Variational Formulations for the l0 Pseudonorm and Application to Sparse Optimization

Jean-Philippe Chancelier, Michel de Lara

Variational Formulations for the l_0 Pseudonorm and Application to Sparse Optimization

Jean-Philippe Chancelier and Michel De Lara*
Université Paris-Est, CERMICS (ENPC)

January 29, 2020

Abstract

The so-called ℓ_0 pseudonorm on $\mathbb{R}^d$ counts the number of nonzero components of a vector. It is used in sparse optimization, either as criterion or in the constraints, to obtain solutions with few nonzero entries. However, the mathematical expression of the ℓ_0 pseudonorm, taking integer values, makes it difficult to handle in optimization problems on $\mathbb{R}^d$. Moreover, the Fenchel conjugacy fails to provide relevant insight. In this paper, we analyze the ℓ_0 pseudonorm by means of a family of so-called Capra conjugacies. For this purpose, to each (source) norm on $\mathbb{R}^d$, we attach a so-called Capra coupling between $\mathbb{R}^d$ and $\mathbb{R}^d$, and sequences of generalized top-k dual norms and k-support dual norms. When we suppose that both the source norm and its dual norm are orthant-strictly monotonic, we obtain three main results. First, we show that the ℓ_0 pseudonorm is Capra biconjugate, that is, a Capra-convex function. Second, we deduce an unexpected consequence: the ℓ_0 pseudonorm coincides, on the unit sphere of the source norm, with a proper convex convex lower semicontinuous (lsc) function on $\mathbb{R}^d$. Third, we establish variational formulations for the ℓ_0 pseudonorm and, with these novel expressions, we provide, on the one hand, a new family of lower and upper bounds for the ℓ_0 pseudonorm, as a ratio between two norms, and, on the other hand, reformulations for exact sparse optimization problems.

Key words: ℓ_0 pseudonorm, orthant-strictly monotonic norm, Fenchel-Moreau conjugacy, generalized k-support dual norm, sparse optimization.
AMS classification: 46N10, 49N15, 46B99, 52A41, 90C46

1 Introduction

The *counting function*, also called *cardinality function* or ℓ_0 *pseudonorm*, counts the number of nonzero components of a vector in $\mathbb{R}^d$. It is used in sparse optimization, either as

*delara@cermics.enpc.fr

criterion or in the constraints, to obtain solutions with few nonzero entries. However, the mathematical expression of the ℓ_0 pseudonorm makes it difficult to handle as such in optimization problems on $\mathbb{R}^d$. This is why most of the literature on sparse optimization resorts to *substitute* problems, obtained either from *estimates* (inequalities) for the ℓ_0 pseudonorm, or from *alternative* sparsity-inducing terms (especially suitable norms).

In this paper, we present variational formulations for the ℓ_0 pseudonorm, suitable for exact sparse optimization. By exact, we mean that we keep the ℓ_0 pseudonorm in the optimization problem, either in the criterion or in the constraints, and that we do not resort to substitutes.

The paper is organized as follows. In Sect. 2, we provide background on properties of norms that prove relevant for the ℓ_0 pseudonorm: we introduce sequences of *generalized top-k and k-support dual norms*, generated from any (source) norm on $\mathbb{R}^d$, the notion of sequences of norms that are *graded with respect to the ℓ_0 pseudonorm*, and a new class of *orthant-strictly monotonic norms*. The material is taken from the companion paper [4].

In Sect. 3, we define a so-called Capra coupling between $\mathbb{R}^d$ and $\mathbb{R}^d$, that depends on any (source) norm on $\mathbb{R}^d$. In our main result, we prove that, when both the source norm and its dual norm are orthant-strictly monotonic, the ℓ_0 pseudonorm is equal to its biconjugate, under the associated Capra conjugacy. The result relies on identities established in the companion paper [3]. A surprising consequence is that the ℓ_0 pseudonorm coincides, on the unit sphere, with a proper convex lsc function on $\mathbb{R}^d$.

In Sect. 4, we deduce, from the expressions for the Capra conjugates and biconjugates of the ℓ_0 pseudonorm, a variational formula for the ℓ_0 pseudonorm which involves the whole sequence of generalized k-support dual norms. Finally, with these novel expressions for the ℓ_0 pseudonorm, we provide, on the one hand, a new family of lower and upper bounds for the ℓ_0 pseudonorm, as a ratio between two norms, and, on the other hand, reformulations for exact sparse optimization problems in Sect. 4.4. The Appendix A gathers background on Fenchel-Moreau conjugacies.

2 Background on relevant norms for the ℓ_0 pseudonorm

We provide background on properties of norms that prove relevant for the ℓ_0 pseudonorm as developed in the companion paper [4]. In §2.1, we define the ℓ_0 pseudonorm and its level sets. In §2.2, we review notions related to dual norms. In §2.3, we recall definitions of orthant-monotonic and orthant-strictly monotonic norms. In §2.4, we introduce graded sequences of norms. In §2.5, we introduce generalized top-k and k-support dual norms and recall some of their properties. In §2.6, we introduce coordinate-k and dual coordinate-k norms and recall some of their properties, established in the companion paper [3].

2.1 The ℓ_0 pseudonorm and its level sets

For any vector $x \in \mathbb{R}^d$, we define its *support* by

$$\operatorname{supp}(x) = \big\{ j \in \{1, \ldots, d\} \,\big|\, x_j \neq 0 \big\} \subset \{1, \ldots, d\} \; . \tag{1}$$

The so-called ℓ_0 *pseudonorm* is the function $\ell_0 : \mathbb{R}^d \to \{0, 1, \ldots, d\}$ defined by

$$\ell_0(x) = |\mathrm{supp}(x)| = \text{number of nonzero components of } x \;, \;\; \forall x \in \mathbb{R}^d \;, \tag{2}$$

where $|K|$ denotes the cardinal of a subset $K \subset \{1, \ldots, d\}$. The ℓ_0 pseudonorm shares three out of the four axioms of a norm: nonnegativity, positivity except for $x = 0$, subadditivity. The axiom of 1-homogeneity does not hold true; in contrast, the ℓ_0 pseudonorm is 0-homogeneous:

$$\ell_0(\rho x) = \ell_0(x) \;, \;\; \forall \rho \in \mathbb{R}\backslash\{0\} \;, \;\; \forall x \in \mathbb{R}^d \;. \tag{3}$$

The ℓ_0 pseudonorm is used in exact sparse optimization problems of the form $\inf_{\ell_0(x) \leq k} f(x)$. Thus, we introduce the *level sets*

$$\ell_0^{\leq k} = \{x \in \mathbb{R}^d \mid \ell_0(x) \leq k\} \;, \;\; \forall k = 0, 1, \ldots, d \;. \tag{4}$$

2.2 Dual norm, $|||\cdot|||$-duality, normal cone

Dual norm. For any norm $|||\cdot|||$ on $\mathbb{R}^d$, we denote the unit sphere and the unit ball of the norm $|||\cdot|||$ by

$$\mathbb{S} = \left\{x \in \mathbb{R}^d \,\middle|\, |||x||| = 1\right\} \;, \tag{5a}$$
$$\mathbb{B} = \left\{x \in \mathbb{R}^d \,\middle|\, |||x||| \leq 1\right\} \;. \tag{5b}$$

Recall that the following expression

$$|||y|||_\star = \sup_{|||x||| \leq 1} \langle x \,, y \rangle \;, \;\; \forall y \in \mathbb{R}^d \tag{6}$$

defines a norm on $\mathbb{R}^d$, called the *dual norm* $|||\cdot|||_\star$. We denote the unit sphere and the unit ball of the dual norm $|||\cdot|||_\star$ by

$$\mathbb{S}_\star = \left\{y \in \mathbb{R}^d \,\middle|\, |||y|||_\star = 1\right\} \;, \tag{7a}$$
$$\mathbb{B}_\star = \left\{y \in \mathbb{R}^d \,\middle|\, |||y|||_\star \leq 1\right\} \;. \tag{7b}$$

$|||\cdot|||$-duality, normal cone. By definition of the dual norm in (6), we have the inequality

$$\langle x \,, y \rangle \leq |||x||| \times |||y|||_\star \;, \;\; \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \;. \tag{8a}$$

We are interested in the case where this inequality is an equality. One says that $y \in \mathbb{R}^d$ is $|||\cdot|||$-*dual* to $x \in \mathbb{R}^d$, denoted by $y \parallel_{|||\cdot|||} x$, if equality holds in Inequality (8a), that is,

$$y \parallel_{|||\cdot|||} x \iff \langle x \,, y \rangle = |||x||| \times |||y|||_\star \;. \tag{8b}$$

It will be convenient to express this notion of $|||\cdot|||$-duality in terms of geometric objects of convex analysis. For this purpose, we recall that the *normal cone* $N_C(x)$ to the (nonempty) closed convex subset $C \subset \mathbb{R}^d$ at $x \in C$ is the closed convex cone defined by [6, p.136]

$$N_C(x) = \left\{y \in \mathbb{R}^d \,\middle|\, \langle y \,, x' - x \rangle \leq 0 \;, \;\; \forall x' \in C\right\} \;. \tag{9}$$

Now, easy computations show that the notion of $|||\cdot|||$-duality can be rewritten in terms of normal cones $N_{\mathbb{B}}$ and $N_{\mathbb{B}_\star}$ as follows:

$$\Big(y \parallel_{|||\cdot|||} x \iff y \in N_{\mathbb{B}}\big(\frac{x}{|||x|||}\big) \iff x \in N_{\mathbb{B}_\star}\big(\frac{y}{|||y|||}\big)\Big) \; , \;\; \forall (x,y) \in \mathbb{R}^d\backslash\{0\} \times \mathbb{R}^d\backslash\{0\} \; . \tag{10}$$

2.3 Orthant-strictly monotonic norms

We recall definitions of orthant-monotonic and orthant-strictly monotonic norms that will prove especially relevant for the ℓ_0 pseudonorm.

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

$$x = (x_1, \ldots, x_d) \Rightarrow |x| = (|x_1|, \ldots, |x_d|) \; . \tag{11}$$

Definition 1 *A norm $|||\cdot|||$ on the space $\mathbb{R}^d$ is called*

- orthant-monotonic *[5] if, for all x, x' in $\mathbb{R}^d$, we have $\big(\, |x| \leq |x'|$ and $x \circ x' \geq 0 \Rightarrow |||x||| \leq |||x'||| \, \big)$, where $|x| \leq |x'|$ means $|x_i| \leq |x'_i|$ for all $i = 1, \ldots, d$, and where $x \circ x' = (x_1 x'_1, \ldots, x_d x'_d)$ is the Hadamard (entrywise) product,*
- orthant-strictly monotonic *[4, Definition 3] if, for all x, x' in $\mathbb{R}^d$, we have $\big(\, |x| < |x'|$ and $x \circ x' \geq 0 \Rightarrow |||x||| < |||x'||| \, \big)$, where $|x| < |x'|$ means that $|x_i| \leq |x'_i|$ for all $i = 1, \ldots, d$, and there exists $j \in \{1, \ldots, d\}$, such that $|x_j| < |x'_j|$.*

We recall properties of orthant-monotonic and orthant-strictly monotonic norms.

Proposition 2 *([4, Proposition 6], [5, Theorem 2.26], [7, Theorem 3.2]) Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$. The following assertions are equivalent.*

1. *The norm $|||\cdot|||$ is orthant-monotonic.*
2. *The dual norm $|||\cdot|||_\star$ is orthant-monotonic.*
3. *The norm $|||\cdot|||$ is* increasing with the coordinate subspaces*, in the sense that, for any $x \in \mathbb{R}^d$ and any $J \subset K \subset \{1, \ldots, d\}$, we have $|||x_J||| \leq |||x_K|||$.*

Proposition 3 *([4, Proposition 8]) Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$. The following assertions are equivalent.*

1. *The norm $|||\cdot|||$ is orthant-strictly monotonic.*
2. *The norm $|||\cdot|||$ is* strictly increasing with the coordinate subspaces *in the sense that[1], for any $x \in \mathbb{R}^d$ and any $J \subsetneq K \subset \{1, \ldots, d\}$, we have $x_J \neq x_K \Rightarrow |||x_J||| < |||x_K|||$.*
3. *For any vector $u \in \mathbb{R}^d\backslash\{0\}$, there exists a vector $v \in \mathbb{R}^d\backslash\{0\}$ such that* $\mathrm{supp}(v) = \mathrm{supp}(u)$*, that $u \circ v \geq 0$, and that v is $|||\cdot|||$-dual to u, that is, $\langle u \, , v \rangle = |||u||| \times |||v|||_\star$.*

[1] By $J \subsetneq K$, we mean that $J \subset K$ and $J \neq K$.

2.4 Strictly increasingly graded sequences of norms

A strictly increasingly graded sequence of norms detects the number of nonzero components of a vector in $\mathbb{R}^d$ when the sequence becomes stationary.

Definition 4 *([4, Definition 19]) We say that a sequence* $\{|||\cdot|||_k\}_{k=1,\dots,d}$ *of norms on* $\mathbb{R}^d$ *is* strictly increasingly graded *with respect to the* ℓ_0 *pseudonorm if, for any* $x \in \mathbb{R}^d$*, one of the three following equivalent statements holds true.*

1. *We have the equivalence, for any* $l = 1, \dots, d$*,*
$$\ell_0(x) = l \iff |||x|||_1 \leq \cdots \leq |||x|||_{l-1} < |||x|||_l = \cdots = |||x|||_d \ . \tag{12a}$$
2. *The sequence* $k \in \{1, \dots, d\} \mapsto |||x|||_k$ *is nondecreasing and we have the equivalence, for any* $l = 1, \dots, d$*,*
$$\ell_0(x) \leq l \iff |||x|||_l = |||x|||_d \quad \Big(\iff |||x|||_l \geq |||x|||_d \Big) \ . \tag{12b}$$
3. *The sequence* $k \in \{1, \dots, d\} \mapsto |||x|||_k$ *is nondecreasing and we have the equality*
$$\ell_0(x) = \min \Big\{ k \in \{1, \dots, d\} \, \Big| \, |||x|||_k = |||x|||_d \Big\} \ . \tag{12c}$$

2.5 Generalized top-k and k-support dual norms

We introduce *generalized top-k and k-support dual norms* that are constructed from a *source norm*, and then we recall some of their properties [4].

Restriction norms. For any $x \in \mathbb{R}^d$ and subset $K \subset \{1, \dots, d\}$, 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 subspace[2]
$$\mathcal{R}_K = \mathbb{R}^K \times \{0\}^{-K} = \big\{ x \in \mathbb{R}^d \, \big| \, x_j = 0 \ , \ \forall j \notin K \big\} \subset \mathbb{R}^d \ , \tag{13}$$
where $\mathcal{R}_\emptyset = \{0\}$.

Definition 5 *For any norm* $|||\cdot|||$ *on* $\mathbb{R}^d$ *and any subset* $K \subset \{1, \dots, d\}$*, we define three norms on the subspace* $\mathcal{R}_K$ *of* $\mathbb{R}^d$*, as defined in* (13)*, as follows.*

- *The* K-restriction norm $|||\cdot|||_K$ *is defined by*
$$|||x|||_K = |||x||| \ , \ \ \forall x \in \mathcal{R}_K \ . \tag{14}$$
- *The* $(\star, K)$*-norm* $|||\cdot|||_{\star,K}$ *is the norm* $\big(|||\cdot|||_\star\big)_K$*, given by the restriction to the subspace* $\mathcal{R}_K$ *of the dual norm* $|||\cdot|||_\star$ *(first dual, then restriction),*
- *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).*

[2]Here, 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 $K \cap (-K) = \emptyset$.

Definition of generalized top-k and k-support dual norms. Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$, that we will call the *source norm*. In [4], we have defined the *generalized top-k norms* by

$$|||x|||^{\mathrm{tn}}_{(k)} = \sup_{|K|\leq k} |||x_K||| \;, \;\; \forall x \in \mathbb{R}^d \;, \;\; \forall k = 0, 1, \ldots, d \;. \tag{15}$$

Now, we do the same but with the dual norm $|||\cdot|||_\star$ in lieu of the source norm $|||\cdot|||$.

Definition 6 *For $k \in \{1, \ldots, d\}$, we call* generalized top-k dual norm *(associated with the source norm $|||\cdot|||$) the norm defined by*

$$|||y|||^{\mathrm{tn}}_{\star,(k)} = \sup_{|K|\leq k} |||y_K|||_\star = \sup_{|K|\leq k} |||y_K|||_{\star,K} \;, \;\; \forall y \in \mathbb{R}^d \;. \tag{16}$$

We call generalized k-support dual norm *the dual norm of the generalized top-k dual norm, denoted by*[3] $|||\cdot|||^{\star\mathrm{sn}}_{\star,(k)}$*:*

$$|||\cdot|||^{\star\mathrm{sn}}_{\star,(k)} = \left(|||\cdot|||^{\mathrm{tn}}_{\star,(k)}\right)_\star \;. \tag{17}$$

We adopt the convention $|||\cdot|||^{\mathrm{tn}}_{\star,(0)} = 0$ (although this is not a norm, but a seminorm). We denote the unit sphere and the unit ball of the generalized k-support dual norm $|||\cdot|||^{\star\mathrm{sn}}_{\star,(k)}$ by

$$\mathbb{S}^{\star\mathrm{sn}}_{\star,(k)} = \left\{ x \in \mathbb{R}^d \,\middle|\, |||x|||^{\star\mathrm{sn}}_{\star,(k)} = 1 \right\} \;, \;\; k = 1, \ldots, d \;, \tag{18a}$$

$$\mathbb{B}^{\star\mathrm{sn}}_{\star,(k)} = \left\{ x \in \mathbb{R}^d \,\middle|\, |||x|||^{\star\mathrm{sn}}_{\star,(k)} \leq 1 \right\} \;, \;\; k = 1, \ldots, d \;. \tag{18b}$$

Examples of generalized top-k and k-support dual norms in the case of ℓ_p source norms. In [4], we have named *top-(k,p) norm* — denoted by $||\cdot||^{\mathrm{tn}}_{k,p}$ — the generalized top-k norm in (15) when the source norm $|||\cdot|||$ is the ℓ_p-norm $||\cdot||_p$, for $p \in [1, \infty]$. Therefore, the generalized top-k dual norm in (16) is the top-(k,q) norm when the source norm $|||\cdot|||$ is the ℓ_p-norm $||\cdot||_p$, for $p \in [1, \infty]$ and with $1/p + 1/q = 1$.

In [9, Definition 21], the authors define the (p,k)*-support norm* for $p \in [1, \infty]$. They show, in [9, Corollary 22], that the dual norm $\left(\left(||\cdot||_p\right)^{\mathrm{tn}}_{(k)}\right)_\star$ of the top-(k,p) norm is the (q,k)-support norm, where $1/p + 1/q = 1$. Therefore, the generalized k-support dual norm in (17) is the (p,k)-support norm — denoted by $||\cdot||^{\mathrm{sn}}_{q,k}$ — when the source norm $|||\cdot|||$ is the ℓ_p-norm $||\cdot||_p$, for $p \in [1, \infty]$.

Table 1 provides a summary [4]. For $y \in \mathbb{R}^d$, ν denotes a permutation of $\{1, \ldots, d\}$ such that $|y_{\nu(1)}| \geq |y_{\nu(2)}| \geq \cdots \geq |y_{\nu(d)}|$.

[3] We use the symbol $\star$ in the superscript to indicate that the generalized k-support dual norm $|||\cdot|||^{\star\mathrm{sn}}_{\star,(k)}$ is a dual norm.

source norm $\vert\vert\vert\cdot\vert\vert\vert$	$\vert\vert\vert x\vert\vert\vert^{\star\text{sn}}_{\star,(k)}$	$\vert\vert\vert y\vert\vert\vert^{\text{tn}}_{\star,(k)}$
$\vert\vert\cdot\vert\vert_p$	(p,k)-support norm $\vert\vert x\vert\vert^{\text{sn}}_{p,k}$	top (k,q)-norm $\vert\vert y\vert\vert^{\text{tn}}_{k,q}$ $= \big(\sum_{l=1}^{k} \vert y_{\nu(l)}\vert^q\big)^{1/q}$, $1/p+1/q=1$
$\vert\vert\cdot\vert\vert_1$	$(1,k)$-support norm ℓ_1-norm $\vert\vert x\vert\vert^{\text{sn}}_{1,k} = \vert\vert x\vert\vert_1$	top (k,∞)-norm ℓ_∞-norm $\vert\vert y\vert\vert^{\text{tn}}_{k,\infty} = \vert y_{\nu(1)}\vert = \vert\vert y\vert\vert_\infty$
$\vert\vert\cdot\vert\vert_2$	$(2,k)$-support norm	top $(k,2)$-norm $\vert\vert y\vert\vert^{\text{tn}}_{k,2} = \sqrt{\sum_{l=1}^{k} \vert y_{\nu(l)}\vert^2}$
$\vert\vert\cdot\vert\vert_\infty$	(∞,k)-support norm	top $(k,1)$-norm $\vert\vert y\vert\vert^{\text{tn}}_{k,1} = \sum_{l=1}^{k} \vert y_{\nu(l)}\vert$

Table 1: Examples of generalized top-k and k-support dual norms generated by the ℓ_p source norms $|||\cdot||| = ||\cdot||_p$ for $p \in [1,\infty]$

2.6 Coordinate-k and dual coordinate-k norms

Definition of coordinate-k and dual coordinate-k norms. Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$, that we will call the *source norm.*

Definition 7 *([3, Definition 3]) 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*

$$|||y|||^{\mathcal{R}}_{(k),\star} = \sup_{|K|\leq k} |||y_K|||_{K,\star} \ , \quad \forall y \in \mathbb{R}^d \ , \tag{19}$$

where the $(K,\star)$-norm $|||\cdot|||_{K,\star}$ is given in Definition 5, and where the notation $\sup_{|K|\leq k}$ is a shorthand for $\sup_{K\subset\{1,\dots,d\},|K|\leq k}$.

We denote the unit sphere and the unit ball of the coordinate-k norm $|||\cdot|||^{\mathcal{R}}_{(k)}$ by

$$\mathbb{S}^{\mathcal{R}}_{(k)} = \left\{x \in \mathbb{R}^d \,\middle|\, |||x|||^{\mathcal{R}}_{(k)} = 1\right\} \ , \tag{20a}$$

$$\mathbb{B}^{\mathcal{R}}_{(k)} = \left\{x \in \mathbb{R}^d \,\middle|\, |||x|||^{\mathcal{R}}_{(k)} \leq 1\right\} \ . \tag{20b}$$

Properties of coordinate-k and dual coordinate-k norms.

Proposition 8 *Let $|||\cdot|||$ be a source norm on $\mathbb{R}^d$.*

Coordinate-k norms are always lower than k-support dual norms, that is,

$$|||x|||^{\mathcal{R}}_{(k)} \leq |||x|||^{\star\text{sn}}_{\star,(k)} \ , \quad \forall x \in \mathbb{R}^d \ , \quad \forall k = 1,\dots,d \ , \tag{21a}$$

whereas dual coordinate-k norms are always greater than generalized top-k dual norms, that is,

$$|||y|||^{\mathcal{R}}_{(k),\star} \geq |||y|||^{\mathrm{tn}}_{\star,(k)} \; , \;\; \forall y \in \mathbb{R}^d \; , \;\; \forall k = 1, \dots, d \; . \tag{21b}$$

If the source norm norm $|||\cdot|||$ is orthant-monotonic, then equalities hold true, that is,

$$|||\cdot||| \textit{is orthant-monotonic} \Rightarrow \forall k = 1, \dots, d \quad \begin{cases} |||\cdot|||^{\mathcal{R}}_{(k)} & = |||\cdot|||^{\star\mathrm{sn}}_{\star,(k)} \; , \\ |||\cdot|||^{\mathcal{R}}_{(k),\star} & = |||\cdot|||^{\mathrm{tn}}_{\star,(k)} \; . \end{cases} \tag{22}$$

Proof. It is easily established that, for any nonempty subset $K \subset \{1, \dots, d\}$, we have the inequality $|||\cdot|||_{K,\star} \leq |||\cdot|||_{\star,K}$ [4, Lemma 2]. From the definition (16) of the generalized top-k dual norm, and the definition (19) of the dual coordinate-k norm, we obtain (21b). By taking the dual norms, we get (21a).

The norms for which the equality $|||\cdot|||_{K,\star} = |||\cdot|||_{\star,K}$ holds true for all nonempty subsets $K \subset \{1, \dots, d\}$, are the orthant-monotonic norms ([5, Theorem 2.26],[7, Theorem 3.2]). Therefore, if the norm $|||\cdot|||$ is orthant-monotonic, we have (22). Indeed, from the definition (16) of the generalized top-k dual norm, the inequality (21b) becomes an equality and so with the inequality (21a), by taking the dual norm.

This ends the proof. □

Here is a property of the coordinate-k norms, proved in [3, Proposition 6], that will be useful in the sequel. It is a "non strict" version of the notion of graded sequence of norm (see Definition 4) as introduced in [4, Definition 19]. We have

$$\ell_0(x) \leq k \Rightarrow |||x|||^{\mathcal{R}}_{(k)} = |||x||| \; , \;\; \forall x \in \mathbb{R}^d \; , \;\; \forall k = 1, \dots, d \; . \tag{23}$$

We finish by results that will be useful for our main Theorem 15.

We recall that the normed space $(\mathbb{R}^d, |||\cdot|||)$ is said to be *strictly convex* if the unit ball $\mathbb{B}$ (of the norm $|||\cdot|||$) is *rotund*, that is, if all points of the unit sphere $\mathbb{S}$ are extreme points of the unit ball $\mathbb{B}$. The normed space $(\mathbb{R}^d, ||\cdot||_p)$, equipped with the ℓ_p-norm $||\cdot||_p$ (for $p \in [1, \infty]$), is strictly convex if and only if $p \in]1, \infty[$.

Proposition 9 *The following statements are equivalent.*

1. *The norm $|||\cdot|||$ is orthant-strictly monotonic and the sequence $\left\{|||\cdot|||^{\mathrm{tn}}_{\star,(j)}\right\}_{j=1,\dots,d}$ of generalized top-k dual norms in Definition 6 is strictly increasingly graded with respect to the ℓ_0 pseudonorm.*

2. *The norm $|||\cdot|||$ is orthant-strictly monotonic and the sequence $\left\{|||\cdot|||^{\mathcal{R}}_{(j),\star}\right\}_{j=0,1,\dots,d}$ of dual coordinate-k norms in Definition 7 is strictly increasingly graded with respect to the ℓ_0 pseudonorm.*

3. *Both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic.*

If the norm $|||\cdot|||$ is orthant-strictly monotonic and if the normed space $\left(\mathbb{R}^d, |||\cdot|||_\star\right)$ is strictly convex, then Item 1, Item 2 and Item 3 hold true.

Proof.

• ($1 \Leftrightarrow 2$) Suppose that Item 1 is satisfied and let us show that Item 2 holds true. Since the norm $|||\cdot|||$ is orthant-strictly monotonic, it is orthant-monotonic, hence $|||\cdot|||^{\mathcal{R}}_{(k)} = |||\cdot|||^{\star\mathrm{sn}}_{\star,(k)}$ and $|||\cdot|||^{\mathcal{R}}_{(k),\star} = |||\cdot|||^{\mathrm{tn}}_{\star,(k)}$ by Proposition 8. Therefore, the sequence $\left\{|||\cdot|||^{\mathrm{tn}}_{\star,(j)}\right\}_{j=1,\ldots,d}$ of generalized top-k dual norms is equal to the sequence $\left\{|||\cdot|||^{\mathcal{R}}_{(j),\star}\right\}_{j=0,1,\ldots,d}$ of dual coordinate-k norms in Definition 7. As a consequence, Item 2 holds true. In the same way, we prove that Item 2 implies Item 1, so that they are equivalent.

• ($1 \Rightarrow 3$) Suppose that Item 1 is satisfied and let us show that Item 3 holds true. To prove that the dual norm $|||\cdot|||_\star$ is orthant-strictly monotonic, we will show that Item 2 in Proposition 3 holds true for $|||\cdot|||_\star$. For this purpose, we consider $y \in \mathbb{R}^d$ and $J \subsetneq K \subset \{1, \ldots, d\}$ such that $y_J \neq y_K$. By definition of the ℓ_0 pseudonorm in (2), we have $j = \ell_0(y_J) < k = \ell_0(y_K)$.

On the one hand, as the norm $|||\cdot|||$ is orthant-strictly monotonic, it is orthant-monotonic, so that the dual norm $|||\cdot|||_\star$ is also orthant-monotonic, by the equivalence between Item 1 and Item 2 in Proposition 2. As a consequence, the norms in the sequence $\left\{|||\cdot|||^{\mathrm{tn}}_{\star,(j)}\right\}_{j=1,\ldots,d}$ are also orthant-monotonic by [4, Proposition 15], and we get that $|||y_J|||^{\mathrm{tn}}_{\star,(k-1)} \leq |||y_K|||^{\mathrm{tn}}_{\star,(k-1)}$, in particular, by the equivalence between Item 1 and Item 3 in Proposition 2.

On the other hand, since, by assumption, the sequence $\left\{|||\cdot|||^{\mathrm{tn}}_{\star,(j)}\right\}_{j=0,1,\ldots,d}$ of dual coordinate-k norms is strictly increasingly graded with respect to the ℓ_0 pseudonorm, we have by (12a) that, on the one hand, $|||y_J|||^{\mathrm{tn}}_{\star,(1)} \leq \cdots \leq |||y_J|||^{\mathrm{tn}}_{\star,(j-1)} < |||y_J|||^{\mathrm{tn}}_{\star,(j)} = \cdots = |||y_J|||^{\mathrm{tn}}_{\star,(d)} = |||y_J|||_\star$, because $j = \ell_0(y_J)$, and, on the other hand, $|||y_K|||^{\mathrm{tn}}_{\star,(1)} \leq \cdots \leq |||y_K|||^{\mathrm{tn}}_{\star,(k-1)} < |||y_K|||^{\mathrm{tn}}_{\star,(k)} = \cdots = |||y_K|||^{\mathrm{tn}}_{\star,(d)} = |||y_K|||_\star$, because $k = \ell_0(y_K)$. Since $j < k$, we deduce that

$$|||y_J|||_\star = |||y_J|||^{\mathrm{tn}}_{\star,(j)} = |||y_J|||^{\mathrm{tn}}_{\star,(k-1)} \leq |||y_K|||^{\mathrm{tn}}_{\star,(k-1)} < |||y_K|||^{\mathrm{tn}}_{\star,(k)} = |||y_K|||_\star \ ,$$

and therefore that $|||y_J|||_\star < |||y_K|||_\star$. Thus, Item 2 in Proposition 3 holds true for $|||\cdot|||_\star$, so that the dual norm $|||\cdot|||_\star$ is orthant-strictly monotonic. Hence, we have shown that Item 3 is satisfied.

• ($3 \Rightarrow 1$) Suppose that Item 3 is satisfied and let us show that Item 1 holds true. Since the dual norm $|||\cdot|||_\star$ is orthant-strictly monotonic it is proved in [4, Proposition 21] that the sequence $\left\{|||\cdot|||^{\mathrm{tn}}_{\star,(j)}\right\}_{j=0,1,\ldots,d}$ is strictly increasingly graded with respect to the ℓ_0 pseudonorm. Hence, Item 1 holds true.

• Finally, suppose that the norm $|||\cdot|||$ is orthant-strictly monotonic and that the normed space $\left(\mathbb{R}^d, |||\cdot|||_\star\right)$ is strictly convex, As the norm $|||\cdot|||$ is orthant-strictly monotonic, it is orthant-monotonic, so that the dual norm $|||\cdot|||_\star$ is also orthant-monotonic, by the equivalence between Item 1 and Item 2 in Proposition 2. As the normed space $\left(\mathbb{R}^d, |||\cdot|||_\star\right)$ is strictly convex, it is proved in [4, Proposition 10] that the dual norm $|||\cdot|||_\star$ is orthant-strictly monotonic. Hence, Item 3 holds true.

This ends the proof. □

3 The Capra conjugacy under orthant-strict monotonicity

We introduce the coupling Capra in §3.1. Then, in §3.2 we recall a formula for the Capra-subdifferential of functions of the ℓ_0 pseudonorm in (2). When both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic, we prove that the Capra-subdifferential of nondecreasing functions of the ℓ_0 pseudonorm is nonempty. Under the same assumptions, we establish in §3.3, relations between the ℓ_0 pseudonorm in (2) and the sequence of generalized top-k dual norms in Definition 6; we show that the ℓ_0 pseudonorm is Capra-convex.

We work on the Euclidian space $\mathbb{R}^d$ (with $d \in \mathbb{N}^*$), equipped with the scalar product $\langle \cdot \, , \cdot \rangle$ (but not necessarily with the Euclidian norm).

3.1 Constant along primal rays coupling (Capra)

Following [3, 2], we introduce the coupling Capra. Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$.

Definition 10 *([3, Definition 8]) We define the* constant along primal rays coupling ¢*, or Capra, between* $\mathbb{R}^d$ *and* $\mathbb{R}^d$ *by*

$$\forall y \in \mathbb{R}^d \, , \quad \begin{cases} ¢(x,y) & = \dfrac{\langle x \, , y \rangle}{|||x|||} \, , \quad \forall x \in \mathbb{R}^d \backslash \{0\} \, , \\ ¢(0,y) & = 0. \end{cases} \tag{24}$$

We stress the point that, in (24), the Euclidian scalar product $\langle x \, , y \rangle$ and the norm term $|||x|||$ need not be related, that is, the norm $|||\cdot|||$ is not necessarily Euclidian.

The coupling Capra has the property of being constant along primal rays, hence the acronym Capra (Constant Along Primal RAys). We introduce the primal *normalization mapping* n, from $\mathbb{R}^d$ towards the unit sphere $\mathbb{S}$ united with $\{0\}$, as follows:

$$n : \mathbb{R}^d \to \mathbb{S} \cup \{0\} \, , \ n(x) = \begin{cases} \frac{x}{|||x|||} & \text{if } x \neq 0 \, , \\ 0 & \text{if } x = 0 \, . \end{cases} \tag{25}$$

We will see below that the Capra-conjugacy, induced by the coupling Capra, shares some relations with the Fenchel conjugacy (see §A.2).

Capra-conjugates and biconjugates. Here are expressions for the Capra-conjugates and biconjugates of a function.

In the whole paper, we use $\overline{\mathbb{R}} = [-\infty, +\infty]$.

Proposition 11 *([3, Proposition 9]) For any function* $g : \mathbb{R}^d \to \overline{\mathbb{R}}$*, the* ¢′*-Fenchel-Moreau conjugate is given by*

$$g^{¢'} = g^{\star'} \circ n \, . \tag{26a}$$

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

$$f^{¢} = \big(\inf \big[f \mid n \big] \big)^{\star} \; , \tag{26b}$$

where the epi-composition $\inf \big[f \mid n \big]$ *[11, p. 27] has here the expression*

$$\inf \big[f \mid n \big](x) = \inf \big\{ f(x') \, \big| \, n(x') = x \big\} = \begin{cases} \inf_{\lambda > 0} f(\lambda x) & \text{if } x \in \mathbb{S} \cup \{0\} \; , \\ +\infty & \text{if } x \not\in \mathbb{S} \cup \{0\} \; , \end{cases} \tag{26c}$$

and the ¢-Fenchel-Moreau biconjugate is given by

$$f^{¢¢'} = \big(f^{¢} \big)^{\star'} \circ n = \big(\inf \big[f \mid n \big] \big)^{\star\star'} \circ n \; . \tag{26d}$$

We observe that the ¢-Fenchel-Moreau conjugate $f^{¢}$ is a closed convex function on $\mathbb{R}^d$ (see §A.2).

Capra-convex functions. We recall that so-called ¢-*convex functions* are all functions of the form $g^{¢'}$, for any $g : \mathbb{R}^d \to \overline{\mathbb{R}}$, or, equivalently, all functions of the form $f^{¢¢'}$, for any $f : \mathbb{R}^d \to \overline{\mathbb{R}}$, or, equivalently, all functions that are equal to their ¢-biconjugate ($f^{¢¢'} = f$) [13, 12, 8].

We recall that a function is closed convex on $\mathbb{R}^d$ if and only if it is either a proper convex lower semi continuous (lsc) function or one of the two constant functions $-\infty$ or $+\infty$ (see §A.2).

Proposition 12 *([3, Proposition 10]) A function is ¢-convex if and only if it is the composition of a closed convex function on $\mathbb{R}^d$ with the normalization mapping* (25)*. More precisely, for any function $h : \mathbb{R}^d \to \overline{\mathbb{R}}$, we have the equivalences*

$$\begin{aligned} & h \text{ is ¢-convex} \\ & \quad \iff h = h^{¢¢'} \\ & \quad \iff h = \big(h^{¢} \big)^{\star'} \circ n \text{ (where } \big(h^{¢} \big)^{\star'} \text{ is a closed convex function)} \\ & \quad \iff \text{there exists a closed convex function } f : \mathbb{R}^d \to \overline{\mathbb{R}} \text{ such that } h = f \circ n \; . \end{aligned}$$

3.2 Capra-subdifferentials related to the ℓ_0 pseudonorm

We recall formulas for the Capra-subdifferential of the ℓ_0 pseudonorm in (2). Then, we provide conditions under which the Capra-subdifferential of a function of the ℓ_0 pseudonorm is not empty.

Following the definition of the subdifferential of a function with respect to a duality in [1], we define the *Capra-subdifferential* of the function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ at $x \in \mathbb{R}^d$ by

$$\partial_{¢} f(x) = \big\{ y \in \mathbb{R}^d \, \big| \, ¢(x', y) \dotplus \big(-f(x') \big) \leq ¢(x, y) \dotplus \big(-f(x) \big) \; , \;\; \forall x' \in \mathbb{R}^d \big\} \tag{28a}$$

$$= \big\{ y \in \mathbb{R}^d \, \big| \, f^{¢}(y) = ¢(x, y) \dotplus \big(-f(x) \big) \big\} \; . \tag{28b}$$

Proposition 13 *([3, Proposition 14]) Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$, with associated sequences $\left\{|||\cdot|||^{\mathcal{R}}_{(j)}\right\}_{j=1,\ldots,d}$ of coordinate-k norms and $\left\{|||\cdot|||^{\mathcal{R}}_{\star(j)}\right\}_{j=1,\ldots,d}$ of dual coordinate-k norms, as in Definition 7, and with associated Capra coupling ¢ in (24).*

Let a function $\varphi : \{0, 1, \ldots, d\} \to \overline{\mathbb{R}}$ and a vector $x \in \mathbb{R}^d$ be given.

- *The Capra-subdifferential, as in (28), of the function $\varphi \circ \ell_0$ at $x = 0$ is given by*

$$\partial_{¢}(\varphi \circ \ell_0)(0) = \bigcap_{j=1,\ldots,d} \big[\varphi(j) \dotplus \big(- \varphi(0)\big)\big] \mathbb{B}^{\mathcal{R}}_{(j),\star} , \tag{29a}$$

  where, by convention $\lambda \mathbb{B}^{\mathcal{R}}_{(j),\star} = \emptyset$, for any $\lambda \in [-\infty, 0[$, and $+\infty \mathbb{B}^{\mathcal{R}}_{(j),\star} = \mathbb{R}^d$.

- *The Capra-subdifferential, as in (28), of the function $\varphi \circ \ell_0$ at $x \neq 0$ is given by the following cases*
  - *if $l = \ell_0(x) \geq 1$ and either $\varphi(l) = -\infty$ or $\varphi \equiv +\infty$, then $\partial_{¢}(\varphi \circ \ell_0)(x) = \mathbb{R}^d$,*
  - *if $l = \ell_0(x) \geq 1$ and $\varphi(l) = +\infty$ and there exists $j \in \{0, 1, \ldots, d\}$ such that $\varphi(j) \neq +\infty$, then $\partial_{¢}(\varphi \circ \ell_0)(x) = \emptyset$,*
  - *if $l = \ell_0(x) \geq 1$ and $-\infty < \varphi(l) < +\infty$, then*

$$y \in \partial_{¢}(\varphi \circ \ell_0)(x) \iff \begin{cases} y \in N_{\mathbb{B}^{\mathcal{R}}_{(l)}}\big(\frac{x}{|||x|||^{\mathcal{R}}_{(l)}}\big) \\ \textit{and} \\ l \in \operatorname{argmax}_{j=0,1,\ldots,d} \big[|||y|||^{\mathcal{R}}_{(j),\star} - \varphi(j)\big] . \end{cases} \tag{29b}$$

Proposition 14 *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$, such that both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic. Let $\varphi : \{0, 1, \ldots, d\} \to \mathbb{R}$ be a nondecreasing function. Then, we have*

$$\partial_{¢}(\varphi \circ \ell_0)(x) \neq \emptyset , \ \forall x \in \mathbb{R}^d .$$

More precisely, $\partial_{¢}(\varphi \circ \ell_0)(0) = \bigcap_{j=1,\ldots,d} \big[\varphi(j) - \varphi(0)\big] \mathbb{B}^{\mathcal{R}}_{(j),\star} \neq \emptyset$ and, when $x \neq 0$, for any $y \in \mathbb{R}^d$ such that $\operatorname{supp}(y) = \operatorname{supp}(x)$, and that $\langle x , y \rangle = |||x||| \times |||y|||_\star$, we have that $\lambda y \in \partial_{¢}(\varphi \circ \ell_0)(x)$ for $\lambda > 0$ large enough.

Proof. Since the norm $|||\cdot|||$ is orthant-strictly monotonic, it is orthant-monotonic, so that we have $|||\cdot|||^{\mathcal{R}}_{(j)} = |||\cdot|||^{\star\mathrm{sn}}_{\star,(j)}$ and $|||\cdot|||^{\mathcal{R}}_{(j),\star} = |||\cdot|||^{\mathrm{tn}}_{\star,(j)}$, for $j = 0, 1, \ldots, d$ by Proposition 8 (with the proper conventions for the case $j = 0$). Therefore, we can translate all the results with generalized top-k and k-support dual norms instead of coordinate-k and dual coordinate-k norms.

When $x = 0$, we have, by (29a), that $\partial_{¢}(\varphi \circ \ell_0)(0) = \bigcap_{j=1,\ldots,d} \big[\varphi(j) - \varphi(0)\big] \mathbb{B}^{\mathcal{R}}_{(j),\star}$ because $\varphi(j) \dotplus \big(-\varphi(0)\big) = \varphi(j) - \varphi(0)$ since the function φ takes finite values. The set $\bigcap_{j=1,\ldots,d} \big[\varphi(j) - \varphi(0)\big] \mathbb{B}^{\mathcal{R}}_{(j),\star}$ is nonempty (it contains 0), because $\varphi(j) - \varphi(0) \geq 0$ for $j = 1, \ldots, d$ since $\varphi : \{0, 1, \ldots, d\} \to \mathbb{R}$ is a nondecreasing function.

From now on, we consider $x \in \mathbb{R}^d \backslash \{0\}$ such that $\ell_0(x) = l \in \{1, \ldots, d\}$, and we will use the characterization (29b) of the subdifferential $\partial_{\c{C}}(\varphi \circ \ell_0)(x)$.

Since the norm $|||\cdot|||$ is orthant-strictly monotonic, by Proposition 3 (equivalence between Item 1 and Item 3), there exists a vector $y \in \mathbb{R}^d$ such that

$$L = \mathrm{supp}(x) = \mathrm{supp}(y) \ \text{ hence } \ \ell_0(y) = \ell_0(x) = l > 1 \ , \tag{30a}$$

$$\langle x \,, y \rangle = |||x||| \times |||y|||_\star \ . \tag{30b}$$

Since both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic, the sequence $\left\{ |||\cdot|||^{\mathcal{R}}_{(j),\star} \right\}_{j=0,1,\ldots,d}$ is strictly increasingly graded with respect to the ℓ_0 pseudonorm, by Proposition 9. Therefore, we have (see Definition 4):

$$|||y|||^{\mathcal{R}}_{(1),\star} \leq \cdots \leq |||y|||^{\mathcal{R}}_{(l-1),\star} < |||y|||^{\mathcal{R}}_{(l),\star} = \cdots = |||y|||^{\mathcal{R}}_{(d),\star} = |||y|||_\star \ . \tag{31}$$

• First, we are going to establish that we have $y \in N_{\mathbb{B}^{\mathcal{R}}_{(l)}}(\frac{x}{|||x|||^{\mathcal{R}}_{(l)}})$, that is, the first of the two conditions in the characterization (29b) of the subdifferential $\partial_{\c{C}}(\varphi \circ \ell_0)(x)$.

On the one hand, because $\ell_0(y) = l$ and by (31), we have that $|||y|||_\star = |||y|||^{\mathcal{R}}_{(l),\star}$. On the other hand, we have $|||x||| = |||x|||^{\mathcal{R}}_{(l)}$ by (23). Hence, from (30b), we get $\langle x \,, y \rangle = |||x|||^{\mathcal{R}}_{(l)} \times |||y|||^{\mathcal{R}}_{(l),\star}$, from which we obtain $y \in N_{\mathbb{B}^{\mathcal{R}}_{(l)}}(\frac{x}{|||x|||^{\mathcal{R}}_{(l)}})$ by (10) as $x \neq 0$. To close this part, notice that, for all $\lambda > 0$, we have that $\lambda y \in N_{\mathbb{B}^{\mathcal{R}}_{(l)}}(\frac{x}{|||x|||^{\mathcal{R}}_{(l)}})$, because this last set is a cone.

• Second, we prove the other of the two conditions in the characterization (29b) of the subdifferential $\partial_{\c{C}}(\varphi \circ \ell_0)(x)$. More precisely, we are going to show that, for λ large enough, $|||\lambda y|||^{\mathcal{R}}_{(l),\star} - \varphi(l) = \sup_{j=0,1,\ldots,d} \left[|||\lambda y|||^{\mathcal{R}}_{(j),\star} - \varphi(j) \right]$. For this purpose, we consider the mapping $\psi :]0, +\infty[\to \mathbb{R}$ defined by

$$\psi(\lambda) = |||\lambda y|||^{\mathcal{R}}_{(l),\star} - \varphi(l) - \sup_{j=0,1,\ldots,d} \left[|||\lambda y|||^{\mathcal{R}}_{(j),\star} - \varphi(j) \right] \ , \ \ \forall \lambda > 0 \ ,$$

and we will show that $\psi(\lambda) = 0$ for λ large enough. We have

$$\begin{aligned}
\psi(\lambda) &= \inf_{j=0,1,\ldots,d} \Big(\lambda \big(|||y|||^{\mathcal{R}}_{(l),\star} - |||y|||^{\mathcal{R}}_{(j),\star} \big) + \varphi(j) - \varphi(l) \Big) \\
&= \inf \Big\{ \lambda |||y|||^{\mathcal{R}}_{(l),\star} + \varphi(0) - \varphi(l), \inf_{j=1,\ldots,l-1} \Big(\lambda \big(|||y|||^{\mathcal{R}}_{(l),\star} - |||y|||^{\mathcal{R}}_{(j),\star} \big) + \varphi(j) - \varphi(l) \Big), \\
&\qquad\qquad (\text{as } |||y|||^{\mathcal{R}}_{(0),\star} = 0 \text{ by convention}) \\
&\qquad \inf_{j=l,\ldots,d} \Big(\lambda \big(|||y|||^{\mathcal{R}}_{(l),\star} - |||y|||^{\mathcal{R}}_{(j),\star} \big) + \varphi(j) - \varphi(l) \Big) \Big\} \\
&= \inf \Big\{ \lambda |||y|||^{\mathcal{R}}_{(l),\star} + \varphi(0) - \varphi(l), \inf_{j=1,\ldots,l-1} \Big(\lambda \big(|||y|||^{\mathcal{R}}_{(l),\star} - |||y|||^{\mathcal{R}}_{(j),\star} \big) + \varphi(j) - \varphi(l) \Big), \\
&\qquad \inf_{j=l,\ldots,d} \big(\varphi(j) - \varphi(l) \big) \Big\} \qquad (\text{as } |||y|||^{\mathcal{R}}_{(j),\star} = |||y|||^{\mathcal{R}}_{(l),\star} \text{ for } j \geq l \text{ by (31)}) \\
&= \inf \Big\{ \lambda |||y|||^{\mathcal{R}}_{(l),\star} + \varphi(0) - \varphi(l), \inf_{j=1,\ldots,l-1} \Big(\lambda \big(|||y|||^{\mathcal{R}}_{(l),\star} - |||y|||^{\mathcal{R}}_{(j),\star} \big) + \varphi(j) - \varphi(l) \Big), 0 \Big\} \ ,
\end{aligned}$$

as $\inf_{j=l,\ldots,d}\big(\varphi(j)-\varphi(l)\big)=0$ because $\varphi:\{0,1,\ldots,d\}\to\mathbb{R}$ is a nondecreasing function.

Let us show that the two first terms in the infimum go to $+\infty$ when $\lambda\to+\infty$. The first term $\lambda|||y|||^{\mathcal{R}}_{(l),\star}+\varphi(0)-\varphi(l)$ goes to $+\infty$ because, by (31), we have $|||y|||^{\mathcal{R}}_{(l),\star}=|||y|||_{\star}>0$ as $y\in\mathbb{R}^d\backslash\{0\}$. The second term $\inf_{j=1,\ldots,l-1}\Big(\lambda\big(|||y|||^{\mathcal{R}}_{(l),\star}-|||y|||^{\mathcal{R}}_{(j),\star}\big)+\varphi(j)-\varphi(l)\Big)$ also goes to $+\infty$ because $\ell_0(y)=l$, so that $|||y|||_{\star}=|||y|||^{\mathcal{R}}_{(l),\star}>|||y|||^{\mathcal{R}}_{(j),\star}$ for $j=1,\ldots,l-1$ by (31). Therefore, $\lim_{\lambda\to+\infty}\psi(\lambda)=\inf\Big\{+\infty,+\infty,0\Big\}=0$, hence $\psi(\lambda)=0$ for λ large enough, and thus $|||\lambda y|||^{\mathcal{R}}_{(l),\star}-\varphi(l)=\sup_{j=0,1,\ldots,d}\big[|||\lambda y|||^{\mathcal{R}}_{(j),\star}-\varphi(j)\big]$, that is, $l\in\operatorname{argmax}_{j=0,1,\ldots,d}\big[|||\lambda y|||^{\mathcal{R}}_{(j),\star}-\varphi(j)\big]$.

Wrapping up the above results, we have shown that, for any vector $y\in\mathbb{R}^d$ such that $\operatorname{supp}(y)=\operatorname{supp}(x)$, and that $\langle x\,,y\rangle=|||x|||\times|||y|||_{\star}$, then, for $\lambda>0$ large enough, λy satisfies the two conditions in the characterization (29b) of the subdifferential $\partial_{\cent}(\varphi\circ\ell_0)(x)$.

This ends the proof. □

3.3 Capra-conjugates and biconjugates related to the ℓ_0 pseudonorm

Our first main result are identities for the Capra conjugates and biconjugates of suitable nondecreasing functions of the ℓ_0 pseudonorm. We will show that, when both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_{\star}$ are orthant-strictly monotonic, then any suitable nondecreasing function of the ℓ_0 pseudonorm is Capra biconjugate, that is, a Capra-convex function.

Theorem 15 *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$ with associated sequence $\Big\{|||\cdot|||^{\mathrm{tn}}_{\star,(j)}\Big\}_{j=1,\ldots,d}$ of generalized top-k dual norms, as in Definition 7, and with associated Capra coupling $\cent$ in* (24).

If the norm $|||\cdot|||$ is orthant-monotonic, then, for any function $\varphi:\{0,1,\ldots,d\}\to\overline{\mathbb{R}}$, we have

$$(\varphi\circ\ell_0)^{\cent}=\sup_{j=0,1,\ldots,d}\Big[|||\cdot|||^{\mathrm{tn}}_{\star,(j)}-\varphi(j)\Big]\,, \tag{32a}$$

with the convention that $|||\cdot|||^{\mathrm{tn}}_{\star,(0)}=0$.

If both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_{\star}$ are orthant-strictly monotonic, then, for any nondecreasing function $\varphi:\{0,1,\ldots,d\}\to\mathbb{R}$, we have

$$(\varphi\circ\ell_0)^{\cent\cent'}=\varphi\circ\ell_0\,. \tag{32b}$$

Proof. In both cases, the norm $|||\cdot|||$ is orthant-monotonic, hence $|||\cdot|||^{\mathcal{R}}_{(k)}=|||\cdot|||^{\star\mathrm{sn}}_{\star,(k)}$ and $|||\cdot|||^{\mathcal{R}}_{(k),\star}=|||\cdot|||^{\mathrm{tn}}_{\star,(k)}$ by Proposition 8.

It is proved in [3, Proposition 11] that $(\varphi\circ\ell_0)^{\cent}=\sup_{j=0,1,\ldots,d}\Big[|||\cdot|||^{\mathcal{R}}_{(j),\star}-\varphi(j)\Big]$. As $|||\cdot|||^{\mathcal{R}}_{(j),\star}=|||\cdot|||^{\mathrm{tn}}_{\star,(j)}$, we obtain (32a).

As, by assumption, both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic, Proposition 14 applies. Therefore, for any vector $x \in \mathbb{R}^d$ and any $y \in \partial_{\text{¢}}(\varphi \circ \ell_0)(x) \neq \emptyset$, we obtain

$$
\begin{aligned}
(\varphi \circ \ell_0)^{\text{¢}\text{¢}'}(x) &\geq \text{¢}(x,y) \dotplus \big(-(\varphi \circ \ell_0)^{\text{¢}}(y) \big) && \text{(by definition (50) of the biconjugate)} \\
&= \text{¢}(x,y) - (\varphi \circ \ell_0)^{\text{¢}}(y) && \text{(because } -\infty < \text{¢}(x,y) < +\infty \text{ by (24))} \\
&= \text{¢}(x,y) - \big(\text{¢}(x,y) - (\varphi \circ \ell_0)(x) \big) \\
& \qquad \text{(by definition (28b) of the Capra-subdifferential } \partial_{\text{¢}}(\varphi \circ \ell_0)(x)) \\
&= (\varphi \circ \ell_0)(x) \; .
\end{aligned}
$$

On the other hand, we have that $(\varphi \circ \ell_0)^{\text{¢}\text{¢}'}(x) \leq (\varphi \circ \ell_0)(x)$ by (51). We conclude that $(\varphi \circ \ell_0)^{\text{¢}\text{¢}'}(x) = (\varphi \circ \ell_0)(x)$, which is (32b).

This ends the proof. □

4 Hidden convexity and variational formulation for the ℓ_0 pseudonorm

From our main result obtained in §3.3 — namely, Theorem 15 which provides conditions under which a suitable nondecreasing function of the ℓ_0 pseudonorm is a Capra-convex function — we will derive two results. We suppose that both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic. In §4.1, we show that any suitable nondecreasing function of the pseudonorm ℓ_0 coincides, on the unit sphere $\mathbb{S} = \big\{ x \in \mathbb{R}^d \,\big|\, |||x||| = 1 \big\}$, with a proper convex lsc function on $\mathbb{R}^d$. In §4.2, we deduce a variational formula for suitable nondecreasing functions of the ℓ_0 pseudonorm, which involves the whole sequence of generalized k-support dual norms.

4.1 Hidden convexity in the ℓ_0 pseudonorm

We will now present a (rather unexpected) consequence of the just established property (Theorem 15) that, under proper assumptions, $(\varphi \circ \ell_0)^{\text{¢}\text{¢}'} = \varphi \circ \ell_0$.

Proposition 16 *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$. Suppose that both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic. Let $\varphi : \{0, 1, \ldots, d\} \to \mathbb{R}_+$ be a nondecreasing function, such that $\varphi(0) = 0$. Then, the following statements hold true.*

- *There exists a proper convex lsc function $\mathcal{L}_0^\varphi : \mathbb{R}^d \to \overline{\mathbb{R}}$ such that the function $\varphi \circ \ell_0$ coincides, on the unit sphere $\mathbb{S} = \big\{ x \in \mathbb{R}^d \,\big|\, |||x||| = 1 \big\}$, with $\mathcal{L}_0^\varphi$:*

$$
(\varphi \circ \ell_0)(x) = \mathcal{L}_0^\varphi(x) \; , \;\; \forall x \in \mathbb{S} \;\; \text{where} \;\; \mathcal{L}_0^\varphi = \big((\varphi \circ \ell_0)^{\text{¢}} \big)^{\star'} \; . \tag{33a}
$$

- *The function $\varphi \circ \ell_0$ can be expressed as the composition of the proper convex lsc function $\mathcal{L}_0^\varphi$ in* (33a) *with the normalization mapping n in* (25)*:*

$$(\varphi \circ \ell_0)(x) = \mathcal{L}_0^\varphi(\frac{x}{|||x|||}) \; , \;\; \forall x \in \mathbb{R}^d \backslash \{0\} \; . \tag{33b}$$

- *The proper convex lsc function $\mathcal{L}_0^\varphi$ is given by*

$$\mathcal{L}_0^\varphi = \Big(\sup_{j=0,1,\ldots,d} \Big[|||\cdot|||^{\mathrm{tn}}_{\star,(j)} - \varphi(j) \Big] \Big)^{\star'} \; . \tag{34a}$$

- *The function $\mathcal{L}_0^\varphi$ is the largest convex lsc function below the integer valued function*

$$\inf_{j=0,1,\ldots,d} \Big[\delta_{\mathbb{B}^{\star\mathrm{sn}}_{\star,(j)}} + \varphi(j) \Big] \; , \tag{34b}$$

that is, below the function $x \in \mathbb{B}^{\star\mathrm{sn}}_{\star,(j)} \backslash \mathbb{B}^{\star\mathrm{sn}}_{\star,(j-1)} \mapsto \varphi(j)$ for $j = 1, \ldots, d$ and $x \in \mathbb{B}^{\star\mathrm{sn}}_{\star,(0)} = \{0\} \mapsto 0$, the function being infinite outside $\mathbb{B}^{\star\mathrm{sn}}_{\star,(d)} = \mathbb{B}$ (the above construction makes sense as $\mathbb{B}^{\star\mathrm{sn}}_{\star,(1)} \subset \cdots \subset \mathbb{B}^{\star\mathrm{sn}}_{\star,(j-1)} \subset \mathbb{B}^{\star\mathrm{sn}}_{\star,(j)} \subset \cdots \subset \mathbb{B}^{\star\mathrm{sn}}_{\star,(d)} = \mathbb{B}$ by (18b)*).*

- *The function $\mathcal{L}_0^\varphi$ is the largest convex lsc function below the integer valued function*

$$\inf_{j=0,1,\ldots,d} \Big[\delta_{\mathbb{S}^{\star\mathrm{sn}}_{\star,(j)}} + \varphi(j) \Big] \; , \tag{34c}$$

that is, below the function $x \in \mathbb{R}^d \mapsto \inf \varphi\big\{j \in \{0, \ldots, d\} \,\big|\, x \in \mathbb{S}^{\star\mathrm{sn}}_{\star,(j)}\big\}$, with the convention that $\mathbb{S}^{\star\mathrm{sn}}_{\star,(0)} = \{0\}$ and that $\inf \emptyset = +\infty$.

- *The proper convex lsc function $\mathcal{L}_0^\varphi$ also has three variational expressions as follows, where Δ_{d+1} is the simplex of $\mathbb{R}^{d+1}$,*

$$\mathcal{L}_0^\varphi(x) = \min_{\substack{(\lambda_0,\lambda_1,\ldots,\lambda_d) \in \Delta_{d+1} \\ x \in \sum_{l=1}^d \lambda_j \mathbb{B}^{\star\mathrm{sn}}_{\star,(l)}}} \sum_{l=1}^d \lambda_j \varphi(j) \; , \;\; \forall x \in \mathbb{R}^d \tag{35a}$$

$$= \min_{\substack{(\lambda_0,\lambda_1,\ldots,\lambda_d) \in \Delta_{d+1} \\ x \in \sum_{l=1}^d \lambda_j \mathbb{S}^{\star\mathrm{sn}}_{\star,(l)}}} \sum_{l=1}^d \lambda_j \varphi(j) \; , \;\; \forall x \in \mathbb{R}^d \tag{35b}$$

$$= \min_{\substack{x^{(1)} \in \mathbb{R}^d, \ldots, x^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d |||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \leq 1 \\ \sum_{j=1}^d x^{(j)} = x}} \sum_{j=1}^d \varphi(j) |||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \; , \;\; \forall x \in \mathbb{R}^d \; . \tag{35c}$$

Proof. As in the beginning of the proof of Theorem 15, we can observe that, since the norm $|||\cdot|||$ is orthant-strictly monotonic, it is orthant-monotonic, so that we have $|||\cdot|||^{\mathcal{R}}_{(j)} = |||\cdot|||^{\star\mathrm{sn}}_{\star,(j)}$ and $|||\cdot|||^{\mathcal{R}}_{(j),\star} = |||\cdot|||^{\mathrm{tn}}_{\star,(j)}$, for $j = 0, 1, \ldots, d$ by Proposition 8 (with the proper conventions for the case $j = 0$).

• The assumptions make it possible to conclude that $(\varphi \circ \ell_0)^{¢¢'} = \varphi \circ \ell_0$, thanks to Theorem 15. We deduce from Proposition 12 that, being ¢-convex, the function $\varphi \circ \ell_0$ coincides, on the sphere $\mathbb{S}$, with the closed convex function $\mathcal{L}_0^{\varphi} : \mathbb{R}^d \to \overline{\mathbb{R}}$ given by (26d), namely $\mathcal{L}_0^{\varphi} = \big((\varphi \circ \ell_0)^{¢}\big)^{\star'}$. Thus, we have proved (33a).

We now show that the closed convex function $\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}$ is proper. Indeed, on the one hand, it is easily seen that the function $(\varphi \circ \ell_0)^{¢}$ takes finite values, from which we deduce that the function $\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}$ never takes the value $-\infty$. On the other hand, from $(\varphi \circ \ell_0)^{¢¢'} = \varphi \circ \ell_0$ we deduce that the function $\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}$ never takes the value $+\infty$ on the unit sphere. Therefore, the$\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}$ is proper.

• The equality (33b) is an easy consequence of the property (3), implying that the function $\varphi \circ \ell_0$ is invariant along any open ray of $\mathbb{R}^d$.

• As $\mathcal{L}_0^{\varphi} = \big((\varphi \circ \ell_0)^{¢}\big)^{\star'}$ by (33a), and as $\big((\varphi \circ \ell_0)^{¢}\big)^{\star'} = \Big(\sup_{j=0,1,\ldots,d} \big[|||\cdot|||^{\mathrm{tn}}_{\star,(j)} - \varphi(j)\big]\Big)^{\star'}$ by (32a), we get (34a).

• We use [3, Proposition 12], and especially Equations (39c) and (39d), to obtain (34b).

• We use [3, Proposition 12], and especially Equations (39e) and (39f), to obtain (34c).

• We use [3, Proposition 12], and especially Equations (39h), (39i) and (39j) to obtain (35a), (35b) and (35c).

This ends the proof. □

4.2 Variational formulation for the ℓ_0 pseudonorm

As a straightforward application of Proposition 16, we obtain our second main result, namely a variational formulation for the ℓ_0 pseudonorm.

Theorem 17 *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$, such that both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic. Let $\varphi : \{0, 1, \ldots, d\} \to \mathbb{R}_+$ be a nondecreasing function, such that $\varphi(0) = 0$. Then, we have the equality*

$$\varphi\big(\ell_0(x)\big) = \frac{1}{|||x|||} \min_{\substack{z^{(1)} \in \mathbb{R}^d, \ldots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d |||z^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \leq |||x||| \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^{d} \varphi(j) |||z^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \;, \;\; \forall x \in \mathbb{R}^d \backslash \{0\} \;, \tag{36}$$

where the sequence of generalized k-support dual norms $\Big\{|||\cdot|||^{\star\mathrm{sn}}_{\star,(j)}\Big\}_{j=1,\ldots,d}$ has been introduced in Definition 6.

When $\ell_0(x) = l \geq 1$, the minimum in (36) *is achieved at* $(z^{(1)}, \ldots, z^{(d)}) \in (\mathbb{R}^d)^d$ *such that* $z^{(j)} = 0$ *for* $j \neq l$ *and* $z^{(l)} = x$.

Proof. Equation (36) derives from (33b) and (35c).

When $\ell_0(x) = l \geq 1$, by (23), we have $|||x||| = |||x|||^{\mathcal{R}}_{(d)} = \ldots = |||x|||^{\mathcal{R}}_{(l)}$. Now, for any $k \in \{1, \ldots, d\}$, we have $|||\cdot|||^{\mathcal{R}}_{(k)} = |||\cdot|||^{\star\mathrm{sn}}_{\star,(k)}$ by Proposition 8, since the norm $|||\cdot|||$ is orthant-strictly monotonic, hence is orthant-monotonic. As a consequence, we have that $|||x||| = |||x|||^{\star\mathrm{sn}}_{\star,(d)} = \ldots = |||x|||^{\star\mathrm{sn}}_{\star,(l)}$. Therefore, $(z^{(1)}, \ldots, z^{(d)}) \in (\mathbb{R}^d)^d$ such that $z^{(j)} = 0$ for $j \neq l$ and $z^{(l)} = x$ is admissible for the minimization problem (36). We deduce that $\varphi(l) = \varphi\big(\ell_0(x)\big) \leq \frac{1}{|||x|||}\varphi(l)|||x|||^{\star\mathrm{sn}}_{\star,(l)} = \varphi(l)$.

This ends the proof. □

As an illustration, Theorem 17 applies when the norm $|||\cdot|||$ is any of the ℓ_p-norms $||\cdot||_p$ on the space $\mathbb{R}^d$, for $p \in]1, \infty[$, giving (see the notations in Table 1)

$$(\varphi \circ \ell_0)(x) = \frac{1}{||x||_p} \min_{\substack{z^{(1)} \in \mathbb{R}^d, \ldots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d ||z^{(j)}||^{\mathrm{sn}}_{p,j} \leq ||x||_p \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d \varphi(j) ||z^{(j)}||^{\mathrm{sn}}_{p,j} \ , \ \ \forall x \in \mathbb{R}^d \backslash \{0\} \ , \ \ \forall p \in]1, \infty[\ . \tag{37}$$

Indeed, when $p \in]1, \infty[$, the ℓ_p-norm $|||\cdot||| = ||\cdot||_p$ is orthant-strictly monotonic, and so is its dual norm $|||\cdot|||_\star = ||\cdot||_q$ where $1/p + 1/q = 1$ as easily seen. When $p = \infty$, the ℓ_∞-norm $|||\cdot||| = ||\cdot||_\infty$ is not orthant-strictly monotonic. When $p = 1$, the ℓ_1-norm $|||\cdot||| = ||\cdot||_1$ is orthant-strictly monotonic, but the dual norm $|||\cdot||| = ||\cdot||_\infty$ is not.

4.3 Upper and lower bounds for the ℓ_0 pseudonorm as norm ratios

The variational formulation obtained in §4.2 yields a family of lower and upper bounds for the ℓ_0 pseudonorm, as a ratio between two norms, the denominator norm being any.

Proposition 18 *Let* $|||\cdot|||$ *be a norm on* $\mathbb{R}^d$, *such that both the norm* $|||\cdot|||$ *and the dual norm* $|||\cdot|||_\star$ *are orthant-strictly monotonic. Let* $\varphi : \{0, 1, \ldots, d\} \to \mathbb{R}_+$ *be a nondecreasing function, such that* $\varphi(1) > \varphi(0) = 0$. *Then, we have the inequalities*

$$\frac{|||x|||^{\star\mathrm{sn}}_{\star,(\varphi)}}{|||x|||} \leq \varphi\big(\ell_0(x)\big) \leq \min_{j=1,\ldots,d} \frac{\varphi(j)|||x|||^{\star\mathrm{sn}}_{\star,(j)}}{|||x|||} \ , \ \ \forall x \in \mathbb{R}^d \backslash \{0\} \ , \tag{38}$$

where the norm $|||\cdot|||^{\star\mathrm{sn}}_{\star,(\varphi)}$ *is characterized*

- *either by its dual norm which has unit ball* $\bigcap_{j=1,\ldots,d} \varphi(j)\mathbb{B}^{\mathrm{tn}}_{\star,(j)}$, *that is,*

$$\mathbb{B}^{\mathrm{tn}}_{\star,(\varphi)} = \bigcap_{j=1,\ldots,d} \varphi(j)\mathbb{B}^{\mathrm{tn}}_{\star,(j)} \ \ \textit{and} \ \ |||\cdot|||^{\star\mathrm{sn}}_{\star,(\varphi)} = \sigma_{\mathbb{B}^{\mathrm{tn}}_{\star,(\varphi)}} \ , \tag{39}$$

  or, equivalently,

$$|||x|||^{\star\mathrm{sn}}_{\star,(\varphi)} = \sup_{j=1,\ldots,d} \frac{|||x|||^{\star\mathrm{sn}}_{\star,(j)}}{\varphi(j)} \ , \ \ \forall x \in \mathbb{R}^d \ , \tag{40}$$

- *or by the inf-convolution*

$$|||\cdot|||^{\star\mathrm{sn}}_{\star,(\varphi)} = \mathop{\square}_{j=1,\ldots,d} \Big(\varphi(j) |||\cdot|||^{\star\mathrm{sn}}_{\star,(j)} \Big) \;, \tag{41}$$

that is,

$$|||x|||^{\star\mathrm{sn}}_{\star,(\varphi)} = \inf_{\substack{z^{(1)}\in\mathbb{R}^d,\ldots,z^{(d)}\in\mathbb{R}^d \\ \sum_{j=1}^d z^{(j)}=x}} \sum_{j=1}^d \varphi(j) |||z^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \;,\;\; \forall x \in \mathbb{R}^d \;. \tag{42}$$

Proof. The proof is a straightforward application of Theorem 17 for the right hand side (upper bound) inequality.

Regarding the left hand side (lower bound) inequality, it follows directly from [3, Proposition 16]. Indeed, the function $\varphi : \{0, 1, \ldots, d\} \to [0, +\infty[$ is such that $\varphi(j) > \varphi(0) = 0$ for all $j = 1, \ldots, d$, because it is a nondecreasing function such that $\varphi(1) > \varphi(0) = 0$.

□

4.4 Applications to sparse optimization

Finally, with the novel expressions for the ℓ_0 pseudonorm obtained in §4.2, we deduce reformulations for exact sparse optimization problems. The following two Propositions are straightforward applications of Theorem 17.

Proposition 19 (Minimization of the pseudonorm ℓ_0 under constraints) *Let $C \subset \mathbb{R}^d$ be such that $0 \notin C$. Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$, such that both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic. Let $\varphi : \{0, 1, \ldots, d\} \to \mathbb{R}_+$ be a nondecreasing function, such that $\varphi(0) = 0$. Then, we have:*

$$\min_{x\in C} \varphi\big(\ell_0(x)\big) = \min_{\substack{x\in C, x^{(1)}\in\mathbb{R}^d,\ldots,x^{(d)}\in\mathbb{R}^d \\ \sum_{j=1}^d |||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \leq 1 \\ \sum_{j=1}^d x^{(j)} = \frac{x}{|||x|||}}} \sum_{j=1}^d \varphi(j) |||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \;, \tag{43a}$$

$$= \min_{\substack{x\in C, z^{(1)}\in\mathbb{R}^d,\ldots,z^{(d)}\in\mathbb{R}^d \\ \sum_{j=1}^d |||z^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \leq |||x||| \\ \sum_{j=1}^d z^{(j)} = x}} \frac{1}{|||x|||} \sum_{j=1}^d \varphi(j) |||z^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \;, \tag{43b}$$

$$= \min_{x\in C} \frac{1}{|||x|||} \underbrace{\min_{\substack{z^{(1)}\in\mathbb{R}^d,\ldots,z^{(d)}\in\mathbb{R}^d \\ \sum_{j=1}^d |||z^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \leq |||x||| \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d \varphi(j) |||z^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)}}_{\textit{convex optimization problem}} \;. \tag{43c}$$

Proposition 20 (Minimization over level sets of the pseudonorm ℓ_0) *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$, such that both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic. Then, we have, for any $k \in \{1, \ldots, d\}$:*

$$\min_{\ell_0(x) \leq k} f(x) = \min_{\substack{x \in \mathbb{R}^d, z^{(1)} \in \mathbb{R}^d, \ldots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d |||z^{(j)}|||_{\star,(j)}^{\star\mathrm{sn}} \leq |||x||| \\ \sum_{j=1}^d z^{(j)} = x \\ \sum_{j=1}^d j |||z^{(j)}|||_{\star,(j)}^{\star\mathrm{sn}} \leq k |||x|||}} f(x) \;, \tag{44a}$$

$$= \min_{\substack{z^{(1)} \in \mathbb{R}^d, \ldots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d |||z^{(j)}|||_{\star,(j)}^{\star\mathrm{sn}} \leq |||\sum_{j=1}^d z^{(j)}||| \\ \sum_{j=1}^d \varphi(j) |||z^{(j)}|||_{\star,(j)}^{\star\mathrm{sn}} \leq k |||\sum_{j=1}^d z^{(j)}|||}} f\Big(\sum_{j=1}^d z^{(j)} \Big) \;. \tag{44b}$$

5 Conclusion

The mathematical expression of the ℓ_0 pseudonorm makes it difficult to handle as such in optimization problems on $\mathbb{R}^d$. In this paper, we have obtained exact variational formulations for the ℓ_0 pseudonorm, suitable for exact sparse optimization. For this purpose, we have introduced notions about norms that were developed in the companion paper [4]: sequences of generalized top-k and k-support dual norms, generated from any (source) norm on $\mathbb{R}^d$; orthant-strictly monotonic norms on $\mathbb{R}^d$, especially relevant for the ℓ_0 pseudonorm, in relation with the concept of strictly increasingly graded sequence of norms.

Our main result is that the ℓ_0 pseudonorm is equal to its biconjugate under the associated conjugacy, when both the source norm and its dual norm are orthant-strictly monotonic. In that case, one says that the ℓ_0 pseudonorm is a Capra-convex function. A surprising consequence is that the ℓ_0 pseudonorm coincides, on the unit sphere of the source norm, with a proper convex lsc function. More generally, this holds true for any function of the ℓ_0 pseudonorm that is nondecreasing, with finite values and which is null at zero.

The reformulations for exact sparse optimization problems that we have obtained make use of new (latent) vectors, in same number than the underlying dimension d. Thus, the algorithmic implementation may be delicate. However, the variational formulation obtained may suggest approximations of the ℓ_0 pseudonorm, or algorithms making use of the partial convexity that our analysis has put to light. Moreover, we have provided expressions for the Capra-subdifferential of suitable functions of the ℓ_0 pseudonorm, which can inspire "gradient-like" algorithms. In all cases, the variational formulation obtained yields a new family of lower and upper bounds for the ℓ_0 pseudonorm, as a ratio between two norms; this may lead to new smooth sparsity inducing terms, proxies for the ℓ_0 pseudonorm.

Acknowledgements. We want to thank Guillaume Obozinski for discussions on first versions of this work.

A Background on Fenchel-Moreau conjugacies

When we manipulate functions with values in $\overline{\mathbb{R}} = [-\infty, +\infty]$, we adopt the following Moreau *lower addition*, that extends the usual addition with

$$(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = -\infty \ . \tag{45}$$

We review general concepts and notations, then we focus on the special case of the Fenchel conjugacy. We denote $\overline{\mathbb{R}} = [-\infty, +\infty]$.

A.1 The general case

Let be given two sets $\mathbb{X}$ ("primal"), $\mathbb{Y}$ ("dual"), together with a *coupling* function

$$c : \mathbb{X} \times \mathbb{Y} \to \overline{\mathbb{R}} \ . \tag{46}$$

With any coupling, we associate *conjugacies* from $\overline{\mathbb{R}}^{\mathbb{X}}$ to $\overline{\mathbb{R}}^{\mathbb{Y}}$ and from $\overline{\mathbb{R}}^{\mathbb{Y}}$ to $\overline{\mathbb{R}}^{\mathbb{X}}$ as follows.

Definition 21 *The* c-Fenchel-Moreau conjugate *of a function* $f : \mathbb{X} \to \overline{\mathbb{R}}$*, with respect to the coupling* c*, is the function* $f^c : \mathbb{Y} \to \overline{\mathbb{R}}$ *defined by*

$$f^c(y) = \sup_{x \in \mathbb{X}} \Big(c(x,y) \dot{+} \big(-f(x) \big) \Big) \ , \ \forall y \in \mathbb{Y} \ . \tag{47}$$

With the coupling c*, we associate the* reverse coupling c' *defined by*

$$c' : \mathbb{Y} \times \mathbb{X} \to \overline{\mathbb{R}} \ , \ c'(y,x) = c(x,y) \ , \ \forall (y,x) \in \mathbb{Y} \times \mathbb{X} \ . \tag{48}$$

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

$$g^{c'}(x) = \sup_{y \in \mathbb{Y}} \Big(c(x,y) \dot{+} \big(-g(y) \big) \Big) \ , \ \forall x \in \mathbb{X} \ . \tag{49}$$

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

$$f^{cc'}(x) = \big(f^c \big)^{c'}(x) = \sup_{y \in \mathbb{Y}} \Big(c(x,y) \dot{+} \big(-f^c(y) \big) \Big) \ , \ \forall x \in \mathbb{X} \ . \tag{50}$$

The biconjugate of a function $f : \mathbb{X} \to \overline{\mathbb{R}}$ satisfies

$$f^{cc'}(x) \leq f(x) \ , \ \forall x \in \mathbb{X} \ . \tag{51}$$

A.2 The Fenchel conjugacy

When the sets $\mathbb{X}$ and $\mathbb{Y}$ are vector spaces equipped with a bilinear form $\langle\,,\rangle$, the corresponding conjugacy is the classical *Fenchel conjugacy*. For any functions $f : \mathbb{X} \to \overline{\mathbb{R}}$ and $g : \mathbb{Y} \to \overline{\mathbb{R}}$, we denote[4]

$$f^{\star}(y) = \sup_{x\in\mathbb{X}} \Big(\langle x\,,y\rangle \dot{+} \big(-f(x) \big) \Big) \;,\;\; \forall y \in \mathbb{Y} \;, \tag{52a}$$

$$g^{\star'}(x) = \sup_{y\in\mathbb{Y}} \Big(\langle x\,,y\rangle \dot{+} \big(-g(y) \big) \Big) \;,\;\; \forall x \in \mathbb{X} \;, \tag{52b}$$

$$f^{\star\star'}(x) = \sup_{y\in\mathbb{Y}} \Big(\langle x\,,y\rangle \dot{+} \big(-f^{\star}(y) \big) \Big) \;,\;\; \forall x \in \mathbb{X} \;. \tag{52c}$$

For any function $h : \mathbb{W} \to \overline{\mathbb{R}}$, its *epigraph* is $\mathrm{epi}h = \big\{(w,t) \in \mathbb{W}\times\mathbb{R} \,\big|\, h(w) \leq t\big\}$, its *effective domain* is $\mathrm{dom}h = \big\{w \in \mathbb{W} \,\big|\, h(w) < +\infty\big\}$. A function $h : \mathbb{W} \to \overline{\mathbb{R}}$ is said to be *proper* if it never takes the value $-\infty$ and that $\mathrm{dom}h \neq \emptyset$. When $\mathbb{W}$ is equipped with a topology, the function $h : \mathbb{W} \to \overline{\mathbb{R}}$ is said to be *lower semi continuous (lsc)* if its epigraph is closed, and is said to be *closed* if h is either *lower semi continuous (lsc)* and nowhere having the value $-\infty$, or is the constant function $-\infty$ [10, p. 15].

It is proved that, when the two vector spaces $\mathbb{X}$ and $\mathbb{Y}$ are *paired* in the sense of convex analysis[5], the Fenchel conjugacy induces a one-to-one correspondence between the closed convex functions on $\mathbb{X}$ and the closed convex functions on $\mathbb{Y}$ [10, Theorem 5]. Here, a function is said to be *convex* if its epigraph is convex. Notice that the set of closed convex functions is the set of proper convex functions united with the two constant functions $-\infty$ and $+\infty$.

References

[1] M. Akian, S. Gaubert, and V. Kolokoltsov. Invertibility of functional Galois connections. *Comptes Rendus Mathematique*, 335(11):883 – 888, 2002.

[2] J.-P. Chancelier and M. De Lara. Hidden convexity in the l_0 pseudonorm, June 2019. Preprint hal-02146454, arXiv:1906.04038.

[3] J.-P. Chancelier and M. De Lara. Constant along primal rays conjugacies and the l_0 pseudonorm, 2020. Preprint hal-02459673.

[4] J.-P. Chancelier and M. De Lara. Orthant-strictly monotonic norms, graded sequences and generalized top-k and k-support norms for sparse optimization, 2020. Preprint hal-02456433, arXiv:2001.10441.

[4] In convex analysis, one does not use the notation $\star'$, but simply the notation $\star$, as it is often the case that $\mathbb{X} = \mathbb{Y}$ in the Euclidian and Hilbertian cases.

[5] That is, $\mathbb{X}$ and $\mathbb{Y}$ are equipped with a bilinear form $\langle\,,\rangle$, and locally convex topologies that are compatible in the sense that the continuous linear forms on $\mathbb{X}$ are the functions $x \in \mathbb{X} \mapsto \langle x\,,y\rangle$, for all $y \in \mathbb{Y}$, and that the continuous linear forms on $\mathbb{Y}$ are the functions $y \in \mathbb{Y} \mapsto \langle x\,,y\rangle$, for all $x \in \mathbb{X}$.

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

[6] J. B. Hiriart-Ururty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I*. Springer-Verlag, Berlin, 1993.

[7] E. Marques de Sà and M.-J. Sodupe. Characterizations of *orthant-monotonic norms. *Linear Algebra and its Applications*, 193:1–9, 1993.

[8] 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.

[9] 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.

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

[11] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer-Verlag, Berlin, 1998.

[12] A. Rubinov. *Abstract convexity and global optimization*, volume 44 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 2000.

[13] I. Singer. *Abstract Convex Analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons, Inc., New York, 1997.