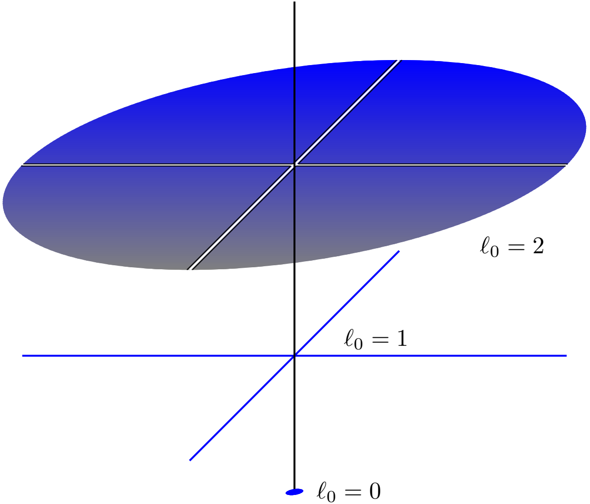
A Suitable Conjugacy for the l_0 Pseudonorm with Application to Exact Sparse Optimization

Jean-Philippe Chancelier and Michel De Lara
CERMICS, École des Ponts ParisTech

PGMO, Paris
4 December 2019

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



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

Only 1-homogeneity is missing

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

- For any vector $x \in \mathbb{R}^d$, we define its pseudonorm $\ell_0(x)$ by

$$\ell_0(x) = |x|_0 = \text{number of nonzero components of } x$$

- The function pseudonorm $\ell_0 : \mathbb{R}^d \to \{0, 1, \ldots, d\}$ satisfies 3 out of 4 axioms of a norm
  - we have $\ell_0(x) \geq 0$ ✓
  - we have $\big(\ell_0(x) = 0 \iff x = 0 \big)$ ✓
  - we have $\ell_0(x + x') \leq \ell_0(x) + \ell_0(x')$ ✓
  - But... 0-homogeneity holds true

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

First, let us have a look at the least squares regression sparse optimization problem

The pseudonorm ℓ_0 is in the constraints

Let a matrix A and a vector z be given (with proper dimensions)

- Letting $k \in \{1, \ldots, d\}$ be a (small) integer, the least squares regression sparse optimization problem is

$$\min_{\ell_0(x) \leq k} \|Ax - z\|^2$$

- In a sense, you try to "explain" the output variable z by at most k components of x

The Fenchel conjugacy is not suitable to handle such sparse constraints

- Indeed, from the easily obtained inequality

$$\sup_{y\in\mathbb{Y}}\Big(\big(-f^\star(y)\big) \dotplus \big(-\delta_X^{(-\star)}(y)\big)\Big) \leq \inf_{x\in\mathbb{X}}\Big(f(x)\dotplus\delta_X(x)\Big) = \inf_{x\in X} f(x)$$

- we deduce the (disappointing) lower bound

$$-\infty = \sup_{y\in\mathbb{R}^d}\Big(-f^\star(y) \dotplus \big(\underbrace{-\delta^{(-\star)}_{\{x\in\mathbb{R}^d \,\mid\, \ell_0(x)\leq k\}}(y)}_{-\sigma_{\mathbb{R}^d}(y)=-\infty}\big)\Big) \leq \inf_{\ell_0(x)\leq k} f(x)$$

Second, let us turn towards the sparse linear regression problem

The pseudonorm ℓ_0 is the objective function

Let a matrix A and a vector z be given (with proper dimensions)

- The sparse linear regression problem is

$$\min_{Ax=z} \ell_0(x)$$

- In a sense, you try to "explain" the output variable z by a vector x with the least number of components

The Fenchel conjugacy is not suitable to handle the pseudonorm ℓ_0

- Indeed, as ℓ_0 is bounded in every direction,
  the computation of the Fenchel conjugate $\ell_0^\star$ easily gives

$$\ell_0^\star = \delta_{\{0\}} = \begin{cases} 0 & \text{on } \{0\} \\ +\infty & \text{on } \mathbb{R}^d \backslash \{0\} \end{cases}$$

- leading to the Fenchel biconjugate $\ell_0^{\star\star'}$

$$\underbrace{\ell_0^{\star\star'}(x) = 0}_{\substack{\text{best convex lsc} \\ \text{lower approximation}}} \quad \leq \ell_0(x)$$

- and to the (once again disappointing) lower bound

$$0 \leq \min_{Ax=z} \ell_0(x)$$

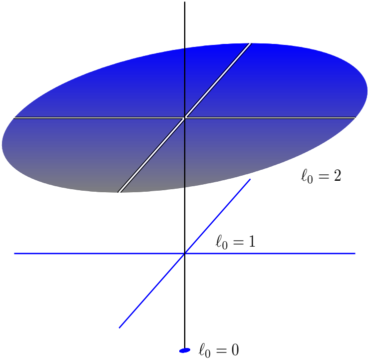
We introduce a family of new conjugacies (Capra) to tackle the pseudonorm ℓ_0

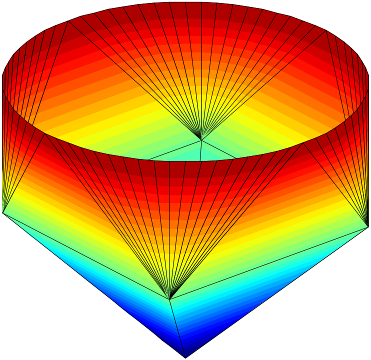
Each conjugacy is function of a (source) norm on $\mathbb{R}^d$, and

- yields lower bound convex programs
  for exact sparse optimization problems
  $\min_{|||x|||^{\mathcal{R}}_{(k)} \leq 1} \big(\inf \big[f \mid n \big] \big)^{\star\star'}(x) \leq \inf_{\ell_0(x) \leq k} f(x)$
- reveals covert convexity in the pseudonorm ℓ_0
  and yields variational formulas for the ℓ_0 pseudonorm

$$\ell_0(x) = \frac{1}{|||x|||} \underbrace{\min_{\substack{z^{(1)} \in \mathbb{R}^d, \ldots, z^{(d)} \in \mathbb{R}^d \\ \sum_{k=1}^d |||z^{(k)}|||^{\star\mathrm{sn}}_{\star,(k)} \leq |||x||| \\ \sum_{k=1}^d z^{(k)} = x}} \sum_{k=1}^d k |||z^{(k)}|||^{\star\mathrm{sn}}_{\star,(k)}}_{\text{convex optimization problem}}$$

The pseudonorm ℓ_0 coincides, on the sphere (circle in $\mathbb{R}^2$), with a convex lsc function





Outline of the presentation

The Capra conjugacy $\cent$ and the pseudonorm ℓ_0

Lower bound convex programs for exact sparse optimization

Variational formulas for the ℓ_0 pseudonorm

Conclusion

Outline of the presentation

The Capra conjugacy $¢$ and the pseudonorm ℓ_0

Lower bound convex programs for exact sparse optimization

Variational formulas for the ℓ_0 pseudonorm

Conclusion

Outline of the presentation

The Capra conjugacy ¢ and the pseudonorm ℓ_0
Background on couplings and Fenchel-Moreau conjugacies
Definition and properties of the Capra conjugacy

Lower bound convex programs for exact sparse optimization
Conjugate of the pseudonorm ℓ_0
Lower bound convex programs

Variational formulas for the ℓ_0 pseudonorm
Biconjugate of the pseudonorm ℓ_0
Covert convexity in the pseudonorm ℓ_0
Variational formulas for the pseudonorm ℓ_0

Conclusion

The Fenchel conjugacy

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

Definition

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

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

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

Fenchel conjugate	Fourier transform
$\sup \to +$	
$+ \to \times$	
$\sup_{x \in \mathbb{X}} \Big(\langle x \, , y \rangle \underset{\cdot}{+} \big(- f(x) \big) \Big)$	$\int_{\mathbb{X}} e^{\langle x \, , y \rangle} f(x) dx$

Background on couplings and Fenchel-Moreau conjugacies

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

$$c : \mathbb{X} \times \mathbb{Y} \to \overline{\mathbb{R}} = [-\infty, +\infty]$$

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

[Martínez-Legaz, 2005]

What are couplings good for?

Couplings are good for providing

- lower bounds for optimization problems with constraints (uses conjugates)
- c-convex lower approximations of functions (uses biconjugates)
- dual representation formulas for c-convex functions (uses biconjugates)

Fenchel-Moreau conjugate

Definition

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

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

Fenchel-Moreau conjugate $(\max, +)$	Kernel transform $(+, \times)$
$\sup_{x \in \mathbb{X}} \Big(c(x,y) \dot{+} \big(-f(x) \big) \Big)$	$\int_{\mathbb{X}} c(x,y) f(x) dx$

The $(-c)$-Fenchel-Moreau conjugate of $h : \mathbb{X} \to \overline{\mathbb{R}}$ is given by

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

Fenchel inequality yields lower bounds

- Conjugacies are special cases of dualites, that make it possible to obtain dual problems

$$\sup_{y \in \mathbb{Y}} \Big(\big(-f^{c}(y) \big) \dot{+} \big(-g^{-c}(y) \big) \Big) \leq \inf_{x \in \mathbb{X}} \Big(f(x) \dot{+} g(x) \Big)$$

- In particular, optimization under constraints $x \in X$ gives

$$\sup_{y \in \mathbb{Y}} \Big(\big(-f^{c}(y) \big) \dot{+} \big(-\delta_X^{-c}(y) \big) \Big) \leq \inf_{x \in X} f(x)$$

where $\delta_X(x) = \begin{cases} 0 & \text{if } x \in X \\ +\infty & \text{if } x \notin X \end{cases}$

- Hence, the issue is to find a coupling c that gives nice expressions for f^{c} and δ_X^{-c}

Fenchel-Moreau biconjugate

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

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

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

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

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

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

So called c-convex functions have dual representations

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

Definition

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

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

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

Example: convex lsc function = supremum of affine functions

Outline of the presentation

The Capra conjugacy ¢ and the pseudonorm ℓ_0
Background on couplings and Fenchel-Moreau conjugacies
Definition and properties of the Capra conjugacy

Lower bound convex programs for exact sparse optimization
Conjugate of the pseudonorm ℓ_0
Lower bound convex programs

Variational formulas for the ℓ_0 pseudonorm
Biconjugate of the pseudonorm ℓ_0
Covert convexity in the pseudonorm ℓ_0
Variational formulas for the pseudonorm ℓ_0

Conclusion

We will display a conjugacy Capra which, as the pseudonorm ℓ_0, is invariant by normalization

- The pseudonorm ℓ_0 is a 0-homogeneous function

$$\forall x \in \mathbb{R}^d , \ \ell_0(\rho x) = \ell_0(x) , \ \forall \rho \neq 0$$

- and, therefore, the pseudonorm ℓ_0 is invariant by normalization

$$\forall x \in \mathbb{R}^d \backslash \{0\} , \ \ell_0(x) = \ell_0(\frac{x}{|||x|||})$$

for any norm $|||\cdot|||$ on $\mathbb{R}^d$

We introduce the coupling Capra

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

Definition

We introduce the coupling Capra $\mathbb{X} \overset{¢}{\longleftrightarrow} \mathbb{Y}$

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

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

Capra = Fenchel coupling after primal normalization

- We denote the unit sphere

$$\mathbb{S} = \left\{ x \in \mathbb{X} \,\middle|\, \|\!|x|\!\| = 1 \right\}$$

and we introduce the primal normalization mapping

$$n : \mathbb{X} \to \mathbb{S} \cup \{0\} \; , \;\; n(x) = \begin{cases} \frac{x}{\|\!|x|\!\|} & \text{if } \; x \neq 0 \\ 0 & \text{if } \; x = 0 \end{cases}$$

- so that the coupling Capra

$$¢(x, y) = \langle n(x) \, , y \rangle \;\; , \;\; \forall x \in \mathbb{X} \, , \;\; \forall y \in \mathbb{Y}$$

appears as the Fenchel coupling after primal normalization

The Capra-subdifferential shares properties with the Rockafellar-Moreau subdifferential

Capra-subdifferential

For any function $f : \mathbb{X} \to \overline{\mathbb{R}}$ and $x \in \mathbb{R}^d$

$$\partial_{¢} f(x) = \{y \in \mathbb{R}^d \mid \quad ¢(x', y) \dotplus \big(- f(x') \big) \leq \quad ¢(x, y) \dotplus \big(- f(x) \big) \ , \ \ \forall x' \in \mathbb{R}^d \}$$

$$y \in \partial_{¢} f(x) \iff f^{¢}(y) = ¢(x, y) \dotplus \big(- f(x) \big)$$

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

$$\partial_{¢} f + \partial_{¢} h \subset \partial_{¢} (f \dotplus h)$$

$$\partial_{¢} f(x) \neq \emptyset \Rightarrow f^{¢¢'}(x) = f(x)$$

The Capra conjugacy shares properties with the Fenchel conjugacy

Proposition

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

$$f^{¢} = \big(\inf\big[f \mid n\big]\big)^{\star} \qquad \textit{where}$$

$$\inf\big[f \mid n\big](x) = \begin{cases} \inf_{\rho>0} f(\rho x) & \textit{if } x \in \mathbb{S} \cup \{0\} \\ +\infty & \textit{if } x \notin \mathbb{S} \cup \{0\} \end{cases}$$

As a consequence, when $\mathbb{X}$ and $\mathbb{Y}$ are paired vector spaces the $¢$-Fenchel-Moreau conjugate $f^{¢}$ is a convex lsc function

The ¢-convex functions are 0-homogeneous and coincide, on the sphere, with a convex lsc function

- The ¢′-Fenchel-Moreau conjugate of $g : \mathbb{Y} \to \overline{\mathbb{R}}$ is given by
$$g^{¢'} = g^{\star} \circ n$$
- The ¢-convex functions are $\{g^{¢'} \mid g \in \overline{\mathbb{R}}^{\mathbb{Y}}\}$, hence
$$g^{¢'}(x) = g^{\star}\big(n(x)\big) = g^{\star}\big(\tfrac{x}{|||x|||}\big) \text{ if } x \neq 0$$

Proposition

When $\mathbb{X}$ and $\mathbb{Y}$ are paired vector spaces, any ¢-convex function coincides, on the sphere $\mathbb{S}$, with a convex lsc function defined on the whole space $\mathbb{X}$

$$¢\textit{-convex function}(x) = \textit{convex lsc function}\Big(\frac{x}{|||x|||}\Big)$$

Outline of the presentation

The Capra conjugacy ¢ and the pseudonorm ℓ_0

Lower bound convex programs for exact sparse optimization

Variational formulas for the ℓ_0 pseudonorm

Conclusion

A first step towards a dual problem for sparse optimization

- From the easily obtained inequality

$$\sup_{y \in \mathbb{Y}} \Big(\big(-f^{\cent}(y)\big) \dotplus \big(-\delta_X^{-\cent}(y)\big) \Big) \leq \inf_{x \in \mathbb{X}} \Big(f(x) \dotplus \delta_X(x) \Big) = \inf_{x \in X} f(x)$$

  we deduce that

$$\sup_{y \in \mathbb{R}^d} \Big(\underbrace{-\big(\inf\left[f \mid n\right]\big)^{\star}(y)}_{\text{concave usc}} \dotplus \underbrace{\big(-\delta^{-\cent}_{\{x \in \mathbb{R}^d \,|\, \ell_0(x) \leq k\}}(y)\big)}_{\text{what is it?}} \Big) \leq \inf_{\ell_0(x) \leq k} f(x)$$

- We denote the level sets of the pseudonorm ℓ_0 by

$$\ell_0^{\leq k} = \left\{ x \in \mathbb{R}^d \,\middle|\, \ell_0(x) \leq k \right\}, \ \ \forall k \in \{0, 1, \ldots, d\}$$

Outline of the presentation

The Capra conjugacy ¢ and the pseudonorm ℓ_0

Background on couplings and Fenchel-Moreau conjugacies

Definition and properties of the Capra conjugacy

Lower bound convex programs for exact sparse optimization

Conjugate of the pseudonorm ℓ_0

Lower bound convex programs

Variational formulas for the ℓ_0 pseudonorm

Biconjugate of the pseudonorm ℓ_0

Covert convexity in the pseudonorm ℓ_0

Variational formulas for the pseudonorm ℓ_0

Conclusion

We reformulate sparsity in terms of coordinate subspaces

- For any $x \in \mathbb{R}^d$ and $K \subset \{1, \ldots, d\}$,
  we denote by $x_K \in \mathbb{R}^d$ the vector which coincides with x,
  except for the components outside of K that vanish

$$x = (1, 2, 3, 4, 5, 6) \to x_{\{2,4,5\}} = (0, 2, 0, 4, 5, 0)$$

- x_K is the orthogonal projection of x
  onto the (coordinate) subspace

$$\mathcal{R}_K = \mathbb{R}^K \times \{0\}^{-K} = \left\{x \in \mathbb{R}^d \,\middle|\, x_j = 0 \;, \;\; \forall j \notin K\right\} \subset \mathbb{R}^d$$

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

$$\ell_0^{\leq k} = \bigcup_{|K| \leq k} \mathcal{R}_K \;, \;\; \forall k = 0, 1, \ldots, d$$

We generate a sequence of coordinate norms from any source norm

For any source norm $|||\cdot|||$, we define

- a sequence $\left\{ |||\cdot|||^{\mathcal{R}}_{(k)} \right\}_{k=1,\ldots,d}$ of coordinate-k norms characterized by the following dual norms
- a sequence $\left\{ |||\cdot|||^{\mathcal{R}}_{(k),\star} \right\}_{k=1,\ldots,d}$ of dual coordinate-k norms by

$$|||\cdot|||^{\mathcal{R}}_{(k),\star} = \left(|||\cdot|||^{\mathcal{R}}_{(k)} \right)_{\star} = \sup_{|K| \leq k} \sigma_{\mathcal{R}_K \cap \mathbb{S}} = \sigma_{\ell_0^{\leq k} \cap \mathbb{S}}$$

$$|||y|||^{\mathcal{R}}_{(k),\star} = \sup_{|K| \leq k} |||y_K|||_{K,\star} \, , \quad \forall y \in \mathbb{R}^d$$

The case of ℓ_p-norms: $|||\cdot||| = \|\cdot\|_p$

For $y \in \mathbb{R}^d$, let ν be a permutation of $\{1, \ldots, d\}$ such that

$$|y_{\nu(1)}| \geq |y_{\nu(2)}| \geq \cdots \geq |y_{\nu(d)}|$$

$|||\cdot|||$	$|||x|||^{\mathcal{R}}_{(k)}$	$|||y|||^{\mathcal{R}}_{(k),\star}$
$\|\cdot\|_1$	$\|x\|_1$	$|y_{\nu(1)}| = \|y\|_\infty$
$\|\cdot\|_2$		$\sqrt{\sum_{l=1}^{k} |y_{\nu(l)}|^2}$
$\|\cdot\|_\infty$		$\sum_{l=1}^{k} |y_{\nu(l)}|$
$\|\cdot\|_p$		$\left(\sum_{l=1}^{k} |y_{\nu(l)}|^q\right)^{1/q}$ $1/p + 1/q = 1$

(More on these norms, and their dual norms, later)

The pseudonorm ℓ_0 and the ¢-coupling

Proposition

The pseudonorm ℓ_0,
the characteristic function $\delta_{\ell_0^{\leq k}}$ of its level sets
and the dual coordinate-k norms $|||\cdot|||^{\mathcal{R}}_{(k),\star}$
are conjugate as follows

$$\delta^{-¢}_{\ell_0^{\leq k}} = \delta^{¢}_{\ell_0^{\leq k}} = |||\cdot|||^{\mathcal{R}}_{(k),\star} \,, \quad k = 0, 1, \ldots, d$$

$$\ell_0^{¢} = \sup_{l=0,1,\ldots,d} \left[|||\cdot|||^{\mathcal{R}}_{(l),\star} - l \right]$$

Where have we gone till now? And what comes next

Fenchel conjugacy	Capra conjugacy
$\delta_{\ell_0^{\leq k}}^{(-\star)} = +\infty$	$\delta_{\ell_0^{\leq k}}^{-¢} = \|\!\|\!\|\cdot\|\!\|\!\|_{(k),\star}^{\mathcal{R}}$
$\ell_0^{\star} = \delta_{\{0\}}$	$\ell_0^{¢} = \sup_{l=0,1,\dots,d} \left[\|\!\|\!\|\cdot\|\!\|\!\|_{(l),\star}^{\mathcal{R}} - l\right]$

- The highly nonconvex constraint $x \in \ell_0^{\leq k}$ cannot be handled by the Fenchel conjugacy because $\delta_{\ell_0^{\leq k}}^{(-\star)} = +\infty$
- We have exhibited a new conjugacy Capra such that $\delta_{\ell_0^{\leq k}}^{-¢} = \|\!\|\!\|\cdot\|\!\|\!\|_{(k),\star}^{\mathcal{R}} < +\infty$, and with which we are ready to obtain lower bound dual problems for exact sparse optimization

Outline of the presentation

The Capra conjugacy ¢ and the pseudonorm ℓ_0
Background on couplings and Fenchel-Moreau conjugacies
Definition and properties of the Capra conjugacy

Lower bound convex programs for exact sparse optimization
Conjugate of the pseudonorm ℓ_0
Lower bound convex programs

Variational formulas for the ℓ_0 pseudonorm
Biconjugate of the pseudonorm ℓ_0
Covert convexity in the pseudonorm ℓ_0
Variational formulas for the pseudonorm ℓ_0

Conclusion

A second step towards a dual problem for sparse optimization

From

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

we deduce that

$$\sup_{y \in \mathbb{R}^d} \Big(-\big(\inf \big[f \mid n \big] \big)^{\star}(y) \dotplus \big(-\underbrace{\delta_{\ell_0^{\leq k}}^{-\cent}(y)}_{\|y\|_{(k),\star}^{\mathcal{R}}} \big) \Big) \leq \inf_{\ell_0(x) \leq k} f(x)$$

Concave dual problem for exact sparse optimization

Theorem

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

$$\sup_{y \in \mathbb{R}^d} \Big(-\big(\inf \big[f \mid n \big] \big)^{\star}(y) - \|\!| y |\!\|^{\mathcal{R}}_{(k),\star} \Big) \leq \inf_{\ell_0(x) \leq k} f(x)$$
$$= \inf_{\ell_0(x) \leq k} \inf \big[f \mid n \big](x)$$

The dual problem is the maximization of a concave usc function (possibly opening the way for numerical computation)

Convex primal problem for exact sparse optimization

Theorem

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

$$\min_{|||x|||_{(k)}^{\mathcal{R}} \leq 1} \left(\inf\left[f \mid n\right]\right)^{\star\star'}(x) \leq \inf_{\ell_0(x) \leq k} f(x) = \inf_{\ell_0(x) \leq k} \inf\left[f \mid n\right](x)$$

The primal problem is the minimization of a convex lsc function on the unit ball of the coordinate-k norm $|||\cdot|||_{(k)}^{\mathcal{R}}$
(possibly opening the way for numerical computation)

Least squares regression sparse optimization

$$
\begin{aligned}
& -\lVert\!| z |\!\rVert^2 + \sup_{y \in \mathbb{R}^d} \left(\left(-\frac{\langle z \,, A \,\cdot \rangle^2}{\lVert\!| A \cdot |\!\rVert^2} \mathbb{I}_{\langle z \,, A \,\cdot \rangle > 0} \dotplus \delta_{\mathbb{S}} \right)^{\star}(y) - \lVert\!| y |\!\rVert_{(k),\star}^{\mathcal{R}} \right) \\
= & \lVert\!| z |\!\rVert^2 + \min_{\lVert\!| x |\!\rVert_{(k)}^{\mathcal{R}} \leq 1} \left(-\frac{\langle z \,, A \,\cdot \rangle^2}{\lVert\!| A \cdot |\!\rVert^2} \mathbb{I}_{\langle z \,, A \,\cdot \rangle > 0} \dotplus \delta_{\mathbb{S}} \right)^{\star\star'}(x) \\
\leq & \inf_{\ell_0(x) \leq k} \lVert\!| z - Ax |\!\rVert^2
\end{aligned}
$$

Where have we gone till now? And what comes next

- Till now, we have used Capra conjugates, and have obtained lower bounds for optimization problems with constraints with any source norm

Fenchel conjugacy	Capra conjugacy
$\delta_{\ell_0^{\leq k}}^{(-\star)} = +\infty$	$\delta_{\ell_0^{\leq k}}^{-\text{¢}} = \lvert\lvert\lvert\cdot\rvert\rvert\rvert_{(k),\star}^{\mathcal{R}}$
$\ell_0^{\star} = \delta_{\{0\}}$	$\ell_0^{\text{¢}} = \sup_{l=0,1,\ldots,d} \big[\lvert\lvert\lvert\cdot\rvert\rvert\rvert_{(l),\star}^{\mathcal{R}} - l\big]$
$\ell_0^{\star\star'} = 0$	$\ell_0^{\text{¢}\text{¢}'} = ???$

- Now, we will study Capra biconjugates, and we will obtain dual representation formulas for so-called c-convex functions for orthant-strictly monotonic source norms

Outline of the presentation

The Capra conjugacy ¢ and the pseudonorm ℓ_0

Lower bound convex programs for exact sparse optimization

Variational formulas for the ℓ_0 pseudonorm

Conclusion

Outline of the presentation

The Capra conjugacy ¢ and the pseudonorm ℓ_0

Background on couplings and Fenchel-Moreau conjugacies

Definition and properties of the Capra conjugacy

Lower bound convex programs for exact sparse optimization

Conjugate of the pseudonorm ℓ_0

Lower bound convex programs

Variational formulas for the ℓ_0 pseudonorm

Biconjugate of the pseudonorm ℓ_0

Covert convexity in the pseudonorm ℓ_0

Variational formulas for the pseudonorm ℓ_0

Conclusion

Biconjugates provide lower c-convex functions

Proposition

$$\delta_{\left\{x\in\mathbb{R}^d \,\middle|\, |||x|||_{(k)}^{\mathcal{R}}=|||x|||\right\}} = \delta_{\ell_0^{\leq k}}^{¢¢'} \leq \delta_{\ell_0^{\leq k}}\ , \quad \forall k \in \{1,\ldots,d\}$$

$$\frac{1}{|||x|||} \min_{\substack{z^{(1)}\in\mathbb{R}^d,\ldots,z^{(d)}\in\mathbb{R}^d \\ \sum_{k=1}^d |||z^{(l)}|||_{(k)}^{\mathcal{R}} \leq |||x||| \\ \sum_{k=1}^d z^{(l)}=x}} \sum_{k=1}^d k|||z^{(l)}|||_{(k)}^{\mathcal{R}} = \ell_0^{¢¢'}(x) \leq \ell_0(x)$$

Our roadmap

- We are going to provide (necessary and) sufficient conditions under which the characteristic functions $\delta_{\ell_0^{\leq k}}$ and the ℓ_0 pseudonorm are ¢-convex, that is,

$$\delta_{\ell_0^{\leq k}}^{¢¢'} = \delta_{\ell_0^{\leq k}}$$

$$\ell_0^{¢¢'} = \ell_0$$

- For this purpose, we introduce the new notions of
  - orthant-strictly monotonic norm
  - graded sequence of norms

Orthant-strictly monotonic norms

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

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
  $\Big(|x| \leq |x'| \text{ and } x \circ x' \geq 0 \Rightarrow |||x||| \leq |||x'||| \Big)$,
  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, 2019]
  if, for all x, x' in $\mathbb{R}^d$, we have
  $\Big(|x| < |x'| \text{ and } x \circ x' \geq 0 \Rightarrow |||x||| < |||x'||| \Big)$,
  where $|x| < |x'|$ means that there exists $j \in \big\{1, \ldots, d\big\}$
  such that $|x_j| < |x_j^{'}|$

Examples of orthant-strictly monotonic norms among the ℓ_p-norms $\|\cdot\|_p$

- All the ℓ_p-norms $\|\cdot\|_p$ on the space $\mathbb{R}^d$, for $p \in [1, \infty]$, are monotonic, hence orthant-monotonic
- All the ℓ_p-norms $\|\cdot\|_p$ on the space $\mathbb{R}^d$, for $p \in [1, \infty[$, are orthant-strictly monotonic
- The ℓ_1-norm $\|\cdot\|_1$ is orthant-strictly monotonic, whereas its dual norm, the ℓ_∞-norm $\|\cdot\|_\infty$, is orthant-monotonic, but not orthant-strictly monotonic

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

Definition

For any source norm $|||\cdot|||$ on $\mathbb{R}^d$, for any $k \in \{1, \dots, d\}$, we call

- generalized top-k dual norm the norm

$$|||y|||^{\mathrm{tn}}_{\star,(k)} = \sup_{|K| \leq k} |||y_K|||_{\star} = \sup_{|K| \leq k} |||y_K|||_{\star,K} \; , \quad \forall y \in \mathbb{R}^d$$

- generalized k-support dual norm the dual norm

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

In the Euclidian case were the source norm is $\|\cdot\|_2$, we recover the original definition of top-k dual norms, used to define the k-support dual norms in [Argyriou, Foygel, and Srebro, 2012]

The case of ℓ_p-norms: $|||\cdot||| = \|\cdot\|_p$

For $y \in \mathbb{R}^d$, let ν be a permutation of $\{1, \ldots, d\}$ such that

$$|y_{\nu(1)}| \geq |y_{\nu(2)}| \geq \cdots \geq |y_{\nu(d)}|$$

$\vert\vert\vert\cdot\vert\vert\vert$	$\vert\vert\vert x\vert\vert\vert^{\star\mathrm{sn}}_{\star,(k)}$	$\vert\vert\vert y\vert\vert\vert^{\mathrm{tn}}_{\star,(k)}$
$\Vert\cdot\Vert_p$	(p, k)-support norm $\vert\vert x\vert\vert^{\mathrm{sn}}_{p,k}$	top (k, q)-norm $\vert\vert y\vert\vert^{\mathrm{tn}}_{k,q}$ $= \big(\sum_{l=1}^{k} \vert y_{\nu(l)}\vert^q\big)^{1/q}$, $1/p + 1/q = 1$
$\Vert\cdot\Vert_1$	$(1, k)$-support norm ℓ_1-norm $\vert\vert x\vert\vert^{\mathrm{sn}}_{1,k} = \Vert x\Vert_1$	top (k, ∞)-norm ℓ_∞-norm $\vert\vert y\vert\vert^{\mathrm{tn}}_{k,\infty} = \vert y_{\nu(1)}\vert = \Vert y\Vert_\infty$
$\Vert\cdot\Vert_2$	$(2, k)$-support norm	top $(k, 2)$-norm $\vert\vert y\vert\vert^{\mathrm{tn}}_{k,2} = \sqrt{\sum_{l=1}^{k} \vert y_{\nu(l)}\vert^2}$
$\Vert\cdot\Vert_\infty$	(∞, k)-support norm	top $(k, 1)$-norm $\vert\vert y\vert\vert^{\mathrm{tn}}_{k,1} = \sum_{l=1}^{k} \vert y_{\nu(l)}\vert$

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

k-coordinate norm		k-support dual norm
$\|\|\|\cdot\|\|\|^{\mathcal{R}}_{(k)}$	$\leq$	$\|\|\|\cdot\|\|\|^{\star\mathrm{sn}}_{\star,(k)}$
dual k-coordinate norm		top-k dual norm
$\|\|\|\cdot\|\|\|^{\mathcal{R}}_{(k),\star} = \sup_{\lvert K\rvert\leq k}\|\|\|\cdot\|\|\|_{K,\star}$	$\geq$	$\sup_{\lvert K\rvert\leq k}\|\|\|\cdot\|\|\|_{\star,K} = \|\|\|\cdot\|\|\|^{\mathrm{tn}}_{\star,(k)}$

Orthant-monotonic source norms generate coordinate norms and duals that are generalized top-k and k-support dual norms

Proposition

If the source norm is orthant monotonic, we have

$$|||\cdot|||_{K,\star} = |||\cdot|||_{\star,K} \; , \;\; \forall K \subset \{1, \ldots, d\}$$

hence, for all $k \in \{1, \ldots, d\}$,

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

We define *graded sequence of norms*

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

when the sequence becomes stationary

Definition

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

$$\ell_0(y) = l \iff |||y|||_1 \leq \cdots \leq |||y|||_{l-1} < |||y|||_l = \cdots = |||y|||_d$$

or, equivalently, $k \in \{1, \dots, d\} \mapsto |||y|||_k$ is nondecreasing and

$$\ell_0(y) \leq l \iff |||y|||_l = |||y|||_d$$

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

$$\underbrace{\left\{ |||\cdot|||^{\mathrm{tn}}_{\star,(l)} \right\}_{l=1,\ldots,d}}_{\textit{generalized top-k dual norm}} = \underbrace{\left\{ |||\cdot|||^{\mathcal{R}}_{(l),\star} \right\}_{l=1,\ldots,d}}_{\textit{dual-k coordinate norm}}$$

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

We establish ¢-convexity of the pseudonorm ℓ_0

Theorem

- *The sequence* $\left\{ |||\cdot|||^{\mathcal{R}}_{(l)} \right\}_{l=1,\ldots,d}$ *of coordinate-k norms is* *decreasingly graded with respect to the* ℓ_0 *pseudonorm iff*

$$\delta^{¢¢'}_{\ell_0^{\leq k}} = \delta_{\ell_0^{\leq k}}$$

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

$$\ell_0^{¢¢'} = \ell_0$$

Proof: conditions for nonempty Capra-subdifferentials

$$\partial_{\cent}\delta_{\ell_0^{\leq k}}(x) = \begin{cases} \emptyset & \text{if } \ell_0(x) = k+1, \dots, d \text{ or } |||x||| < |||x|||^{\mathcal{R}}_{(k)} \\ N_{\mathbb{B}^{\mathcal{R}}_{(k)}}\big(\frac{x}{|||x|||^{\mathcal{R}}_{(k)}}\big) & \text{if } \ell_0(x) = 1, \dots, k \text{ and } |||x||| = |||x|||^{\mathcal{R}}_{(k)} \\ \{0\} & \text{if } \ell_0(x) = 0 \end{cases}$$

$$\partial_{\cent}\ell_0(x) = \begin{cases} N_{\mathbb{B}^{\mathcal{R}}_{(l)}}\big(\frac{x}{|||x|||^{\mathcal{R}}_{(k)}}\big) \cap \\ \Big\{ y \in \mathbb{R}^d \mid |||y|||^{\mathcal{R}}_{(l),\star} - l = \sup_{k=0,1,\dots,d} \big[|||y|||^{\mathcal{R}}_{(k),\star} - k \big] \\ \quad \text{and } |||y|||^{\mathcal{R}}_{(l),\star} = |||y_L|||_{L,\star} \text{ where } L = \mathrm{supp}(x) \Big\} \\ \quad \text{if } l = \ell_0(x) \geq 1 \\ \bigcap_{k=1,\dots,d} k\mathbb{B}^{\mathcal{R}}_{(k),\star} \qquad \text{if } x = 0 \end{cases}$$

Outline of the presentation

The Capra conjugacy ¢ and the pseudonorm ℓ_0
Background on couplings and Fenchel-Moreau conjugacies
Definition and properties of the Capra conjugacy

Lower bound convex programs for exact sparse optimization
Conjugate of the pseudonorm ℓ_0
Lower bound convex programs

Variational formulas for the ℓ_0 pseudonorm
Biconjugate of the pseudonorm ℓ_0
Covert convexity in the pseudonorm ℓ_0
Variational formulas for the pseudonorm ℓ_0

Conclusion

The pseudonorm ℓ_0 coincides, on the sphere, with a convex lsc function defined on the whole space

Proposition

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

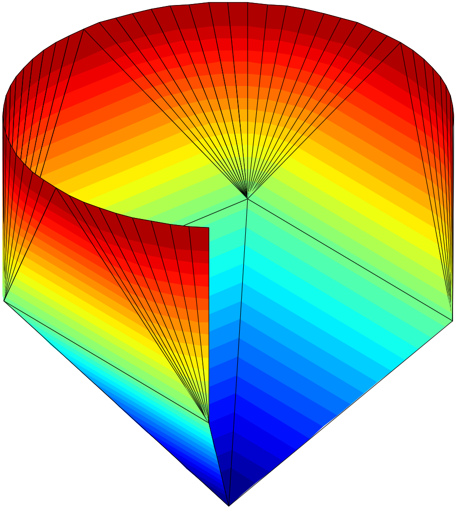
$$\ell_0(x) = \mathcal{L}_0\Big(\frac{x}{|||x|||}\Big) \; , \quad \forall x \in \mathbb{R}^d \backslash \{0\}$$

$$\textit{where} \;\; \mathcal{L}_0 = \underbrace{\Big(\sup_{l=0,1,\ldots,d} \big[|||\cdot|||^{\mathrm{tn}}_{\star,(l)} - l \big] \Big)^\star}_{\textit{convex lsc on } \mathbb{R}^d}$$

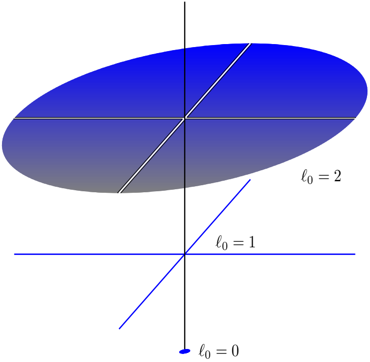
$$\begin{aligned}
\text{Proof: } \ell_0(x) &= \ell_0^{¢¢'}(x) \\
&= \sup_{y \in \mathbb{R}^d} \Big(¢(x,y) \dotplus \big(-\ell_0^{¢}(y) \big) \Big) \\
&= \sup_{y \in \mathbb{R}^d} \Big(\frac{\langle x \, , y \rangle}{\|x\|} \dotplus \big(- \sup_{l=0,1,\ldots,d} \big[|||y|||^{\mathrm{tn}}_{\star,(l)} - l \big] \big) \Big) = \mathcal{L}_0\Big(\frac{x}{|||x|||}\Big)
\end{aligned}$$

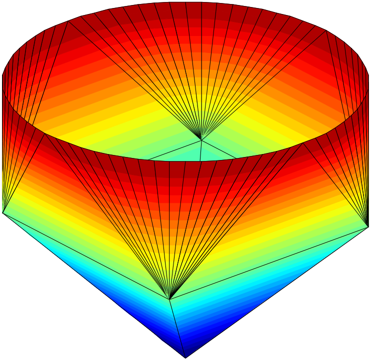
Covert convexity in the pseudonorm ℓ_0

Here is graph of the convex lsc function $\mathcal{L}_0$ such that $\ell_0 = \mathcal{L}_0$ on the circle



The pseudonorm ℓ_0 coincides, on the sphere (circle on $\mathbb{R}^2$), with a convex lsc function





What is the convex lsc function $\mathcal{L}_0$?

Proposition

In dimension $d = 2$, the function $\mathcal{L}_0$ is given by

$$\mathcal{L}_0(x_1, x_2) = \begin{cases} +\infty & \text{if } x_1^2 + x_2^2 > 1 \,, & (1) \\ 1 & \text{if } (x_1, x_2) \in \{(1,0), (0,1), (-1,0), (0,-1)\} \,, & (2) \\ 2 & \text{if } x_1^2 + x_2^2 = 1 \text{ and } (x_1, x_2) \notin \{(1,0), (0,1), (-1,0), (0,-1)\} \,, & (3) \end{cases}$$

and, for any (x_1, x_2) such that $x_1^2 + x_2^2 < 1$ by

$$\mathcal{L}_0(x_1, x_2) = \begin{cases} |x_1| + |x_2| & \text{if } |x_1| + |x_2| \le 1 \,, & (4) \\ \dfrac{|x_1| + |x_2| - 2 + \sqrt{2}}{\sqrt{2} - 1} & \text{if } \begin{cases} (\sqrt{2} - 1)|x_1| + |x_2| < 1 < |x_1| + |x_2| \\ \text{or} \\ |x_1| + (\sqrt{2} - 1)|x_2| < 1 < |x_1| + |x_2| \,, \end{cases} & (5) \\ \dfrac{3 - |x_2|}{2} + \dfrac{x_1^2}{2(1 - |x_2|)} & \text{if } (\sqrt{2} - 1)|x_1| + |x_2| \ge 1 \text{ and } |x_2| > |x_1| \,, & (6) \\ \dfrac{3 - |x_1|}{2} + \dfrac{x_2^2}{2(1 - |x_1|)} & \text{if } |x_1| + (\sqrt{2} - 1)|x_2| \ge 1 \text{ and } |x_1| > |x_2| \,. & (7) \end{cases}$$

Outline of the presentation

The Capra conjugacy ¢ and the pseudonorm ℓ_0

Background on couplings and Fenchel-Moreau conjugacies

Definition and properties of the Capra conjugacy

Lower bound convex programs for exact sparse optimization

Conjugate of the pseudonorm ℓ_0

Lower bound convex programs

Variational formulas for the ℓ_0 pseudonorm

Biconjugate of the pseudonorm ℓ_0

Covert convexity in the pseudonorm ℓ_0

Variational formulas for the pseudonorm ℓ_0

Conclusion

Variational formula for the ℓ_0 pseudonorm

Theorem

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

$$\ell_0(x) = \frac{1}{|||x|||} \underbrace{\min_{\substack{z^{(1)}\in\mathbb{R}^d,\ldots,z^{(d)}\in\mathbb{R}^d \\ \sum_{k=1}^d |||z^{(k)}|||_{\star,(k)}^{\star\mathrm{sn}} \leq |||x||| \\ \sum_{k=1}^d z^{(k)} = x}} \sum_{k=1}^d k |||z^{(k)}|||_{\star,(k)}^{\star\mathrm{sn}}}_{\textit{convex optimization problem}}$$

The case of ℓ_p-norms: $|||\cdot||| = \|\cdot\|_p$ for $p \in]1, \infty[$

$$\ell_0(x) = \frac{1}{\|x\|_p} \min_{\substack{z^{(1)} \in \mathbb{R}^d, \dots, z^{(d)} \in \mathbb{R}^d \\ \sum_{k=1}^d ||z^{(k)}||_{p,k}^{\mathrm{sn}} \le \|x\|_p \\ \sum_{k=1}^d z^{(k)} = x}} \sum_{k=1}^d k ||z^{(k)}||_{p,k}^{\mathrm{sn}}$$

With any norm, we have an inequality

$$\ell_0(x) \geq \frac{1}{|||x|||} \min_{\substack{z^{(1)}\in\mathbb{R}^d,\ldots,z^{(d)}\in\mathbb{R}^d \\ \sum_{k=1}^d |||z^{(k)}|||_{(k)}^{\mathcal{R}} \leq |||x||| \\ \sum_{k=1}^d z^{(k)} = x}} \sum_{k=1}^{d} k |||z^{(k)}|||_{(k)}^{\mathcal{R}}$$

In the case of the ℓ_1-norm, $|||\cdot||| = \|\cdot\|_1$,
we obtain the trivial inequality $x \neq 0 \Rightarrow \ell_0(x) \geq 1$...

Minimization of the pseudonorm ℓ_0 under constraints

Proposition

Let $C \subset \mathbb{R}^d$ be such that $0 \notin C$
If both the norm $\|\!|\cdot|\!\|$ and the dual norm $\|\!|\cdot|\!\|_\star$ are orthant-strictly monotonic, we have

$$\min_{x \in C} \ell_0(x) = \min_{\substack{x \in C, z^{(1)} \in \mathbb{R}^d, \ldots, z^{(d)} \in \mathbb{R}^d \\ \sum_{k=1}^d \|\!|z^{(k)}|\!\|_{\star,(k)}^{\star \mathrm{sn}} \leq \|\!|x|\!\| \\ \sum_{k=1}^d z^{(k)} = x}} \frac{1}{\|\!|x|\!\|} \sum_{k=1}^d k \|\!|z^{(k)}|\!\|_{\star,(k)}^{\star \mathrm{sn}}$$

$$= \min_{x \in C} \frac{1}{\|\!|x|\!\|} \underbrace{\min_{\substack{z^{(1)} \in \mathbb{R}^d, \ldots, z^{(d)} \in \mathbb{R}^d \\ \sum_{k=1}^d \|\!|z^{(k)}|\!\|_{\star,(k)}^{\star \mathrm{sn}} \leq \|\!|x|\!\| \\ \sum_{k=1}^d z^{(k)} = x}} \sum_{k=1}^d k \|\!|z^{(k)}|\!\|_{\star,(k)}^{\star \mathrm{sn}}}_{\textit{convex optimization problem}}$$

Minimization over level sets of the pseudonorm ℓ_0

Proposition

Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ and $k \in \{1, \ldots, d\}$
If both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$
are orthant-strictly monotonic, we have

$$\min_{\ell_0(x) \leq k} f(x) = \min_{\substack{x \in \mathbb{R}^d, z^{(1)} \in \mathbb{R}^d, \ldots, z^{(d)} \in \mathbb{R}^d \\ \sum_{k=1}^d |||z^{(k)}|||_{\star,(k)}^{\star\mathrm{sn}} \leq |||x||| \\ \sum_{k=1}^d z^{(k)} = x \\ \sum_{k=1}^d k|||z^{(k)}|||_{\star,(k)}^{\star\mathrm{sn}} \leq k|||x|||}} f(x)$$

$$= \min_{\substack{z^{(1)} \in \mathbb{R}^d, \ldots, z^{(d)} \in \mathbb{R}^d \\ \sum_{k=1}^d |||z^{(k)}|||_{\star,(k)}^{\star\mathrm{sn}} \leq |||\sum_{k=1}^d z^{(k)}||| \\ \sum_{k=1}^d k|||z^{(k)}|||_{\star,(k)}^{\star\mathrm{sn}} \leq k|||\sum_{k=1}^d z^{(k)}|||}} f\left(\sum_{k=1}^d z^{(k)}\right)$$

Outline of the presentation

The Capra conjugacy ¢ and the pseudonorm ℓ_0

Lower bound convex programs for exact sparse optimization

Variational formulas for the ℓ_0 pseudonorm

Conclusion

Conclusion (1/3)

- We have dealt with sparse optimization in an exact way (and *not* with *susbstitute* convex formulations)
- Using generalized convexity with an original coupling Capra, Fenchel after primal normalization with a (source) norm, we have displayed a suitable conjugacy for the pseudonorm ℓ_0

Conclusion (2/3)

Without any assumption on the (source) norm
and on the objective function to be minimized,
we have obtained a lower bound
for any k-sparse optimization problem, which is

- a usc concave dual maximization problem
  involving the dual coordinate-k norm
  (always)
- a lsc convex primal minimization problem
  on the unit ball of the coordinate-k norm
  (under a mild assumption)

Conclusion (3/3)

With proper assumptions on the (source) norm
(related to orthant-strict monotonicity and rotundity),
we have

- produced graded sequences of generalized top-k dual norms, suitable for "difference of convex" (DC) optimization methods
- revealed covert convexity in the pseudonorm ℓ_0
- yielded variational formulas for the ℓ_0 pseudonorm involving generalized k-support dual norms and convex parts

Open questions

- Are the lower bounds accurate?
- Do the lower bound convex programs provide good approximate solutions?
- Are variational formulas for the ℓ_0 pseudonorm computationaly tractable?
- Do Capra-subdifferentials formulas pave the way for suitable algorithms for sparse optimization?

Towards a Capra-subdifferential descent method?

- For any $y \in \partial_{¢} f(x)$, we have

$$¢(x', y) \dotplus \big(- f(x') \big) \leq ¢(x, y) \dotplus \big(- f(x) \big)$$

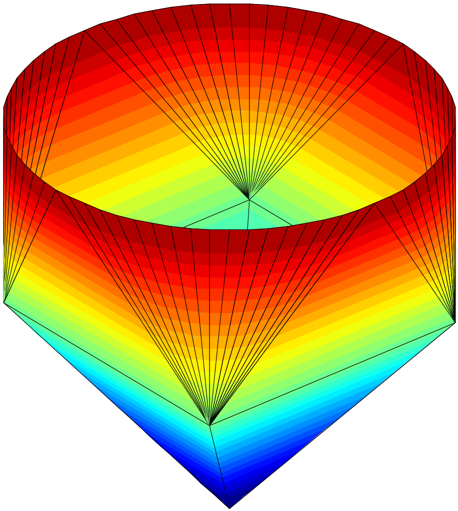
  so that an algorithm to find a minimum of f over $\ell_0^{\leq k}$ would exploit the inequality

$$(f \dotplus \delta_{\ell_0^{\leq k}})(x^{(j+1)}) \dotplus \Big(- (f \dotplus \delta_{\ell_0^{\leq k}})(x^{(j)}) \Big) \leq \Big\langle \frac{x^{(j)}}{|||x^{(j)}|||} - \frac{x^{(j+1)}}{|||x^{(j+1)}|||} , \underbrace{y^{(j+1)}}_{\in \partial_{¢}(f \dotplus \delta_{\ell_0^{\leq k}})(x^{(j+1)})} \Big\rangle$$

- Starting from $(x^{(j)}, y^{(j)}) \in \mathbb{R}^d \times \mathbb{R}^d$, find $(x^{(j+1)}, y^{(j+1)}) \in \mathbb{R}^d \times \mathbb{R}^d$ such that

$$\begin{cases} y^{(j+1)} \in \partial_{¢}(f \dotplus \delta_{\ell_0^{\leq k}})(x^{(j+1)}) \\ \Big\langle \frac{x^{(j)}}{|||x^{(j)}|||} - \frac{x^{(j+1)}}{|||x^{(j+1)}|||} , y^{(j+1)} \Big\rangle \leq \cdots \leq 0 \end{cases}$$

Thank you :-)



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. Orthant-strictly monotonic norms, graded sequences and generalized top-k and k-support norms for sparse optimization, 2019. preprint.

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

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