



HAL
open science

The Capra-subdifferential of the l_0 pseudonorm

Adrien Le Franc, Jean-Philippe Chancelier, Michel de Lara

► **To cite this version:**

Adrien Le Franc, Jean-Philippe Chancelier, Michel de Lara. The Capra-subdifferential of the l_0 pseudonorm. Optimization, 2022, pp.1-23. 10.1080/02331934.2022.2145172 . hal-03505168v3

HAL Id: hal-03505168

<https://hal.science/hal-03505168v3>

Submitted on 25 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Capra-subdifferential of the ℓ_0 pseudonorm

Adrien Le Franc* Jean-Philippe Chancelier* Michel De Lara*

January 25, 2023

Abstract

The ℓ_0 pseudonorm counts the nonzero coordinates of a vector. It is often used in optimization problems to enforce the sparsity of the solution. However, this function is nonconvex and noncontinuous, and optimization problems formulated with ℓ_0 — be it in the objective function or in the constraints — are hard to solve in general. Recently, a new family of coupling functions — called Capra (constant along primal rays) — has proved to induce relevant generalized Fenchel-Moreau conjugacies to handle the ℓ_0 pseudonorm. In particular, under a suitable choice of source norm on \mathbb{R}^d — used in the definition of the Capra coupling — the function ℓ_0 is Capra-subdifferentiable, hence is Capra-convex. In this article, we give explicit formulations for the Capra-subdifferential of the ℓ_0 pseudonorm, when the source norm is a ℓ_p norm with $p \in [1, \infty]$. We illustrate our results with graphical visualizations of the Capra-subdifferential of ℓ_0 for the Euclidean source norm.

Keywords Generalized subdifferential; ℓ_0 pseudonorm; Sparsity; Capra-coupling

1 Introduction

The ℓ_0 pseudonorm is a function which counts the number of nonzero elements of a vector. This function appears in numerous optimization problems to enforce the sparsity of the solution. As this function is nonconvex and noncontinuous, the powerful framework of convex analysis is unadapted to address such problems, unless considering a convex relaxation of the function ℓ_0 . In a recent series of works [1, 3, 2], it was shown that conjugacies induced by the so-called Capra (constant along primal rays) coupling are well-suited to handle the ℓ_0 pseudonorm. In particular, the authors show in [2] that — for a large class of source norms (that encompasses the ℓ_p norms for $p \in]1, \infty[$) employed in the definition of the Capra coupling — the ℓ_0 pseudonorm is equal to its Capra-biconjugate, meaning that it is a Capra-convex function. They also provide formulae for the Capra-subdifferential of ℓ_0 in [3], and prove that this subdifferential is nonempty for the same class of source norms that guarantee the Capra-convexity of ℓ_0 in [2].

*CERMICS, École des Ponts ParisTech, France.

The formulation of the Capra-subdifferential of the ℓ_0 pseudonorm in [3] involves the so-called coordinate- k and dual coordinate- k norms, defined by variational expressions, and is not readily computable. The main contribution of this article is to derive explicit formulations to compute the Capra-subdifferential of the ℓ_0 pseudonorm for all ℓ_p source norms with $p \in [1, \infty]$. Subsequently, we comment on the domain of these subdifferentials, and extend previous results by showing that, in the extreme cases where $p \in \{1, \infty\}$, the ℓ_0 pseudonorm is not Capra-convex. We also provide graphical illustrations of the Capra-subdifferential of ℓ_0 , and compare it with other notions of generalized subdifferentials for ℓ_0 found in [5]. With the Capra-subdifferential, we can naturally derive “polyhedral-like” [9, p. 114] lower bounds for the ℓ_0 pseudonorm, that is, lower bounds that are the maximum of a finite number of so-called *Capra-affine* functions.

The paper is organized as follows. First, we recall background notions on Capra-couplings in §2. Second, we derive explicit formulations for the Capra-subdifferential of ℓ_0 in §3. Finally, we provide illustrative visualizations and discuss the positioning of the Capra-subdifferential of ℓ_0 with respect to other notions of subdifferentials in §4.

2 Background on the Capra coupling and the ℓ_0 pseudonorm

For any pair of integers $i \leq j$, we denote $\llbracket i, j \rrbracket = \{i, i+1, \dots, j-1, j\}$. We work on the Euclidean space \mathbb{R}^d , where $d \in \mathbb{N}^*$, equipped with the canonical scalar product $\langle \cdot, \cdot \rangle$, and with a norm $\|\cdot\|$ that we call the *source norm*. We stress the point that $\|\cdot\|$ can be *any* norm, and is not required to be the Euclidean norm. We denote the unit sphere and the unit ball of the norm $\|\cdot\|$ by, respectively,

$$\mathbb{S} = \{x \in \mathbb{R}^d \mid \|x\| = 1\} \quad \text{and} \quad \mathbb{B} = \{x \in \mathbb{R}^d \mid \|x\| \leq 1\}, \quad (1)$$

or, more explicitly, by $\mathbb{S}_{\|\cdot\|}$ and $\mathbb{B}_{\|\cdot\|}$ when needed.

First, we recall the definition of the so-called Capra coupling and of the resulting Capra conjugacy in §2.1. Second, we review the main results relating Capra conjugacies and the ℓ_0 pseudonorm in §2.2.

2.1 Capra conjugacies

We start by recalling the definition of the Capra coupling.

Definition 2.1 ([3], Definition 4.1) *Let $\|\cdot\|$ be a norm on \mathbb{R}^d . We define the coupling $\zeta : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ between \mathbb{R}^d and \mathbb{R}^d , that we call the Capra coupling, by*

$$\forall y \in \mathbb{R}^d, \quad \zeta(x, y) = \begin{cases} \frac{\langle x, y \rangle}{\|x\|}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases} \quad (2)$$

A coupling function such as the Capra coupling \mathfrak{c} given in Definition 2.1 gives rise to generalized Fenchel-Moreau conjugacies [9, 6], that we briefly recall. Let us introduce the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$ and consider a function $f : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$. The \mathfrak{c} -Fenchel-Moreau conjugate of f is the function $f^\mathfrak{c} : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ defined by

$$f^\mathfrak{c}(y) = \sup_{x \in \mathbb{R}^d} \left(\mathfrak{c}(x, y) - f(x) \right), \quad \forall y \in \mathbb{R}^d, \quad (3a)$$

and the \mathfrak{c} -Fenchel-Moreau biconjugate of f is the function $f^{\mathfrak{c}\mathfrak{c}'} : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ defined by

$$f^{\mathfrak{c}\mathfrak{c}'}(x) = \sup_{y \in \mathbb{R}^d} \left(\mathfrak{c}(x, y) - f^\mathfrak{c}(y) \right), \quad \forall x \in \mathbb{R}^d. \quad (3b)$$

Moreover, we have the inequality

$$f^{\mathfrak{c}\mathfrak{c}'}(x) \leq f(x), \quad \forall x \in \mathbb{R}^d, \quad (3c)$$

and following [6], we say that the function f is Capra-convex iff we have an equality in (3c). Lastly, Capra conjugacies also induce a notion of Capra-subdifferential. The Capra-subdifferential of f is the set-valued mapping $\partial_\mathfrak{c}f : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ defined by

$$y \in \partial_\mathfrak{c}f(x) \iff f^\mathfrak{c}(y) = \mathfrak{c}(x, y) - f(x), \quad (4a)$$

and which takes closed and convex set values [2, Proposition 1]. We say that [9, Definition 10.1] the function f is *Capra-subdifferentiable at* $x \in \mathbb{R}^d$ when $\partial_\mathfrak{c}f(x) \neq \emptyset$, and we introduce the domain of $\partial_\mathfrak{c}f$, defined as the set

$$\text{dom}(\partial_\mathfrak{c}f) = \{x \in \mathbb{R}^d \mid \partial_\mathfrak{c}f(x) \neq \emptyset\}. \quad (4b)$$

Observe that if we replace the Capra coupling \mathfrak{c} with the scalar product $\langle \cdot, \cdot \rangle$ in (3) and (4), we retrieve well-known notions of standard convex analysis. We refer to [3] for a more complete introduction to Capra conjugacies.

2.2 Capra-convexity and Capra-subdifferentiability of the ℓ_0 pseudonorm

We define the *support* of a vector $x \in \mathbb{R}^d$ by $\text{supp}(x) = \{j \in \{1, \dots, d\} \mid x_j \neq 0\}$. The ℓ_0 *pseudonorm* is the function $\ell_0 : \mathbb{R}^d \rightarrow \{0, 1, \dots, d\}$ defined by

$$\ell_0(x) = |\text{supp}(x)|, \quad \forall x \in \mathbb{R}^d, \quad (5)$$

where $|K|$ denotes the cardinality of a subset $K \subseteq \{1, \dots, d\}$. We recall the main results relating the Capra coupling \mathfrak{c} of Definition 2.1 and the ℓ_0 pseudonorm. To ease the reading, we gather the required background notions on norms in Appendix A.

First, we recall that, under a suitable choice of source norm, the ℓ_0 pseudonorm is Capra-subdifferentiable everywhere on \mathbb{R}^d , hence is a Capra-convex function.

Theorem 2.2 (from [2], Theorem 1 and Proposition 2) *Let $\|\cdot\|$ be the source norm employed for the Capra coupling $\dot{\varsigma}$ in Definition 2.1. If both the norm $\|\cdot\|$ and the dual norm $\|\cdot\|_\star$ are orthant-strictly monotonic (see Definition A.1), then we have that*

$$\partial_{\dot{\varsigma}} \ell_0(x) \neq \emptyset, \quad \forall x \in \mathbb{R}^d.$$

As a consequence, we have that

$$\ell_0^{\dot{\varsigma}\dot{\varsigma}'} = \ell_0.$$

Second, a generic formula for the Capra-subdifferential of ℓ_0 is given in [3]. To state this last result, we introduce the sets

$$Y_l = \{y \in \mathbb{R}^d \mid l \in \arg \max_{j \in \llbracket 0, d \rrbracket} (\|y\|_{(j),\star}^{\mathcal{R}} - j)\}, \quad \forall l \in \llbracket 0, d \rrbracket, \quad (6)$$

where $\{\|\cdot\|_{(j),\star}^{\mathcal{R}}\}_{j \in \llbracket 1, d \rrbracket}$ are the dual coordinate- k norms associated with the source norm $\|\cdot\|$, whose expressions are given in Definition A.3. Also, for a nonempty closed convex set $C \subseteq \mathbb{R}^d$ and $x \in \mathbb{R}^d$, we denote by $\mathcal{N}_C(x)$ the *normal cone of C at x* , whose definition (37) and properties are recalled in Appendix A.

Proposition 2.3 (from [3], Proposition 4.7 and [2], Proposition 1) *Let $\|\cdot\|$ be the source norm employed for the Capra coupling $\dot{\varsigma}$ in Definition 2.1. Let $\{\|\cdot\|_{(j)}^{\mathcal{R}}\}_{j \in \llbracket 1, d \rrbracket}$ and $\{\|\cdot\|_{(j),\star}^{\mathcal{R}}\}_{j \in \llbracket 1, d \rrbracket}$ be the associated sequences of coordinate- k and dual coordinate- k norms, as in Definition A.3, and let $\{\mathbb{B}_{(j)}^{\mathcal{R}}\}_{j \in \llbracket 1, d \rrbracket}$ and $\{\mathbb{B}_{(j),\star}^{\mathcal{R}}\}_{j \in \llbracket 1, d \rrbracket}$ be the corresponding sequences of unit balls for these norms. The Capra-subdifferential of the function ℓ_0 is the closed convex set given by*

- if $x = 0$,

$$\partial_{\dot{\varsigma}} \ell_0(0) = \bigcap_{j \in \llbracket 1, d \rrbracket} j \mathbb{B}_{(j),\star}^{\mathcal{R}}, \quad (7a)$$

- if $x \neq 0$ and $\ell_0(x) = l$,

$$\partial_{\dot{\varsigma}} \ell_0(x) = \mathcal{N}_{\mathbb{B}_{(l)}^{\mathcal{R}}} \left(\frac{x}{\|x\|_{(l)}^{\mathcal{R}}} \right) \cap Y_l. \quad (7b)$$

3 Capra-subdifferential of ℓ_0 for the ℓ_p source norms

The main contribution of this article is the following Theorem 3.1. It provides explicit formulas for the Capra-subdifferential of the ℓ_0 pseudonorm, as introduced in (4a) and as characterized in Proposition 2.3 for the ℓ_p source norms $\|\cdot\| = \|\cdot\|_p$ when $p \in [1, \infty]$.

We need to introduce the following norms and notations. For $y \in \mathbb{R}^d$, if ν is a permutation of $\llbracket 1, d \rrbracket$ such that $|y_{\nu(1)}| \geq \dots \geq |y_{\nu(d)}|$, the top (k, q) -norm $\|\cdot\|_{(k,q)}^{\text{tn}}$, for $k \in \llbracket 1, d \rrbracket$, is given by

$$\|y\|_{(k,q)}^{\text{tn}} = \left(\sum_{i=1}^k |y_{\nu(i)}|^q \right)^{\frac{1}{q}}, \quad \text{if } q \in [1, \infty[, \quad \text{and} \quad \|y\|_{(k,\infty)}^{\text{tn}} = \|y\|_\infty, \quad (8)$$

and the (p, k) -support norm $\|\cdot\|_{(p,k)}^{\text{sn}}$ is the dual norm of the top (k, q) -norm $\|\cdot\|_{(k,q)}^{\text{tn}}$, as defined in [7, §8.1]. Besides, for any $x \in \mathbb{R}^d$ and subset $K \subseteq \{1, \dots, d\}$, 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_K is the orthogonal projection of x onto the subspace¹

$$\mathcal{R}_K = \mathbb{R}^K \times \{0\}^{-K} = \{x \in \mathbb{R}^d \mid x_j = 0, \forall j \notin K\} \subseteq \mathbb{R}^d, \quad (9)$$

where $\mathcal{R}_\emptyset = \{0\}$.

Theorem 3.1 *Let the source norm $\|\cdot\| = \|\cdot\|_p$, where $p \in [1, \infty]$.*

- *If $p = 1$, the ℓ_0 pseudonorm is not Capra-convex, as its Capra-biconjugate is*

$$\ell_0^{\check{C}\check{C}'} : x \mapsto \begin{cases} 0, & \text{if } x = 0, \\ 1, & \text{if } x \neq 0. \end{cases} \quad (10)$$

Moreover, ℓ_0 is only Capra-subdifferentiable over $\text{dom}(\partial_{\check{C}}\ell_0) = \{x \in \mathbb{R}^d \mid \ell_0(x) \leq 1\}$. Over this domain, the Capra-subdifferential of ℓ_0 is given by

$$\partial_{\check{C}}\ell_0(0) = \mathbb{B}_{\|\cdot\|_\infty} \quad \text{and} \quad \partial_{\check{C}}\ell_0(x) = \mathcal{N}_{\mathbb{B}_{\|\cdot\|_1}}\left(\frac{x}{\|x\|_1}\right) \cap \{y \in \mathbb{R}^d \mid \|y\|_\infty \geq 1\}, \quad \text{if } \ell_0(x) = 1. \quad (11)$$

- *If $p \in]1, \infty[$, the ℓ_0 pseudonorm is Capra-convex and Capra-subdifferentiable everywhere, meaning that $\text{dom}(\partial_{\check{C}}\ell_0) = \mathbb{R}^d$. Its Capra-subdifferential is given by*

$$\partial_{\check{C}}\ell_0(0) = \mathbb{B}_{\|\cdot\|_\infty}, \quad (12a)$$

and at $x \neq 0$, denoting $l = \ell_0(x)$, $L = \text{supp}(x)$, and $q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$, by

$$y \in \partial_{\check{C}}\ell_0(x) \iff \begin{cases} y_L \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_p}}\left(\frac{x}{\|x\|_p}\right), \\ |y_j| \leq \min_{i \in L} |y_i|, \quad \forall j \notin L, \\ |y_{\nu(k+1)}|^q \geq (\|y\|_{(k,q)}^{\text{tn}} + 1)^q - (\|y\|_{(k,q)}^{\text{tn}})^q, \quad \forall k \in \llbracket 0, l-1 \rrbracket, \\ |y_{\nu(l+1)}|^q \leq (\|y\|_{(l,q)}^{\text{tn}} + 1)^q - (\|y\|_{(l,q)}^{\text{tn}})^q \quad (\text{when } l \neq d), \end{cases} \quad (12b)$$

where, for any $y \in \mathbb{R}^d$, ν denotes a permutation of $\llbracket 1, d \rrbracket$ such that $|y_{\nu(1)}| \geq \dots \geq |y_{\nu(d)}|$.

- *If $p = \infty$, the ℓ_0 pseudonorm is not Capra-convex, as its Capra-biconjugate is*

$$\ell_0^{\check{C}\check{C}'} : x \mapsto \begin{cases} 0, & \text{if } x = 0, \\ \frac{\|x\|_1}{\|x\|_\infty}, & \text{if } x \neq 0. \end{cases} \quad (13)$$

Moreover, the function ℓ_0 is only Capra-subdifferentiable over the domain

$$\text{dom}(\partial_{\check{C}}\ell_0) = \{x \in \mathbb{R}^d \mid \exists \lambda > 0, x_k \in \{-\lambda, 0, \lambda\}, \forall k \in \llbracket 1, d \rrbracket\} = \bigcup_{\lambda > 0} \{-\lambda, 0, \lambda\}^d. \quad (14)$$

¹Here, following notation from Game Theory, we have denoted by $-K$ the complementary subset of K in $\{1, \dots, d\}$: $K \cup (-K) = \{1, \dots, d\}$ and $K \cap (-K) = \emptyset$.

Over this domain, the Capra-subdifferential of ℓ_0 is given by

$$\partial_{\zeta} \ell_0(0) = \mathbb{B}_{\|\cdot\|_{\infty}}, \quad (15a)$$

and, at $x \in \cup_{\lambda>0} \{-\lambda, 0, \lambda\}^d \setminus \{0\}$, denoting $l = \ell_0(x)$, $L = \text{supp}(x)$, by

$$y \in \partial_{\zeta} \ell_0(x) \iff \begin{cases} y_L \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_{\infty}}}\left(\frac{x}{\|x\|_{\infty}}\right), \\ |y_j| \leq \min_{i \in L} |y_i|, \quad \forall j \notin L, \\ |y_{\nu(k+1)}| \geq 1, \quad \forall k \in \llbracket 0, l-1 \rrbracket, \\ |y_{\nu(l+1)}| \leq 1 \text{ (when } l \neq d), \end{cases} \quad (15b)$$

where, for any $y \in \mathbb{R}^d$, ν denotes a permutation of $\llbracket 1, d \rrbracket$ such that $|y_{\nu(1)}| \geq \dots \geq |y_{\nu(d)}|$.

We proceed in three steps to prove Theorem 3.1. First, in §3.1, we provide an explicit description of the set Y_l in (6) (that appears in (7b)). Second, in §3.2, we provide an explicit expression for the normal cone $\mathcal{N}_{\mathbb{B}_{\binom{\mathcal{R}}{l}}}$ in (7b). Third, in §3.3, we apply both results to the generic formulation of the Capra-subdifferential of ℓ_0 given in (7), and wrap up the proof of Theorem 3.1.

We will need the following properties of the coordinate- k and dual coordinate- k norms of Definition A.3.

Proposition 3.2 (from [3], Table 1) *Let the source norm $\|\cdot\|$ be a ℓ_p norm with $p \in [1, \infty]$, and let $q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$. The coordinate- k and dual coordinate- k norms in Definition A.3 are given, for $k \in \llbracket 1, d \rrbracket$, by*

$$\|\cdot\|_{\binom{\mathcal{R}}{(k), \star}} = \|\cdot\|_{(k, q)}^{\text{tn}} \quad \text{and} \quad \|\cdot\|_{\binom{\mathcal{R}}{(k)}} = \|\cdot\|_{(p, k)}^{\text{sn}}. \quad (16)$$

3.1 Description of the sets Y_l

We derive explicit descriptions of the sets Y_l in (6) for the ℓ_p source norms $\|\cdot\| = \|\cdot\|_p$, when $p \in [1, \infty]$. We start with two preliminary results on the top (k, q) -norm $\|\cdot\|_{(k, q)}^{\text{tn}}$, whose expression is given in (8). We state our first preliminary result in Lemma 3.3.

Lemma 3.3 *Let $y \in \mathbb{R}^d$, $q \in [1, \infty[$ and $k \in \llbracket 0, d-1 \rrbracket$. We have that*

$$\|y\|_{(k+1, q)}^{\text{tn}} - 1 \leq \|y\|_{(k, q)}^{\text{tn}} \implies \|y\|_{(k+j, q)}^{\text{tn}} - j \leq \|y\|_{(k, q)}^{\text{tn}}, \quad \forall j \in \llbracket 1, d-k \rrbracket. \quad (17)$$

Moreover, the same result holds if inequalities are strict in (17).

Proof. Let $y \in \mathbb{R}^d$ and ν a permutation of $\llbracket 1, d \rrbracket$ such that $|y_{\nu(1)}| \geq \dots \geq |y_{\nu(d)}|$. Let also $q \in [1, \infty[$, $k \in \llbracket 0, d-1 \rrbracket$ and $j \in \llbracket 1, d-k \rrbracket$. We use the shorthand notation

$$y_{k, q}^{\Sigma} = \sum_{i=1}^k |y_{\nu(i)}|^q, \quad (18)$$

so that, from Proposition 3.2, we have that $\|y\|_{(k,q)}^{\text{tn}} = (y_{k,q}^\Sigma)^{\frac{1}{q}}$.

First, we prove the inequality

$$(y_{k,q}^\Sigma + j|y_{\nu(k+1)}|^q)^{\frac{1}{q}} - (y_{k,q}^\Sigma)^{\frac{1}{q}} \leq j \left[(y_{k,q}^\Sigma + |y_{\nu(k+1)}|^q)^{\frac{1}{q}} - (y_{k,q}^\Sigma)^{\frac{1}{q}} \right]. \quad (19)$$

Indeed, we have that

$$\frac{1}{j}(y_{k,q}^\Sigma + j|y_{\nu(k+1)}|^q)^{\frac{1}{q}} + (1 - \frac{1}{j})(y_{k,q}^\Sigma)^{\frac{1}{q}} \leq \left(\frac{1}{j}(y_{k,q}^\Sigma + j|y_{\nu(k+1)}|^q) + (1 - \frac{1}{j})y_{k,q}^\Sigma \right)^{\frac{1}{q}},$$

by concavity of the function $x \mapsto x^{\frac{1}{q}}$ on \mathbb{R}_+ for $q \geq 1$,

$$\begin{aligned} \implies (y_{k,q}^\Sigma + j|y_{\nu(k+1)}|^q)^{\frac{1}{q}} + (j-1)(y_{k,q}^\Sigma)^{\frac{1}{q}} &\leq j \left(y_{k,q}^\Sigma + |y_{\nu(k+1)}|^q \right)^{\frac{1}{q}}, \\ \implies (y_{k,q}^\Sigma + j|y_{\nu(k+1)}|^q)^{\frac{1}{q}} - (y_{k,q}^\Sigma)^{\frac{1}{q}} &\leq j \left[(y_{k,q}^\Sigma + |y_{\nu(k+1)}|^q)^{\frac{1}{q}} - (y_{k,q}^\Sigma)^{\frac{1}{q}} \right]. \end{aligned}$$

Second, we prove the implication in (17) in its nonstrict inequality version. Let us assume that $\|y\|_{(k+1,q)}^{\text{tn}} - 1 \leq \|y\|_{(k,q)}^{\text{tn}}$. By definition of $y_{k,q}^\Sigma$ in (18) and since $|y_{\nu(k+1)}| \geq |y_{\nu(k+2)}| \geq \dots \geq |y_{\nu(k+j)}|$, we have that

$$\begin{aligned} \|y\|_{(k+j,q)}^{\text{tn}} - \|y\|_{(k,q)}^{\text{tn}} &\leq (y_{k,q}^\Sigma + j|y_{\nu(k+1)}|^q)^{\frac{1}{q}} - (y_{k,q}^\Sigma)^{\frac{1}{q}}, \\ &\leq j \left[(y_{k,q}^\Sigma + |y_{\nu(k+1)}|^q)^{\frac{1}{q}} - (y_{k,q}^\Sigma)^{\frac{1}{q}} \right], \quad (\text{from (19)}) \\ &= j \left[\|y\|_{(k+1,q)}^{\text{tn}} - \|y\|_{(k,q)}^{\text{tn}} \right], \\ &\quad (\text{from the expression of } \|\cdot\|_{(k,q)}^{\text{tn}} \text{ in (8) and by (18)}) \\ &\leq j, \quad (\text{by the assumption that } \|y\|_{(k+1,q)}^{\text{tn}} - \|y\|_{(k,q)}^{\text{tn}} \leq 1) \end{aligned}$$

which proves that $\|y\|_{(k+j,q)}^{\text{tn}} - j \leq \|y\|_{(k,q)}^{\text{tn}}$. The proof of the strict inequality version of (17) is analogous.

This ends the proof. \square

We state our second preliminary result in Lemma 3.4.

Lemma 3.4 *Let $y \in \mathbb{R}^d$, $q \in [1, \infty[$ and $k \in \llbracket 0, d-1 \rrbracket$. We have that*

$$\|y\|_{(k+1,q)}^{\text{tn}} - 1 \leq \|y\|_{(k,q)}^{\text{tn}} \iff |y_{\nu(k+1)}|^q \leq (\|y\|_{(k,q)}^{\text{tn}} + 1)^q - (\|y\|_{(k,q)}^{\text{tn}})^q. \quad (20)$$

Moreover, the same result holds if inequalities are strict or replaced with equalities in (20).

Proof. For $y \in \mathbb{R}^d$ and $k \in \llbracket 0, d-1 \rrbracket$, we have that

$$\begin{aligned} \|y\|_{(k+1,q)}^{\text{tn}} - 1 \leq \|y\|_{(k,q)}^{\text{tn}} &\iff \left(\sum_{i=1}^k |y_{\nu(i)}|^q + |y_{\nu(k+1)}|^q \right)^{\frac{1}{q}} - 1 \leq \|y\|_{(k,q)}^{\text{tn}}, \\ &\quad (\text{from the expression of } \|\cdot\|_{(k,q)}^{\text{tn}} \text{ in (8)}) \\ &\iff \sum_{i=1}^k |y_{\nu(i)}|^q + |y_{\nu(k+1)}|^q \leq (\|y\|_{(k,q)}^{\text{tn}} + 1)^q, \\ &\quad (\text{as the function } x \mapsto x^q \text{ is nondecreasing on } \mathbb{R}_+) \end{aligned}$$

so that finally, by definition (8) of $\|\cdot\|_{(k,q)}^{\text{tn}}$, we get

$$\|y\|_{(k+1,q)}^{\text{tn}} - 1 \leq \|y\|_{(k,q)}^{\text{tn}} \iff |y_{\nu(k+1)}|^q \leq (\|y\|_{(k,q)}^{\text{tn}} + 1)^q - (\|y\|_{(k,q)}^{\text{tn}})^q.$$

The proof of the strict inequality and equality versions of (17) is analogous. \square

We now provide explicit expressions of the sets Y_l in (6) for the ℓ_p source norms $\|\cdot\| = \|\cdot\|_p$, when $p \in [1, \infty]$.

Proposition 3.5 *Let the source norm be the ℓ_p norm $\|\cdot\| = \|\cdot\|_p$, where $p \in [1, \infty]$, and let $q \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. For $l \in \llbracket 0, d \rrbracket$, the set Y_l in (6) is given by*

- if $p = 1$,

$$Y_l = \begin{cases} \mathbb{B}_{\|\cdot\|_\infty} & \text{if } l = 0, \\ \{y \in \mathbb{R}^d \mid \|y\|_\infty \geq 1\} & \text{if } l = 1, \\ \emptyset & \text{else,} \end{cases} \quad (21a)$$

- if $p \in]1, \infty]$, for $y \in \mathbb{R}^d$ and ν a permutation of $\llbracket 1, d \rrbracket$ such that $|y_{\nu(1)}| \geq \dots \geq |y_{\nu(d)}|$,

$$y \in Y_l \iff \begin{cases} |y_{\nu(k+1)}|^q \geq (\|y\|_{(k,q)}^{\text{tn}} + 1)^q - (\|y\|_{(k,q)}^{\text{tn}})^q, \quad \forall k \in \llbracket 0, l-1 \rrbracket, \\ |y_{\nu(l+1)}|^q \leq (\|y\|_{(l,q)}^{\text{tn}} + 1)^q - (\|y\|_{(l,q)}^{\text{tn}})^q \quad (\text{when } l \neq d). \end{cases} \quad (21b)$$

Proof. We consider the case $p = 1$. When the source norm is $\|\cdot\| = \|\cdot\|_1$, we have that for $k \in \llbracket 1, d \rrbracket$, $\|\cdot\|_{(k),\star}^{\mathcal{R}} = \|\cdot\|_\infty$ [3, Table 1], and that $\|\cdot\|_{(0),\star}^{\mathcal{R}} = 0$, following the convention introduced in [3, §3.2]. Therefore, from the expression of Y_l in (6), we get that

$$Y_l = \{y \in \mathbb{R}^d \mid l \in \arg \max_{j \in \llbracket 0, d \rrbracket} (\|y\|_\infty \mathbf{1}_{j \neq 0} - j)\} = \begin{cases} \mathbb{B}_{\|\cdot\|_\infty} & \text{if } l = 0, \\ \{y \in \mathbb{R}^d \mid \|y\|_\infty \geq 1\} & \text{if } l = 1, \\ \emptyset & \text{else,} \end{cases}$$

hence (21a).

Next, we consider $p \in]1, \infty]$, and proceed in two steps to prove the equivalence in (21b).

In the first step (\Leftarrow), we take $y \in \mathbb{R}^d$ and we consider the two following cases that correspond to the right-hand side in (21b).

- If $|y_{\nu(k+1)}|^q \geq (\|y\|_{(k,q)}^{\text{tn}} + 1)^q - (\|y\|_{(k,q)}^{\text{tn}})^q$, $\forall k \in \llbracket 0, l-1 \rrbracket$,

then we get that

$$\begin{aligned} \|y\|_{(k+1,q)}^{\text{tn}} - 1 &\geq \|y\|_{(k,q)}^{\text{tn}}, \quad \forall k \in \llbracket 0, l-1 \rrbracket, && \text{(from (20))} \\ \implies \|y\|_{(k+1,q)}^{\text{tn}} - (k+1) &\geq \|y\|_{(k,q)}^{\text{tn}} - k, \quad \forall k \in \llbracket 0, l-1 \rrbracket, \\ \implies l &\in \arg \max_{j \in \llbracket 0, l \rrbracket} (\|y\|_{(j,q)}^{\text{tn}} - j). \end{aligned}$$

- If $l \neq d$ and $|y_{\nu(l+1)}|^q \leq (\|y\|_{(l,q)}^{\text{tn}} + 1)^q - (\|y\|_{(l,q)}^{\text{tn}})^q$,

then we get that

$$\|y\|_{(l+1,q)}^{\text{tn}} - 1 \leq \|y\|_{(l,q)}^{\text{tn}}, \quad (\text{from (20)})$$

$$\implies \|y\|_{(l+j,q)}^{\text{tn}} - j \leq \|y\|_{(l,q)}^{\text{tn}}, \quad \forall j \in \llbracket 1, d-l \rrbracket, \quad (\text{from (17)})$$

$$\implies \|y\|_{(l+j,q)}^{\text{tn}} - (l+j) \leq \|y\|_{(l,q)}^{\text{tn}} - l, \quad \forall j \in \llbracket 1, d-l \rrbracket,$$

$$\implies l \in \arg \max_{j \in \llbracket l, d \rrbracket} (\|y\|_{(j,q)}^{\text{tn}} - j).$$

Therefore, if the vector y satisfies both of the above assumptions, we get that $l \in \arg \max_{j \in \llbracket 0, d \rrbracket} (\|y\|_{(j,q)}^{\text{tn}} - j)$, and hence that $y \in Y_l$ by (6). This concludes the first step.

In the second step (\implies), we proceed by contraposition, assuming that either one of the two inequalities in the right-hand side of (21b) is not satisfied.

- If $\exists k \in \llbracket 0, l-1 \rrbracket$, $|y_{\nu(k+1)}|^q < (\|y\|_{(k,q)}^{\text{tn}} + 1)^q - (\|y\|_{(k,q)}^{\text{tn}})^q$,

then we get that

$$\exists k \in \llbracket 0, l-1 \rrbracket, \quad \|y\|_{(k+1,q)}^{\text{tn}} - 1 < \|y\|_{(k,q)}^{\text{tn}}, \quad (\text{from (20) with strict inequality})$$

$$\implies \exists k \in \llbracket 0, l-1 \rrbracket, \quad \|y\|_{(k+j,q)}^{\text{tn}} - j < \|y\|_{(k,q)}^{\text{tn}}, \quad \forall j \in \llbracket 1, d-k \rrbracket, \quad (\text{from (17)})$$

$$\implies \exists k \in \llbracket 0, l-1 \rrbracket, \quad \|y\|_{(k+j,q)}^{\text{tn}} - (k+j) < \|y\|_{(k,q)}^{\text{tn}} - k, \quad \forall j \in \llbracket 1, d-k \rrbracket,$$

$$\implies \exists k \in \llbracket 0, l-1 \rrbracket, \quad \|y\|_{(l,q)}^{\text{tn}} - l < \|y\|_{(k,q)}^{\text{tn}} - k, \quad (\text{as } l \in \{k+j \mid j \in \llbracket 1, d-k \rrbracket\} \text{ since } k < l)$$

$$\implies \exists k \in \llbracket 0, d \rrbracket, \quad \|y\|_{(l,q)}^{\text{tn}} - l < \|y\|_{(k,q)}^{\text{tn}} - k, \quad (\text{as } \llbracket 0, l-1 \rrbracket \subset \llbracket 0, d \rrbracket)$$

$$\implies l \notin \arg \max_{j \in \llbracket 0, d \rrbracket} (\|y\|_{(j,q)}^{\text{tn}} - j).$$

- If $l \neq d$ and $|y_{\nu(l+1)}|^q > (\|y\|_{(l,q)}^{\text{tn}} + 1)^q - (\|y\|_{(l,q)}^{\text{tn}})^q$,

then we get that

$$\|y\|_{(l+1,q)}^{\text{tn}} - 1 > \|y\|_{(l,q)}^{\text{tn}}, \quad (\text{from (20) with strict inequality})$$

$$\implies \|y\|_{(l+1,q)}^{\text{tn}} - (l+1) > \|y\|_{(l,q)}^{\text{tn}} - l,$$

$$\implies l \notin \arg \max_{j \in \llbracket 0, d \rrbracket} (\|y\|_{(j,q)}^{\text{tn}} - j).$$

In either case, we get that $y \notin Y_l$ by (6), which concludes the second step. We have finally proved the equivalence in (21b).

This ends the proof. □

3.2 Expression of the normal cone $\mathcal{N}_{\mathbb{B}_{(p,l)}^{\text{sn}}}$

We now turn to giving a description of the normal cone $\mathcal{N}_{\mathbb{B}_{(p,l)}^{\text{sn}}}$ in (7b) for the ℓ_p source norms $\|\cdot\| = \|\cdot\|_p$, when $p \in [1, \infty]$. We start with the following Lemma 3.6.

Lemma 3.6 *Let the source norm be the ℓ_p norm $\|\cdot\| = \|\cdot\|_p$, where $p \in [1, \infty]$. Let $x \in \mathbb{R}^d$, $l = \ell_0(x)$, $L = \text{supp}(x)$. If $l \in \llbracket 1, d \rrbracket$, we have that*

$$\left\| \frac{x}{\|x\|_{(p,l)}^{\text{sn}}} \right\|_p = 1, \quad (22a)$$

$$y \in \mathcal{N}_{\mathbb{B}_{(p,l)}^{\text{sn}}} \left(\frac{x}{\|x\|_{(p,l)}^{\text{sn}}} \right) \iff \|y\|_{(l,q)}^{\text{tn}} = \left\langle \frac{x}{\|x\|_{(p,l)}^{\text{sn}}}, y_L \right\rangle, \quad (22b)$$

$$y \in \mathcal{N}_{\mathbb{B}_{(p,l)}^{\text{sn}}} \left(\frac{x}{\|x\|_{(p,l)}^{\text{sn}}} \right) \implies \|y\|_{(l,q)}^{\text{tn}} \leq \|y_L\|_q. \quad (22c)$$

Proof. Let $q \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Let $x \in \mathbb{R}^d$, $l = \ell_0(x)$ and $L = \text{supp}(x)$. As we assume that $l \in \llbracket 1, d \rrbracket$, we have that $x \neq 0$ and we set $x' = \frac{x}{\|x\|_{(p,l)}^{\text{sn}}}$.

First, we prove (22a). We have that $l \geq 1$ and $\ell_0(x') = \ell_0(x) = l$. Thus, using [3, Proposition 3.5], we obtain that $\|x'\| = \|x'\|_{(l,q)}^{\mathcal{R}}$. Thus, from Proposition 3.2 we deduce that $\|x'\|_p = \|x'\|_{(p,l)}^{\text{sn}} = 1$.

Second, we prove (22b). We have the equivalence

$$\begin{aligned} y \in \mathcal{N}_{\mathbb{B}_{(p,l)}^{\text{sn}}}(x') &\iff \|x'\|_{(p,l)}^{\text{sn}} \|y\|_{(l,q)}^{\text{tn}} = \langle x', y \rangle && \text{(by definition (38) of the normal cone)} \\ &\iff \|y\|_{(l,q)}^{\text{tn}} = \langle x', y_L \rangle. && \text{(from } \|x'\|_{(p,l)}^{\text{sn}} = 1 \text{ and } L = \text{supp}(x')) \end{aligned}$$

Third, we prove (22c). We have that

$$\begin{aligned} y \in \mathcal{N}_{\mathbb{B}_{(p,l)}^{\text{sn}}}(x') &\iff \|y\|_{(l,q)}^{\text{tn}} = \langle x', y_L \rangle_{\mathbb{R}^l} && \text{(from (22b))} \\ &\implies \|y\|_{(l,q)}^{\text{tn}} \leq \|y_L\|_q. && \text{(from the Hölder inequality and (22a))} \end{aligned}$$

This ends the proof. \square

We now provide an explicit expression of the normal cone in (7b) for the ℓ_p source norms $\|\cdot\| = \|\cdot\|_p$, when $p \in [1, \infty]$.

Proposition 3.7 *Let the source norm be the ℓ_p norm $\|\cdot\| = \|\cdot\|_p$, where $p \in [1, \infty]$. Let $x \in \mathbb{R}^d$, $l = \ell_0(x)$ and $L = \text{supp}(x)$. If $l \in \llbracket 1, d \rrbracket$, we have that*

$$y \in \mathcal{N}_{\mathbb{B}_{(p,l)}^{\text{sn}}} \left(\frac{x}{\|x\|_{(p,l)}^{\text{sn}}} \right) \iff \begin{cases} y_L \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_p}} \left(\frac{x}{\|x\|_p} \right), \\ |y_j| \leq \min_{i \in L} |y_i|, \quad \forall j \notin L. \end{cases} \quad (23)$$

Proof. Let $q \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Let $x \in \mathbb{R}^d$, $l = \ell_0(x)$ and $L = \text{supp}(x)$, and let us set $x' = \frac{x}{\|x\|_{(p,l)}^{\text{sn}}}$. Let $y \in \mathbb{R}^d$, and let us set $I = \text{supp}(y)$.

First, we prove that

$$y \in \mathcal{N}_{\mathbb{B}_{(p,l)}^{\text{sn}}}(x') \implies \|y\|_{(l,q)}^{\text{tn}} = \|y_L\|_q. \quad (24)$$

We consider two cases. In the first case, we assume that $|I| = |\text{supp}(y)| \leq |\text{supp}(x)| = |L| = l$. Since the vector y has at most l nonzero coordinates, we get that $\|y\|_{(l,q)}^{\text{tn}} = \|y_I\|_q$ from the expression (8) of $\|\cdot\|_{(l,q)}^{\text{tn}}$. It follows that

$$\begin{aligned} y \in \mathcal{N}_{\mathbb{B}_{(p,l)}^{\text{sn}}}(x') &\implies \|y_I\|_q \leq \|y_L\|_q, && \text{(from (22c))} \\ &\implies \|y_L\|_q = \|y_I\|_q = \|y\|_{(l,q)}^{\text{tn}}. \\ &\text{(from } \|y_L\|_q \leq \|y\|_q = \|y_I\|_q, \text{ because } |I| = |\text{supp}(y)|) \end{aligned}$$

In the second case, we assume that $|I| = |\text{supp}(y)| > |\text{supp}(x)| = |L| = l$. Since the vector y has more than l nonzero coordinates, we get that $\|y\|_{(l,q)}^{\text{tn}} \geq \|y_L\|_q$ from the expression (8) of $\|\cdot\|_{(l,q)}^{\text{tn}}$. Combined with (22c), we deduce that $y \in \mathcal{N}_{\mathbb{B}_{(p,l)}^{\text{sn}}}(x') \implies \|y\|_{(l,q)}^{\text{tn}} = \|y_L\|_q$. Gathering the conclusions of both cases, we obtain (24).

Second, we prove (23). Observing that $\|x'\|_p = 1$ from (22a), we have that

$$\begin{aligned} y \in \mathcal{N}_{\mathbb{B}_{(p,l)}^{\text{sn}}}(x') &\iff \begin{cases} \|x'\|_p \|y_L\|_q = \langle x', y_L \rangle, \\ \|y\|_{(l,q)}^{\text{tn}} = \|y_L\|_q, \end{cases} && (\implies \text{ from (22b), (24); } \iff \text{ from (22b)}) \\ &\iff \begin{cases} y_L \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_p}(\frac{x}{\|x\|_p})}, \\ |y_j| \leq \min_{i \in L} |y_i|, \quad \forall j \notin L, \end{cases} \end{aligned}$$

by definition (38) of the normal cone, observing that $x' = \frac{x}{\|x\|_p}$ from (22a), and by the expression of $\|\cdot\|_{(l,q)}^{\text{tn}}$ in Proposition 3.2.

This ends the proof. □

3.3 Proof of Theorem 3.1

For the proof of Theorem 3.1, we use the following result that is essentially an application of [9, Proposition 10.1] to the special case of Capra conjugacies.

Fact 1 *Let $\|\cdot\|$ be a norm on \mathbb{R}^d and ζ be the Capra coupling as in Definition 2.1, inducing the definitions of the Capra-biconjugate (3b) and the Capra-subdifferential (4a). We have that*

$$\ell_0^{\zeta\zeta'}(x) \neq \ell_0(x) \implies \partial_{\zeta}\ell_0(x) = \emptyset, \quad \forall x \in \mathbb{R}^d. \quad (25)$$

We now turn to the proof of Theorem 3.1.

Proof. The proof is structured as follows: for the source norms $\|\cdot\| = \|\cdot\|_p$ with $p \in [1, \infty]$,

- (i) first, we give the expression of the Capra-biconjugate $\ell_0^{\zeta\zeta'}$,
- (ii) second, we give the expression of the Capra-subdifferential $\partial_{\zeta}\ell_0(x)$ at $x = 0$,
- (iii) third, we give the expression of the Capra-subdifferential $\partial_{\zeta}\ell_0(x)$ at $x \neq 0$,

(iv) fourth, we give the domain of the Capra-subdifferential $\partial_{\dot{\zeta}}\ell_0$.

(i) For $p \in]1, \infty[$, the norm $\|\cdot\|_p$ and its dual norm $\|\cdot\|_q$ (with $\frac{1}{p} + \frac{1}{q} = 1$) are orthant-strictly monotonic, following Definition A.1, so that $\ell_0^{\dot{\zeta}\dot{\zeta}'}$ = ℓ_0 , from Theorem 2.2. Turning to the case $p \in \{1, \infty\}$, we recall that, from [3, Proposition 4.4] and Proposition 3.2, if $q \in \{1, \infty\}$ is such that $\frac{1}{p} + \frac{1}{q} = 1$, then we get that

$$\ell_0^{\dot{\zeta}}(y) = \max_{j=1, \dots, d} (\|y\|_{(j,q)}^{\text{tn}} - j)^+, \quad \forall y \in \mathbb{R}^d. \quad (26)$$

First, we consider $p = 1$. From (26) and (8), we have that

$$\ell_0^{\dot{\zeta}}(y) = \max_{j \in \llbracket 1, d \rrbracket} (\|y\|_{\infty} - j)^+ = (\|y\|_{\infty} - 1)^+, \quad \forall y \in \mathbb{R}^d.$$

Thus, by definition of the Capra-biconjugate in (3b), we have that $\ell_0^{\dot{\zeta}\dot{\zeta}'}(0) = 0$, and that for any $x \in \mathbb{R}^d \setminus \{0\}$,

$$\begin{aligned} \ell_0^{\dot{\zeta}\dot{\zeta}'}(x) &= \sup_{y \in \mathbb{R}^d} \left(\frac{\langle x, y \rangle}{\|x\|_1} - (\|y\|_{\infty} - 1)^+ \right) \\ &= \max \left(\sup_{\|y\|_{\infty} \leq 1} \frac{\langle x, y \rangle}{\|x\|_1}, 1 + \sup_{\|y\|_{\infty} \geq 1} \frac{\langle x, y \rangle}{\|x\|_1} - \|y\|_{\infty} \right) \\ &= 1, \end{aligned}$$

since $\sup_{\|y\|_{\infty} \leq 1} \langle x, y \rangle = \|x\|_1$, by $\|\cdot\|_1 = (\|\cdot\|_{\infty})^*$, and $\langle x, y \rangle \leq \|x\|_1 \|y\|_{\infty}$, by Hölder's inequality. This proves (10).

Second, we consider $p = \infty$. From (26) and (8), for $y \in \mathbb{R}^d$ and ν a permutation of $\llbracket 1, d \rrbracket$ such that $|y_{\nu(1)}| \geq \dots \geq |y_{\nu(d)}|$, we have that

$$\ell_0^{\dot{\zeta}}(y) = \max_{j \in \llbracket 1, d \rrbracket} \left(\left(\sum_{k=1}^j |y_{\nu(k)}| \right) - j \right)^+ = \sum_{k=1}^d (|y_{\nu(k)}| - 1) \mathbf{1}_{|y_{\nu(k)}| \geq 1} = \sum_{k=1}^d (|y_k| - 1) \mathbf{1}_{|y_k| \geq 1}, \quad \forall y \in \mathbb{R}^d.$$

Thus, by definition of the Capra-biconjugate in (3b), we have that $\ell_0^{\dot{\zeta}\dot{\zeta}'}(0) = 0$, and that for any $x \in \mathbb{R}^d \setminus \{0\}$,

$$\begin{aligned} \ell_0^{\dot{\zeta}\dot{\zeta}'}(x) &= \sup_{y \in \mathbb{R}^d} \left(\frac{\langle x, y \rangle}{\|x\|_{\infty}} - \sum_{k=1}^d (|y_k| - 1) \mathbf{1}_{|y_k| \geq 1} \right) \\ &= \sum_{k=1}^d \sup_{y_k \in \mathbb{R}} \left(\frac{x_k y_k}{\|x\|_{\infty}} - (|y_k| - 1) \mathbf{1}_{|y_k| \geq 1} \right) \\ &= \sum_{k=1}^d \max \left(\sup_{|y_k| \leq 1} \frac{x_k y_k}{\|x\|_{\infty}}, 1 + \sup_{|y_k| \geq 1} \frac{x_k y_k}{\|x\|_{\infty}} - |y_k| \right) \\ &= \sum_{k=1}^d \frac{|x_k|}{\|x\|_{\infty}} = \frac{\|x\|_1}{\|x\|_{\infty}}, \end{aligned}$$

since, using similar arguments as above, $\sup_{|y_k| \leq 1} x_k y_k = |x_k|$, and

$$1 + \sup_{|y_k| \geq 1} \frac{x_k y_k}{\|x\|_\infty} - |y_k| \leq 1 + \sup_{|y_k| \geq 1} \frac{|x_k y_k|}{\|x\|_\infty} - |y_k| = 1 + \sup_{|y_k| \geq 1} \left(\frac{|x_k|}{\|x\|_\infty} - 1 \right) |y_k| = \frac{|x_k|}{\|x\|_\infty}.$$

This proves (13).

(ii) Let us recall that $\partial_{\zeta} \ell_0(0) = \bigcap_{j \in \llbracket 1, d \rrbracket} j \mathbb{B}_{(j), \star}^{\mathcal{R}}$, from (7a). If $p = 1$, from (16) we get that $\|\cdot\|_{(j), \star}^{\mathcal{R}} = \|\cdot\|_\infty$, $\forall j \in \llbracket 1, d \rrbracket$. We deduce that $\partial_{\zeta} \ell_0(0) = \mathbb{B}_{\|\cdot\|_\infty}$. We now assume that $p \in]1, \infty[$. From (16), we get that $\|\cdot\|_{(j), \star}^{\mathcal{R}} = \|\cdot\|_{(j, q)}^{\text{tn}}$, $\forall j \in \llbracket 1, d \rrbracket$ (with $\frac{1}{p} + \frac{1}{q} = 1$, $q \in [1, \infty[$). For $j = 1$, from (8), we get that $\|\cdot\|_{(1, q)}^{\text{tn}} = \|\cdot\|_\infty$, hence that $\mathbb{B}_{(1, q)}^{\text{tn}} = \mathbb{B}_{\|\cdot\|_\infty}$. Letting $j > 1$, we prove the inclusion $\mathbb{B}_{\|\cdot\|_\infty} \subseteq j \mathbb{B}_{(j, q)}^{\text{tn}}$. Indeed, we have that

$$\begin{aligned} y \in \mathbb{B}_{\|\cdot\|_\infty} &\implies |y_{\nu(1)}|^q \leq 1, && \text{(where } |y_{\nu(1)}| = \|y\|_\infty) \\ &\implies \sum_{i=1}^j |y_{\nu(i)}|^q \leq \sum_{i=1}^j 1 = j, && \text{(where } |y_{\nu(1)}| \geq \dots \geq |y_{\nu(d)}|) \\ &\implies \left(\sum_{i=1}^j |y_{\nu(i)}|^q \right)^{\frac{1}{q}} \leq j^{\frac{1}{q}}, \\ &\implies \|y\|_{(j, q)}^{\text{tn}} \leq j && \text{(by definition of } \|\cdot\|_{(j, q)}^{\text{tn}} \text{ in (8) and from } j \geq j^{\frac{1}{q}}) \\ &\implies y \in j \mathbb{B}_{(j, q)}^{\text{tn}}. \end{aligned}$$

We conclude that $\partial_{\zeta} \ell_0(0) = \bigcap_{j \in \llbracket 1, d \rrbracket} \mathbb{B}_{(j, q)}^{\text{tn}} = \mathbb{B}_{\|\cdot\|_\infty} \cap \left(\bigcap_{j \in \llbracket 2, d \rrbracket} \mathbb{B}_{(j, q)}^{\text{tn}} \right) = \mathbb{B}_{\|\cdot\|_\infty}$.

(iii) Let us recall that, for $x \neq 0$ and $\ell_0(x) = l$, we have that $\partial_{\zeta} \ell_0(x) = \mathcal{N}_{\mathbb{B}_{(p, l)}^{\text{sn}}} \left(\frac{x}{\|x\|_{(p, l)}^{\text{sn}}} \right) \cap Y_l$, from (7b). If $p \in]1, \infty[$, the expressions of $\partial_{\zeta} \ell_0(x)$ in (12b) and in (15b) are obtained combining Proposition 3.7 and Proposition 3.5. If $p = 1$, for $l \geq 2$, $Y_l = \emptyset$ from (21a) and thus $\partial_{\zeta} \ell_0(x) = \emptyset$. We now turn to the case $l = 1$, denoting $L = \text{supp}(x) = \{k\}$ where $k \in \llbracket 1, d \rrbracket$. We have that

$$\begin{aligned} y \in \mathcal{N}_{\mathbb{B}_{(1, 1)}^{\text{sn}}} \left(\frac{x}{\|x\|_{(1, 1)}^{\text{sn}}} \right) &\iff \begin{cases} y_L \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_1}} \left(\frac{x}{\|x\|_1} \right), \\ |y_j| \leq \min_{i \in L} |y_i|, \quad \forall j \notin L, \end{cases} && \text{(from Proposition 3.7)} \\ &\iff \begin{cases} \|x\|_1 \|y_L\|_\infty = \langle x, y_L \rangle, \\ \|y_L\|_\infty = \|y\|_\infty, \end{cases} && \text{(from (38) and by definition of } \|\cdot\|_\infty) \\ &\iff \|x\|_1 \|y\|_\infty = \langle x, y \rangle, && \text{(from } \langle x, y \rangle = x_k y_k \text{ and } \|x\|_1 = |x_k|) \\ &\iff y \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_1}} \left(\frac{x}{\|x\|_1} \right), && \text{(from (38))} \end{aligned}$$

therefore, we deduce from (21a) the expression of $\partial_{\zeta} \ell_0(x)$ in (11).

(iv) For $p \in]1, \infty[$, the norm $\|\cdot\|_p$ and its dual norm $\|\cdot\|_q$ (with $\frac{1}{p} + \frac{1}{q} = 1$) are orthant-strictly monotonic, following Definition A.1, so that $\text{dom}(\partial_{\zeta} \ell_0) = \mathbb{R}^d$, from Theorem 2.2. We now turn to the case $p \in \{1, \infty\}$.

First, we consider $p = 1$. Given the expression of $\ell_0^{\dot{c}c'}$ in (10), $\ell_0^{\dot{c}c'}(x) = \ell_0(x) \iff \ell_0(x) \leq 1$. We deduce from Fact 1 that $\text{dom}(\partial_{\dot{c}}\ell_0) \subseteq \{x \in \mathbb{R}^d \mid \ell_0(x) \leq 1\}$. We prove the reciprocal inclusion. We already know from (11) that $\partial_{\dot{c}}\ell_0(0) \neq \emptyset$. Let $x \in \mathbb{R}^d$ be such that $\ell_0(x) = 1$. There exists $k \in \llbracket 1, d \rrbracket$ such that $\text{supp}(x) = \{k\}$. Let us introduce $y \in \mathbb{R}^d$ such that $\text{supp}(y) = \{k\}$ with $y_k \in \{-1, 1\}$ and $x_k y_k = |x_k|$. It follows that $\|y\|_\infty = 1$ and $\|x\|_1 \|y\|_\infty = x_k y_k = \langle x, y \rangle$ and thus that $y \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_1}}(\frac{x}{\|x\|_1})$, from (38). We deduce from (11) that $y \in \partial_{\dot{c}}\ell_0(x)$, hence that $\partial_{\dot{c}}\ell_0(x) \neq \emptyset$. This proves the reciprocal inclusion, and we conclude that $\text{dom}(\partial_{\dot{c}}\ell_0) = \{x \in \mathbb{R}^d \mid \ell_0(x) \leq 1\}$.

Second, we consider $p = \infty$. Given the expression of $\ell_0^{\dot{c}c'}$ in (13), we have that $\ell_0^{\dot{c}c'}(0) = \ell_0(0)$, and for $x \neq 0$,

$$\begin{aligned} \ell_0^{\dot{c}c'}(x) = \ell_0(x) &\iff l = \sum_{i=1}^l \frac{|x_{\nu(i)}|}{|x_{\nu(1)}|}, & (\text{where } \ell_0(x) = l \text{ and } |x_{\nu(1)}| \geq \dots \geq |x_{\nu(d)}|) \\ &\iff \begin{cases} |x_{\nu(k)}| = |x_{\nu(1)}|, & \forall k \in \llbracket 1, l \rrbracket, \\ x_{\nu(k)} = 0, & \forall k \in \llbracket l+1, d \rrbracket \text{ (when } l \neq d). \end{cases} \end{aligned}$$

We deduce that $\ell_0^{\dot{c}c'}(x) = \ell_0(x) \iff x \in \cup_{\lambda>0} \{-\lambda, 0, \lambda\}^d$, and thus from Fact 1 that $\text{dom}(\partial_{\dot{c}}\ell_0) \subseteq \cup_{\lambda>0} \{-\lambda, 0, \lambda\}^d$. We prove the reciprocal inclusion. We already know from (15a) that $\partial_{\dot{c}}\ell_0(0) \neq \emptyset$. Let $x \in \cup_{\lambda>0} \{-\lambda, 0, \lambda\}^d$ be such that $x \neq 0$, and ν be a permutation of $\llbracket 1, d \rrbracket$ such that $|x_{\nu(1)}| \geq \dots \geq |x_{\nu(d)}|$. Let us introduce $y \in \mathbb{R}^d$ such that $|y_{\nu(k)}| = 1$, $\forall k \in \llbracket 1, l \rrbracket$ and $|y_{\nu(k)}| = 0$, $\forall k \in \llbracket l+1, d \rrbracket$. It follows that, denoting $L = \text{supp}(x)$, $\|x\|_\infty \|y_L\|_1 = \lambda l = \langle x, y_L \rangle$, and thus that $y_L \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_\infty}}(\frac{x}{\|x\|_\infty})$, from (38). We deduce from (15b) that $y \in \partial_{\dot{c}}\ell_0(x)$, hence that $\partial_{\dot{c}}\ell_0(x) \neq \emptyset$. This proves the reciprocal inclusion, and we conclude that $\text{dom}(\partial_{\dot{c}}\ell_0) = \cup_{\lambda>0} \{-\lambda, 0, \lambda\}^d$.

This ends the proof. □

4 Graphical representations and discussion

First, we provide graphical representations of the Capra-subdifferential of the ℓ_0 pseudonorm in §4.1. Second, we compare our expression of $\partial_{\dot{c}}\ell_0$ with other notions of generalized subdifferential for the ℓ_0 pseudonorm and illustrate one of its applications in §4.2.

4.1 Visualization with the ℓ_2 source norm

We detail the Capra-subdifferential of ℓ_0 for the ℓ_2 source norm $\|\cdot\| = \|\cdot\|_2$. According to Theorem 3.1, we have that

$$\partial_{\dot{c}}\ell_0(0) = \mathbb{B}_{\|\cdot\|_\infty}, \tag{27a}$$

and for $x \neq 0$, $y \in \mathbb{R}^d$, denoting $l = \ell_0(x)$, $L = \text{supp}(x)$, and ν a permutation of $\llbracket 1, d \rrbracket$ such that $|y_{\nu(1)}| \geq \dots \geq |y_{\nu(d)}|$,

$$y \in \partial_{\zeta} \ell_0(x) \iff \begin{cases} y_L = \lambda x, \quad \lambda \geq 0, \\ |y_j| \leq \min_{i \in L} |y_i|, \quad \forall j \notin L, \\ |y_{\nu(k+1)}|^2 \geq (\|y\|_{(k,2)}^{\text{tn}} + 1)^2 - (\|y\|_{(k,2)}^{\text{tn}})^2, \quad \forall k \in \llbracket 0, l-1 \rrbracket, \\ |y_{\nu(l+1)}|^2 \leq (\|y\|_{(l,2)}^{\text{tn}} + 1)^2 - (\|y\|_{(l,2)}^{\text{tn}})^2 \quad (\text{when } l \neq d). \end{cases} \quad (27b)$$

We illustrate in Figure 1 the Capra-subdifferentials obtained with (27) in the two-dimensional case where $\ell_0 : \mathbb{R}^2 \rightarrow \{0, 1, 2\}$. In Figure 1a, we display the Capra-subdifferential of ℓ_0 at three typical points, covering the three possible cases in \mathbb{R}^2 , with $\ell_0(x) = 0$ (green color), $\ell_0(x) = 1$ (red color), and $\ell_0(x) = 2$ (blue color). Then, using the same colors, we display in Figure 1b the Capra-subdifferential of ℓ_0 at all points in \mathbb{R}^2 .

4.2 Discussion

First, we compare the Capra-subdifferential of the ℓ_0 pseudonorm given in Theorem 3.1 with other notions of subdifferentials. We recall that, for ℓ_0 , the standard subdifferential of convex analysis obtained with the Fenchel conjugacy is given by (see [3, Table 3])

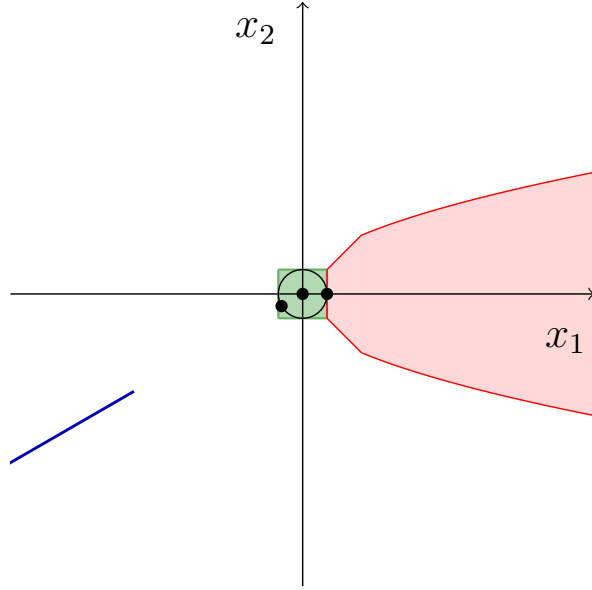
$$\partial \ell_0(0) = \{0\} \text{ and } \partial \ell_0(x) = \emptyset, \quad \forall x \in \mathbb{R}^d \setminus \{0\}. \quad (28)$$

We also recall further notions of generalized subdifferentials obtained for the ℓ_0 pseudonorm. We refer the reader to [5] for the definitions of the Fréchet, viscosity, proximal, Clarke and limiting subdifferentials, where the author establishes that all these notions coincide for the ℓ_0 pseudonorm, and are equal to the set-valued mapping

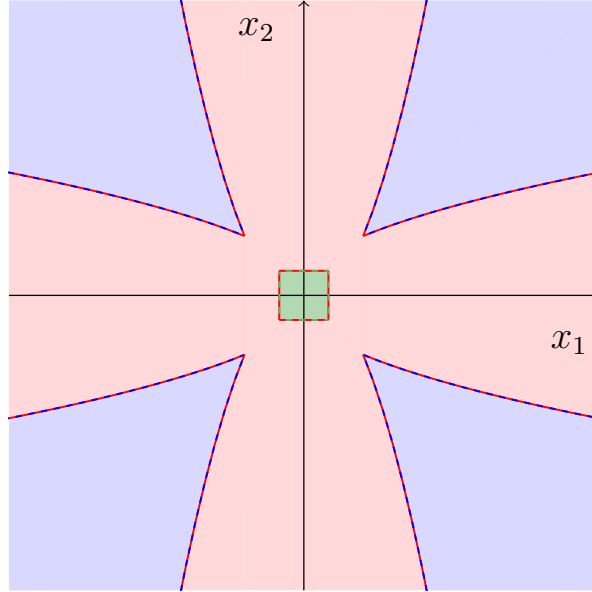
$$\mathcal{D} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d, \quad x \mapsto \{y \in \mathbb{R}^d \mid y_L = 0\}, \quad (29)$$

where $L = \text{supp}(x)$, from [5, Theorems 1, 2]. We deduce that the Capra-subdifferential of the ℓ_0 pseudonorm is significantly different from previous notions of generalized subdifferentials of ℓ_0 , summarized by (28) and (29). In particular, whereas $\{y \in \mathbb{R}^d \mid y_L = 0\}$ is a vector subspace, the Capra-subdifferential $\partial_{\zeta} \ell_0(x)$ is a closed convex set, but not a vector subspace. However, we recall that the Capra-subdifferential of ℓ_0 is related to the standard subdifferential of \mathcal{L}_0 — a proper closed convex function first introduced in [1, §4.1] for the Euclidean source norm, then generalized in [2, Equation (19), Proposition 3] — that “factorizes” the ℓ_0 pseudonorm, in the sense that $\ell_0 = \mathcal{L}_0 \circ n$, where $n : \mathbb{R}^d \rightarrow \mathbb{S}_{\|\cdot\|} \cup \{0\}$ is the normalization mapping such that $\zeta(\cdot, \cdot) = \langle n(\cdot), \cdot \rangle$ in (2). Indeed, by application of [2, Item (c), Proposition 3], when the source norm $\|\cdot\|$ is a ℓ_p norm with $p \in]1, \infty[$, the Capra-subdifferential of ℓ_0 and the standard subdifferential of \mathcal{L}_0 coincide on the unit sphere, that is,

$$p \in]1, \infty[\text{ and } \|x\|_p = 1 \implies \partial_{\zeta} \ell_0(x) = \partial \mathcal{L}_0(x). \quad (30)$$



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



(b) $\partial_{\dot{\zeta}} \ell_0(0) \cup \left\{ \bigcup_{\ell_0(x)=1} \partial_{\dot{\zeta}} \ell_0(x) \right\} \cup \left\{ \bigcup_{\ell_0(x)=2} \partial_{\dot{\zeta}} \ell_0(x) \right\}$

Figure 1: Capra-subdifferential of the ℓ_0 pseudonorm in \mathbb{R}^2 with the ℓ_2 source norm $\|\cdot\| = \|\cdot\|_2$, illustrated for three typical points (Figure 1a) and for all points in \mathbb{R}^2 (Figure 1b)

It follows that, for these ℓ_p source norms, Equation (12b) — which provides explicit formulas for $\partial_{\dot{\zeta}}\ell_0(x)$ — then also gives, on the unit sphere $\mathbb{S}_{\|\cdot\|_p}$, explicit formulas for the standard subdifferential of the proper closed convex function \mathcal{L}_0 .

Second, we argue that, since the ℓ_0 pseudonorm displays the Capra-convex properties stated in Theorem 3.1, the Capra-subdifferential is relevant to obtain lower approximations of the ℓ_0 pseudonorm. We recall that nonconvex continuous approximations of the ℓ_0 pseudonorm have gained a lot of interest in the field of sparse optimization, especially due to applications in machine learning [10, 8, 11]. The lower approximation of ℓ_0 that we propose next can be seen as a generalization of polyhedral lower approximations obtained for a proper, lower semicontinuous and convex function: here, the maximum of a finite number of affine functions now translates into “polyhedral-like” [9, p. 114] functions that are the maximum of a finite number of so-called *Capra-affine* functions, that is, functions of the form $x \mapsto \dot{\zeta}(x, y) - z$ for fixed $y \in \mathbb{R}^d$ and $z \in \overline{\mathbb{R}}$.

Let the source norm $\|\cdot\|$ be a ℓ_p norm, with $p \in]1, \infty[$, and let $\{x_i\}_{i \in I}$ and $\{y_i\}_{i \in I}$ be two collections of points such that for $i \in I$, $x_i \in \mathbb{R}^d$ and $y_i \in \partial_{\dot{\zeta}}\ell_0(x_i)$. By definition of the Capra-biconjugate in (3b), we have that

$$\max_{i \in I} \left(\dot{\zeta}(x, y_i) - \ell_0^{\dot{\zeta}}(y_i) \right) \leq \sup_{y \in \mathbb{R}^d} \left(\dot{\zeta}(x, y) - \ell_0^{\dot{\zeta}}(y) \right) = \ell_0^{\dot{\zeta}\dot{\zeta}'}(x), \quad \forall x \in \mathbb{R}^d. \quad (31)$$

Therefore, we deduce from (3c) that the function

$$\underline{\ell}_0 : \mathbb{R}^d \rightarrow \mathbb{R}, \quad x \mapsto \max_{i \in I} \left(\dot{\zeta}(x, y_i) - \ell_0^{\dot{\zeta}}(y_i) \right) \quad (32)$$

gives a lower bound for ℓ_0 . Moreover, by definition of the Capra-subdifferential in (4a), we have that for, $i \in I$,

$$\dot{\zeta}(x_i, y_i) - \ell_0^{\dot{\zeta}}(y_i) = \ell_0(x_i), \quad (33)$$

so that this lower bound is exact (tight) at the points in $\{x_i\}_{i \in I}$, in the sense that $\underline{\ell}_0(x_i) = \ell_0(x_i)$. Thus, we can tighten the inequality in (31) by enlarging the collections $\{x_i\}_{i \in I}$ and $\{y_i\}_{i \in I}$. We provide an example of such a lower approximation of ℓ_0 in Figure 2b, using the ℓ_2 source norm $\|\cdot\| = \|\cdot\|_2$. By definition of $\underline{\ell}_0$ in (32) and of the Capra coupling in (3), it is straightforward to see that $\underline{\ell}_0$ is constant along rays, so that we only give its representation on $\mathbb{S} \cup \{0\}$ (orange color). Observe that, at the sample points $\{x_i\}_{i \in I}$ (black dots), $\underline{\ell}_0$ takes the same values as ℓ_0 (blue color, Figure 2a).

5 Conclusion

We have derived explicit formulations for the Capra-subdifferential of the ℓ_0 pseudonorm for the ℓ_p source norms with $p \in [1, \infty]$. With these formulations, it is now possible to compute elements in such Capra-subdifferentials, that we have illustrated by a graphical representation. On top of that, we have extended previous knowledge on ℓ_0 , establishing

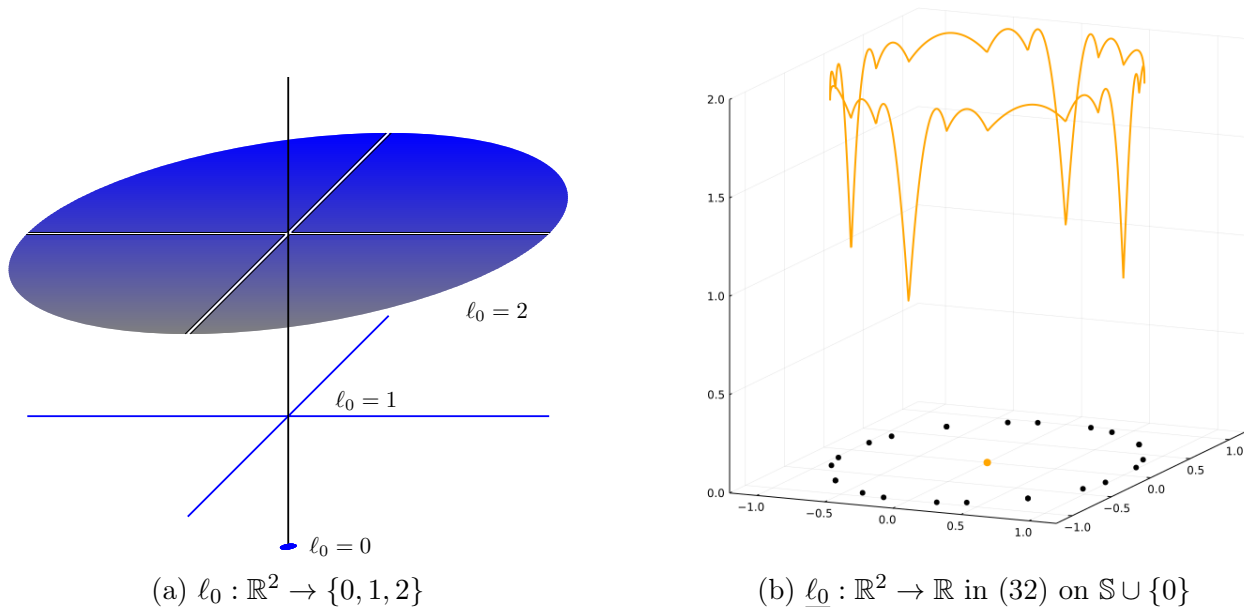


Figure 2: The ℓ_0 pseudonorm in \mathbb{R}^2 (blue color, Figure 2a) and a “polyhedral-like” [9, p. 114] lower bound $\underline{\ell}_0$ as in (32) represented on $\mathbb{S} \cup \{0\}$ (orange color, Figure 2b) obtained for the ℓ_2 source norm $\|\cdot\| = \|\cdot\|_2$ with points $\{x_i\}_{i \in I}$ sampled on \mathbb{S} (black dots, Figure 2b)

that it is neither Capra-convex nor Capra-subdifferentiable everywhere in the limit cases where $p \in \{1, \infty\}$.

The formulation that we obtain differs drastically from previous notions of generalized subdifferential for the ℓ_0 pseudonorm. Whereas most other notions coincide, the Capra-subdifferential enriches this collection and is an interesting tool to deal with the function ℓ_0 , in the spirit of the usual notion of subdifferential for proper lower semicontinuous convex functions.

Acknowledgement. We thank the two anonymous referees that helped us to improve the quality of this paper.

A Background on norms

For any norm $\|\cdot\|$ on \mathbb{R}^d , we introduce derived norms and some of their properties.

Dual norms and normal cones

The following expression

$$\|y\|_* = \sup_{\|x\| \leq 1} \langle x, y \rangle, \quad \forall y \in \mathbb{R}^d \quad (34)$$

defines a norm on \mathbb{R}^d , called the *dual norm* $\|\cdot\|_\star$. In line with our notations for the norm $\|\cdot\|$ in (1), we denote the unit sphere and the unit ball of the dual norm $\|\cdot\|_\star$ by

$$\mathbb{S}_\star = \{y \in \mathbb{R}^d \mid \|y\|_\star = 1\}, \quad (35a)$$

$$\mathbb{B}_\star = \{y \in \mathbb{R}^d \mid \|y\|_\star \leq 1\}. \quad (35b)$$

Note that by definition of the dual norm in (34), we have the inequality

$$\langle x, y \rangle \leq \|x\| \times \|y\|_\star, \quad \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d. \quad (36)$$

Equality cases in the above inequality can be characterized in term of geometric objects of convex analysis. For this purpose, we recall that the *normal cone* $\mathcal{N}_C(x)$ to the nonempty closed convex subset $C \subseteq \mathbb{R}^d$ at $x \in C$ is the closed convex cone defined by [4, Definition 5.2.3]

$$\mathcal{N}_C(x) = \{y \in \mathbb{R}^d \mid \langle x' - x, y \rangle \leq 0, \quad \forall x' \in C\}. \quad (37)$$

Now, easy computations show that for any $(x, y) \in \mathbb{R}^d \setminus \{0\} \times \mathbb{R}^d \setminus \{0\}$, we have the equivalence

$$\langle x, y \rangle = \|x\| \times \|y\|_\star \iff y \in \mathcal{N}_{\mathbb{B}}\left(\frac{x}{\|x\|}\right) \iff x \in \mathcal{N}_{\mathbb{B}_\star}\left(\frac{y}{\|y\|_\star}\right). \quad (38)$$

Orthant strict monotonicity

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

Definition A.1 (from [2], Definition 5) *A norm $\|\cdot\|$ on the space \mathbb{R}^d is called orthant-strictly monotonic if, for all x, x' in \mathbb{R}^d , we have*

$$(|x| < |x'| \text{ and } x \circ x' \geq 0) \implies \|x\| < \|x'\|, \quad (39)$$

where $|x| < |x'|$ means that $|x_i| \leq |x'_i|$ for all $i = 1, \dots, d$, and that there exists $j \in \{1, \dots, d\}$, such that $|x_j| < |x'_j|$; and $x \circ x' = (x_1x'_1, \dots, x_dx'_d)$ is the Hadamard (entrywise) product.

Restriction norms, coordinate- k and dual coordinate- k norms

We start by introducing restriction norms and their dual.

Definition A.2 ([3], Definition 3.1) *For any norm $\|\cdot\|$ on \mathbb{R}^d and any subset $K \subseteq \{1, \dots, d\}$, we define two norms on the subspace \mathcal{R}_K of \mathbb{R}^d , as defined in (9), as follows.*

- The K -restriction norm $\|\cdot\|_K$ is defined by

$$\|x\|_K = \|x\|, \quad \forall x \in \mathcal{R}_K. \quad (40)$$

- The (K, \star) -norm $\|\cdot\|_{K, \star}$ is the norm $(\|\cdot\|_K)_\star$, given by the dual norm (on the subspace \mathcal{R}_K) of the restriction norm $\|\cdot\|_K$ to the subspace \mathcal{R}_K (first restriction, then dual).

With these norms, we define the coordinate- k and dual coordinate- k norms.

Definition A.3 ([3], Definition 3.2) For $k \in \{1, \dots, d\}$, we call coordinate- k norm the norm $\|\cdot\|_{(k)}^{\mathcal{R}}$ whose dual norm is the dual coordinate- k norm, denoted by $\|\cdot\|_{(k),\star}^{\mathcal{R}}$, with expression

$$\|y\|_{(k),\star}^{\mathcal{R}} = \sup_{|K| \leq k} \|y_K\|_{K,\star}, \quad \forall y \in \mathbb{R}^d, \quad (41)$$

where the (K, \star) -norm $\|\cdot\|_{K,\star}$ is given in Definition A.2, and where the notation $\sup_{|K| \leq k}$ is a shorthand for $\sup_{K \subseteq \{1, \dots, d\}, |K| \leq k}$.

Also, following [3, §3.2], we extend the dual coordinate- k norms in Definition A.3 with the convention $\|\cdot\|_{(0),\star}^{\mathcal{R}} = 0$, although this is not a norm on \mathbb{R}^d but a seminorm.

References

- [1] J.-P. Chancelier and M. De Lara. Hidden convexity in the l_0 pseudonorm. *Journal of Convex Analysis*, 28(1):203–236, 2021.
- [2] J.-P. Chancelier and M. De Lara. Capra-convexity, convex factorization and variational formulations for the l_0 pseudonorm. *Set-Valued and Variational Analysis*, 30:597–619, 2022.
- [3] J.-P. Chancelier and M. De Lara. Constant along primal rays conjugacies and the l_0 pseudonorm. *Optimization*, 71(2):355–386, 2022. doi: 10.1080/02331934.2020.1822836.
- [4] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Springer, 2004.
- [5] H. Y. Le. Generalized subdifferentials of the rank function. *Optimization Letters*, 7(4): 731–743, 2013.
- [6] J. E. Martínez-Legaz. Generalized convex duality and its economic applications. In S. 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.
- [7] A. M. McDonald, M. Pontil, and D. Stamos. New perspectives on k-support and cluster norms. *Journal of Machine Learning Research*, 17(155):1–38, 2016.
- [8] Q. Qu, J. Sun, and J. Wright. Finding a sparse vector in a subspace: linear sparsity using alternating directions. *IEEE Transactions on Information Theory*, 62(10):5855–5880, Oct 2016. doi: 10.1109/tit.2016.2601599.
- [9] I. Singer. *Abstract Convex Analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons, Inc., New York, 1997. ISBN 0-471-16015-6.

- [10] E. Soubies, L. Blanc-Féraud, and G. Aubert. A continuous exact l_0 penalty (CEL0) for least squares regularized problem. *SIAM Journal on Imaging Sciences*, 8(3):1607–1639, 2015.
- [11] F. Wen, L. Chu, P. Liu, and R. C. Qiu. A survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning. *IEEE Access*, 6:69883–69906, 2018.