection ϱ as

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

- so that the coupling Capra

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

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

The Capra conjugacy shares properties with the Fenchel conjugacy

Proposition

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

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

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

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

$$g^{¢'} = g^{\star'} \circ \varrho$$

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

Proposition

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

$$\iff h = h^{¢¢'}$$

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

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

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

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

[Chancelier and De Lara, 2022b]

Theorem

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

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

and, as a consequence,

$$\boxed{\ell_0^{¢¢'} = \ell_0}$$

and thus

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

Variational formulas for the ℓ_0 pseudonorm

Proposition

$$\ell_0(x) = \frac{1}{\|x\|} \min_{\substack{x^{(1)}\in\mathbb{R}^d,\dots,x^{(d)}\in\mathbb{R}^d \\ \sum_{\ell=1}^d \|x^{(\ell)}\|_{(\ell)}^{\top\star} \leq \|x\| \\ \sum_{\ell=1}^d x^{(\ell)} = x}} \sum_{\ell=1}^{d} \ell \|x^{(\ell)}\|_{(\ell)}^{\top\star} \ , \ \ \forall x \in \mathbb{R}^d$$

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

Conclusion

- We have proposed systematic ways to design unit balls that enhance sparsity at a given threshold
- The corresponding norms originally appeared related to generalized Capra-convexity of the ℓ_0 pseudonorm, as well as the property of orthant-strict monotonicity
- For classic ℓ_∞, ℓ_2 and ℓ_1 source norms, we have a complete description of the corresponding sparsity-inducing unit balls

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*, (3):743–769, 2023.

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

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.

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 :-)