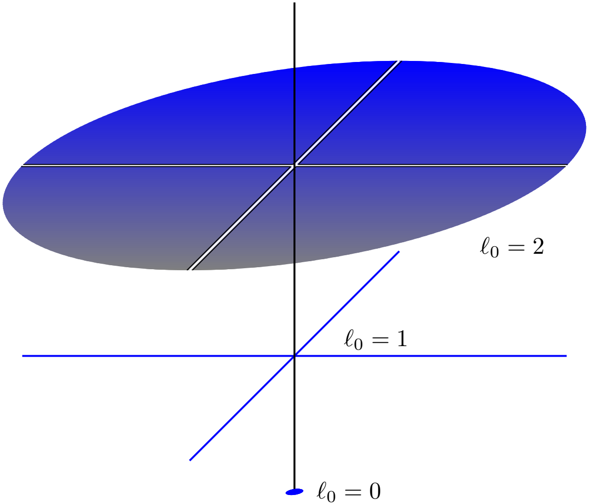
Hidden Convexity in the ℓ_0 Pseudonorm

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

NCSU Nonlinear Analysis Seminar
19 October 2022

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



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

Only 1-homogeneity is missing, whereas 0-homogeneity holds true

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

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

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

- The function ℓ_0 pseudonorm : $\mathbb{R}^d \to \{0, 1, \ldots, d\}$ 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

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

- We denote the level sets of the ℓ_0 pseudonorm by

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

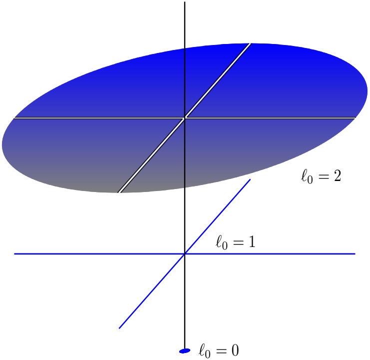
Fenchel *versus* E-Capra conjugacies for the ℓ_0 pseudonorm

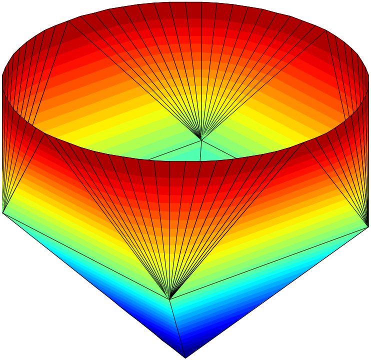
Fenchel conjugacy	E-Capra conjugacy
$\delta^{\star}_{\ell_0^{\leq k}} = \delta_{\{0\}}, \; k \neq 0$	$\delta^{\cent}_{\ell_0^{\leq k}} = \|\cdot\|^{\mathrm{tn}}_{2,k}$
$\ell_0^{\star} = \delta_{\{0\}}$	$\ell_0^{\cent} = \sup_{l=0,1,\dots,d} \left[\|\cdot\|^{\mathrm{tn}}_{2,l} - l\right]$
$\ell_0^{\star\star'} = 0$	$\ell_0^{\cent\cent'} = \ell_0$

where, for any subset $W \subset \mathbb{R}^d$,
the *characteristic function* δ_W of the set W is given by

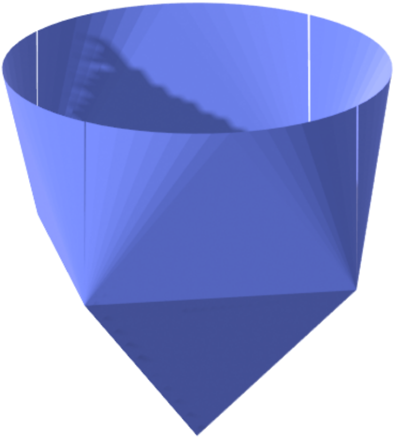
$$\delta_W(w) = 0 \;\text{ if }\; w \in W \;, \;\; \delta_W(w) = +\infty \;\text{ if }\; w \not\in W$$

The ℓ_0 pseudonorm coincides, on the Euclidean unit sphere with a proper convex lsc function $\mathcal{L}_0$





This function $\mathcal{L}_0$ is the best convex lower approximation of the ℓ_0 pseudonorm on the Euclidean unit ball



Variational formulas for the ℓ_0 pseudonorm

Proposition

[Chancelier and De Lara, 2021]

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

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

Outline of the presentation

Background on one-sided linear couplings

The Euclidean Capra conjugacy

Extension: constant along primal rays conjugacies

Conclusion

Outline of the presentation

Background on one-sided linear couplings

Background on couplings and Fenchel-Moreau conjugacies

One-sided linear couplings (and hidden convexity)

The Euclidean CAPRA conjugacy

Extension: constant along primal rays conjugacies

CAPRA conjugacies

Best convex approximations of 0-homogeneous functions

The case of norms

Conclusion

Outline of the presentation

Background on one-sided linear couplings

Background on couplings and Fenchel-Moreau conjugacies

One-sided linear couplings (and hidden convexity)

The Euclidean CAPRA conjugacy

Extension: constant along primal rays conjugacies

CAPRA conjugacies

Best convex approximations of 0-homogeneous functions

The case of norms

Conclusion

The Fenchel conjugacy

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\} = [-\infty, +\infty]$$

Definition

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

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

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

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

Background on couplings and Fenchel-Moreau conjugacies

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

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

  We also use the notation $\mathbb{X} \overset{c}{\leftrightarrow} \mathbb{Y}$ for a coupling
- The Moreau lower addition extends the usual addition with

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

- The Moreau upper addition extends the usual addition with

$$(+\infty) \dotplus (-\infty) = (-\infty) \dotplus (+\infty) = +\infty$$

Fenchel-Moreau conjugate

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

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 $(+, \times)$
$\sup_{x \in \mathbb{X}} \Big(c(x,y) \dot{+} \big(-f(x)\big) \Big)$	$\int_{\mathbb{X}} c(x,y) f(x) dx$

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, hence a tool for duality in optimization
  (uses biconjugates)
- dual representation formulas for c-convex functions (uses biconjugates and subdifferentials)

[Martínez-Legaz, 2005]

"Fenchel-like" inequality yields a lower bound

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

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

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

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

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

Example: $\star$-convex functions
$=$ closed convex functions [Rockafellar, 1974, p. 15]
$=$ proper convex lsc or $\equiv -\infty$ or $\equiv +\infty$
$=$ suprema of affine functions

Subdifferential of a conjugacy

For any function $f : \mathbb{X} \to \overline{\mathbb{R}}$ and $x \in \mathbb{X}$, there are three possibilities for the c-subdifferential

$$
\begin{aligned}
y \in \mathbb{Y}\,,\ \ y \in \partial_c f(x) &\iff f^c(y) = c(x,y) \dotplus (-f(x)) \\
y \in \mathbb{Y}\,,\ \ y \in \partial^c f(x) &\iff f(x) = c(x,y) \dotplus (-f^c(y)) \\
y \in \mathbb{Y}\,,\ \ y \in \partial_c^c f(x) &\iff c(x,y) = f(x) \dotplus (-f^c(y))
\end{aligned}
$$

$$\partial^c f(x) \neq \emptyset \Rightarrow f^{cc'}(x) = f(x)$$

If $-\infty < c < +\infty$ and $x \in \mathrm{dom} f$, we have

$$
\begin{aligned}
&\partial_c f(x) = \partial^c f(x) = \partial_c^c f(x) \\
=&\{y \in \mathbb{Y} \mid c(x',y) - f(x') \leq c(x,y) - f(x)\,,\ \ \forall x' \in \mathbb{X}\}
\end{aligned}
$$

Dual problems: perturbation scheme

- Set $\mathbb{W}$, function $h : \mathbb{W} \to \overline{\mathbb{R}}$ and original minimization problem

$$\inf_{w \in \mathbb{W}} h(w)$$

- Embedding/perturbation scheme given by a nonempty set $\mathbb{X}$, an element $\overline{x} \in \mathbb{X}$ and a function $H : \mathbb{W} \times \mathbb{X} \to \overline{\mathbb{R}}$ such that

$$h(w) = H(w, \overline{x}) \; , \; \; \forall w \in \mathbb{W}$$

- Value function

$$\varphi(x) = \inf_{w \in \mathbb{W}} H(w, x) \; , \; \; \forall x \in \mathbb{X}$$

- Original minimization problem

$$\varphi(\overline{x}) = \inf_{w \in \mathbb{W}} H(w, \overline{x}) = \inf_{w \in \mathbb{W}} h(w)$$

Dual problems: conjugacy, weak and strong duality

- Coupling $\mathbb{X} \overset{c}{\leftrightarrow} \mathbb{Y}$, and Lagrangian $\mathcal{L} : \mathbb{W} \times \mathbb{Y} \to \overline{\mathbb{R}}$ given by

$$\mathcal{L}(w, y) = \inf_{x \in \mathbb{X}} \left\{ H(w, x) \dotplus (-c(x, y)) \right\}$$

- Dual maximization problem

$$-\varphi^{c}(y) = -\sup_{x \in \mathbb{X}} \left\{ c(x, y) \dotplus (-\inf_{w \in \mathbb{W}} H(w, x)) \right\} = \inf_{w \in \mathbb{W}} \mathcal{L}(w, y)$$

$$\varphi^{cc'}(\overline{x}) = \sup_{y \in \mathbb{Y}} \{ c(\overline{x}, y) \dotplus \inf_{w \in \mathbb{W}} \mathcal{L}(w, y) \}$$

- Weak duality always holds true

$$\varphi^{cc'}(\overline{x}) = \sup_{y \in \mathbb{Y}} \{ c(\overline{x}, y) \dotplus \inf_{w \in \mathbb{W}} \mathcal{L}(w, y) \} \leq \inf_{w \in \mathbb{W}} h(w) = \varphi(\overline{x})$$

- Strong duality holds true when φ is c-convex at $\overline{x}$, that is,

$$\varphi^{cc'}(\overline{x}) = \sup_{y \in \mathbb{Y}} \{ c(\overline{x}, y) \dotplus \inf_{w \in \mathbb{W}} \mathcal{L}(w, y) \} = \inf_{w \in \mathbb{W}} h(w) = \varphi(\overline{x})$$

Outline of the presentation

Background on one-sided linear couplings
Background on couplings and Fenchel-Moreau conjugacies
One-sided linear couplings (and hidden convexity)

The Euclidean Capra conjugacy

Extension: constant along primal rays conjugacies
Capra conjugacies
Best convex approximations of 0-homogeneous functions
The case of norms

Conclusion

One-sided linear couplings

- We consider two vector spaces $\mathbb{X}$ and $\mathbb{Y}$ paired by a bilinear form $\langle \cdot \, , \, \cdot \rangle$
- We suppose given a mapping $\theta : \mathbb{W} \to \mathbb{X}$, where $\mathbb{W}$ is any set

Definition

We define the one-sided linear coupling (OSL)

$$\mathbb{W} \overset{\star_\theta}{\longleftrightarrow} \mathbb{Y}$$

between $\mathbb{W}$ and $\mathbb{Y}$ by

$$\star_\theta(w, y) = \langle \theta(w) \, , \, y \rangle \ , \ \forall w \in \mathbb{W} \ , \ \forall y \in \mathbb{Y}$$

OSL-couplings induce conjugacies that share nice properties with the classic Fenchel conjugacy

Proposition

[Chancelier and De Lara, 2021]
For any functions $h : \mathbb{W} \to \overline{\mathbb{R}}$ and $g : \mathbb{Y} \to \overline{\mathbb{R}}$,
the Fenchel-Moreau conjugates are given by

$$h^{\star_\theta} = \big(\inf\big[h \mid \theta\big]\big)^{\star}$$
$$g^{\star'_\theta} = g^{\star} \circ \theta$$

where, for all $x \in \mathbb{X}$,

$$\inf\big[h \mid \theta\big](x) = \inf\big\{h(w) \,\big|\, w \in \mathbb{W},\ \theta(w) = x\big\}$$

OSL-subdifferentials share properties with the Rockafellar-Moreau subdifferential

Definition

For any function $h : \mathbb{W} \to \overline{\mathbb{R}}$ and $w \in \mathbb{W}$, the $\star_\theta$-subdifferential is

$$\partial_{\star_\theta} h(w) = \{y \in \mathbb{Y} \mid \quad \langle \theta(w'), \, y \rangle - h(w') \\ \leq \quad \langle \theta(w), \, y \rangle - h(w) \, , \ \forall w' \in \mathbb{W}\}$$

The following properties are satisfied

$$\partial_{\star_\theta} h(w) \text{ is a closed convex subset of } \mathbb{Y}$$

$$y \in \partial_{\star_\theta} h(w) \iff h^{\star_\theta}(y) = \langle \theta(w), \, y \rangle - h(w)$$

$$w \in \arg\min h \iff 0 \in \partial_{\star_\theta} h(w)$$

$$\partial_{\star_\theta} h + \partial_{\star_\theta} k \subset \partial_{\star_\theta} (h \dotplus k)$$

$$w \in \mathrm{dom} h \, , \quad \partial_{\star_\theta} h(w) \neq \emptyset \Rightarrow h^{\star_\theta \star_\theta'}(w) = h(w)$$

The $\star_\theta$-convex functions are characterized by a *convex factorization* property (hidden convexity)

$$\star_\theta\text{-convex function} = \underbrace{\text{closed convex function}}_{\text{proper convex lsc or } \equiv -\infty \text{ or } \equiv +\infty} \circ \ \theta$$

Proposition

[Chancelier and De Lara, 2021]

$$\star_\theta \text{-convexity of the function } h : \mathbb{W} \to \overline{\mathbb{R}}$$

$$\iff h = h^{\star_\theta \star_\theta{}'}$$

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

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

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

such that $h = f \circ \theta$

Concave dual problem

Proposition

For any function $h : \mathbb{W} \to \overline{\mathbb{R}}$, and nonempty set $W \subset \mathbb{W}$, we have the following lower bound

$$\sup_{y \in \mathbb{Y}} \overbrace{\Big(\big(-\big(\inf \left[h \mid \theta \right]\big)^{\star}(y) \big) \dot{+} \big(-\sigma_{-\theta(W)}(y) \big) \Big)}^{\text{concave usc function}}$$

$$\leq \inf_{x \in \theta(W)} \inf \left[h \mid \theta \right](x) = \inf_{w \in W} h(w)$$

Perturbation scheme

- Functions $k : \mathbb{W} \to \overline{\mathbb{R}}$, $h : \mathbb{W} \to \overline{\mathbb{R}}$ $\star_\theta$-convex, and original minimization problem

$$\inf_{w \in \mathbb{W}} \Big\{ k(w) \dot{+} h(w) \Big\} = \inf_{w \in \mathbb{W}} \Big\{ k(w) \dot{+} h^{\star_\theta \star'}\big(\theta(w)\big) \Big\}$$

because $h = h^{\star_\theta \star_\theta'} = h^{\star_\theta \star'} \circ \theta$

- Embedding/perturbation scheme $H : \mathbb{W} \times \mathbb{X} \to \overline{\mathbb{R}}$ given by

$$H(w,x) = k(w) \dot{+} h^{\star_\theta \star'}\big(\theta(w) + x\big) \ , \ \ \forall (w,x) \in \mathbb{W} \times \mathbb{X}$$

- Value function

$$\varphi(x) = \inf_{w \in \mathbb{W}} \Big\{ k(w) \dot{+} h^{\star_\theta \star'}\big(\theta(w) + x\big) \Big\} \ , \ \ \forall x \in \mathbb{X}$$

Lagrangian and dual problem

- Lagrangian $\mathcal{L} : \mathbb{W} \times \mathbb{Y} \to \overline{\mathbb{R}}$ given, for any $(w, y) \in \mathbb{W} \times \mathbb{Y}$, by

$$\mathcal{L}(w, y) = \inf_{x \in \mathbb{X}} \Big\{ k(w) \dotplus h^{\star\theta\star'}\big(\theta(w) + x\big) - \langle x, \ y \rangle \Big\}$$
$$= k(w) \dotplus \langle \theta(w), \ y \rangle \dotplus \big(-h^{\star\theta}(y)\big)$$

- Dual maximization problem

$$\varphi^{\star\star'}(0) = \sup_{y \in \mathbb{Y}} \inf_{w \in \mathbb{W}} \mathcal{L}(w, y) = \sup_{y \in \mathbb{Y}} \Big\{ \big(-k^{-\star\theta}(y)\big) \, \underset{\cdot}{+} \, \big(-h^{\star\theta}(y)\big) \Big\}$$

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

$$\varphi(0) = \inf_{w \in \mathbb{W}} \sup_{y \in \mathbb{Y}} \mathcal{L}(w, y) = \inf_{w \in \mathbb{W}} \Big\{ k(w) \dotplus h(w) \Big\}$$

- Existence of a saddle point? Algorithms?

Our roadmap (1/2)

- Introduce the Euclidean-CAPRA coupling (E-Capra), a particular one-sided linear coupling
- Show how the Euclidean-CAPRA coupling proves suitable to analyze the ℓ_0 pseudonorm
  - E-Capra-convexity
  - hidden convexity
  - best convex lower approximation on the unit ball
  - E-Capra-subdifferential (thanks to Adrien Le Franc)
  - variational formulas
  - difference of convex (DC) formulas with graded sequences of induced norms
  - concave dual problems in sparse optimization
  - duality

Our roadmap (2/2)

- Introduce a subclass of one-sided linear couplings, the constant along primal rays (CAPRA) couplings, depending on a source norm, and more generally on a 1-homogeneous nonnegative function
  - relevant classes of norms
  - relevant classes of functions
  - matrix functions and norms

Outline of the presentation

Background on one-sided linear couplings

Background on couplings and Fenchel-Moreau conjugacies

One-sided linear couplings (and hidden convexity)

The Euclidean Capra conjugacy

Extension: constant along primal rays conjugacies

Capra conjugacies

Best convex approximations of 0-homogeneous functions

The case of norms

Conclusion

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

Definition

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

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

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 $\mathbb{S}_2$ and the pointed unit sphere $\mathbb{S}_2^{(0)}$ by

$$\mathbb{S}_2 = \left\{x \in \mathbb{R}^d \;\middle|\; \|x\|_2 = 1\right\} , \quad \mathbb{S}_2^{(0)} = \mathbb{S}_2 \cup \{0\}$$

- and we define the primal normalization mapping n as

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

- so that the coupling E-Capra

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

  appears as the Fenchel coupling after primal normalization
- hence, the coupling E-Capra is one-sided linear

The E-Capra conjugacy shares properties with the Fenchel conjugacy

Proposition

[Chancelier and De Lara, 2021]
For any function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$,
the ¢-Fenchel-Moreau conjugate is given by

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

$$\inf \big[f \mid n\big](x) = \begin{cases} \inf_{\rho>0} f(\rho x) & \text{if } x \in \mathbb{S}_2^{(0)} \\ +\infty & \text{if } x \notin \mathbb{S}_2^{(0)} \end{cases}$$

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

Proposition

[Chancelier and De Lara, 2021]

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

$$\iff h = h^{¢¢'}$$

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

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

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

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

The ℓ_0 pseudonorm is E-CAPRA-convex

We recall the top-$(2,k)$ norms $\|\cdot\|_{2,k}^{\mathrm{tn}}$

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

$$\begin{aligned}\|y\|_{2,k}^{\mathrm{tn}} &= \sqrt{\sum_{l=1}^{k} |y_{\nu(l)}|^2}\ , \quad |y_{\nu(1)}| \geq |y_{\nu(2)}| \geq \cdots \geq |y_{\nu(d)}| \\ &= \sup_{|K| \leq k} \|y_K\|_2\end{aligned}$$

where $y_K \in \mathbb{R}^d$ is the vector which coincides with y, except for the components outside of $K \subset \{1, \ldots, d\}$ that vanish

The ℓ_0 pseudonorm and the E-Capra-coupling

Theorem

[Chancelier and De Lara, 2021]
The ℓ_0 pseudonorm,
the characteristic functions $\delta_{\ell_0^{\leq k}}$ of its level sets
and the top-$(2,k)$ norm norms $\|\cdot\|_{2,k}^{\mathrm{tn}}$ are related by

$$\delta_{\ell_0^{\leq k}}^{-\cent} = \delta_{\ell_0^{\leq k}}^{\cent} = \|\cdot\|_{2,k}^{\mathrm{tn}} \ , \quad k = 0, 1, \ldots, d$$

$$\ell_0^{\cent} = \sup_{l=0,1,\ldots,d} \left[\|\cdot\|_{2,l}^{\mathrm{tn}} - l \right]$$

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

The ℓ_0 pseudonorm displays hidden convexity

The ℓ_0 pseudonorm displays a convex factorization property

Theorem

[Chancelier and De Lara, 2021]
As the ℓ_0 pseudonorm is E-Capra-convex, we get that

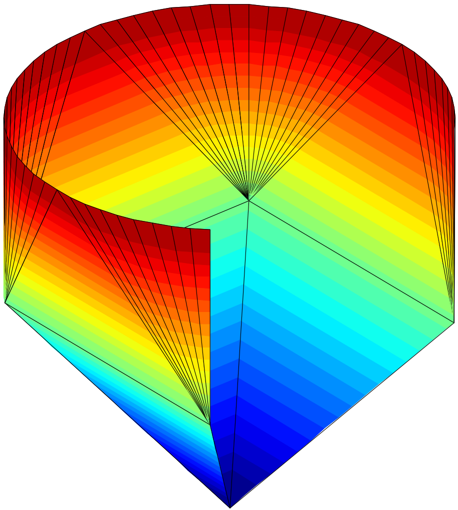
$$\ell_0 = \ell_0^{\cent\cent'} = \ell_0^{\cent\star'} \circ n = \underbrace{\left(\ell_0^{\cent}\right)^{\star'}}_{\text{convex lsc function } \mathcal{L}_0} \circ n$$

that is,

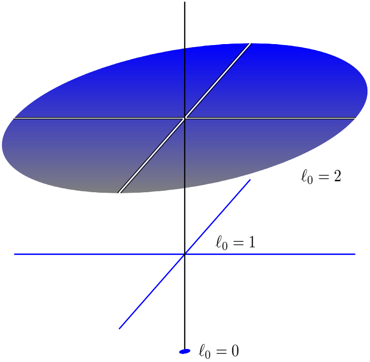
$$\ell_0(x) = \mathcal{L}_0(x) \; , \; \forall x \in \mathbb{S}_2$$

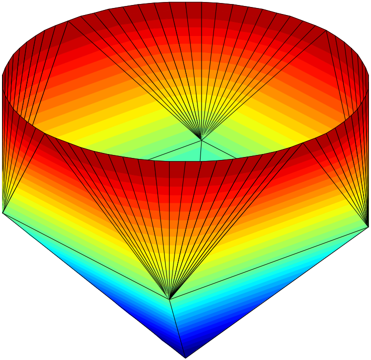
Hidden convexity in the ℓ_0 pseudonorm

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



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





Best convex lower approximation on the unit ball

Best convex lower approximation of the ℓ_0 pseudonorm on the unit ball

Theorem

(work in progress)

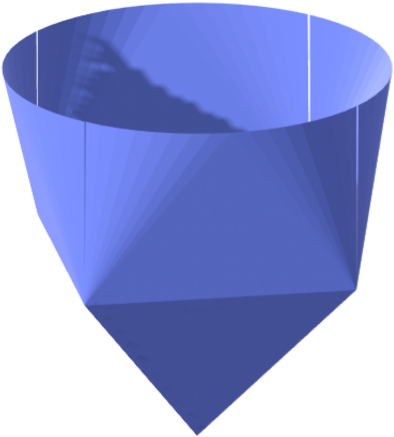
$$\ell_0 = \underbrace{\left(\ell_0^{\cent}\right)^{\star\prime}}_{\text{convex lsc function } \mathcal{L}_0} \circ n$$

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

$$\underbrace{\mathcal{L}_0(x)}_{\text{best convex lsc function}} \leq \ell_0(x) \; , \quad \forall x \in \mathbb{B}_2$$

on the unit ball $\mathbb{B}_2 = \left\{x \in \mathbb{R}^d \;\middle|\; \|x\|_2 \leq 1\right\}$

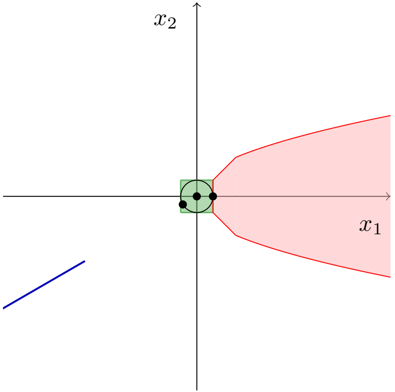
Best convex lower approximation of the ℓ_0 pseudonorm on the Euclidean unit ball



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

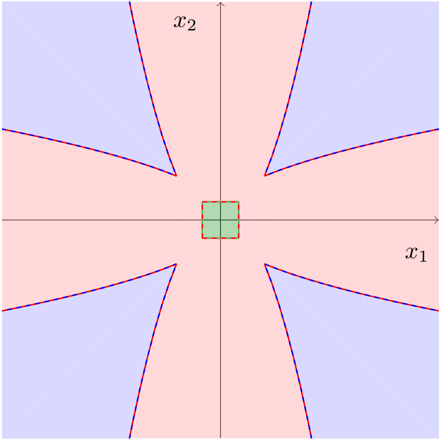
Capra-subdifferential of the ℓ_0 pseudonorm on $\mathbb{R}^2$

Illustration at three points (black dots)



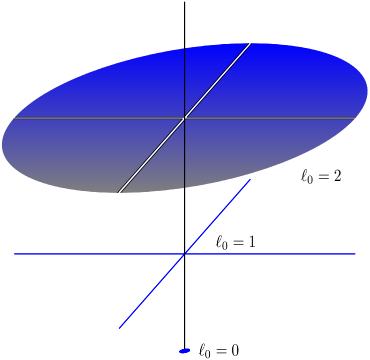
$$\partial_{\cent}\ell_0(0,0)\ ,\ \ \partial_{\cent}\ell_0(1,0)\ ,\ \ \partial_{\cent}\ell_0(-\tfrac{\sqrt{3}}{2},-\tfrac{1}{2})$$

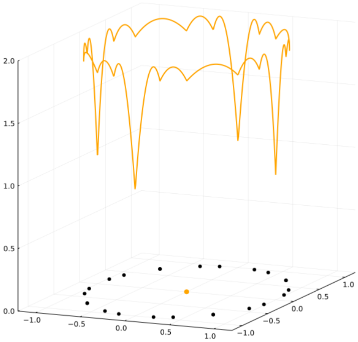
Capra-subdifferential of the ℓ_0 pseudonorm on $\mathbb{R}^2$



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

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





Variational formulas

We recall the $(2,k)$-support norms $\|\cdot\|^{\mathrm{sn}}_{2,k}$

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

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

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

Proposition

[Chancelier and De Lara, 2021]

- The proper convex lsc function $\mathcal{L}_0$ has epigraph

$$\mathrm{epi}\,\mathcal{L}_0 = \overline{\mathrm{co}}\Big(\bigcup_{l=0}^{d} \mathbb{B}^{\star\mathrm{sn}}_{(l)} \times [l, +\infty[\Big)$$

- $\mathcal{L}_0$ is the largest proper convex lsc function below

$$L_0(x) = \begin{cases} 0 & \text{if } x = 0, \\ l & \text{if } x \in \mathbb{B}^{\star\mathrm{sn}}_{(l)} \backslash \mathbb{B}^{\star\mathrm{sn}}_{(l-1)}\ , \quad l = 1, \ldots, d \\ +\infty & \text{if } x \notin \mathbb{B}^{\star\mathrm{sn}}_{(d)} = \mathbb{B} \end{cases}$$

- $\mathcal{L}_0$ has the variational expression

$$\mathcal{L}_0(x) = \min_{\substack{x^{(1)} \in \mathbb{R}^d, \ldots, x^{(d)} \in \mathbb{R}^d \\ \sum_{l=1}^{d} \left\|x^{(l)}\right\|^{\star\mathrm{sn}}_{(l)} \leq 1 \\ \sum_{l=1}^{d} x^{(l)} = x}} \sum_{l=1}^{d} l \left\|x^{(l)}\right\|^{\star\mathrm{sn}}_{(l)}\ , \quad \forall x \in \mathbb{R}^d$$

Variational formulas for the ℓ_0 pseudonorm

Proposition

[Chancelier and De Lara, 2021]

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

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

Difference of convex (DC) formulas
with graded sequences of induced norms

Difference of convex (DC) formulas

Well-known formulas

$$\ell_0(y) = \min \left\{ k \in [\![1, d]\!] \;\middle|\; \|y\|_{2,k}^{\mathrm{tn}} = \|y\|_2 \right\}$$
$$\forall y \in \mathbb{R}^d$$
$$\ell_0(x) = \min \left\{ k \in [\![1, d]\!] \;\middle|\; \|x\|_{2,k}^{\mathrm{sn}} = \|x\|_2 \right\}$$
$$\forall x \in \mathbb{R}^d$$

Lower bound convex programs for exact sparse optimization

Concave dual problem for exact 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

$$\sup_{y \in \mathbb{R}^d} \Big(-\big(\inf\big[f \mid n\big]\big)^{\star}(y) \dotplus \big(- \underbrace{\delta_{\ell_0^{\leq k}}^{-\cent}(y)}_{\|y\|_{2,k}^{\mathrm{tn}}} \big) \Big) \leq \inf_{\ell_0(x) \leq k} f(x)$$

Proposition

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

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

$$= \inf_{\ell_0(x) \leq k} \inf\big[f \mid n\big](x)$$

Convex primal problem for exact sparse optimization

Proposition

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

$$\min_{\|x\|_{2,k}^{\mathrm{sn}} \leq 1} \left(\inf\left[f \mid n\right]\right)^{\star\star'}(x) \leq \inf_{\ell_0(x) \leq k} f(x) = \inf_{\ell_0(x) \leq k} \inf\left[f \mid n\right](x)$$

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

Duality

Perturbation scheme

- Functions $k : \mathbb{R}^d \to \overline{\mathbb{R}}$, $\varphi : \{0, 1, \ldots, d\} \to \overline{\mathbb{R}}$ nondecreasing (ex: identity, $\delta_{\{0,1,\ldots,k\}}$) and original minimization problem

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

  because $\varphi \circ \ell_0 = \big(\varphi \circ \ell_0\big)^{¢¢'} = \big(\varphi \circ \ell_0\big)^{¢\star'} \circ n$
  [Chancelier and De Lara, 2022c]
- Embedding/perturbation scheme $H : \mathbb{R}^d \times \mathbb{R}^d \to \overline{\mathbb{R}}$ given by

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

- Value function

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

Lagrangian and dual problem

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

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

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

- Dual maximization problem

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

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

$$\varphi(0) = \inf_{w \in \mathbb{R}^d} \sup_{y \in \mathbb{R}^d} \mathcal{L}(w, y) = \inf_{w \in \mathbb{R}^d} \Big\{ k(w) \dotplus \varphi\big(\ell_0(w)\big) \Big\}$$

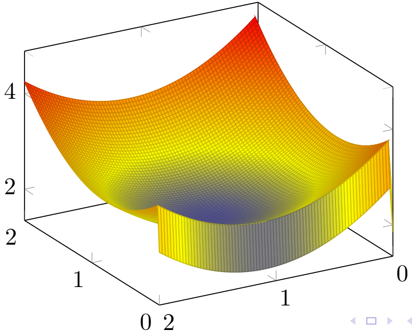
Numerics

A toy example

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

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

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



The toy example as a min-max problem

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

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

with

$$\ell_0^{¢}(y) = \sup_{k=1,\ldots,d} \left[\|y\|_{2,k}^{\mathrm{tn}} - k \right]_+$$

Generalized primal-dual proximal splitting

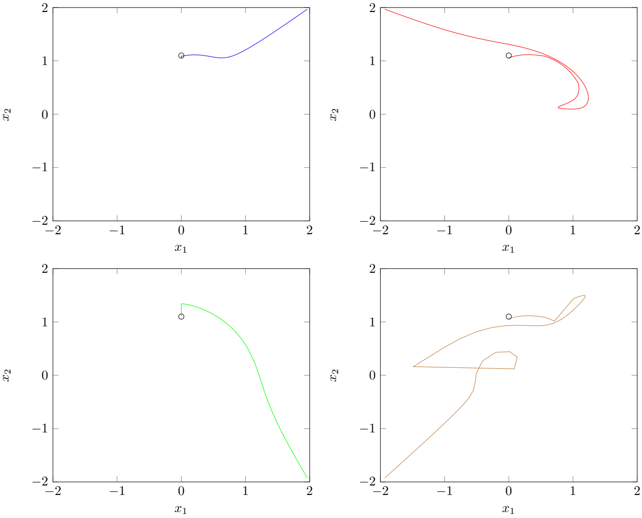
GPDPS Algorithm [Clason, Mazurenko, and Valkonen, 2020]

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

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

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

GPDPS convergence, varying the starting point



Outline of the presentation

Background on one-sided linear couplings

- Background on couplings and Fenchel-Moreau conjugacies
- One-sided linear couplings (and hidden convexity)

The Euclidean CAPRA conjugacy

Extension: constant along primal rays conjugacies

- CAPRA conjugacies
- Best convex approximations of 0-homogeneous functions
- The case of norms

Conclusion

	Norm Euclidean	Norm orthant-strictly monotonic	Norm any
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, 2022b]	
$\varphi \circ \ell_0$ $\varphi : \mathbb{N} \to \overline{\mathbb{R}}$ nondecreasing		¢-convex ($(\varphi \circ \ell_0)^{¢¢'} = \varphi \circ \ell_0$) [Chancelier and De Lara, 2022c] hidden convexity [Chancelier and De Lara, 2022c] variational formula [Chancelier and De Lara, 2022c] subdifferential [Chancelier and De Lara, 2022c]	
$\varphi \circ \ell_0$ $\varphi : \mathbb{N} \to \overline{\mathbb{R}}$ any			$(\varphi \circ \ell_0)^{¢¢'}$ [Chancelier and De Lara, variational inequalit [Chancelier and De Lar subdifferential [Chancelier and De Lara,
$F \circ$ support			
0-homogeneous			

Outline of the presentation

Background on one-sided linear couplings
Background on couplings and Fenchel-Moreau conjugacies
One-sided linear couplings (and hidden convexity)

The Euclidean Capra conjugacy

Extension: constant along primal rays conjugacies
Capra conjugacies
Best convex approximations of 0-homogeneous functions
The case of norms

Conclusion

Homogeneous functions

Definition

We say that a function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ is

- 0-homogeneous if $f(\rho x) = f(x)\ ,\ \ \forall \rho \in \mathbb{R} \setminus \{0\}\ ,\ \ \forall x \in \mathbb{R}^d$
  Example: the *pseudonorm* ℓ_0
- 1-homogeneous if $f(\rho x) = \rho f(x)\ ,\ \ \forall \rho \in \mathbb{R}\ ,\ \ \forall x \in \mathbb{R}^d$
- absolutely 1-homogeneous if
  $f(\rho x) = |\rho| f(x)\ ,\ \ \forall \rho \in \mathbb{R} \setminus \{0\}\ ,\ \ \forall x \in \mathbb{R}^d$
  Examples: norms

$$\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{+\infty\} = [0, +\infty]$$

For any nonnegative 1-homogeneous function $\nu : \mathbb{R}^d \to \overline{\mathbb{R}}_+$, one has that $\nu(0) \in \{0, +\infty\}$

Normalization mapping

Definition

For any nonnegative 1-homogeneous function $\nu : \mathbb{R}^d \to \overline{\mathbb{R}}_+$, the primal normalization mapping $n_\nu : \mathbb{R}^d \to \mathbb{S}_\nu^{(0)}$ is defined by

$$n_\nu : x \in \mathbb{R}^d \mapsto \begin{cases} \frac{x}{\nu(x)} \,, & \text{if } 0 < \nu(x) < +\infty \\ 0 \,, & \text{else} \end{cases}$$

where the unit "sphere" $\mathbb{S}_\nu$ and the pointed unit "sphere" $\mathbb{S}_\nu^{(0)}$ are

$$\mathbb{S}_\nu = \left\{ x \in \mathbb{R}^d \,\middle|\, \nu(x) = 1 \right\} \,, \quad \mathbb{S}_\nu^{(0)} = \mathbb{S}_\nu \cup \{0\}$$

and the unit "ball" $\mathbb{B}_\nu$ is

$$\mathbb{B}_\nu = \left\{ x \in \mathbb{R}^d \,\middle|\, \nu(x) \le 1 \right\}$$

Capra-couplings

Definition

Let $\nu : \mathbb{R}^d \to \overline{\mathbb{R}}_+$ be a nonnegative 1-homogeneous function
The Capra coupling $\cent_\nu : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, between $\mathbb{R}^d$ and itself, associated with ν, is the function

$$\cent_\nu : (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto \langle n_\nu(x), \ y \rangle = \begin{cases} \frac{\langle x, y \rangle}{\nu(x)} \ , & \text{if } 0 < \nu(x) < +\infty \\ 0 \ , & \text{else} \end{cases}$$

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

Special case: $\nu = |||\cdot|||$ (source) norm

The $\cent_\nu$-subdifferential shares properties with the Rockafellar-Moreau subdifferential

Definition

For any function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ and $x \in \mathbb{R}^d$, the $\cent_\nu$-subdifferential is

$$\begin{aligned} \partial_{\cent_\nu} f(x) = \{y \in \mathbb{R}^d \mid \quad & \cent_\nu(x', y) - f(x') \\ \leq \quad & \cent_\nu(x, y) - f(x) \, , \ \forall x' \in \mathbb{R}^d\} \end{aligned}$$

- The $\cent_\nu$-subdifferential $\partial_{\cent_\nu} f(x)$ is a closed convex set
- $y \in \partial_{\cent_\nu} f(x) \iff f^{\cent_\nu}(y) = \cent_\nu(x, y) - f(x)$
- $x \in \arg\min f \iff 0 \in \partial_{\cent_\nu} f(x)$
- $\partial_{\cent_\nu} f + \partial_{\cent_\nu} h \subset \partial_{\cent_\nu} (f \dot{+} h)$
- $x \in \mathrm{dom} f$ and $\partial_{\cent_\nu} f(x) \neq \emptyset \Rightarrow f^{\cent_\nu \cent_\nu{}'}(x) = f(x)$

The $¢_\nu$-conjugacy shares properties with the Fenchel conjugacy

Proposition

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

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

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

As a consequence, the $¢_\nu$-Fenchel-Moreau conjugate $f^{¢_\nu}$ is a closed convex function

The $\cent_\nu$-convex functions are 0-homogeneous and coincide, on the "sphere", with a closed convex function

- The $\cent'_\nu$-Fenchel-Moreau conjugate of $g : \mathbb{R}^d \to \overline{\mathbb{R}}$ is given by
$$g^{\cent'_\nu} = g^\star \circ n_\nu$$
- The $\cent_\nu$-convex functions are $\left\{ g^{\cent'_\nu} \,\middle|\, g : \mathbb{R}^d \to \overline{\mathbb{R}} \right\}$, hence
$$g^{\cent'_\nu}(x) = g^\star\big(n_\nu(x)\big)$$
and therefore $\cent_\nu$-convex functions are 0-homogeneous

Proposition

Any $\cent_\nu$-convex function coincides, on the unit "sphere" $\mathbb{S}_\nu$, with a closed convex function defined on $\mathbb{R}^d$

$$\cent_\nu\text{-convex function} = \text{closed convex function} \circ n_\nu$$

Outline of the presentation

Background on one-sided linear couplings
- Background on couplings and Fenchel-Moreau conjugacies
- One-sided linear couplings (and hidden convexity)

The Euclidean CAPRA conjugacy

Extension: constant along primal rays conjugacies
- CAPRA conjugacies
- Best convex approximations of 0-homogeneous functions
- The case of norms

Conclusion

Fenchel conjugates for 0-homogeneous functions

For any 0-homogeneous function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$,

$$f^\star = \delta_{\{0\}} - \inf_{x \in \mathbb{R}^d} f(x)$$

$$f^{\star\star'} = \inf_{x \in \mathbb{R}^d} f(x)$$

Best convex lower approximations of 0-homogeneous functions (thanks to Thomas Bittar)

Proposition

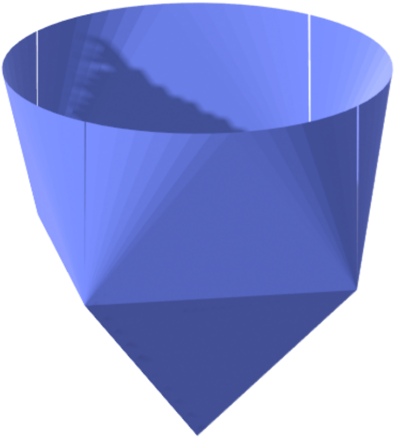
Let $\nu : \mathbb{R}^d \to \mathbb{R}_+$ be a normalization function, with unit "ball" $\mathbb{B}_\nu$ and let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be a 0-homogeneous function

- The function $f^{\cent_\nu \star'}$ is the tightest closed convex function below f on the unit "ball" $\mathbb{B}_\nu$, where

$$f^{\cent_\nu} = \left(f \dotplus \delta_{\mathbb{B}_\nu}\right)^\star = \left(f \dotplus \delta_{\mathbb{S}_\nu^{(0)}}\right)^\star$$

- If $f(0) = 0$, the function $\sigma_{\partial_{\cent_\nu} f(0)}$ is the tightest closed convex positively 1-homogeneous function below f on the unit "ball" $\mathbb{B}_\nu$

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



Best convex and norm lower approximations of the ℓ_0 pseudonorm on the ℓ_p unit "balls"

1-homogeneous function ν	Best convex lower approximation of the ℓ_0 pseudonorm	Best norm lower approximation of the ℓ_0 pseudonorm
$\|\cdot\|_p$ $0 < p \leq 1$	$\|\cdot\|_1 + \delta_{\mathbb{B}_1}$	ℓ_1-norm $\|\cdot\|_1$
$\|\cdot\|_p$ $1 < p < \infty$	not a norm	ℓ_1-norm $\|\cdot\|_1$
$\|\cdot\|_\infty$	$\|\cdot\|_1 + \delta_{\mathbb{B}_\infty}$	ℓ_1-norm $\|\cdot\|_1$

Outline of the presentation

Background on one-sided linear couplings
Background on couplings and Fenchel-Moreau conjugacies
One-sided linear couplings (and hidden convexity)

The Euclidean Capra conjugacy

Extension: constant along primal rays conjugacies
Capra conjugacies
Best convex approximations of 0-homogeneous functions
The case of norms

Conclusion

Generalized coordinate, top and support norms

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 ℓ_0 pseudonorm 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 $|||\cdot|||$ on $\mathbb{R}^d$, we define

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

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

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

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

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

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

source norm $|||\cdot|||$	$|||\cdot|||^{\mathcal{R}}_{(k)}$	$|||\cdot|||^{\mathcal{R}}_{(k),\star}$
$\|\cdot\|_p$	(p, k)-support norm $\|x\|^{\mathrm{sn}}_{p,k}$	top (k, q)-norm $\|y\|^{\mathrm{tn}}_{k,q}$ $= \left(\sum_{j=1}^{k} |y_{\mu(j)}|^q\right)^{1/q}$, $1/p + 1/q = 1$
$\|\cdot\|_1$	$(1, k)$-support norm ℓ_1-norm $\|x\|^{\mathrm{sn}}_{1,k} = \|x\|_1$	top (k, ∞)-norm ℓ_∞-norm $\|y\|^{\mathrm{tn}}_{k,\infty} = |y_{\mu(1)}| = \|y\|_\infty$
$\|\cdot\|_2$	$(2, k)$-support norm	top $(k, 2)$-norm $\|y\|^{\mathrm{tn}}_{k,2} = \sqrt{\sum_{j=1}^{k} |y_{\mu(j)}|^2}$
$\|\cdot\|_\infty$	(∞, k)-support norm	top $(k, 1)$-norm $\|y\|^{\mathrm{tn}}_{k,1} = \sum_{j=1}^{k} |y_{\mu(j)}|$

Concave dual problem for exact sparse optimization

Proposition

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

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

The dual problem is the maximization of a concave usc function

Convex primal problem for exact sparse optimization

Proposition

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

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

The primal problem is the minimization of a closed convex function on the unit ball of the coordinate-k norm $|||\cdot|||^{\mathcal{R}}_{(k)}$

Fenchel *versus* CAPRA conjugacies for ℓ_0

[Chancelier and De Lara, 2022a], [Chancelier and De Lara, 2022c]

Fenchel conjugacy	CAPRA conjugacy
$\delta_{\ell_0^{\leq k}}^{(-\star)} = \delta_{\{0\}},\ k \neq 0$	$\delta_{\ell_0^{\leq k}}^{-\cent_{\vert\vert\vert\cdot\vert\vert\vert}} = \vert\vert\vert\cdot\vert\vert\vert_{(k),\star}^{\mathcal{R}}$
$\ell_0^{\star} = \delta_{\{0\}}$	$\ell_0^{\cent_{\vert\vert\vert\cdot\vert\vert\vert}} = \sup_{l=0,1,\ldots,d} \left[\vert\vert\vert\cdot\vert\vert\vert_{(l),\star}^{\mathcal{R}} - l \right]$
$\delta_{\ell_0^{\leq k}}^{\star\star'} = 0$	$\delta_{\ell_0^{\leq k}}^{\cent_{\vert\vert\vert\cdot\vert\vert\vert}\cent_{\vert\vert\vert\cdot\vert\vert\vert}'} \leq \delta_{\ell_0^{\leq k}}$
$\ell_0^{\star\star'} = 0$	$\ell_0^{\cent_{\vert\vert\vert\cdot\vert\vert\vert}\cent_{\vert\vert\vert\cdot\vert\vert\vert}'} \leq \ell_0$

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

Definition

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

Support and top norms induced by the ℓ_p-norms $\|\cdot\|_p$

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

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

source norm $|||\cdot|||$	$|||x|||^{\star\mathrm{sn}}_{\star,(k)}$	$|||y|||^{\mathrm{tn}}_{\star,(k)}$
$\|\cdot\|_p$	(p,k)-support norm $\|x\|^{\mathrm{sn}}_{p,k}$	top (k,q)-norm $\|y\|^{\mathrm{tn}}_{k,q}$ $= \left(\sum_{l=1}^{k} |y_{\mu(l)}|^q\right)^{1/q}$, $1/p + 1/q = 1$
$\|\cdot\|_1$	$(1,k)$-support norm ℓ_1-norm $\|x\|^{\mathrm{sn}}_{1,k} = \|x\|_1$	top (k,∞)-norm ℓ_∞-norm $\|y\|^{\mathrm{tn}}_{k,\infty} = |y_{\mu(1)}| = \|y\|_\infty$
$\|\cdot\|_2$	$(2,k)$-support norm	top $(k,2)$-norm $\|y\|^{\mathrm{tn}}_{k,2} = \sqrt{\sum_{l=1}^{k} |y_{\mu(l)}|^2}$
$\|\cdot\|_\infty$	(∞,k)-support norm	top $(k,1)$-norm $\|y\|^{\mathrm{tn}}_{k,1} = \sum_{l=1}^{k} |y_{\mu(l)}|$

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

k-coordinate norm		k-support dual norm
$\lvert\lvert\lvert\cdot\rvert\rvert\rvert^{\mathcal{R}}_{(k)}$	$\leq$	$\lvert\lvert\lvert\cdot\rvert\rvert\rvert^{\star\mathrm{sn}}_{\star,(k)}$

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

Orthant-strictly monotonic norms and CAPRA-convexity

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, 2022b]
  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 $j \in \{1, \ldots, d\}$
  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

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

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

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

We establish $\cent_{|||\cdot|||}$-convexity of the ℓ_0 pseudonorm

Proposition

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

$$\delta_{\ell_0^{\leq k}}^{\cent_{|||\cdot|||}\cent_{|||\cdot|||}'} = \delta_{\ell_0^{\leq k}}$$

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

$$\ell_0^{\cent_{|||\cdot|||}\cent_{|||\cdot|||}'} = \ell_0$$

Capra-subdifferentiability properties of the ℓ_0 pseudonorm

- $\{|||\cdot|||^{\mathcal{R}}_{(j)}\}_{j=1,\dots,d}$ and $\{|||\cdot|||^{\mathcal{R}}_{(j),\star}\}_{j=1,\dots,d}$, associated coordinate-k and dual coordinate-k norms
- $\{\mathbb{B}^{\mathcal{R}}_{(j)}\}_{j=1,\dots,d}$ and $\{\mathbb{B}^{\mathcal{R}}_{(j),\star}\}_{j=1,\dots,d}$, corresponding unit balls

Proposition

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

$$\text{if } x = 0, \quad \partial_{\cent_{|||\cdot|||}} \ell_0(0) = \bigcap_{j=1,\dots,d} j\mathbb{B}^{\mathcal{R}}_{(j),\star}$$

$$\text{if } x \neq 0 \text{ and } \ell_0(x) = l, \quad \partial_{\cent_{|||\cdot|||}} \ell_0(x) = \mathcal{N}_{\mathbb{B}^{\mathcal{R}}_{(l)}}\Big(\frac{x}{|||x|||^{\mathcal{R}}_{(l)}}\Big) \cap Y_l$$

where $\quad Y_l = \Big\{ y \in \mathbb{Y} \,\Big|\, l \in \arg\max_{j=0,\dots,d} \big(|||y|||^{\mathcal{R}}_{(j),\star} - j\big) \Big\}, \ \forall l = 0, \dots, d$

Capra-subdifferentiability properties of the ℓ_0 pseudonorm

Proposition

[Chancelier and De Lara, 2022c]
If both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic, we have that

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

that is, the pseudonorm ℓ_0 is $\textsc{Capra}$-subdifferentiable on $\mathbb{R}^d$

Fenchel *versus* Capra conjugacies for ℓ_0

[Chancelier and De Lara, 2022a], [Chancelier and De Lara, 2022c]
If the source norm is orthant-strictly monotonic, we have that

Fenchel conjugacy	Capra conjugacy
$\delta_{\ell_0^{\leq k}}^{(-\star)} = \delta_{\{0\}}, \; k \neq 0$	$\delta_{\ell_0^{\leq k}}^{-\cent_{\vert\vert\vert\cdot\vert\vert\vert}} = \vert\vert\vert\cdot\vert\vert\vert_{(k),\star}^{\mathcal{R}} = \vert\vert\vert\cdot\vert\vert\vert_{\star,(k)}^{\mathrm{tn}}$
$\ell_0^{\star} = \delta_{\{0\}}$	$\ell_0^{\cent_{\vert\vert\vert\cdot\vert\vert\vert}} = \sup_{l=0,1,\ldots,d} \big[\vert\vert\vert\cdot\vert\vert\vert_{(l),\star}^{\mathcal{R}} - l \big] = \sup_{l=0,1,\ldots,d} \big[\vert\vert\vert\cdot\vert\vert\vert_{\star,(l)}^{\star\mathrm{sn}} - l \big]$
$\delta_{\ell_0^{\leq k}}^{\star\star'} = 0$	$\delta_{\ell_0^{\leq k}}^{\cent_{\vert\vert\vert\cdot\vert\vert\vert}\cent_{\vert\vert\vert\cdot\vert\vert\vert}'} = \delta_{\ell_0^{\leq k}}$
$\ell_0^{\star\star'} = 0$	$\ell_0^{\cent_{\vert\vert\vert\cdot\vert\vert\vert}\cent_{\vert\vert\vert\cdot\vert\vert\vert}'} = \ell_0$

Outline of the presentation

Background on one-sided linear couplings

The Euclidean CAPRA conjugacy

Extension: constant along primal rays conjugacies

Conclusion

Conclusion (1/2)

- Sparsity is, by nature, indifferent to magnitude, which is reflected in the support mapping being 0-homogeneous
- But the Fenchel conjugacy is not a suitable tool to analyze 0-homogeneous functions

Conclusion (2/2)

We have proposed the Capra coupling $¢(x, y) = \frac{\langle x, y \rangle}{\nu(x)}$
and, with the Capra-conjugacy, we have obtained

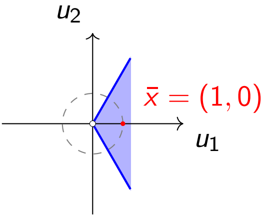
- Capra-convexity
  (by displaying nonempty Capra-subdifferential)
- hidden convexity
- best convex lower approximation on the unit ball
- E-Capra-subdifferential (thanks to Adrien Le Franc)
- variational formulas
- difference of convex (DC) formulas
  with graded sequences of induced norms
- concave dual problems in sparse optimization
- duality

Perspectives

- Tackle open theoretical questions
  - duality gap between lower bound convex program and original sparse optimization problem
  - Conditions for $\partial_{\cent_\nu} f + \partial_{\cent_\nu} h \supset \partial_{\cent_\nu}(f \dotplus h)$ (with ex-PhD student Adrien Le Franc)
- Matrix functions and norms
  - Rank-based norms and suitable matrix norms for $\textsc{Capra}$-conjugacy of the rank function (with ENPC students Paul Barbier and Valentin Paravy)
  - formula “à la Lewis” $(F \circ \sigma)^{\cent_\nu} = F^{\cent_\nu} \circ \sigma$ for $\textsc{Capra}$-conjugacy
- Algorithms with $\textsc{Capra}$-couplings (with ex-PhD student Adrien Le Franc)
  - Mirror descent, Bregman divergence
  - $\textsc{Capra}$-convex sparse optimization problems

An example where the subdifferential of the sum. . .

$$|||\cdot||| = \ell_2$$

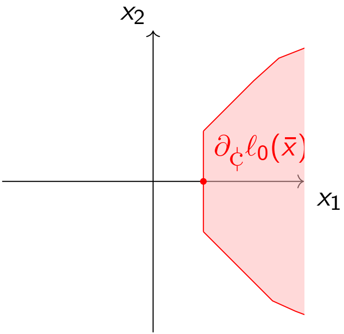


$$\bar{x} \in \operatorname*{arg\,min}_{K} \ell_0 \implies 0 \in \partial_{\text{ç}}(\ell_0 + \delta_K)(\bar{x})$$

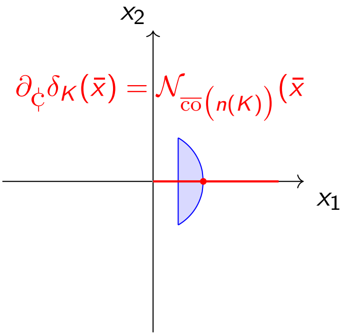
(a property of one-sided linear couplings)

...is not the sum of the subdifferentials (Adrien Le Franc)

Let $y' \in \partial_{\c{C}} \ell_0(\bar{x})$ and $y'' \in \partial_{\c{C}} \delta_K(\bar{x})$



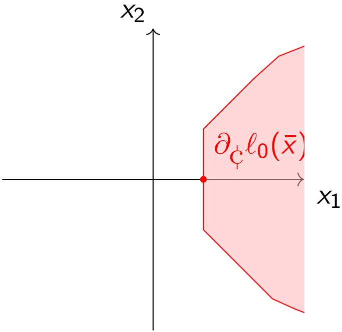
$y'_1 \geq 1$



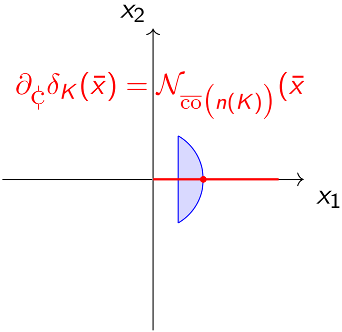
$y''_1 \geq 0$

...is not the sum of the subdifferentials (Adrien Le Franc)

Let $y' \in \partial_{\text{Ç}} \ell_0(\bar{x})$ and $y'' \in \partial_{\text{Ç}} \delta_K(\bar{x})$



$y'_1 \geq 1$

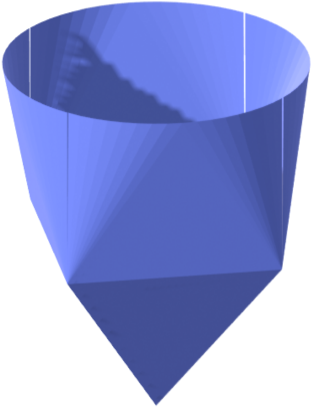


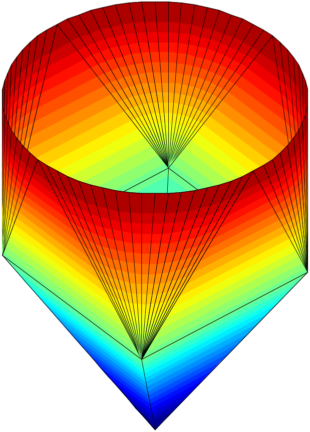
$y''_1 \geq 0$

$0 \notin \partial_{\text{Ç}} \ell_0(\bar{x}) + \partial_{\text{Ç}} \delta_K(\bar{x})$ hence

$$\partial_{\text{Ç}} \ell_0(\bar{x}) + \partial_{\text{Ç}} \delta_K(\bar{x}) \subsetneq \partial_{\text{Ç}} \big(\ell_0 + \delta_K\big)(\bar{x})$$

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. Hidden convexity in the l_0 pseudonorm. *Journal of Convex Analysis*, 28(1):203–236, 2021.

Jean-Philippe Chancelier and Michel De Lara. Constant along primal rays conjugacies and the l_0 pseudonorm. *Optimization*, 71(2):355–386, 2022a. doi: 10.1080/02331934.2020.1822836.

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

Jean-Philippe Chancelier and Michel De Lara. Capra-convexity, convex factorization and variational formulations for the l_0 pseudonorm. *Set-Valued and Variational Analysis*, 30:597–619, 2022c.

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 l0 pseudonorm. Preprint hal-03505168, 2022.

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.

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