The Capra-subdifferential of the l0 pseudonorm

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

The Capra-subdifferential of the ℓ_0 pseudonorm

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

December 30, 2021

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 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 ℓ_0, 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

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 [2, 3, 1], 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 [1] 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 [1].

*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 when $p \in \{1, \infty\}$, the ℓ_0 pseudonorm is not Capra-convex. We also illustrate the Capra-subdifferential of ℓ_0 that we find, and compare it with other notions of generalized subdifferentials for ℓ_0 found in [5]. We argue that the Capra subdifferential is a is well-suited notion of generalized subdifferential to handle ℓ_0, since it is based on a conjugacy. To illustrate the subsequent opportunities allowed by conjugacies, we show how to derive polyhedral-like lower bounds on ℓ_0, based on the knowledge of its Capra subdifferential.

The article is organized as follows. First, we recall background notions on the Capra coupling 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 $[\![i, j]\!] = \{i, i+1, \dots, j-1, j\}$. 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,

$$\mathbb{S} = \left\{ x \in \mathbb{R}^d \,\middle|\, |||x||| = 1 \right\} \quad \text{and} \quad \mathbb{B} = \left\{ x \in \mathbb{R}^d \,\middle|\, |||x||| \leq 1 \right\}, \tag{1}$$

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 conjugacies 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 1 ([3], Definition 4.1) *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$. We define the* coupling $\mathcal{c} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *between $\mathbb{R}^d$ and $\mathbb{R}^d$, that we call the* Capra *coupling, by*

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

A coupling function such as the Capra coupling $\dot{\varsigma}$ given in Definition 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 $f : \mathbb{R}^d \to \overline{\mathbb{R}}$. The $\dot{\varsigma}$-Fenchel-Moreau conjugate of f is the function $f^{\dot{\varsigma}} : \mathbb{R}^d \to \overline{\mathbb{R}}$ defined by

$$f^{\dot{\varsigma}}(y) = \sup_{x \in \mathbb{R}^d} \Big(\dot{\varsigma}(x, y) - f(x) \Big) \ , \ \ \forall y \in \mathbb{R}^d \ , \tag{3a}$$

and the $\dot{\varsigma}$-Fenchel-Moreau biconjugate of f is the function $f^{\dot{\varsigma}\dot{\varsigma}'} : \mathbb{R}^d \to \overline{\mathbb{R}}$ defined by

$$f^{\dot{\varsigma}\dot{\varsigma}'}(x) = \sup_{y \in \mathbb{R}^d} \Big(\dot{\varsigma}(x, y) - f^{\dot{\varsigma}}(y) \Big) \ , \ \ \forall x \in \mathbb{R}^d \ . \tag{3b}$$

Moreover, we have the inequality

$$f^{\dot{\varsigma}\dot{\varsigma}'}(x) \leq f(x) \ , \ \ \forall x \in \mathbb{R}^d \ , \tag{3c}$$

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 $\partial_{\dot{\varsigma}} f : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ defined by

$$y \in \partial_{\dot{\varsigma}} f(x) \iff f^{\dot{\varsigma}}(y) = \dot{\varsigma}(x, y) - f(x) \ . \tag{4}$$

Observe that if we replace the Capra coupling $\dot{\varsigma}$ 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 $\mathrm{supp}(x) = \big\{ j \in \{1, \dots, d\} \,\big|\, x_j \neq 0 \big\}$. The ℓ_0 *pseudonorm* is the function $\ell_0 : \mathbb{R}^d \to \{0, 1, \dots, d\}$ defined by

$$\ell_0(x) = \big| \mathrm{supp}(x) \big| \ , \ \ \forall x \in \mathbb{R}^d \ , \tag{5}$$

where $|K|$ denotes the cardinality of a subset $K \subseteq \{1, \dots, d\}$. We recall the main results relating the Capra coupling $\dot{\varsigma}$ of Definition 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 judicious choice of source norm, the ℓ_0 pseudonorm is Capra-subdifferentiable everywhere on $\mathbb{R}^d$, hence is a Capra-convex function.

Theorem 2 (from [1], Theorem 8 and Proposition 7) *Let $|||\cdot|||$ be the source norm employed for the Capra coupling $\dot{\varsigma}$ in Definition 1. If both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic (see Definition 12), then we have that*

$$\partial_{\dot{\varsigma}} \ell_0(x) \neq \emptyset \ , \ \ \forall x \in \mathbb{R}^d \ . \tag{6}$$

As a consequence, we have that

$$\ell_0^{\dot{\varsigma}\dot{\varsigma}'} = \ell_0 \ . \tag{7}$$

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

$$Y_l = \left\{ y \in \mathbb{R}^d \,\middle|\, l \in \operatorname*{arg\,max}_{j \in [\![0,d]\!]} \left(|||y|||^{\mathcal{R}}_{(j),\star} - j \right) \right\} , \quad \forall l \in [\![0,d]\!] , \tag{8}$$

where $\{|||\cdot|||^{\mathcal{R}}_{(j),\star}\}_{j\in[\![1:d]\!]}$ are the dual coordinate-k norms associated with the source norm $|||\cdot|||$, whose expressions are given in Definition 14. 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 3 (from [3], Proposition 4.7 and [1], Proposition 6) *Let $|||\cdot|||$ be the source norm employed for the Capra coupling ç in Definition 1. Let $\{|||\cdot|||^{\mathcal{R}}_{(j)}\}_{j\in[\![1:d]\!]}$ and $\{|||\cdot|||^{\mathcal{R}}_{(j),\star}\}_{j\in[\![1:d]\!]}$ be the associated sequences of coordinate-k and dual coordinate-k norms, as in Definition 14, and let $\{\mathbb{B}^{\mathcal{R}}_{(j)}\}_{j\in[\![1:d]\!]}$ and $\{\mathbb{B}^{\mathcal{R}}_{(j),\star}\}_{j\in[\![1:d]\!]}$ 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 $x = 0$,*

$$\partial_{\text{ç}} \ell_0(0) = \bigcap_{j\in[\![1,d]\!]} j\mathbb{B}^{\mathcal{R}}_{(j),\star} , \tag{9a}$$

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

$$\partial_{\text{ç}} \ell_0(x) = \mathcal{N}_{\mathbb{B}^{\mathcal{R}}_{(l)}}\left(\frac{x}{|||x|||^{\mathcal{R}}_{(l)}}\right) \cap Y_l . \tag{9b}$$

3 Capra-subdifferential of ℓ_0 for the ℓ_p source norms

The main contribution of this article is the following Theorem 4. It provides explicit formulas for the Capra-subdifferential of the ℓ_0 pseudonorm, as introduced in (4) and as characterized in Theorem 2 for the ℓ_p source norms $|||\cdot||| = \|\cdot\|_p$, when $p \in [1,\infty]$.

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

$$\|y\|^{\text{tn}}_{(k,q)} = \left(\sum_{i=1}^{k} |y_{\nu(i)}|^q \right)^{\frac{1}{q}} , \quad \text{if} \quad q \in [1,\infty[\ , \quad \text{and} \quad \|y\|^{\text{tn}}_{(k,\infty)} = \|y\|_\infty , \tag{10}$$

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

$$\mathcal{R}_K = \mathbb{R}^K \times \{0\}^{-K} = \left\{ x \in \mathbb{R}^d \,\middle|\, x_j = 0 \ , \ \forall j \notin K \right\} \subseteq \mathbb{R}^d , \tag{11}$$

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

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

Theorem 4 *Let the source norm $|||\cdot||| = \|\cdot\|_p$, where $p \in [1, \infty]$.*

• *If $p = 1$, the ℓ_0 pseudonorm is not* Capra*-convex, and only* Capra*-subdifferentiable at $x = 0$. Its* Capra*-subdifferential is given by*

$$\partial_{\acute{\varsigma}} \ell_0(0) = \mathbb{B}_{\|\cdot\|_\infty} \quad \text{and} \quad \partial_{\acute{\varsigma}} \ell_0(x) = \emptyset \ , \ \ \forall x \in \mathbb{R}^d \setminus \{0\} \ , \tag{12}$$

and its Capra*-biconjugate is*

$$\ell_0^{\acute{\varsigma}\acute{\varsigma}'} \ : \ x \mapsto \begin{cases} 0 \ , & \text{if } u = 0 \ , \\ 1 \ , & \text{if } u \neq 0 \ . \end{cases} \tag{13}$$

• *If $p \in]1, \infty[$, the ℓ_0 pseudonorm is* Capra*-convex and* Capra*-subdifferentiable everywhere on $\mathbb{R}^d$. Its* Capra*-subdifferential is given by*

$$\partial_{\acute{\varsigma}} \ell_0(0) = \mathbb{B}_{\|\cdot\|_\infty} \ , \tag{14a}$$

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

$$y \in \partial_{\acute{\varsigma}} \ell_0(x) \iff \begin{cases} y_L \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_p}}(\frac{x}{\|x\|_p}) \ , \\ |y_j| \leq \min_{i \in L} |y_i| \ , \ \ \forall j \notin L \ , \\ |y_{\nu(k+1)}|^q \geq \big(\|y\|_{(k,q)}^{\mathrm{tn}} + 1\big)^q - \big(\|y\|_{(k,q)}^{\mathrm{tn}}\big)^q \ , \ \ \forall k \in [\![0, l-1]\!] \ , \\ |y_{\nu(l+1)}|^q \leq \big(\|y\|_{(l,q)}^{\mathrm{tn}} + 1\big)^q - \big(\|y\|_{(l,q)}^{\mathrm{tn}}\big)^q \ , \end{cases} \tag{14b}$$

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

• *If $p = \infty$, the ℓ_0 pseudonorm is not* Capra*-convex, and not* Capra*-subdifferentiable everywhere on $\mathbb{R}^d$. Its* Capra*-subdifferential is given by*

$$\partial_{\acute{\varsigma}} \ell_0(0) = \mathbb{B}_{\|\cdot\|_\infty} \ , \tag{15a}$$

and at $x \neq 0$, denoting $l = \ell_0(x)$, $L = supp(x)$, by

$$y \in \partial_{\acute{\varsigma}} \ell_0(x) \iff \begin{cases} y_L \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_\infty}}(\frac{x}{\|x\|_\infty}) \ , \\ |y_j| \leq \min_{i \in L} |y_i| \ , \ \ \forall j \notin L \ , \\ |y_{\nu(k+1)}| \geq 1 \ , \ \ \forall k \in [\![0, l-1]\!] \ , \\ |y_{\nu(l+1)}| \leq 1 \ , \end{cases} \tag{15b}$$

where, for any $y \in \mathbb{R}^d$, ν denotes a permutation of $[\![1, d]\!]$ such that $|y_{\nu(1)}| \geq \ldots \geq |y_{\nu(d)}|$, and its Capra*-biconjugate is*

$$\ell_0^{\acute{\varsigma}\acute{\varsigma}'} \ : \ x \mapsto \begin{cases} 0 \ , & \text{if } u = 0 \ , \\ \frac{\|x\|_1}{\|x\|_\infty} \ , & \text{if } u \neq 0 \ . \end{cases} \tag{16}$$

We proceed in four steps to prove Theorem 4, starting from the generic formulation of the Capra-subdifferential of ℓ_0 in (9). First, in §3.1, we provide an explicit formulation for the set Y_l in (9b), defined in (8). Second, in §3.2, we provide an explicit formulation for the normal cone $\mathcal{N}_{\mathbb{B}^{\mathcal{R}}_{(l)}}$ in (9b). Third, in §3.3 we gather both results to provide an explicit formulation of the Capra-subdifferential of the ℓ_0 pseudonorm. In addition, we comment on the Capra-convexity of ℓ_0 for the ℓ_p source norms, when $p \in \{1, \infty\}$. Finally, in §3.4, we wrap up the proof Theorem 4.

We will need the following properties of the coordinate-k and dual coordinate-k norms of Definition 14.

Proposition 5 (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 14 are given, for $k \in [\![1, d]\!]$, by*

$$|||\cdot|||^{\mathcal{R}}_{(k),\star} = \|\cdot\|^{\mathrm{tn}}_{(k,q)} \quad \textit{and} \quad |||\cdot|||^{\mathcal{R}}_{(k)} = \|\cdot\|^{\mathrm{sn}}_{(p,k)} \ , \tag{17}$$

3.1 Explicit formulation of the sets Y_l

We derive explicit formulations of the sets Y_l in (8) 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 $\|\cdot\|^{\mathrm{tn}}_{(k,q)}$, whose expression is given in (10). We state our first preliminary result in Lemma 6.

Lemma 6 *Let $y \in \mathbb{R}^d$, $q \in [1, \infty[$ and $k \in [\![0, d-1]\!]$. We have that*

$$\|y\|^{\mathrm{tn}}_{(k+1,q)} - \|y\|^{\mathrm{tn}}_{(k,q)} \leq 1 \implies \|y\|^{\mathrm{tn}}_{(k+j,q)} - \|y\|^{\mathrm{tn}}_{(k,q)} \leq j \ , \ \ \forall j \in [\![1, d-k]\!] \ . \tag{18}$$

Moreover, the same result holds if inequalities are strict in (18).

Proof. Let $y \in \mathbb{R}^d$ and ν denote a permutation of $[\![1, d]\!]$ such that $|y_{\nu(1)}| \geq \ldots \geq |y_{\nu(d)}|$, let be $q \in [1, \infty[$, $k \in [\![0, d-1]\!]$ and $j \in [\![1, d-k]\!]$. We denote

$$y^{\Sigma}_{k,q} = \sum_{i=1}^{k} |y_{\nu(i)}|^q \ , \tag{19}$$

so that, from Proposition 5, we have that $\|y\|^{\mathrm{tn}}_{(k,q)} = \big(y^{\Sigma}_{k,q}\big)^{\frac{1}{q}}$.

First, we prove the inequality

$$\big(y^{\Sigma}_{k,q} + j|y_{\nu(k+1)}|^q\big)^{\frac{1}{q}} - \big(y^{\Sigma}_{k,q}\big)^{\frac{1}{q}} \leq j\Big[\big(y^{\Sigma}_{k,q} + |y_{\nu(k+1)}|^q\big)^{\frac{1}{q}} - \big(y^{\Sigma}_{k,q}\big)^{\frac{1}{q}}\Big] \ . \tag{20}$$

Indeed, we have that

$$\frac{1}{j}\big(y^{\Sigma}_{k,q} + j|y_{\nu(k+1)}|^q\big)^{\frac{1}{q}} + \big(1 - \frac{1}{j}\big)\big(y^{\Sigma}_{k,q}\big)^{\frac{1}{q}} \leq \Big(\frac{1}{j}\big(y^{\Sigma}_{k,q} + j|y_{\nu(k+1)}|^q\big) + \big(1 - \frac{1}{j}\big)y^{\Sigma}_{k,q}\Big)^{\frac{1}{q}} \ ,$$

by concavity of the function $x \mapsto x^{\frac{1}{q}}$ (as $q \geq 1$),

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

Second, we prove the implication in (18) in its nonstrict inequality version. Let us assume that $\|y\|_{(k+1,q)}^{\mathrm{tn}} - \|y\|_{(k,q)}^{\mathrm{tn}} \leq 1$. Then, we have that

$$\begin{aligned}
\|y\|_{(k+j,q)}^{\mathrm{tn}} - \|y\|_{(k,q)}^{\mathrm{tn}} &\leq \left(y_{k,q}^{\Sigma} + j|y_{\nu(k+1)}|^q\right)^{\frac{1}{q}} - \left(y_{k,q}^{\Sigma}\right)^{\frac{1}{q}} \,, \\
&\qquad \text{(from (19) and } |y_{\nu(k+1)}| \geq |y_{\nu(k+2)}| \geq \ldots \geq |y_{\nu(k+j)}|\text{)} \\
&\leq j\Big[\left(y_{k,q}^{\Sigma} + |y_{\nu(k+1)}|^q\right)^{\frac{1}{q}} - \left(y_{k,q}^{\Sigma}\right)^{\frac{1}{q}}\Big] \,, \qquad \text{(from (20))} \\
&= j\Big[\|y\|_{(k+1,q)}^{\mathrm{tn}} - \|y\|_{(k,q)}^{\mathrm{tn}}\Big] \,,
\end{aligned}$$

from the expression of $\|\cdot\|_{(k,q)}^{\mathrm{tn}}$ in (10) and by (19)

$$\leq j \,. \qquad \text{(by assumption } \|y\|_{(k+1,q)}^{\mathrm{tn}} - \|y\|_{(k,q)}^{\mathrm{tn}} \leq 1\text{)}$$

The proof of the strict inequality version of (18) is analogous. □

We state our second preliminary result in Lemma 7.

Lemma 7 *Let $y \in \mathbb{R}^d$, $q \in [1, \infty[$ and $k \in [\![0, d-1]\!]$. We have that*

$$\|y\|_{(k+1,q)}^{\mathrm{tn}} - 1 \leq \|y\|_{(k,q)}^{\mathrm{tn}} \iff |y_{\nu(k+1)}|^q \leq \left(\|y\|_{(k,q)}^{\mathrm{tn}} + 1\right)^q - \left(\|y\|_{(k,q)}^{\mathrm{tn}}\right)^q \,. \tag{21}$$

Moreover, the same result holds if inequalities are strict or replaced with equalities in (21).

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

$$\begin{aligned}
\|y\|_{(k+1,q)}^{\mathrm{tn}} - 1 \leq \|y\|_{(k,q)}^{\mathrm{tn}} &\iff \Big(\sum_{i=1}^{k} |y_{\nu(i)}|^q + |y_{\nu(k+1)}|^q\Big)^{\frac{1}{q}} - 1 \leq \|y\|_{(k,q)}^{\mathrm{tn}} \,, \\
&\qquad \text{(from the expression of } \|\cdot\|_{(k,q)}^{\mathrm{tn}} \text{ in (10))} \\
&\iff \sum_{i=1}^{k} |y_{\nu(i)}|^q + |y_{\nu(k+1)}|^q \leq \left(\|y\|_{(k,q)}^{\mathrm{tn}} + 1\right)^q \,, \\
&\qquad \text{(as the function } x \mapsto x^q \text{ is nondecreasing on } \mathbb{R}_+\text{)}
\end{aligned}$$

so that finally, by definition (10) of $\|\cdot\|_{(k,q)}^{\mathrm{tn}}$, we get

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

The proof of the strict inequality and equality versions of (18) is analogous. □

We now provide explicit formulations of the sets Y_l in (8) for the ℓ_p source norms $|||\cdot||| = \|\cdot\|_p$, when $p \in [1, \infty]$.

Proposition 8 *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 $l \in [\![0,d]\!]$, let the set Y_l be as in (8). We have that*

- *if $p=1$,*

$$Y_l = \begin{cases} \mathbb{R}^d & \text{if } l=0\ , \\ \emptyset & \text{else,} \end{cases} \tag{22a}$$

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

$$y \in Y_l \iff \begin{cases} |y_{\nu(k+1)}|^q \geq \left(\|y\|^{\mathrm{tn}}_{(k,q)}+1\right)^q - \left(\|y\|^{\mathrm{tn}}_{(k,q)}\right)^q\ , \ \ \forall k \in [\![0,l-1]\!]\ , \\ |y_{\nu(l+1)}|^q \leq \left(\|y\|^{\mathrm{tn}}_{(l,q)}+1\right)^q - \left(\|y\|^{\mathrm{tn}}_{(l,q)}\right)^q\ . \end{cases} \tag{22b}$$

Proof. We consider the case $p=1$. When the source norm is $|||\cdot||| = \|\cdot\|_1$, we have, for $k \in [\![1,d]\!]$, $|||\cdot|||^{\mathcal{R}}_{(k),\star} = \|\cdot\|_\infty$ [3, Table 1]. Therefore, from the expression of Y_l in (8), we get that

$$Y_l = \left\{ y \in \mathbb{R}^d \,\middle|\, l \in \operatorname*{arg\,max}_{j \in [\![0,d]\!]} \left(\|y\|_\infty - j\right) \right\} = \begin{cases} \mathbb{R}^d & \text{if } l=0\ , \\ \emptyset & \text{else,} \end{cases}$$

hence (22a).

Next, we consider $p \in]1,\infty]$, and proceed in two steps to prove the equivalence in (22b).

In the first step ($\Longleftarrow$), we take $y \in \mathbb{R}^d$ and we consider the two following cases.

- If $|y_{\nu(k+1)}|^q \geq \left(\|y\|^{\mathrm{tn}}_{(k,q)}+1\right)^q - \left(\|y\|^{\mathrm{tn}}_{(k,q)}\right)^q\ , \ \ \forall k \in [\![0,l-1]\!]\ ,$

then, we get that

$$\begin{aligned} & \|y\|^{\mathrm{tn}}_{(k+1,q)} - 1 \geq \|y\|^{\mathrm{tn}}_{(k,q)}\ , \ \ \forall k \in [\![0,l-1]\!]\ , && \text{(from (21))} \\ \Longrightarrow\ & \|y\|^{\mathrm{tn}}_{(k+1,q)} - (k+1) \geq \|y\|^{\mathrm{tn}}_{(k,q)} - k\ , \ \ \forall k \in [\![0,l-1]\!]\ , \\ \Longrightarrow\ & l \in \operatorname*{arg\,max}_{j \in [\![0,l]\!]} \left(\|y\|^{\mathrm{tn}}_{(j,q)} - j\right)\ . \end{aligned}$$

- If $|y_{\nu(l+1)}|^q \leq \left(\|y\|^{\mathrm{tn}}_{(l,q)}+1\right)^q - \left(\|y\|^{\mathrm{tn}}_{(l,q)}\right)^q\ ,$

then, we get that

$$\begin{aligned} & \|y\|^{\mathrm{tn}}_{(l+1,q)} - 1 \leq \|y\|^{\mathrm{tn}}_{(l,q)}\ , && \text{(from (21))} \\ \Longrightarrow\ & \|y\|^{\mathrm{tn}}_{(l+j,q)} - j \leq \|y\|^{\mathrm{tn}}_{(l,q)}\ , \ \ \forall j \in [\![1,d-l]\!]\ , && \text{(from (18))} \\ \Longrightarrow\ & \|y\|^{\mathrm{tn}}_{(l+j,q)} - (l+j) \leq \|y\|^{\mathrm{tn}}_{(l,q)} - l\ , \ \ \forall j \in [\![1,d-l]\!]\ , \\ \Longrightarrow\ & l \in \operatorname*{arg\,max}_{j \in [\![l,d]\!]} \left(\|y\|^{\mathrm{tn}}_{(j,q)} - j\right)\ . \end{aligned}$$

Therefore, $y \in Y_l$, which concludes the first step.

In the second step ($\Longrightarrow$), we proceed by contraposition, assuming that either one of the two assumptions above is not satisfied.

• If $\exists k \in [\![0, l-1]\!]\ ,\ \ |y_{\nu(k+1)}|^q < \big(\|y\|^{\mathrm{tn}}_{(k,q)} + 1\big)^q - \big(\|y\|^{\mathrm{tn}}_{(k,q)}\big)^q\ ,$

then, we get that

$$\begin{aligned} & \exists k \in [\![0, l-1]\!]\ ,\ \ \|y\|^{\mathrm{tn}}_{(k+1,q)} - 1 < \|y\|^{\mathrm{tn}}_{(k,q)}\ , && \text{(from (21))} \\ \implies\ & \exists k \in [\![0, l-1]\!]\ ,\ \ \|y\|^{\mathrm{tn}}_{(k+1,q)} - (k+j) < \|y\|^{\mathrm{tn}}_{(k,q)} - k\ ,\ \ \forall j \in [\![1, d-k]\!]\ , && \text{(from (18))} \\ & \qquad \implies\ l \notin \operatorname*{arg\,max}_{j \in [\![0,l]\!]} \big(\|y\|^{\mathrm{tn}}_{(j,q)} - j\big)\ ; \end{aligned}$$

• If $|y_{\nu(l+1)}|^q > \big(\|y\|^{\mathrm{tn}}_{(l,q)} + 1\big)^q - \big(\|y\|^{\mathrm{tn}}_{(l,q)}\big)^q\ ,$

then, we get that

$$\begin{aligned} & \|y\|^{\mathrm{tn}}_{(l+1,q)} - 1 > \|y\|^{\mathrm{tn}}_{(l,q)}\ , && \text{(from (21))} \\ \implies\ & \|y\|^{\mathrm{tn}}_{(l+1,q)} - (l+1) > \|y\|^{\mathrm{tn}}_{(l,q)} - l\ , \\ & \qquad \implies\ l \notin \operatorname*{arg\,max}_{j \in [\![l,d]\!]} \big(\|y\|^{\mathrm{tn}}_{(j,q)} - j\big)\ . \end{aligned}$$

In either case, $y \notin Y_l$, which concludes the second step, and finally proves the equivalence in (22b). □

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

We turn to the explicit formulation of the normal cone $\mathcal{N}_{\mathbb{B}^{\mathrm{sn}}_{(p,l)}}$ in (9b) for the ℓ_p source norms $\|\!|\cdot|\!\| = \|\cdot\|_p$, when $p \in [1, \infty]$. We start with the following Lemma 9.

Lemma 9 *Let the source norm be the ℓ_p norm $\|\!|\cdot|\!\| = \|\cdot\|_p$, where $p \in [1, \infty]$. Let $x \in \mathbb{R}^d$, $l = \ell_0(x)$, $L = supp(x)$. If $l \in [\![1, d]\!]$, we have that $x' = \frac{x}{\|x\|^{\mathrm{sn}}_{(p,l)}}$ satisfies*

$$\|x'\|_p = 1\ , \tag{23a}$$

$$y \in \mathcal{N}_{\mathbb{B}^{\mathrm{sn}}_{(p,l)}}(x') \iff \|y\|^{\mathrm{tn}}_{(l,q)} = \langle x',\, y_L \rangle\ , \tag{23b}$$

$$y \in \mathcal{N}_{\mathbb{B}^{\mathrm{sn}}_{(p,l)}}(x') \implies \|y\|^{\mathrm{tn}}_{(l,q)} \leq \|y_L\|_q\ . \tag{23c}$$

Proof. Let $q \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$.

First, we prove (23a). From [3, Proposition 3.5], $l \geq 1$ and $\ell_0(x') = l$ give that $\|\!|x'|\!\| = \|\!|x'|\!\|^{\mathcal{R}}_{(l)}$. Thus, from Proposition 5 we deduce that $\|x'\|_p = \|x'\|^{\mathrm{sn}}_{(p,l)} = 1$.

Second, we prove (23b). We have the equivalence

$$\begin{aligned} y \in \mathcal{N}_{\mathbb{B}^{\mathrm{sn}}_{(p,l)}}(x') & \iff \big\|x'\big\|^{\mathrm{sn}}_{(p,l)} \|y\|^{\mathrm{tn}}_{(l,q)} = \big\langle x',\, y \big\rangle && \text{(by definition (37) of the normal cone)} \\ & \iff \|y\|^{\mathrm{tn}}_{(l,q)} = \big\langle x',\, y_L \big\rangle\ . && \text{(from } \|x'\|^{\mathrm{sn}}_{(p,l)} = 1 \text{ and } L = \mathrm{supp}(x')) \end{aligned}$$

Third, we prove (23c). We have that

$$\begin{aligned} y \in \mathcal{N}_{\mathbb{B}^{\mathrm{sn}}_{(p,l)}}(x') & \iff \|y\|^{\mathrm{tn}}_{(l,q)} = \big\langle x',\, y_L \big\rangle_{\mathbb{R}^l}\ , && \text{(from (23b))} \\ & \implies \|y\|^{\mathrm{tn}}_{(l,q)} \leq \|y_L\|_q\ . && \text{(from the Hölder inequality and (23a))} \end{aligned}$$

□

We now provide an explicit expression of the normal cone in (9b) for the ℓ_p source norms $|||\cdot||| = \|\cdot\|_p$, when $p \in [1, \infty]$.

Proposition 10 *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 $L = supp(x)$. If $l \in [\![1, d]\!]$, we have that*

$$y \in \mathcal{N}_{\mathbb{B}^{\mathrm{sn}}_{(p,l)}}\Big(\frac{x}{\|x\|^{\mathrm{sn}}_{(p,l)}}\Big) \iff \begin{cases} y_L \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_p}}(\frac{x}{\|x\|_p}) \,, \\ |y_j| \leq \min_{i \in L} |y_i| \,, \ \ \forall j \notin L \,. \end{cases} \tag{24}$$

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 $L = \mathrm{supp}(x)$, and let us set $x' = \frac{x}{\|x\|^{\mathrm{sn}}_{(p,l)}}$. Let $y \in \mathbb{R}^d$, and let us set $I = \mathrm{supp}(y)$.

First, we prove that

$$y \in \mathcal{N}_{\mathbb{B}^{\mathrm{sn}}_{(p,l)}}(x') \implies \|y\|^{\mathrm{tn}}_{(l,q)} = \|y_L\|_q \,. \tag{25}$$

We consider two cases. In the first case, we assume $|I| = |\mathrm{supp}(y)| \leq |\mathrm{supp}(x)| = |L| = l$. Since y has at most l nonzero coordinates, from the expression (10) of $\|\cdot\|^{\mathrm{tn}}_{(l,q)}$, we get that $\|y\|^{\mathrm{tn}}_{(l,q)} = \|y_I\|_q$. It follows that

$$\begin{aligned} y \in \mathcal{N}_{\mathbb{B}^{\mathrm{sn}}_{(p,l)}}(x') &\implies \|y_I\|_q \leq \|y_L\|_q \,, && \text{(from (23c))} \\ &\implies \|y_L\|_q = \|y_I\|_q = \|y\|^{\mathrm{tn}}_{(l,q)} \,. \\ &\qquad \text{(from } \|y_L\|_q \leq \|y\|_q = \|y_I\|_q \text{, because } |I| = |\mathrm{supp}(y)|) \end{aligned}$$

In the second case, we assume that $|I| = |\mathrm{supp}(y)| > |\mathrm{supp}(x)| = |L| = l$. Since y has more than l nonzero coordinates, from the expression (10) of $\|\cdot\|^{\mathrm{tn}}_{(l,q)}$, we get that $\|y\|^{\mathrm{tn}}_{(l,q)} \geq \|y_L\|_q$. Combined with (23c), we have that $y \in \mathcal{N}_{\mathbb{B}^{\mathrm{sn}}_{(p,l)}}(x') \implies \|y\|^{\mathrm{tn}}_{(l,q)} = \|y_L\|_q$. Gathering the conclusions of both cases, we obtain (25).

Second, we prove (24). Observing that $\|x'\|_p = 1$ from (23a), we have that

$$\begin{aligned} y \in \mathcal{N}_{\mathbb{B}^{\mathrm{sn}}_{(p,l)}}(x') &\iff \begin{cases} \|x'\|_p \|y_L\|_q = \langle x', \, y_L \rangle \,, \\ \|y\|^{\mathrm{tn}}_{(l,q)} = \|y_L\|_q \,, \end{cases} && (\implies \text{from (23b), (25)}; \ \impliedby \text{from (23b)}) \\ &\iff \begin{cases} y_L \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_p}}(\frac{x}{\|x\|_p}) \,, \\ |y_j| \leq \min_{i \in L} |y_i| \,, \ \ \forall j \notin L \,, \end{cases} \end{aligned}$$

by definition (37) of the normal cone, observing that $x' = \frac{x}{\|x\|_p}$ from (23a), and by the expression of of $\|\cdot\|^{\mathrm{tn}}_{(l,q)}$ in Proposition 5. This ends the proof. □

3.3 Capra-subdifferential and Capra-convexity of the pseudonorm ℓ_0

We gather the explicit expressions of the Capra-subdifferential of the pseudonorm ℓ_0, for the ℓ_p source norms $\|\!|\cdot|\!\| = \|\cdot\|_p$, when $p \in [1,\infty]$.

Proposition 11 *Let the source norm $\|\!|\cdot|\!\| = \|\cdot\|_p$, where $p \in [1,\infty]$, let $q \in [1,\infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$, and let the associated* Capra *coupling $\dot{\varsigma}$ be as in Definition 1.*

- *If $p = 1$, $\mathrm{dom}\big(\partial_{\dot{\varsigma}}\ell_0\big) = \{0\}$ and $\partial_{\dot{\varsigma}}\ell_0$ is given by* (12).
- *If $p \in]1,\infty[$, $\mathrm{dom}\big(\partial_{\dot{\varsigma}}\ell_0\big) = \mathbb{R}^d$, and $\partial_{\dot{\varsigma}}\ell_0$ is given by* (14a)–(14b).
- *If $p = \infty$, $\{0\} \subsetneq \mathrm{dom}\big(\partial_{\dot{\varsigma}}\ell_0\big) \subsetneq \mathbb{R}^d$, and $\partial_{\dot{\varsigma}}\ell_0$ is given by* (15a)–(15b).

Proof.

First, we prove that $\partial_{\dot{\varsigma}}\ell_0(0) = \mathbb{B}_{\|\cdot\|_\infty}$ for all values of $p \in [1,\infty]$. Let us recall that, from (9a), $\partial_{\dot{\varsigma}}\ell_0(0) = \bigcap_{j\in[\![1,d]\!]} j\mathbb{B}^{\mathcal{R}}_{(j),\star}$. If $p = 1$, from (17) we get that $\|\!|\cdot|\!\|^{\mathcal{R}}_{(j),\star} = \|\cdot\|_\infty$, $\forall j \in [\![1,d]\!]$. We deduce that $\partial_{\dot{\varsigma}}\ell_0(0) = \mathbb{B}_{\|\cdot\|_\infty}$. We now assume that $p \in]1,\infty]$. From (17), we get that $\|\!|\cdot|\!\|^{\mathcal{R}}_{(j),\star} = \|\cdot\|^{\mathrm{tn}}_{(j,q)}$, $\forall j \in [\![1,d]\!]$, where $q \in [1,\infty[$. For $j = 1$, from (10), we get that $\|\cdot\|^{\mathrm{tn}}_{(1,q)} = \|\cdot\|_\infty$, hence that $\mathbb{B}^{\mathrm{tn}}_{(1,q)} = \mathbb{B}_{\|\cdot\|_\infty}$. Letting $j > 1$, we prove an inclusion:

$$
\begin{aligned}
y \in \mathbb{B}_{\|\cdot\|_\infty} &\implies |y_{\nu(1)}|^q \le 1 \ , && \text{(where } |y_{\nu(1)}| = \|y\|_\infty\text{)}\\
&\implies \sum_{i=1}^{j} |y_{\nu(1)}|^q \le \sum_{i=1}^{j} 1 = j \ , && \text{(where } |y_{\nu(1)}| \ge \ldots \ge |y_{\nu(d)}|\text{)}\\
&\implies \Big(\sum_{i=1}^{j} |y_{\nu(1)}|^q\Big)^{\frac{1}{q}} \le j^{\frac{1}{q}} \ ,\\
&\implies \|y\|^{\mathrm{tn}}_{(j,q)} \le j && \text{(by definition of } \|\cdot\|^{\mathrm{tn}}_{(j,q)} \text{ in (10) and from } j \ge j^{\frac{1}{q}}\text{)}\\
&\implies y \in j\mathbb{B}^{\mathrm{tn}}_{(j,q)} \ ,
\end{aligned}
$$

so that $\mathbb{B}_{\|\cdot\|_\infty} \subseteq j\mathbb{B}^{\mathrm{tn}}_{(j,q)}$. We conclude that $\partial_{\dot{\varsigma}}\ell_0(0) = \bigcap_{j\in[\![1,d]\!]} \mathbb{B}^{\mathrm{tn}}_{(j,q)} = \mathbb{B}_{\|\cdot\|_\infty} \cap \Big(\bigcap_{j\in[\![2,d]\!]} \mathbb{B}^{\mathrm{tn}}_{(j,q)}\Big) = \mathbb{B}_{\|\cdot\|_\infty}$.

Second, we prove the expression of $\partial_{\dot{\varsigma}}\ell_0(x)$ at $x \neq 0$ for all values of $p \in [1,\infty]$. Let us recall that, from (9b), $\partial_{\dot{\varsigma}}\ell_0(x) = \mathcal{N}_{\mathbb{B}^{\mathcal{R}}_{(l)}}\big(\frac{x}{\|\!|x|\!\|^{\mathcal{R}}_{(l)}}\big) \cap Y_l$. If $p = 1$, from (22a), $Y_l = \emptyset$, hence $\partial_{\dot{\varsigma}}\ell_0(x) = \emptyset$, as stated in (12). If $p \in]1,\infty]$, the expressions of $\partial_{\dot{\varsigma}}\ell_0(x)$ in (14b) and in (15b) are obtained combining Proposition 10 and Proposition 8.

Third, we prove our claims on $\mathrm{dom}\big(\partial_{\dot{\varsigma}}\ell_0\big)$ for all values of $p \in [1,\infty]$. The case $p = 1$ is clear from (12). If $p \in]1,\infty[$, the norm $\|\cdot\|_p$ and the dual norm $\|\cdot\|_q$ are orthant-strictly monotonic, following Definition 12, so that ℓ_0 is Capra-subdifferentiable on $\mathbb{R}^d$, from Theorem 2. We now consider the case $p = \infty$. Let us for example take $x \in \mathbb{R}^d$ defined, for some $i \in [\![1,d]\!]$ and $\varepsilon \in]0,1[$, by

$$
\begin{cases} x_j = 1 \ , \ \ \forall j \neq i \ , \\ x_i = \varepsilon \ . \end{cases} \tag{26}
$$

We prove that $\partial_{\dot{\complement}}\ell_0(x) = \emptyset$, by contradiction. Let $y \in \partial_{\dot{\complement}}\ell_0(x)$. From (15b), $y \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_\infty}}(\frac{x}{\|x\|_\infty})$, so that $\langle x' - x,\, y\rangle \leq 0$ for any $x' \in \mathbb{B}_{\|\cdot\|_\infty}$. In particular, for $j \neq i$, taking $x' = x - e_j$, where $\{e_k\}_{k\in[\![1,d]\!]}$ is the canonical basis of $\mathbb{R}^d$, we obtain that $y_j \geq 0$. Moreover,

$$
\begin{aligned}
& y \in \mathcal{N}_{\mathbb{B}_{\|\cdot\|_\infty}}(\frac{x}{\|x\|_\infty}) \\
& \quad \implies \|x\|_\infty \|y\|_1 = \langle x,\, y\rangle \ , && \text{(by (38))} \\
& \quad \implies \sum_{j\neq i} y_j + |y_i| = \sum_{j\neq i} y_j + \varepsilon y_i \ , && \text{(from the definition (26) of } x\text{)} \\
& \quad \implies y_i = 0 \ . && \text{(from } \varepsilon \in]0,1[\text{)}
\end{aligned}
$$

However, since $\ell_0(x) = d$, we deduce from (15b) that we must have $y_i \geq 1$. We arrive at a contradiction, and conclude that $\partial_{\dot{\complement}}\ell_0(x) = \emptyset$, and therefore that $\mathrm{dom}\big(\partial_{\dot{\complement}}\ell_0\big) \subsetneq \mathbb{R}^d$. Finally, if $\varepsilon = 1$ in (26), it is straightforward to check in (15b) that $(1,\dots,1) \in \partial_{\dot{\complement}}\ell_0(x)$, so that $\{0\} \subsetneq \mathrm{dom}\big(\partial_{\dot{\complement}}\ell_0\big)$. This ends the proof. □

3.4 Proof of Theorem 4

Proof. Results about Capra-subdifferentiability are taken from Proposition 11, and the Capra-convexity of ℓ_0 for $p \in]1,\infty[$ is taken from Theorem 2. Thus, we only need to prove that ℓ_0 is not Capra-convex for $p \in \{1,\infty\}$ by proving the expressions of the $\dot{\complement}$-biconjugate in (13) and (16). To this end, we recall that, from [3, Proposition 4.4] and Proposition 5, if $q \in \{1,\infty\}$ is such that $\frac{1}{p} + \frac{1}{q} = 1$, then we get that

$$
\ell_0^{\dot{\complement}}(y) = \max_{j=1,\dots,d} \big(\|y\|^{\mathrm{tn}}_{(j,q)} - j\big)^+ \ , \ \ \forall y \in \mathbb{R}^d \ . \tag{27}
$$

First, we consider $p = 1$. From (27) and (10), we have that

$$
\begin{aligned}
\ell_0^{\dot{\complement}}(y) &= \max_{j\in[\![1,d]\!]} \big(\|y\|_\infty - j\big)^+ = \big(\|y\|_\infty - 1\big)^+ \ , \ \ \forall y \in \mathbb{R}^d \ , \text{ and thus,} \\
\ell_0^{\dot{\complement}\dot{\complement}'}(x) &= \sup_{y\in\mathbb{R}^d} \Big(\frac{\langle x,\, y\rangle}{\|x\|_1} - \big(\|y\|_\infty - 1\big)^+ \Big) \ , \ \ \forall x \in \mathbb{R}^d \ , && \text{(from (3b))} \\
&= \max\Big(\sup_{\|y\|_\infty \leq 1} \frac{\langle x,\, y\rangle}{\|x\|_1} \ , \ 1 + \sup_{\|y\|_\infty \geq 1} \frac{\langle x,\, y\rangle}{\|x\|_1} - \|y\|_\infty \Big) \ , \ \ \forall x \in \mathbb{R}^d \ , \\
&= 1 \ , \ \ \forall x \in \mathbb{R}^d \setminus \{0\} \ ,
\end{aligned}
$$

since $\sup_{\|y\|_\infty\leq 1} \langle x,\, y\rangle = \|x\|_1$, by $\|\cdot\|_1 = \big(\|\cdot\|_\infty\big)^*$, and $\langle x,\, y\rangle \leq \|x\|_1 \|y\|_\infty$, by Hölder's inequality. This proves (13).

Second, we consider $p = \infty$. From (27) and (10), for $y \in \mathbb{R}^d$ and ν a permutation of $[\![1,d]\!]$ such

that $|y_{\nu(1)}| \geq \ldots \geq |y_{\nu(d)}|$, we have that

$$\ell_0^{\dot{\complement}}(y) = \max_{j\in[\![1,d]\!]} \Big(\sum_{k=1}^{j} |y_{\nu(k)}| - j\Big)^+ = \sum_{k=1}^{d} (|y_{\nu(k)}| - 1)\mathbf{1}_{|y_{\nu(k)}|\geq 1} \ , \ \ \forall y \in \mathbb{R}^d \ , \text{ and thus,}$$

$$\begin{aligned}
\ell_0^{\dot{\complement}\dot{\complement}'}(x) &= \sup_{y\in\mathbb{R}^d} \Big(\frac{\langle x,\, y\rangle}{\|x\|_\infty} - \sum_{k=1}^{d} (|y_{\nu(k)}| - 1)\mathbf{1}_{|y_{\nu(k)}|\geq 1}\Big) \ , \ \ \forall x \in \mathbb{R}^d \ , && \text{(from (3b))} \\
&= \sum_{k=1}^{d} \sup_{y_k\in\mathbb{R}} \Big(\frac{x_k y_k}{\|x\|_\infty} - (|y_k| - 1)\mathbf{1}_{|y_k|\geq 1}\Big) \ , \ \ \forall x \in \mathbb{R}^d \ , \\
&= \sum_{k=1}^{d} \max\Big(\sup_{|y_k|\leq 1} \frac{x_k y_k}{\|x\|_\infty} \ , \ \ 1 + \sup_{|y_k|\geq 1} \frac{x_k y_k}{\|x\|_\infty} - |y_k|\Big) \ , \ \ \forall x \in \mathbb{R}^d \ , \\
&= \sum_{k=1}^{d} \frac{|x_k|}{\|x\|_\infty} = \frac{\|x\|_1}{\|x\|_\infty} \ , \ \ \forall x \in \mathbb{R}^d \ ,
\end{aligned}$$

since, using similar arguments as above, $\sup_{|y_k|\leq 1} x_k y_k = |x_k|$, and $1 + \sup_{|y_k|\geq 1} \frac{x_k y_k}{\|x\|_\infty} - |y_k| \leq 1 + \sup_{|y_k|\geq 1} \frac{|x_k y_k|}{\|x\|_\infty} - |y_k| = 1 + \sup_{|y_k|\geq 1} \big(\frac{|x_k|}{\|x\|_\infty} - 1\big)|y_k| = \frac{|x_k|}{\|x\|_\infty}$. This proves (16), and ends the proof. □

4 Graphical visualizations 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 $\partial_{\dot{\complement}}\ell_0$ 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 Proposition 11, we have that

$$\partial_{\dot{\complement}}\ell_0(0) = \mathbb{B}_{\|\cdot\|_\infty} \ , \tag{28a}$$

and for $x \neq 0$, $y \in \mathbb{R}^d$, denoting $l = \ell_0(x)$, $L = \mathrm{supp}(x)$, and ν a permutation of $[\![1, d]\!]$ such that $|y_{\nu(1)}| \geq \ldots \geq |y_{\nu(d)}|$,

$$y \in \partial_{\dot{\complement}}\ell_0(x) \iff \begin{cases} y_L = \lambda x \ , \ \ \lambda \geq 0 \ , \\ |y_j| \leq \min_{i\in L} |y_i| \ , \ \ \forall j \notin L \ , \\ |y_{\nu(k+1)}|^2 \geq \big(\|y\|^{\mathrm{tn}}_{(k,2)} + 1\big)^2 - \big(\|y\|^{\mathrm{tn}}_{(k,2)}\big)^2 \ , \ \ \forall k \in [\![0, l-1]\!] \ , \\ |y_{\nu(l+1)}|^2 \leq \big(\|y\|^{\mathrm{tn}}_{(l,2)} + 1\big)^2 - \big(\|y\|^{\mathrm{tn}}_{(l,2)}\big)^2 \ . \end{cases} \tag{28b}$$

We illustrate in Figure 1 the Capra-subdifferentials obtained with (28) in the two-dimensional case where $\ell_0 : \mathbb{R}^2 \to \{0, 1, 2\}$. In Figure 1a, we display the Capra-subdifferential of ℓ_0 at three 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 Proposition 11 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

$$\partial \ell_0(0) = \{0\} \ , \quad \text{and} \quad \partial \ell_0(x) = \emptyset \ , \ \ \forall x \in \mathbb{R}^d \setminus \{0\} \ , \tag{29}$$

following [3, Table 3]. We also recall further notions of generalized subdifferentials obtained for the ℓ_0 pseudonorm. We refer 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

$$\mathcal{D} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d \ , \ \ x \mapsto \left\{ y \in \mathbb{R}^d \,\middle|\, y_L = 0 \right\} \ , \tag{30}$$

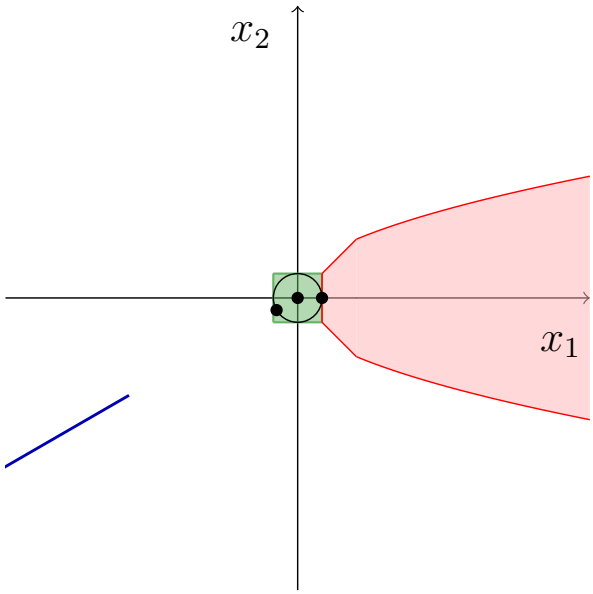
where $L = \mathrm{supp}(x)$, from [5, Theorems 1, 2]. We deduce that the Capra-subdifferential of the ℓ_0 pseudonorm is significantly different form previous notions of generalized subdifferentials of ℓ_0, summarized by (30). In particular, whereas $\left\{ y \in \mathbb{R}^d \,\middle|\, y_L = 0 \right\}$ is a vector subspace, the Capra-subdifferential $\partial_{\dot{\varsigma}} \ell_0(x)$ is a closed convex set, but not a vector subspace.

Second, we argue that since the ℓ_0 pseudonorm displays the Capra-convex properties recalled in Theorem 4, the Capra-subdifferential is a natural object to deal with the function ℓ_0. As an example, we show how to obtain lower approximations of the ℓ_0 pseudonorm from the knowledge of its Capra-subdifferential. 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]. 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. 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 $y_i \in \partial_{\dot{\varsigma}} \ell_0(x_i)$. By definition of the Capra biconjugate in (3b), we have that

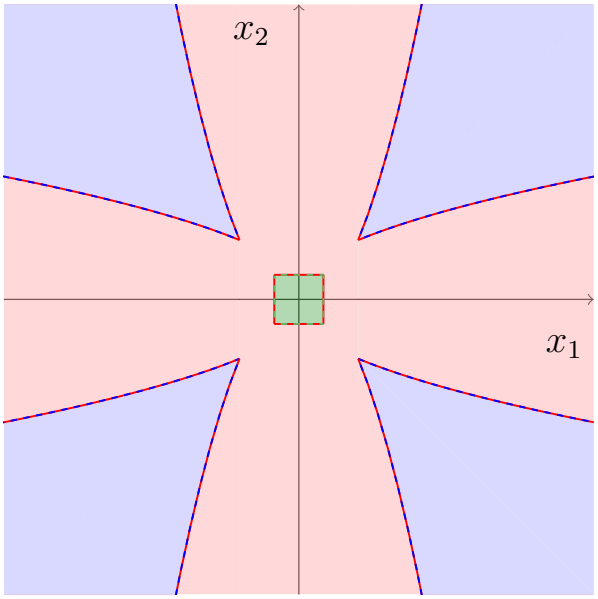
$$\max_{i \in I} \left(\dot{\varsigma}(x, y_i) - \ell_0^{\dot{\varsigma}}(y_i) \right) \leq \sup_{y \in \mathbb{R}^d} \left(\dot{\varsigma}(x, y) - \ell_0^{\dot{\varsigma}}(y) \right) = \ell_0^{\dot{\varsigma}\dot{\varsigma}'}(x) \ , \ \ \forall x \in \mathbb{R}^d \ . \tag{31}$$

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

$$\underline{\ell_0} : \mathbb{R}^d \to \mathbb{R} \ , \ \ x \mapsto \max_{i \in I} \left(\dot{\varsigma}(x, y_i) - \ell_0^{\dot{\varsigma}}(y_i) \right) \tag{32}$$

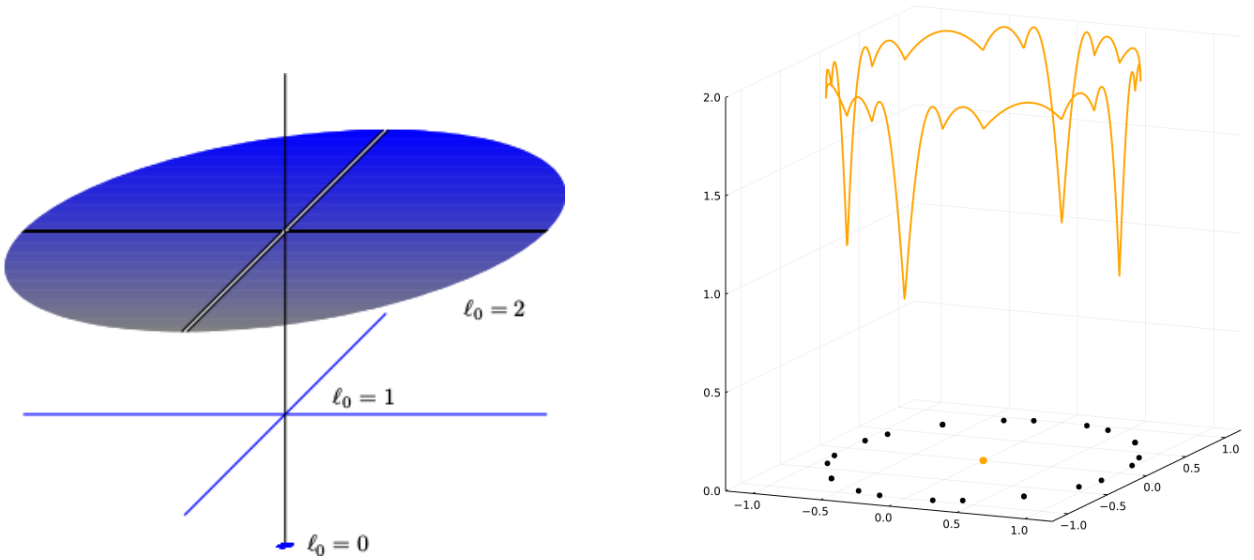


(a) $\partial_{\dot{C}} \ell_0(0,0)$, $\partial_{\dot{C}} \ell_0(1,0)$, $\partial_{\dot{C}} \ell_0(-\frac{\sqrt{3}}{2}, -\frac{1}{2})$



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

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 points (Figure 1a) and for all points in $\mathbb{R}^2$ (Figure 1b)



(a) $\ell_0 : \mathbb{R}^2 \to \{0, 1, 2\}$ (b) $\underline{\ell_0} : \mathbb{R}^2 \to \mathbb{R}$ in (32) on $\mathbb{S} \cup \{0\}$

Figure 2: The ℓ_0 pseudonorm in $\mathbb{R}^2$ (blue color, Figure 2a) and a polyhedral-like 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)

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

$$\mathfrak{c}(x_i, y_i) - \ell_0^{\mathfrak{C}}(y_i) = \ell_0(x_i) \ , \tag{33}$$

so that this lower bound is exact 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 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.

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

$$|||y|||_\star = \sup_{|||x|||\leq 1} \langle x,\, y\rangle \ , \ \ \forall y \in \mathbb{R}^d \tag{34}$$

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

$$\mathbb{S}_\star = \big\{y \in \mathbb{R}^d \,\big|\, |||y|||_\star = 1\big\} \ , \tag{35a}$$

$$\mathbb{B}_\star = \big\{y \in \mathbb{R}^d \,\big|\, |||y|||_\star \leq 1\big\} \ . \tag{35b}$$

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

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

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]

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

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

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

Orthant strict monotonicity

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

Definition 12 (from [1], 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*

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

where $|x| < |x'|$ means that $|x_i| \leq |x_i'|$ for all $i = 1, \ldots, d$, and that there exists $j \in \{1, \ldots, d\}$, such that $|x_j| < |x_j'|$; and $x \circ x' = (x_1 x_1', \ldots, x_d x_d')$ is the Hadamard (entrywise) product.

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

We start by introducing restriction norms and their dual.

Definition 13 ([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* (11)*, as follows.*

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

$$|||x|||_K = |||x||| \ , \ \ \forall x \in \mathcal{R}_K \ . \tag{40}$$

- *The $(K, \star)$-norm $|||\cdot|||_{K,\star}$ is the norm $\big(|||\cdot|||_K\big)_\star$, 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 14 ([3], Definition 3.2) *For $k \in \{1, \dots, d\}$, we call* coordinate-k norm *the norm $|||\cdot|||^{\mathcal{R}}_{(k)}$ whose dual norm is the* dual coordinate-k norm*, denoted by $|||\cdot|||^{\mathcal{R}}_{(k),\star}$, with expression*

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

where the $(K, \star)$-norm $|||\cdot|||_{K,\star}$ is given in Definition 13, 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 14 with the convention $|||\cdot|||^{\mathcal{R}}_{(0),\star} = 0$, although this is not a norm on $\mathbb{R}^d$ but a seminorm.

References

[1] 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.

[2] Jean-Philippe Chancelier and Michel De Lara. Hidden convexity in the l0 pseudonorm. *Journal of Convex Analysis*, 28, 2021.

[3] Jean-Philippe Chancelier and Michel De Lara. Constant along primal rays conjugacies and the l0 pseudonorm. *Optimization*, 0(0):1–32, 2020.

[4] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of convex analysis*. Springer, 2004.

[5] Hai Yen Le. Generalized subdifferentials of the rank function. *Optimization Letters*, 7(4):731–743, 2013.

[6] Juan Enrique Martínez-Legaz. Generalized convex duality and its economic applicatons. In *Handbook of generalized convexity and generalized monotonicity*, pages 237–292. Springer, 2005.

[7] Andrew M. McDonald, Massimiliano Pontil, and Dimitris Stamos. New perspectives on k-support and cluster norms. *The Journal of Machine Learning Research*, 17(1):5376–5413, 2016.

[8] Qing Qu, Ju Sun, and John Wright. Finding a sparse vector in a subspace: Linear sparsity using alternating directions. *IEEE Transactions on Information Theory*, 62(10):5855–5880, Oct 2016.

[9] Ivan Singer. *Abstract convex analysis*. John Wiley & Sons, 1997.

[10] Emmanuel Soubies, Laure Blanc-Féraud, and Gilles Aubert. A continuous exact ℓ_0 penalty (cel0) for least squares regularized problem. *SIAM Journal on Imaging Sciences*, 8(3):1607–1639, 2015.