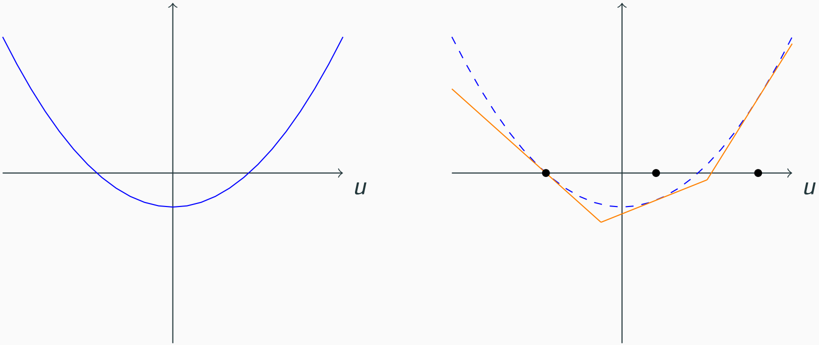
Perspectives of Capra-convexity in sparse optimization

BrainPOP seminar - 14th February 2021

Adrien Le Franc

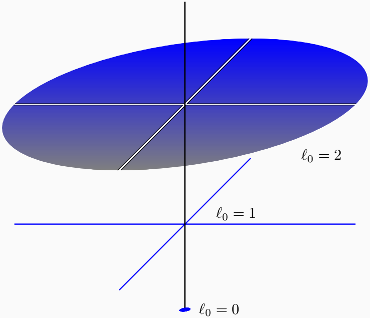
Postdoctoral researcher at LAAS - CNRS

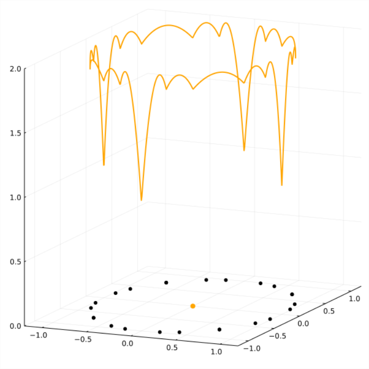
Fenchel conjugate, subdifferential, and polyhedral approximate of a convex lower semicontinuous function



$$f(u) \geq \max_{\substack{v_i \in \partial f(u_i) \\ i \in I}} \Big(\langle u \,, v_i \rangle + \big(-f^\star(v_i) \big) \Big)$$

Beyond convex lower semicontinuous functions...





$$\ell_0(u) \geq \max_{\substack{v_i \in \partial_{¢} \ell_0(u_i) \\ i \in I}} \left(¢(u, v_i) + \left(-\ell_0^{¢}(v_i)\right)\right)$$

Outline of the presentation

1. **Background notions on generalized convexity**
2. **The Capra coupling and the ℓ_0 pseudonorm**
3. **Perspectives in sparse optimization**
4. **Conclusion**

Outline of the section

1. **Background notions on generalized convexity**
2. The Capra coupling and the ℓ_0 pseudonorm
3. Perspectives in sparse optimization
4. Conclusion

Couplings and generalized Fenchel-Moreau conjugacies

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\} \quad \text{and} \quad (+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = -\infty$$

Definition

Two sets $\mathbb{U}$ (“Primal”) and $\mathbb{V}$ (“Dual”) paired by
a **coupling function** $c : \mathbb{U} \times \mathbb{V} \to \overline{\mathbb{R}}$
give rise to the ***c*-Fenchel-Moreau conjugacy**

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

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

Example: two vector spaces $\mathbb{U}$ and $\mathbb{V}$ paired with a bilinear form $\langle \cdot \, , \cdot \rangle$
give rise to the classic Fenchel conjugacy $f \in \overline{\mathbb{R}}^{\mathbb{U}} \mapsto f^\star \in \overline{\mathbb{R}}^{\mathbb{V}}$

Generalized c-biconjugate and c-convexity

- We also introduce the **c'-Fenchel-Moreau conjugacy**

$$g \in \overline{\mathbb{R}}^{\mathbb{V}} \mapsto g^{c'} \in \overline{\mathbb{R}}^{\mathbb{U}} \; , \;\; g^{c'}(u) = \sup_{v \in \mathbb{V}} \Big(\underbrace{c(u,v)}_{=c'(v,u)} \dot{+} \big(-g(v)\big) \Big) \; , \;\; \forall u \in \mathbb{U}$$

- This gives rise to the **c-Fenchel-Moreau biconjugate**

$$f \in \overline{\mathbb{R}}^{\mathbb{U}} \mapsto f^{cc'} \in \overline{\mathbb{R}}^{\mathbb{U}} \; , \;\; f^{cc'}(u) = \big(f^{c}\big)^{c'}(u) \; , \;\; \forall u \in \mathbb{U}$$

Definition

A function $f \in \overline{\mathbb{R}}^{\mathbb{U}}$ is **c-convex** if $f = f^{cc'}$, that is

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

Example: a proper function $f \in \overline{\mathbb{R}}^{\mathbb{U}}$ is $\langle \, , \rangle$-convex iff f is convex and lsc

Generalized c-subdifferential

Definition

The ***c*-subdifferential** of a function $f \in \overline{\mathbb{R}}^{\mathbb{U}}$ at $u \in \mathbb{U}$ with respect to the coupling c is the subset $\partial_c f(u) \subseteq \mathbb{V}$ defined equivalently either by

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

or by

$$v \in \partial_c f(u) \iff c(u,v) \dot{+} \big(-f(u)\big) \geq c(u',v) \dot{+} \big(-f(u')\big) \;, \;\; \forall u' \in \mathbb{U}$$

Outline of the section

1. **Background notions on generalized convexity**
2. **The Capra coupling and the ℓ_0 pseudonorm**
   - Capra-convexity of ℓ_0
   - Capra-subdifferential of the ℓ_0 pseudonorm
3. **Perspectives in sparse optimization**
4. **Conclusion**

Outline of the section

1. **Background notions on generalized convexity**

2. **The Capra coupling and the ℓ_0 pseudonorm**

   Capra-convexity of ℓ_0

   Capra-subdifferential of the ℓ_0 pseudonorm

3. **Perspectives in sparse optimization**

4. **Conclusion**

The Capra coupling

Definition (Chancelier and De Lara [2020])

Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$ called the **source norm**

we define the **Capra** *coupling* $\mathbb{R}^d \overset{\cent}{\longleftrightarrow} \mathbb{R}^d$ by

$$\forall v \in \mathbb{R}^d \ , \ \ \cent(u,v) = \begin{cases} \frac{\langle u \,, v \rangle}{|||u|||} & \text{if } u \neq 0 \\ 0 & \text{if } u = 0 \end{cases}$$

The coupling Capra is **Constant Along Primal RAys (Capra)**

The ℓ_0 pseudonorm

$$\ell_0(u) = \left|\left\{j \in \{1,\dots,d\} \,\middle|\, u_j \neq 0\right\}\right| \; , \;\; \forall u \in \mathbb{R}^d$$

Proposition (Chancelier and De Lara [2021])

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

$$\ell_0 = \ell_0^{¢¢'}$$

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

that is, the pseudonorm ℓ_0 is **Capra-convex** *and* **Capra-subdifferentiable** *everywhere on $\mathbb{R}^d$*

Examples: $\begin{cases} ||(0,1)||_\infty = ||(1,1)||_\infty = 1 \text{ hence } \ell_\infty \text{ is not OSM} \\ \ell_2 \text{ is OSM} \end{cases}$

Outline of the section

Explicit formulations for the Capra-subdifferential of ℓ_0

Proposition (Le Franc et al. [2021])

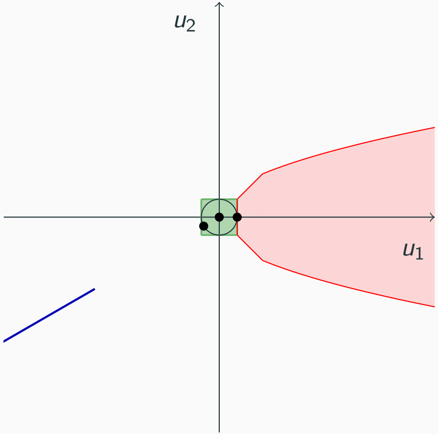
For the source norms $|||\cdot||| = \ell_p$, $p \in]1, \infty[$, *we have that*

$$\partial_{¢}\ell_0(0) = \mathbb{B}_{\ell_\infty}$$

and if $u \neq 0$, $l = \ell_0(u)$, $L = supp(u)$, *for* $v \in \mathbb{R}^d$, $|v_{\nu(1)}| \geq \ldots \geq |v_{\nu(d)}|$, $||v||^{\mathrm{tn}}_{(k,q)} = \left(|v_{\nu(1)}|^q + \ldots + |v^q_{\nu(k)}|\right)^{\frac{1}{q}}$ *and* $\frac{1}{p} + \frac{1}{q} = 1$

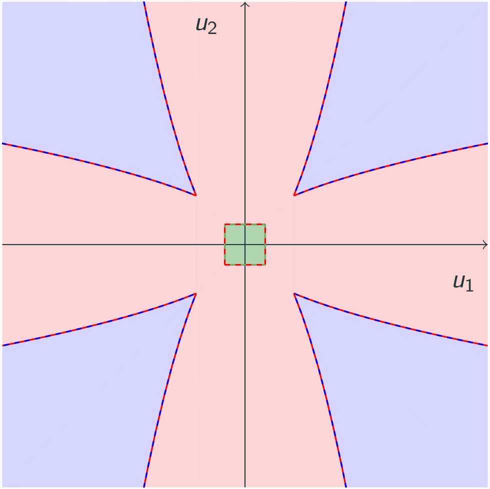
$$v \in \partial_{¢}\ell_0(u) \iff \begin{cases} v_L \in \mathcal{N}_{\mathbb{B}_{||\cdot||_p}}\left(\frac{u}{||u||_p}\right) \\ |v_j| \leq \min_{i\in L} |v_i| \,, \ \forall j \notin L \\ |v_{\nu(k+1)}|^q \geq \left(||v||^{\mathrm{tn}}_{(k,q)} + 1\right)^q - \left(||v||^{\mathrm{tn}}_{(k,q)}\right)^q \,, \ \forall k < l \\ |v_{\nu(l+1)}|^q \leq \left(||v||^{\mathrm{tn}}_{(l,q)} + 1\right)^q - \left(||v||^{\mathrm{tn}}_{(l,q)}\right)^q \end{cases}$$

Examples of sets $\partial_{\c{C}}\ell_0(u)$ in $\mathbb{R}^2$ with the source norm $|||\cdot||| = \ell_2$



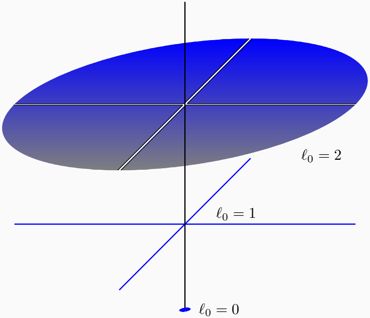
$$\partial_{\c{C}}\ell_0(0,0)\ ,\ \ \partial_{\c{C}}\ell_0(1,0)\ ,\ \ \partial_{\c{C}}\ell_0\left(-\frac{\sqrt{3}}{2},-\frac{1}{2}\right)$$

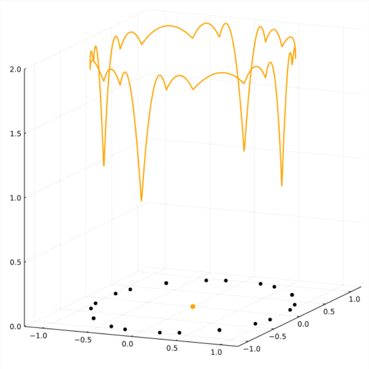
Vizualisation of $\partial_{¢}\ell_0$ in $\mathbb{R}^2$ with the source norm $|||\cdot||| = \ell_2$



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

Polyhedral-like approximation of ℓ_0





$$\ell_0(u) \geq \max_{\substack{v_i \in \partial_{\textcent} \ell_0(u_i) \\ i \in I}} \left(\textcent(u, v_i) + \left(-\ell_0^{\textcent}(v_i)\right)\right)$$

Outline of the section

1. Background notions on generalized convexity
2. The Capra coupling and the ℓ_0 pseudonorm
3. **Perspectives in sparse optimization**
   - Capra-convex sparse optimization problems
   - Numerical methods ?
   - Global optimality conditions ?
4. Conclusion

A key property

we see that if we introduce

$$\mathbb{S}^{(0)} = \mathbb{S} \cup \{0\}$$

and the **normalization mapping**

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

the Capra coupling is **One-sided linear (OSL)**

$$¢(u,v) = \langle n(u) \, , v \rangle \; , \quad \forall (u,v) \in (\mathbb{R}^d)^2$$

Outline of the section

Capra-convex sparse optimization problems

We consider problems of type

$$\min_{u \in U} \ell_0(u)$$

and we look for constraint sets $U \subseteq \mathbb{U}$ for which
we have a **Capra-convex (sparse) optimization problem**

Definition

We say that **the set $U \subseteq \mathbb{U}$ is Capra-convex** if the indicator function **δ_U is a Capra-convex function**

Which sets are Capra-convex ?

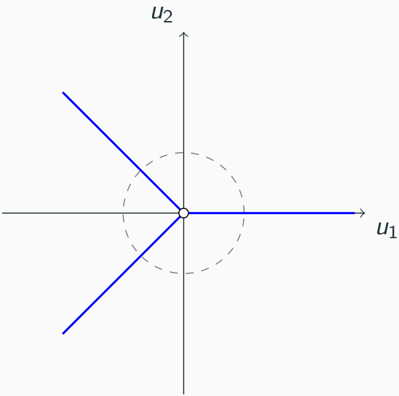
Characterization of Capra-convex sets

Proposition (Le Franc [2021])

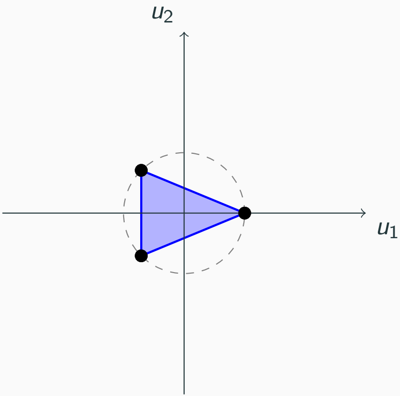
Let the source norm $|||\cdot||| = ||\cdot||_p$, $p \in]1, \infty[$ and $U \subseteq \mathbb{U}$ be a nonempty set

$$U \text{ is Capra-convex} \iff \begin{cases} U \text{ is a cone} \\ U \cup \{0\} \text{ is closed} \\ U \cap \{0\} = \overline{\mathrm{co}}\big(n(U)\big) \cap \{0\} \end{cases}$$

Example with $\lvert\lvert\lvert\cdot\rvert\rvert\rvert = \ell_2$: a non Capra-convex cone

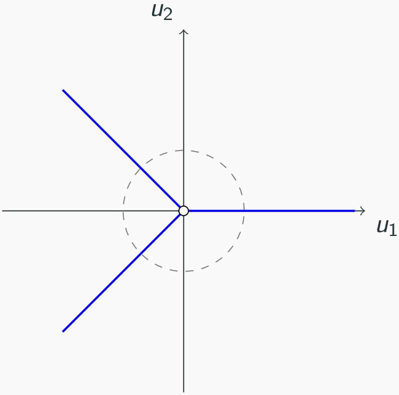


K

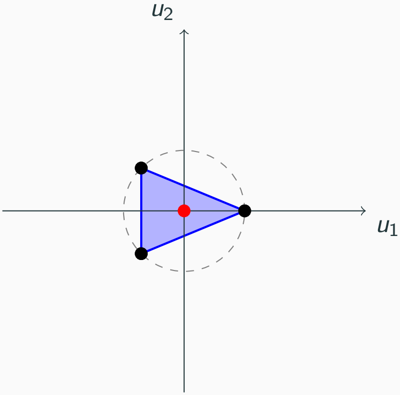


$\overline{\mathrm{co}}(n(K))$

Example with $|||\cdot||| = \ell_2$: a non Capra-convex cone



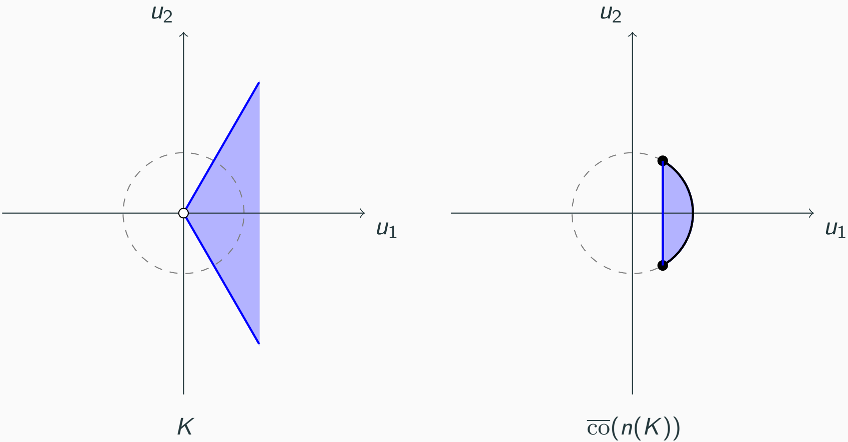
K



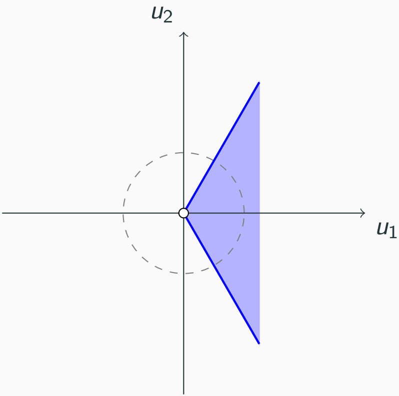
$\overline{\mathrm{co}}(n(K))$

$$K \cap \{0\} \neq \overline{\mathrm{co}}\big(n(K)\big) \cap \{0\}$$

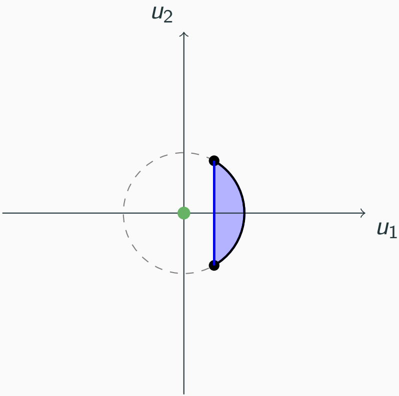
Example with $|||\cdot||| = \ell_2$: a Capra-convex cone



Example with $\|\!|\cdot|\!\| = \ell_2$: a Capra-convex cone



K



$\overline{\mathrm{co}}(n(K))$

$$K \cap \{0\} = \overline{\mathrm{co}}\big(n(K)\big) \cap \{0\}$$

Outline of the section

1. **Background notions on generalized convexity**
2. **The Capra coupling and the ℓ_0 pseudonorm**
3. **Perspectives in sparse optimization**

   Capra-convex sparse optimization problems

   Numerical methods ?

   Global optimality conditions ?

4. **Conclusion**

The standard Bregman divergence

Definition

Let $\mathbb{U}$ and $\mathbb{V}$ be two vector spaces paired by a bilinear form $\langle \cdot \,, \cdot \rangle$ let $\kappa \in \overline{\mathbb{R}}^{\mathbb{U}}$ be a proper, closed, convex and differentiable (divergence generating) function.
We define the **Bregman divergence associated with κ** by

$$D_{\kappa}(u, u') = \kappa(u) - \kappa(u') - \langle u - u' \,, \nabla\kappa(u') \rangle \,, \quad \forall (u, u') \in \mathbb{U} \times \mathrm{dom}(\nabla\kappa)$$

D_{κ} is not a distance, but if κ is **strongly convex**

- $D_{\kappa}(u, u') \geq 0 \,, \ \forall (u, u') \in \mathbb{U} \times \mathrm{dom}(\nabla\kappa)$
- $D_{\kappa}(u, u') = 0 \iff u = u'$
- We have a "triangular inequality" that makes mirror descent work

The Bregman divergence with couplings

Definition

Let $\mathbb{U}$ and $\mathbb{V}$ be two sets and a **finite coupling** $\mathbb{U} \overset{c}{\longleftrightarrow} \mathbb{V}$
let $\kappa \in \overline{\mathbb{R}}^{\mathbb{U}}$ be a proper c-convex (divergence generating) function.
We define the **c-Bregman divergence associated with κ** by

$$D^c_\kappa(u, u', v') = \kappa(u) - \kappa(u') - c(u, v') + c(u', v') \; , \quad \forall (u, u') \in \mathbb{U} \times \mathrm{dom}(\partial_c \kappa) \; , \; \forall v' \in \partial_c \kappa(u')$$

If moreover

- $\mathbb{V}$ is a vector space
- The coupling c is **OSL**
- The function κ is **c-strongly convex**

We retrieve some properties of the original Bregman divergence

The mirror descent algorithm with OSL couplings

We consider the optimization problem

$$\min_{u \in U} f(u)$$

- We initialize three sequences by

$$u_0 \in U$$
$$v_0 \in \partial_c \big(\kappa + \delta_U\big)(u_0)$$
$$v_0^f \in \partial_c f(u_0)$$

- We run $N \in \mathbb{N}$ steps with a step size $\alpha_n > 0$ and the update rules

$$u_{n+1} \in \operatorname*{arg\,min}_{u \in U} \Big(\kappa(u) + c(u, \alpha_n v_n^f - v_n)\Big)$$
$$v_{n+1} = v_n - \alpha_n v_n^f$$
$$v_{n+1}^f \in \partial_c f(u_{n+1})$$

Convergence result

Theorem (Le Franc [2021])

Under a suitable choice of divergence generating function κ we can bound the optimality gap by

$$\min_{0 \leq n \leq N-1} \Big(f(u_n) - \inf_{u \in U} f(u) \Big) \leq \frac{R^2 + \frac{G^2}{4} \sum_{n=0}^{N-1} \alpha_n^2}{\sum_{n=0}^{N-1} \alpha_n}$$

- R and G are constant values determined by the problem and by κ
- We retrieve the same convergence rule as in the standard mirror descent algorithm

Numerical perspectives

Solving $\min_{u \in K} \ell_0(u)$?

- We need to identify suitable **divergence generating functions** κ such that $\kappa + \delta_K$ **is Capra-strongly convex**
- We need to make sure that we can **compute efficiently**

$$u_{n+1} \in \operatorname*{arg\,min}_{u \in K} \big(\kappa(u) + ¢(u, \alpha_n v_n^f - v_n) \big)$$

Not that simple...

Outline of the section

1. **Background notions on generalized convexity**

2. **The Capra coupling and the ℓ_0 pseudonorm**

3. **Perspectives in sparse optimization**

   Capra-convex sparse optimization problems

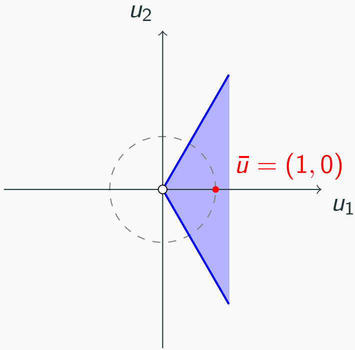
   Numerical methods ?

   Global optimality conditions ?

4. **Conclusion**

An example where the subdifferential of the sum...

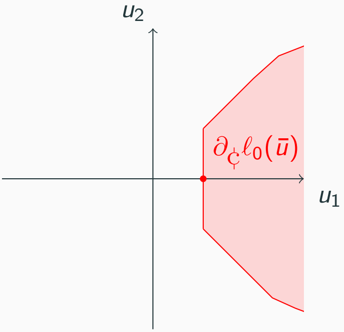
$$|||\cdot||| = \ell_2$$



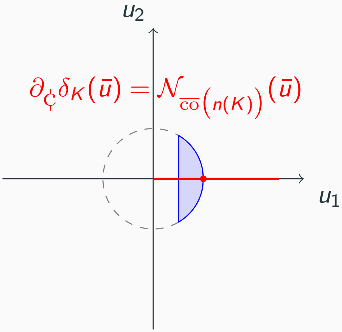
$\bar{u} \in \arg\min_K \ell_0 \implies 0 \in \partial_{¢}(\ell_0 + \delta_K)(\bar{u})$ (a property of OSL couplings)

...is not the sum of the subdifferentials

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



$v_1' \geq 1$



$v_1'' \geq 0$

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

Outline of the section

1. Background notions on generalized convexity
2. The Capra coupling and the ℓ_0 pseudonorm
3. Perspectives in sparse optimization
4. **Conclusion**

Conclusion

Background notions:

- Standard tools of convex analysis (Fenchel conjugacies, subdifferentials...) extend well to general couplings
- The Capra coupling is OSL and well suited to handle ℓ_0

Novelties:

- We have explicited **the Capra-subdifferential of ℓ_0**
- We have identified **Capra-convex sets**
- We have introduced an extention of **the mirror descent algorithm with OSL couplings**

Perspectives: still a long way for numerical applications but they could raise from the mirror descent algorithm

Miscellaneous...

Figure 7: `https://en.wikipedia.org/wiki/Capra_(genus)`

References I

Jean-Philippe Chancelier and Michel De Lara. Constant along primal rays conjugacies and the l0 pseudonorm. *Optimization*, 0(0):1–32, 2020. URL https://doi.org/10.1080/02331934.2020.1822836.

Jean-Philippe Chancelier and Michel De Lara. Capra-convexity, convex factorization and variational formulations for the ℓ_0 pseudonorm. *Set-Valued and Variational Analysis*, pages 1–23, 2021.

Adrien Le Franc. Subdifferentiability in convex and stochastic optimization applied to renewable power systems. 2021.

Adrien Le Franc, Jean-Philippe Chancelier, and Michel De Lara. The Capra-subdifferential of the l0 pseudonorm. 2021. URL https://hal.archives-ouvertes.fr/hal-03505168.