Capra-Convexity, Convex Factorization and Variational Formulations for the l0 Pseudonorm

Jean-Philippe Chancelier, Michel de Lara

Capra-Convexity, Convex Factorization and Variational Formulations for the l_0 Pseudonorm

Jean-Philippe Chancelier and Michel De Lara
CERMICS, Ecole des Ponts, Marne-la-Vallée, France

August 5, 2021

Abstract

The so-called l0 pseudonorm, or cardinality function, counts the number of nonzero components of a vector. In this paper, we analyze the l0 pseudonorm by means of so-called Capra (constant along primal rays) conjugacies, for which the underlying source norm and its dual norm are both orthant-strictly monotonic (a notion that we formally introduce and that encompasses the lp norms, but for the extreme ones). We obtain three main results. First, we show that the l0 pseudonorm is equal to its Capra-biconjugate, that is, is a Capra-convex function. Second, we deduce an unexpected consequence, that we call convex factorization: the l0 pseudonorm coincides, on the unit sphere of the source norm, with a proper convex lower semicontinuous function. Third, we establish a variational formulation for the l0 pseudonorm by means of generalized top-k dual norms and k-support dual norms (that we formally introduce).

Key words: l0 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 objective function or in the constraints, to obtain solutions with few nonzero entries. However, the mathematical expression of the ℓ_0 pseudonorm makes it difficult to be handled as such in optimization problems. This is why most of the literature on sparse optimization resorts to *surrogate* problems, obtained either from *lower approximations* for the ℓ_0 pseudonorm, or from *alternative* sparsity-inducing terms (especially suitable norms). The literature on sparsity-inducing norms is huge, and we just point out a very succint part of it. We refer the reader to [16] that provides a brief tour of the literature dealing with least squares minimization constrained by k-sparsity, and to [9] for a survey of the rank function of a matrix, that shares many properties with the ℓ_0 pseudonorm. We refer the reader to [2] for the support norm, to [23] (and references therein) for top norms, and to [13] for generalizations.

Our approach to tackle the ℓ_0 pseudonorm uses so-called CAPRA (constant along primal rays) conjugacies, introduced in [4]. More precisely, in [4], we presented the class of couplings CAPRA (dependending on an underlying source norm) and we established expressions for CAPRA-conjugates and biconjugates, and CAPRA-subdifferentials of nondecreasing functions of the ℓ_0 pseudonorm. In [5], we introduced the coupling E-CAPRA related to the Euclidean norm and we showed that the ℓ_0 pseudonorm is E-CAPRA-convex and displays hidden convexity in the following sense. The ℓ_0 pseudonorm satisfies a *convex factorization property*: it can be written as the composition of a proper convex lower semi continuous (lsc) function with the normalization mapping that maps any nonzero vector onto the Euclidean unit sphere, hence it coincides with a proper convex lsc function on the Euclidean unit sphere.

In this paper, we go beyond the two above papers in several directions. We generalize the results of [5] by showing that not only the ℓ_0 pseudonorm but any nondecreasing function of the ℓ_0 pseudonorm is CAPRA-convex and displays hidden convexity (convex factorization property), and not only for the Euclidean norm but for a class of norms that encompasses it (including the ℓ_p-norms for $p \in]1, \infty[$). Moreover, we extend the hidden convexity property to subdifferentials. Indeed, we show that the CAPRA-subdifferential of a nondecreasing function of the ℓ_0 pseudonorm coincides, on the unit sphere, with the Rockafellar-Moreau subdifferential of the associated convex lsc function (in the convex factorization property). We also add the result that the CAPRA-subdifferential is a closed convex set. Whereas, in [4], we obtained CAPRA-convex lower bounds (inequalities) for nondecreasing functions of the ℓ_0 pseudonorm, we now obtain identities. Whereas we obtained an expression for the CAPRA-subdifferential of a nondecreasing function of the ℓ_0 pseudonorm, we now prove that it is not empty.

The paper is organized as follows. In Sect. 2, we provide background on the ℓ_0 pseudonorm and on CAPRA-conjugacies. We also introduce a new class of orthant-strictly monotonic norms, as well as sequences of generalized top-k and k-support dual norms. We show that nondecreasing functions of the ℓ_0 pseudonorm are CAPRA-convex. In Sect. 3, we show that any nondecreasing function of the ℓ_0 pseudonorm coincides, when restricted to the unit sphere, with a proper convex lsc function. Then, we deduce variational formulations for nondecreasing functions of the ℓ_0 pseudonorm which involve the sequence of generalized k-support dual norms. Appendix A gathers background on properties of norms that are relevant for the ℓ_0 pseudonorm, Appendix B reproduces [4, Proposition 4.5] to make easier the reading of proofs, and Appendix C gathers background on the Fenchel conjugacy.

2 Capra-convexity of the ℓ_0 pseudonorm with orthant-strictly monotonic norms

In §2.1, we provide background on the ℓ_0 pseudonorm and on the family of CAPRA conjugacies (introduced in [4]). Then, in §2.2, we introduce norms that are especially relevant for the ℓ_0 pseudonorm, like orthant-strictly monotonic norms. Finally, in §2.3, we prove that the ℓ_0 pseudonorm is CAPRA-convex when the underlying norm and its dual norm are both orthant-strictly monotonic.

2.1 Background on the ℓ_0 pseudonorm and the Capra conjugacy

We work on the Euclidean space $\mathbb{R}^d$ (where d is a nonzero integer), equipped with the scalar product $\langle \cdot , \cdot \rangle$ (but not necessarily with the Euclidean norm). We use the notation $[\![j, k]\!] = \{j, j+1, \ldots, k-1, k\}$ for any two integers j, k such that $j \leq k$.

Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$, that we will call the *source norm*. We denote the unit sphere $\mathbb{S}$ and the unit ball $\mathbb{B}$ of the source norm $|||\cdot|||$ by

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

For any vector $x \in \mathbb{R}^d$, $\mathrm{supp}(x) = \left\{ j \in [\![1, d]\!] \,\middle|\, x_j \neq 0 \right\} \subset [\![1, d]\!]$ is the support of x. The so-called ℓ_0 *pseudonorm* is the function $\ell_0 : \mathbb{R}^d \to [\![0, d]\!]$ 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 cardinality of a subset $K \subset [\![1, d]\!]$. 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. By contrast, the ℓ_0 pseudonorm is 0-homogeneous:

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

Following [4], we introduce the coupling CAPRA.

Definition 1 *([4, Definition 4.1]) The* constant along primal rays coupling $¢ : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, *or* CAPRA, *between $\mathbb{R}^d$ and itself, is the function*

$$¢ : (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto \begin{cases} \frac{\langle x , y \rangle}{|||x|||} \ , & x \neq 0 \ , \\ 0 \ , & \textit{else.} \end{cases} \tag{4}$$

In the case of the Capra coupling, the primal and dual space are the same space $\mathbb{R}^d$ but the Capra coupling is not symmetric in the primal and dual variables. To stress the point, we use the letter x for a primal vector and the letter y for a dual vector. We also underline that, in (4), the Euclidean scalar product $\langle x , y \rangle$ and the norm term $|||x|||$ need not be related, that is, the norm $|||\cdot|||$ is not necessarily the Euclidean norm.

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* $n : \mathbb{R}^d \to \mathbb{S} \cup \{0\}$, from $\mathbb{R}^d$ towards the unit sphere $\mathbb{S}$ united with $\{0\}$, as follows:

$$n : x \in \mathbb{R}^d \mapsto \begin{cases} \frac{x}{|||x|||} & x \neq 0 \,, \\ 0 \,, & \text{else.} \end{cases} \tag{5}$$

Now, we introduce notions and notation from generalized convexity [22, 21, 12]. As we manipulate functions with values in $\overline{\mathbb{R}} = [-\infty, +\infty]$, we adopt the Moreau *lower and upper additions* [15] that extend the usual addition with $(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = -\infty$ or with $(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = +\infty$. The *¢-Fenchel-Moreau conjugate* of a function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$, with respect to the coupling ¢, is the function $f^{¢} : \mathbb{R}^d \to \overline{\mathbb{R}}$ defined by

$$f^{¢}(y) = \sup_{x \in \mathbb{R}^d} \Big(¢(x, y) \dot{+} \big(-f(x)\big) \Big) \,, \ \ \forall y \in \mathbb{R}^d \,. \tag{6a}$$

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

$$f^{¢¢'}(x) = \sup_{y \in \mathbb{R}^d} \Big(¢(x, y) \dot{+} \big(-f^{¢}(y)\big) \Big) \,, \ \ \forall x \in \mathbb{R}^d \,. \tag{6b}$$

The biconjugate of a function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ satisfies the inequality

$$f^{¢¢'}(x) \leq f(x) \,, \ \ \forall x \in \mathbb{R}^d \,. \tag{6c}$$

When the coupling ¢ is replaced by the Euclidean scalar product $\langle \cdot \,, \cdot \rangle$, we recover well-known expressions of the Fenchel conjugacy (see Appendix C).

2.2 Relevant norms for the ℓ_0 pseudonorm

In §2.2.1, we recall the notions of restriction norms and of generalized coordinate-k and dual coordinate-k norms. In §2.2.2, we introduce two new families of norms, the generalized top-k and k-support dual norms. Finally, in §2.2.3, we define a new class of orthant-strictly monotonic norms.

2.2.1 Restriction norms, generalized coordinate-k norms, dual coordinate-k norms

For any subset $K \subset [\![1, d]\!]$, we define the subspace

$$\mathcal{R}_K = \big\{ x \in \mathbb{R}^d \,\big|\, x_j = 0 \,, \ \ \forall j \notin K \big\} \subset \mathbb{R}^d \tag{7}$$

with $\mathcal{R}_\emptyset = \{0\}$, and then three norms on the subspace $\mathcal{R}_K$ of $\mathbb{R}^d$, as follows.

- The *K-restriction norm* $|||\cdot|||_K$ is defined by $|||x|||_K = |||x|||$, for any $x \in \mathcal{R}_K$.

- The $(\star, K)$-norm $|||\cdot|||_{\star,K}$ is the norm $\left(|||\cdot|||_{\star}\right)_{K}$, given by the restriction to the subspace $\mathcal{R}_K$ of the dual norm $|||\cdot|||_{\star}$ (first dual, as recalled in definition (27) of a dual norm, then restriction),
- The $(K, \star)$-norm $|||\cdot|||_{K,\star}$ is the norm $\left(|||\cdot|||_{K}\right)_{\star}$, 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).

For any $x \in \mathbb{R}^d$ and subset $K \subset [\![1, d]\!]$, we denote by $x_K \in \mathcal{R}_K \subset \mathbb{R}^d$ the vector which coincides with x, except for the components outside of K that vanish (this definition is valid for $K = \emptyset$, giving $x_\emptyset = 0 \in \mathcal{R}_\emptyset = \{0\}$).

Definition 2 *([4, Definition 3.2]) For $k \in [\![1, d]\!]$, the expression (where the notation $\sup_{|K|\leq k}$ is a shorthand for $\sup_{K\subset[\![1,d]\!],|K|\leq k}$)*

$$|||y|||^{\mathcal{R}}_{(k),\star} = \sup_{|K|\leq k} |||y_K|||_{K,\star} \ , \ \ \forall y \in \mathbb{R}^d \tag{8}$$

defines a norm on $\mathbb{R}^d$, called the generalized dual coordinate-k norm $|||\cdot|||^{\mathcal{R}}_{(k),\star}$. *Its dual norm is the* generalized coordinate-k norm, *denoted by* $|||\cdot|||^{\mathcal{R}}_{(k)}$.

We denote the unit sphere $\mathbb{S}^{\mathcal{R}}_{(k),\star}$ and the unit ball $\mathbb{B}^{\mathcal{R}}_{(k),\star}$ by: $\forall k \in [\![1, d]\!]$,

$$\mathbb{S}^{\mathcal{R}}_{(k),\star} = \left\{ y \in \mathbb{R}^d \,\middle|\, |||y|||^{\mathcal{R}}_{(k),\star} = 1 \right\} \ , \ \ \mathbb{B}^{\mathcal{R}}_{(k),\star} = \left\{ y \in \mathbb{R}^d \,\middle|\, |||y|||^{\mathcal{R}}_{(k),\star} \leq 1 \right\} . \tag{9}$$

We give examples of generalized coordinate-k and dual coordinate-k norms in [4, Table 1].

2.2.2 Generalized top-k and k-support dual norms

We introduce two new families of norms, that we call generalized top-k and k-support dual norms.

Definition 3 *For $k \in [\![1, d]\!]$, the expression*

$$|||y|||^{\mathrm{tn}}_{\star,(k)} = \sup_{|K|\leq k} |||y_K|||_{\star} = \sup_{|K|\leq k} |||y_K|||_{\star,K} \ , \ \ \forall y \in \mathbb{R}^d \tag{10}$$

defines a norm on $\mathbb{R}^d$, called the generalized top-k dual norm *(associated with the source norm $|||\cdot|||$). Its dual norm*

$$|||\cdot|||^{\star\mathrm{sn}}_{\star,(k)} = \left(|||\cdot|||^{\mathrm{tn}}_{\star,(k)}\right)_{\star} \ , \ \ \forall k \in [\![1, d]\!] \tag{11}$$

is called generalized k-support dual norm. *It has unit sphere $\mathbb{S}^{\star\mathrm{sn}}_{\star,(k)}$ and unit ball $\mathbb{B}^{\star\mathrm{sn}}_{\star,(k)}$ given by: $\forall k \in [\![1, d]\!]$,*

$$\mathbb{S}^{\star\mathrm{sn}}_{\star,(k)} = \left\{ x \in \mathbb{R}^d \,\middle|\, |||x|||^{\star\mathrm{sn}}_{\star,(k)} = 1 \right\} \ , \ \ \mathbb{B}^{\star\mathrm{sn}}_{\star,(k)} = \left\{ x \in \mathbb{R}^d \,\middle|\, |||x|||^{\star\mathrm{sn}}_{\star,(k)} \leq 1 \right\} . \tag{12}$$

We use the symbol $\star$ in the superscript in Equation (11) to indicate that the generalized k-support dual norm $|||\cdot|||^{\star\mathrm{sn}}_{\star,(k)}$ is a dual norm. To stress the point, we use the letter x for a primal vector, like in $|||x|||^{\star\mathrm{sn}}_{\star,(k)}$, and the letter y for a dual vector, like in $|||y|||^{\mathrm{tn}}_{\star,(k)}$. We also adopt the conventions $|||\cdot|||^{\mathrm{tn}}_{\star,(0)} = 0$ and $|||\cdot|||^{\star\mathrm{sn}}_{\star,(0)} = 0$, although these are not norms but seminorms.

We now give examples of generalized top-k and k-support dual norms in the case of ℓ_p source norm. We recall that the ℓ_p-norms $\|\cdot\|_p$ on the space $\mathbb{R}^d$ are defined by $\|x\|_p = \big(\sum_{i=1}^d |x_i|^p\big)^{\frac{1}{p}}$ for $p \in [1,\infty[$, and by $\|x\|_\infty = \sup_{i\in[\![1,d]\!]} |x_i|$. It is well-known that the dual norm of the norm $\|\cdot\|_p$ is the ℓ_q-norm $\|\cdot\|_q$, where q is such that $1/p + 1/q = 1$ (with the extreme cases $q = \infty$ when $p = 1$, and $q = 1$ when $p = \infty$).

We start with a Lemma, whose proof is easy. For any $y \in \mathbb{R}^d$, we denote by $|y| = (|y_1|,\ldots,|y_d|)$ the vector of $\mathbb{R}^d$ with components $|y_i|$, $i \in [\![1,d]\!]$. Letting $y \in \mathbb{R}^d$ and ν be a permutation of $[\![1,d]\!]$ such that $|y_{\nu(1)}| \geq |y_{\nu(2)}| \geq \cdots \geq |y_{\nu(d)}|$, we denote $y^{\downarrow} = \big(|y_{\nu(1)}|, |y_{\nu(2)}|,\ldots,|y_{\nu(d)}|\big)$.

Lemma 4 *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$. Then, if the norm $|||\cdot|||$ is permutation invariant and monotonic — that is, for any x, x' in $\mathbb{R}^d$, we have $|x| \leq |x'| \Rightarrow |||x||| \leq |||x'|||$, where $|x| \leq |x'|$ means $|x_i| \leq |x'_i|$ for all $i \in [\![1,d]\!]$ — we have that $|||y|||^{\mathrm{tn}}_{\star,(k)} = |||y^{\downarrow}_{[\![1,k]\!]}|||_\star$, where $y^{\downarrow}_{[\![1,k]\!]} \in \mathbb{R}^d$ is given by $(y^{\downarrow})_{[\![1,k]\!]}$, for all $y \in \mathbb{R}^d$.*

We first present examples of generalized top-k dual norms as in (10) (see the second column of Table 1). When the norm $|||\cdot|||$ is the Euclidean norm $\|\cdot\|_2$ of $\mathbb{R}^d$, the generalized top-k dual norm is known under different names: the top-$(k,2)$ norm in [23, p. 8], or the 2-k-symmetric gauge norm [14] or the Ky Fan vector norm [17]. Indeed, in all these cases, the norm of a vector y is obtained with a subvector of size k having the k largest absolute values of the components, because the assumptions of Lemma 4 are satisfied. More generally, when the norm $|||\cdot|||$ is the ℓ_p-norm $\|\cdot\|_p$, for $p \in [1,\infty]$, the assumptions of Lemma 4 are also satisfied, as ℓ_p-norms are permutation invariant and monotonic. Therefore, we obtain that the corresponding generalized top-k dual norm $\big(\|\cdot\|_p\big)^{\mathrm{tn}}_{\star,(k)}$ has the expression $\big(\|\cdot\|_p\big)^{\mathrm{tn}}_{\star,(k)}(y) = \sup_{|K|\leq k} \|y_K\|_q = \|y^{\downarrow}_{[\![1,k]\!]}\|_q$, for all $y \in \mathbb{R}^d$, and where $1/p + 1/q = 1$. Notice that $\big(\|\cdot\|_p\big)^{\mathrm{tn}}_{\star,(k)}$ is expressed in function of q, which can be misleading (this phenomenon is manifest in Table 1). When the source norm $|||\cdot|||$ is the ℓ_p-norm $||\cdot||_p$, the generalized top-k dual norm $|||\cdot|||^{\mathrm{tn}}_{\star,(k)}$ in (10) is called the*top-(q,k) norm* and is denoted by $||\cdot||^{\mathrm{tn}}_{q,k}$ (we invert the indices in the naming convention of [23, p. 5, p. 8], where top-$(k,1)$ and top-$(k,2)$ were used). Notice that $||\cdot||^{\mathrm{tn}}_{\infty,k} = \|\cdot\|_\infty$ for all $k \in [\![1,d]\!]$.

Now, we turn to examples of generalized k-support dual norms as in (11) (see the third column of Table 1). When the norm $|||\cdot|||$ is the Euclidean norm $\|\cdot\|_2$ of $\mathbb{R}^d$, the generalized k-support norm is the so-called k-support norm [2]. More generally, in [13, Definition 21], the authors define the k-support p-norm or *(p,k)-support norm* for $p \in [1,\infty]$. They show, in [13, Corollary 22], that the dual norm $\big((\|\cdot\|_p)^{\mathrm{tn}}_{(k)}\big)_\star$ of the above top-(p,k) norm is the (q,k)-support norm, where $1/p + 1/q = 1$. Therefore, the generalized k-support dual norm in (11) is the (p,k)-support norm — denoted by $||\cdot||^{\mathrm{sn}}_{p,k}$ — when the source norm $|||\cdot|||$ is the

source norm $\vert\vert\vert\cdot\vert\vert\vert$	$\vert\vert\vert\cdot\vert\vert\vert^{\mathrm{tn}}_{\star,(k)}$, $k \in [\![1,d]\!]$	$\vert\vert\vert\cdot\vert\vert\vert^{\star\mathrm{sn}}_{\star,(k)}$, $k \in [\![1,d]\!]$
$\vert\vert\cdot\vert\vert_p$	top-(q,k) norm $\vert\vert y\vert\vert^{\mathrm{tn}}_{q,k}$ $\vert\vert y\vert\vert^{\mathrm{tn}}_{q,k} = \big(\sum_{l=1}^{k} \vert y_{\nu(l)}\vert^q\big)^{\frac{1}{q}}$	(p,k)-support norm $\vert\vert x\vert\vert^{\mathrm{sn}}_{p,k}$ no analytic expression
$\vert\vert\cdot\vert\vert_1$	top-(∞,k) norm ℓ_∞-norm $\vert\vert y\vert\vert^{\mathrm{tn}}_{\infty,k} = \vert\vert y\vert\vert_\infty$, $\forall k \in [\![1,d]\!]$	(1,k)-support norm ℓ_1-norm $\vert\vert x\vert\vert^{\mathrm{sn}}_{1,k} = \vert\vert x\vert\vert_1$, $\forall k \in [\![1,d]\!]$
$\vert\vert\cdot\vert\vert_2$	top-(2,k) norm $\vert\vert y\vert\vert^{\mathrm{tn}}_{2,k} = \sqrt{\sum_{l=1}^{k} \vert y_{\nu(l)}\vert^2}$ $\vert\vert y\vert\vert^{\mathrm{tn}}_{2,1} = \vert\vert y\vert\vert_\infty$	(2,k)-support norm $\vert\vert x\vert\vert^{\mathrm{sn}}_{2,k}$ no analytic expression (computation [2, Prop. 2.1]) $\vert\vert x\vert\vert^{\mathrm{sn}}_{2,1} = \vert\vert x\vert\vert_1$
$\vert\vert\cdot\vert\vert_\infty$	top-(1,k) norm $\vert\vert y\vert\vert^{\mathrm{tn}}_{1,k} = \sum_{l=1}^{k} \vert y_{\nu(l)}\vert$ $\vert\vert y\vert\vert^{\mathrm{tn}}_{1,1} = \vert\vert y\vert\vert_\infty$	(∞,k)-support norm $\vert\vert x\vert\vert^{\mathrm{sn}}_{\infty,k} = \max\{\frac{\vert\vert x\vert\vert_1}{k}, \vert\vert x\vert\vert_\infty\}$ $\vert\vert x\vert\vert^{\mathrm{sn}}_{1,1} = \vert\vert x\vert\vert_1$

Table 1: Examples of generalized top-k and k-support dual norms generated by the ℓ_p source norms $|||\cdot||| = ||\cdot||_p$ for $p \in [1,\infty]$, where $1/p + 1/q = 1$. For $y \in \mathbb{R}^d$, ν denotes a permutation of $\{1,\ldots,d\}$ such that $|y_{\nu(1)}| \geq |y_{\nu(2)}| \geq \cdots \geq |y_{\nu(d)}|$.

ℓ_p-norm $||\cdot||_p$, for $p \in [1,\infty]$. The formula $||x||^{\mathrm{sn}}_{\infty,k} = \max\{||x||_1/k, ||x||_\infty\}$ can be found in [3, Exercise IV.1.18, p. 90].

2.2.3 Orthant-monotonic and orthant-strictly monotonic norms

We recall the definition of orthant-monotonic norms and we introduce the new definition of orthant-strictly monotonic norms, that will prove especially relevant for the ℓ_0 pseudonorm.

Definition 5 *A norm $|||\cdot|||$ on the space $\mathbb{R}^d$ is called*

- orthant-monotonic *[8, Definition 2.6] 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 $i \in [\![1,d]\!]$, and where $x \circ x' = (x_1x'_1, \ldots, x_dx'_d)$ is the Hadamard (entrywise) product,*
- orthant-strictly monotonic *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 $i \in [\![1,d]\!]$, and there exists $j \in [\![1,d]\!]$, such that $|x_j| < |x'_j|$.*

All the ℓ_p-norms $\|\cdot\|_p$ on the space $\mathbb{R}^d$, for $p \in [1,\infty[$, are strictly monotonic, hence orthant-strictly monotonic. By contrast, the ℓ_∞-norm $\|\cdot\|_\infty$ is orthant-monotonic but not orthant-strictly monotonic.

2.3 Capra-convexity of the ℓ_0 pseudonorm

The main result of this section is Theorem 8 which states that, when both the source norm $|||\cdot|||$ and its dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic, then any nondecreasing function of the ℓ_0 pseudonorm is equal to its CAPRA-biconjugate, that is, is a CAPRA-convex function. This considerably generalizes the result in [5, Theorem 3.5], which was established for the Euclidean norm and only for the ℓ_0 pseudonorm.

The proof of Theorem 8 relies on Proposition 7, which establishes the nonemptiness of a suitable CAPRA-subdifferential. In [4, Equation (32)], we define the CAPRA-*subdifferential* of the function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ at $x \in \mathbb{R}^d$ by

$$\partial_{\cent} f(x) = \big\{ y \in \mathbb{R}^d \,\big|\, f^{\cent}(y) = \cent(x,y) \dotplus \big(-f(x)\big) \big\} \;, \tag{13}$$

where $f^{\cent}(y)$ has been defined in (6a).

Proposition 6 *Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be any function. For all $x \in \mathbb{R}^d$, the* CAPRA*-subdifferential $\partial_{\cent} f(x)$ is a closed convex set.*

Proof. We prove that $\partial_{\cent} f(x)$, as in (13), is a closed convex set. Let $x \in \mathbb{R}^d$.

By definition (6a) of $f^{\cent}$, the CAPRA-subdifferential (13) can be written as

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

where we use the usual addition because $-\infty < \cent(x,y) < +\infty$ by (4).

As a consequence, when $f(x) = -\infty$, we get that $\partial_{\cent} f(x) = \mathbb{R}^d$, which is closed and convex. In the case where $f(x) = +\infty$, we have that $\partial_{\cent} f(x) = \emptyset$ if f is not identically $+\infty$, and that $\partial_{\cent} f(x) = \mathbb{R}^d$ otherwise; in either cases, the CAPRA-subdifferential is closed and convex. Now, suppose that $f(x) \in \mathbb{R}$. By definition (6a) of $f^{\cent}$, the CAPRA-subdifferential (13) can be written as $\partial_{\cent} f(x) = \big\{ y \in \mathbb{R}^d \,\big|\, f^{\cent}(y) \leq \cent(x,y) - f(x) \big\}$ where the function $f^{\cent}$ is a Fenchel conjugate by [4, Equation (30b)], hence is closed convex (see the background material in Appendix C), and the function $g_x : \mathbb{R}^d \ni y \mapsto \cent(x,y) - f(x)$ is affine. As a consequence, $\partial_{\cent} f(x) = \big\{ y \in \mathbb{R}^d \,\big|\, f^{\cent}(y) - g_x(y) \leq 0 \big\}$ is a closed convex set. □

It follows that the CAPRA-subdifferential of the ℓ_0 pseudonorm is a (possibly empty) closed convex set (by contrast, it is shown in [9, Section 8] that all the generalized [Fenchel] subdifferentials [proximal, Fréchet, viscosity, limiting, Clarke] of the rank function coincide and define a vector space). We now provide conditions under which the CAPRA-subdifferential of any nondecreasing function of the ℓ_0 pseudonorm is not empty.

Proposition 7 *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$ with associated* CAPRA *coupling $\cent$ as in* (4)*. If both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic, and if $\varphi : [\![0,d]\!] \to \mathbb{R}$ is a nondecreasing function, then*

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

More precisely, when $x = 0$*, we have that* $\partial_{\dot{C}}(\varphi \circ \ell_0)(0) = \bigcap_{j \in [\![1,d]\!]} \big[\varphi(j) - \varphi(0)\big] \mathbb{B}^{\mathcal{R}}_{(j),\star} \neq \emptyset$*, where the unit ball* $\mathbb{B}^{\mathcal{R}}_{(k),\star}$ *is defined in (9). When* $x \neq 0$*, there exists* $y \in \mathbb{R}^d$ *satisfying* $\mathrm{supp}(y) = \mathrm{supp}(x)$ *and* $\langle x, y \rangle = |\!|\!| x |\!|\!| \times |\!|\!| y |\!|\!|_\star$*, and for all such* $y \in \mathbb{R}^d$ *we have that* $\lambda y \in \partial_{\dot{C}}(\varphi \circ \ell_0)(x)$ *for* $\lambda > 0$ *large enough.*

Proof. The proof relies on results established in Appendix A.

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 $j \in [\![0,d]\!]$ by (32) in Proposition 14 in Appendix A.3 (with the convention that these are the null seminorms in the case $j = 0$). Therefore, we can translate all the results with generalized top-k and k-support dual norms (Definition 3) instead of coordinate-k and dual coordinate-k norms (Definition 2).

When $x = 0$, we have by [4, Equation (39) in Proposition 4.7] that

$$\partial_{\dot{C}}(\varphi \circ \ell_0)(0) = \bigcap_{j \in [\![1,d]\!]} \big[\varphi(j) \dot{+} \big(-\varphi(0)\big)\big] \mathbb{B}^{\mathcal{R}}_{(j),\star} = \bigcap_{j \in [\![1,d]\!]} \big[\varphi(j) - \varphi(0)\big] \mathbb{B}^{\mathcal{R}}_{(j),\star} \ni 0 \; ,$$

because $\varphi(j) \dot{+} \big(-\varphi(0)\big) = \varphi(j) - \varphi(0) \geq 0$ since $\varphi : [\![0,d]\!] \to \mathbb{R}$ is a nondecreasing function.

From now on, we consider $x \in \mathbb{R}^d \setminus \{0\}$ such that $\ell_0(x) = l \in [\![1,d]\!]$, and we will use the following equivalence, established in [4, Equation (40) in Proposition 4.7]

$$y \in \partial_{\dot{C}}(\varphi \circ \ell_0)(x) \iff \begin{cases} y \in N_{\mathbb{B}^{\mathcal{R}}_{(l)}}\big(\frac{x}{|\!|\!| x |\!|\!|^{\mathcal{R}}_{(l)}}\big) \\ \text{and } \; l \in \arg\max_{j \in [\![0,d]\!]} \big[|\!|\!| y |\!|\!|^{\mathcal{R}}_{(j),\star} - \varphi(j)\big] \; , \end{cases} \tag{14}$$

where the normal cone $N_{\mathbb{B}^{\mathcal{R}}_{(l)}}\big(\frac{x}{|\!|\!| x |\!|\!|^{\mathcal{R}}_{(l)}}\big)$ is defined in (29).

Since the norm $|\!|\!| \cdot |\!|\!|$ is orthant-strictly monotonic, we know by Item (***b***) in Proposition 13 (Appendix A.2) that there exists a vector $y \in \mathbb{R}^d$ such that

$$L = \mathrm{supp}(x) = \mathrm{supp}(y) \;\text{ hence }\; \ell_0(y) = \ell_0(x) = l > 1 \; , \tag{15a}$$

$$\langle x, y \rangle = |\!|\!| x |\!|\!| \times |\!|\!| y |\!|\!|_\star \; . \tag{15b}$$

Since the dual norm $|\!|\!| \cdot |\!|\!|_\star$ is orthant-strictly monotonic, we know by (33) in Proposition 15 (Appendix A.3) that

$$|\!|\!| y |\!|\!|^{\mathcal{R}}_{(1),\star} < \cdots < |\!|\!| y |\!|\!|^{\mathcal{R}}_{(l-1),\star} < |\!|\!| y |\!|\!|^{\mathcal{R}}_{(l),\star} = \cdots = |\!|\!| y |\!|\!|^{\mathcal{R}}_{(d),\star} = |\!|\!| y |\!|\!|_\star \; . \tag{16}$$

We now show that $y \in \partial_{\dot{C}}(\varphi \circ \ell_0)(x)$.

First, we are going to establish that $y \in N_{\mathbb{B}^{\mathcal{R}}_{(l)}}\big(\frac{x}{|\!|\!| x |\!|\!|^{\mathcal{R}}_{(l)}}\big)$, that is, the first of the two conditions in the characterization (14) of the subdifferential $\partial_{\dot{C}}(\varphi \circ \ell_0)(x)$.

On the one hand, because $\ell_0(y) = l$ and by (16), we have that $|\!|\!| y |\!|\!|_\star = |\!|\!| y |\!|\!|^{\mathcal{R}}_{(l),\star}$. On the other hand, because $\ell_0(x) = l$ we have that $|\!|\!| x |\!|\!| = |\!|\!| x |\!|\!|^{\mathcal{R}}_{(l)}$ by [4, Equation (25a)]. Hence, from (15b), we get that $\langle x, y \rangle = |\!|\!| x |\!|\!|^{\mathcal{R}}_{(l)} \times |\!|\!| y |\!|\!|^{\mathcal{R}}_{(l),\star}$, from which we obtain $y \in N_{\mathbb{B}^{\mathcal{R}}_{(l)}}\big(\frac{x}{|\!|\!| x |\!|\!|^{\mathcal{R}}_{(l)}}\big)$ by property (30) of the normal cone as $x \neq 0$. To close this part, notice that, for all $\lambda > 0$, we have that $\lambda y \in N_{\mathbb{B}^{\mathcal{R}}_{(l)}}\big(\frac{x}{|\!|\!| x |\!|\!|^{\mathcal{R}}_{(l)}}\big)$, because this last set is a cone.

Second, we prove the other of the two conditions in the characterization (14) of the subdifferential $\partial_{\cent}(\varphi \circ \ell_0)(x)$. More precisely, we are going to show that, for λ large enough, $|||\lambda y|||^{\mathcal{R}}_{(l),\star} - \varphi(l) = \sup_{j\in[\![0,d]\!]} \big[|||\lambda y|||^{\mathcal{R}}_{(j),\star} - \varphi(j)\big]$. For this purpose, we consider the mapping $\psi :]0, +\infty[\to \mathbb{R}$ defined by

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

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

$$\begin{aligned}
\psi(\lambda) &= \inf_{j\in[\![0,d]\!]} \Big(\lambda\big(|||y|||^{\mathcal{R}}_{(l),\star} - |||y|||^{\mathcal{R}}_{(j),\star}\big) + \varphi(j) - \varphi(l)\Big) \\
&= \inf \Big\{ \lambda|||y|||^{\mathcal{R}}_{(l),\star} + \varphi(0) - \varphi(l), \inf_{j\in[\![1,l-1]\!]} \Big(\lambda\big(|||y|||^{\mathcal{R}}_{(l),\star} - |||y|||^{\mathcal{R}}_{(j),\star}\big) + \varphi(j) - \varphi(l)\Big), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\text{as } |||y|||^{\mathcal{R}}_{(0),\star} = 0 \text{ by convention}) \\
&\qquad \inf_{j\in[\![l,d]\!]} \Big(\lambda\big(|||y|||^{\mathcal{R}}_{(l),\star} - |||y|||^{\mathcal{R}}_{(j),\star}\big) + \varphi(j) - \varphi(l)\Big) \Big\} \\
&= \inf \Big\{ \lambda|||y|||^{\mathcal{R}}_{(l),\star} + \varphi(0) - \varphi(l), \inf_{j\in[\![1,l-1]\!]} \Big(\lambda\big(|||y|||^{\mathcal{R}}_{(l),\star} - |||y|||^{\mathcal{R}}_{(j),\star}\big) + \varphi(j) - \varphi(l)\Big), \\
&\qquad \inf_{j\in[\![l,d]\!]} \big(\varphi(j) - \varphi(l)\big) \Big\} \qquad\qquad (\text{as } |||y|||^{\mathcal{R}}_{(j),\star} = |||y|||^{\mathcal{R}}_{(l),\star} \text{ for } j \geq l \text{ by (16)}) \\
&= \inf \Big\{ \lambda|||y|||^{\mathcal{R}}_{(l),\star} + \varphi(0) - \varphi(l), \inf_{j\in[\![1,l-1]\!]} \Big(\lambda\big(|||y|||^{\mathcal{R}}_{(l),\star} - |||y|||^{\mathcal{R}}_{(j),\star}\big) + \varphi(j) - \varphi(l)\Big), 0 \Big\} \ ,
\end{aligned}$$

as $\inf_{j\in[\![l,d]\!]} \big(\varphi(j) - \varphi(l)\big) = 0$ because $\varphi : [\![0,d]\!] \to \mathbb{R}$ 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 (16), we have that $|||y|||^{\mathcal{R}}_{(l),\star} = |||y|||_\star > 0$ as $y \in \mathbb{R}^d \setminus \{0\}$ since $\ell_0(y) = l \geq 1$. The second term $\inf_{j\in[\![1,l-1]\!]} \Big(\lambda\big(|||y|||^{\mathcal{R}}_{(l),\star} - |||y|||^{\mathcal{R}}_{(j),\star}\big) + \varphi(j) - \varphi(l)\Big)$ also goes to $+\infty$ because $\ell_0(y) = l \geq 1$, so that $|||y|||_\star = |||y|||^{\mathcal{R}}_{(l),\star} > |||y|||^{\mathcal{R}}_{(j),\star}$ for $j \in [\![1, l-1]\!]$ again by (16). Therefore, we deduce that $\psi(\lambda) = 0$ for λ large enough, and thus $|||\lambda y|||^{\mathcal{R}}_{(l),\star} - \varphi(l) = \sup_{j\in[\![0,d]\!]} \big[|||\lambda y|||^{\mathcal{R}}_{(j),\star} - \varphi(j)\big]$, that is, $l \in \arg\max_{j\in[\![0,d]\!]} \big[|||\lambda y|||^{\mathcal{R}}_{(j),\star} - \varphi(j)\big]$.

Wrapping up the above results, we have shown that, for any vector $y \in \mathbb{R}^d$ such that $\mathrm{supp}(y) = \mathrm{supp}(x)$, and that $\langle x, y\rangle = |||x||| \times |||y|||_\star$, then, for $\lambda > 0$ large enough, λy satisfies the two conditions in the characterization (14) of the subdifferential $\partial_{\cent}(\varphi \circ \ell_0)(x)$. Hence, we get that $\lambda y \in \partial_{\cent}(\varphi \circ \ell_0)(x)$.

This ends the proof. $\square$

Theorem 8 *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$ with associated sequence $\big\{|||\cdot|||^{\mathrm{tn}}_{\star,(j)}\big\}_{j\in[\![1,d]\!]}$ of generalized top-k dual norms, as in Definition 3, and with associated* CAPRA *coupling $\cent$ as in (4).*
If the norm $|||\cdot|||$ is orthant-monotonic, then, for any function $\varphi : [\![0,d]\!] \to \mathbb{R}$, we have that (we recall the convention that $|||\cdot|||^{\mathrm{tn}}_{\star,(0)} = 0$)

$$(\varphi \circ \ell_0)^{\cent} = \sup_{j\in[\![0,d]\!]} \Big[|||\cdot|||^{\mathrm{tn}}_{\star,(j)} - \varphi(j)\Big] \ . \tag{17}$$

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

$$(\varphi \circ \ell_0)^{¢¢'} = \varphi \circ \ell_0 \; . \tag{18}$$

Proof. Suppose that the norm $\|\!|\cdot|\!\|$ is orthant-monotonic. Then, by (32) in Proposition 14 (Appendix A.3), we get that $\|\!|\cdot|\!\|^{\mathcal{R}}_{(k)} = \|\!|\cdot|\!\|^{\star\mathrm{sn}}_{\star,(k)}$ and $\|\!|\cdot|\!\|^{\mathcal{R}}_{(k),\star} = \|\!|\cdot|\!\|^{\mathrm{tn}}_{\star,(k)}$. Moreover, it is proved in [4, Equation (33) in Proposition 4.4] that $(\varphi \circ \ell_0)^{¢} = \sup_{j \in [\![0,d]\!]} \Big[\|\!|\cdot|\!\|^{\mathcal{R}}_{(j),\star} - \varphi(j) \Big]$. As $\|\!|\cdot|\!\|^{\mathcal{R}}_{(j),\star} = \|\!|\cdot|\!\|^{\mathrm{tn}}_{\star,(j)}$, we obtain (17).

Suppose that both the norm $\|\!|\cdot|\!\|$ and the dual norm $\|\!|\cdot|\!\|_\star$ are orthant-strictly monotonic. Then, Proposition 7 applies. Therefore, for any vector $x \in \mathbb{R}^d$ and any $y \in \partial_{¢}(\varphi \circ \ell_0)(x) \neq \emptyset$, we obtain

$$
\begin{aligned}
(\varphi \circ \ell_0)^{¢¢'}(x) &\geq ¢(x,y) \dot{+} \big(-(\varphi \circ \ell_0)^{¢}(y) \big) && \text{(by definition (6b) of the biconjugate)} \\
&= ¢(x,y) - (\varphi \circ \ell_0)^{¢}(y) \\
&\quad \text{(because } -\infty < ¢(x,y) < +\infty \text{ by (4), so that the usual addition applies)} \\
&= ¢(x,y) - \big(¢(x,y) - (\varphi \circ \ell_0)(x) \big)
\end{aligned}
$$

by definition (13) of the CAPRA-subdifferential $\partial_{¢}(\varphi \circ \ell_0)(x)$, and where again we can use the usual addition

$$= (\varphi \circ \ell_0)(x) \; .$$

On the other hand, we have that $(\varphi \circ \ell_0)^{¢¢'}(x) \leq (\varphi \circ \ell_0)(x)$ by (6c). We conclude that $(\varphi \circ \ell_0)^{¢¢'}(x) = (\varphi \circ \ell_0)(x)$, which is (18).

This ends the proof. □

Our proof of Proposition 7, hence of Theorem 8, uses the property that the nondecreasing function $\varphi : [\![0,d]\!] \to \mathbb{R}$ takes finite values. What happens for a function φ with values in the extended real numbers? A typical case of $\varphi : [\![0,d]\!] \to \overline{\mathbb{R}}$ is that of the characteristic function of a subset $K \subset [\![1,d]\!]$: $\delta_K(j) = 0$ if $j \in K$, and $\delta_K(j) = +\infty$ if $j \notin K$. Regarding CAPRA-convexity of a nondecreasing function of the ℓ_0 pseudonorm taking infinite values, we suspect that a proof would rely on different assumptions than those of Proposition 7 and Theorem 8. As an indication, the comment after the proof of [4, Corollary 4.6] points out that, when the normed space $\big(\mathbb{R}^d, \|\!|\cdot|\!\|\big)$ is strictly convex (that is, when the unit ball $\mathbb{B}$ is rotund), then $\delta_{[\![0,k]\!]} \circ \ell_0$ is $¢$-convex for $k \in [\![0,d]\!]$. As the normed space $\big(\mathbb{R}^d, \|\cdot\|_p\big)$, equipped with the ℓ_p-norm $\|\cdot\|_p$ (for $p \in [1,\infty]$), is strictly convex if and only if $p \in]1,\infty[$, we get that the characteristic functions of the level sets of the ℓ_0 pseudonorm are $¢$-convex when the source norm $\|\!|\cdot|\!\| = \|\cdot\|_p$ for $p \in]1,\infty[$.

3 Convex factorization and variational formulation for the ℓ_0 pseudonorm

In this section, we suppose that both the norm $\|\!|\cdot|\!\|$ and the dual norm $\|\!|\cdot|\!\|_\star$ are orthant-strictly monotonic. In §3.1, we show that any nonnegative nondecreasing function of the pseudonorm ℓ_0 coincides, on the unit sphere, with a proper convex lsc function on $\mathbb{R}^d$, and we provide various expressions for this latter function. In §3.2, we deduce a variational formula for nonnegative nondecreasing functions of the ℓ_0 pseudonorm.

3.1 Convex factorization and hidden convexity in the ℓ_0 pseudonorm

We now present a (rather unexpected) consequence of the just established property that any nondecreasing function of the pseudonorm ℓ_0 is CAPRA-convex (Theorem 8). Indeed, we prove that any nonnegative nondecreasing function of the pseudonorm ℓ_0 coincides, on the unit sphere $\mathbb{S} = \{x \in \mathbb{R}^d \mid |||x||| = 1\}$, with a proper convex lsc function on $\mathbb{R}^d$, and that this property extends to subdifferentials. We also provide various expressions for the underlying proper convex lsc function.

In this part, we make use of the Fenchel conjugacy, denoted by * in (37a), and of the reverse Fenchel conjugacy $^{\star'}$ in (37b) (see Appendix C). In the case of the Fenchel conjugacy, the primal and dual space are the same space $\mathbb{R}^d$ *and* the scalar product coupling is *symmetric* in the primal and dual variables. Hence, we could use the notation * in (37b) instead of $^{\star'}$. By contrast, in the case of the Capra conjugacy, the primal and dual space are the same space $\mathbb{R}^d$ *but* the Capra coupling is *not symmetric* in the primal and dual variables. Hence, $¢' \neq ¢$ and we cannot use the notation $^{¢¢}$, but we have to use the notation $^{¢¢'}$. For the sake of consistency, we maintain the notations $^{\star'}$ and $^{\star\star'}$.

Proposition 9 *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$ with associated sequence $\{|||\cdot|||^{\mathrm{tn}}_{\star,(j)}\}_{j\in[\![1,d]\!]}$ of generalized top-k dual norms, and sequence $\{|||\cdot|||^{\star\mathrm{sn}}_{\star,(j)}\}_{j\in[\![1,d]\!]}$ of generalized k-support dual norms, and with associated* CAPRA *coupling $¢$. Suppose that both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_{\star}$ are orthant-strictly monotonic. Let $\varphi : [\![0,d]\!] \to \mathbb{R}_+$ be a nonnegative nondecreasing function, such that $\varphi(0) = 0$. We define the function $\mathcal{L}_0^{\varphi} : \mathbb{R}^d \to \overline{\mathbb{R}}$ by*

$$\mathcal{L}_0^{\varphi} = \big((\varphi \circ \ell_0)^{¢}\big)^{\star'} \; . \tag{19}$$

Then, the following statements hold true.

(a) *The function $\mathcal{L}_0^{\varphi} : \mathbb{R}^d \to \overline{\mathbb{R}}$ is proper convex lsc.*

(b) *The function $\varphi \circ \ell_0$ coincides, on the unit sphere $\mathbb{S} = \{x \in \mathbb{R}^d \mid |||x||| = 1\}$, with the function $\mathcal{L}_0^{\varphi}$, that is,*

$$(\varphi \circ \ell_0)(s) = \mathcal{L}_0^{\varphi}(s) \; , \;\; \forall s \in \mathbb{S} \; . \tag{20a}$$

(c) *The* CAPRA*-subdifferential of the function $\varphi \circ \ell_0$ coincides, on the unit sphere $\mathbb{S}$, with the (Rockafellar-Moreau) subdifferential of the function $\mathcal{L}_0^{\varphi}$, that is,*

$$\partial_{¢}(\varphi \circ \ell_0)(s) = \partial\mathcal{L}_0^{\varphi}(s) \; , \;\; \forall s \in \mathbb{S} \; . \tag{20b}$$

(d) *Convex factorization property. The function* $\varphi \circ \ell_0$ *can be expressed as the composition of the proper convex lsc function* $\mathcal{L}_0^{\varphi}$ *with the normalization mapping* n*, that is,*

$$\varphi \circ \ell_0 = \mathcal{L}_0^{\varphi} \circ n \tag{20c}$$

or, equivalently,

$$(\varphi \circ \ell_0)(x) = \mathcal{L}_0^{\varphi}\Big(\frac{x}{|||x|||}\Big) \; , \;\; \forall x \in \mathbb{R}^d \setminus \{0\} \; . \tag{20d}$$

(e) *The function* $\mathcal{L}_0^{\varphi}$ *is given by*

$$\mathcal{L}_0^{\varphi} = \Big(\sup_{j \in [\![0,d]\!]} \Big[|||\cdot|||_{\star,(j)}^{\mathrm{tn}} - \varphi(j) \Big] \Big)^{\star'} \; . \tag{21a}$$

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

$$\mathbb{R}^d \ni x \mapsto \inf_{j \in [\![0,d]\!]} \Big[\delta_{\mathbb{B}_{\star,(j)}^{\star\mathrm{sn}}}(x) + \varphi(j) \Big] \; , \tag{21b}$$

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

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

$$\mathbb{R}^d \ni x \mapsto \inf_{j \in [\![0,d]\!]} \Big[\delta_{\mathbb{S}_{\star,(j)}^{\star\mathrm{sn}}}(x) + \varphi(j) \Big] \; , \tag{21c}$$

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

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

$$\mathcal{L}_0^{\varphi}(x) = \min_{\substack{(\lambda_0,\lambda_1,\ldots,\lambda_d) \in \Delta_{d+1} \\ x \in \sum_{l=1}^d \lambda_j \mathbb{B}_{\star,(l)}^{\star\mathrm{sn}}}} \sum_{l=1}^d \lambda_j \varphi(j) \; , \;\; \forall x \in \mathbb{R}^d \tag{22a}$$

$$= \min_{\substack{(\lambda_0,\lambda_1,\ldots,\lambda_d) \in \Delta_{d+1} \\ x \in \sum_{l=1}^d \lambda_j \mathbb{S}_{\star,(l)}^{\star\mathrm{sn}}}} \sum_{l=1}^d \lambda_j \varphi(j) \; , \;\; \forall x \in \mathbb{R}^d \tag{22b}$$

$$= \min_{\substack{x^{(1)} \in \mathbb{R}^d, \ldots, x^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d |||x^{(j)}|||_{\star,(j)}^{\star\mathrm{sn}} \leq 1 \\ \sum_{j=1}^d x^{(j)} = x}} \sum_{j=1}^d \varphi(j) |||x^{(j)}|||_{\star,(j)}^{\star\mathrm{sn}} \; , \;\; \forall x \in \mathbb{R}^d \; . \tag{22c}$$

Proof. As in the beginning of the proof of Proposition 7, we make the observation that, since the norm $\left|\!\left|\!\left|\cdot\right|\!\right|\!\right|$ is orthant-strictly monotonic, it is orthant-monotonic, so that we have $\left|\!\left|\!\left|\cdot\right|\!\right|\!\right|^{\mathcal{R}}_{(j)} = \left|\!\left|\!\left|\cdot\right|\!\right|\!\right|^{\star\mathrm{sn}}_{\star,(j)}$ and $\left|\!\left|\!\left|\cdot\right|\!\right|\!\right|^{\mathcal{R}}_{(j),\star} = \left|\!\left|\!\left|\cdot\right|\!\right|\!\right|^{\mathrm{tn}}_{\star,(j)}$, for $j \in [\![0, d]\!]$ by (32) in Proposition 14 in Appendix A.3 (with the convention that these are the null seminorms in the case $j = 0$).

(***a***) As the Fenchel conjugacy induces a one-to-one correspondence between the closed convex functions on $\mathbb{R}^d$ and themselves [18, Theorem 5], the function $\mathcal{L}_0^{\varphi} = \big((\varphi \circ \ell_0)^{¢}\big)^{\star'}$ in (19) is closed convex. We now show that it is proper. Indeed, on the one hand, it is easily seen, by the very definition (6a), 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 $-\infty$, by (37b). On the other hand, we have $(\varphi \circ \ell_0)^{¢¢'} = \big((\varphi \circ \ell_0)^{¢}\big)^{\star'} \circ n$ by [4, Equation (30d)], and $(\varphi \circ \ell_0)^{¢¢'} \leq \varphi \circ \ell_0$ by (6c), from which we deduce that, for any $x \in \mathbb{S}$, we have $\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}(x) = (\varphi \circ \ell_0)^{¢¢'}(x) \leq (\varphi \circ \ell_0)(x) < +\infty$ since $\varphi : [\![0, d]\!] \to \mathbb{R}_+$. As a consequence, the function $\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}$ is proper.

(***b***) The assumptions make it possible to conclude that $(\varphi \circ \ell_0)^{¢¢'} = \varphi \circ \ell_0$, thanks to Theorem 8. We deduce from [4, Proposition 4.3] that, being ¢-convex, the function $\varphi \circ \ell_0$ coincides, on the unit sphere $\mathbb{S}$, with the closed convex function $\mathcal{L}_0^{\varphi} : \mathbb{R}^d \to \overline{\mathbb{R}}$ given by [4, Equation (30d)] namely $\mathcal{L}_0^{\varphi} = \big((\varphi \circ \ell_0)^{¢}\big)^{\star'}$. Thus, we have proved (20a).

(***c***) Let $s \in \mathbb{S}$. We prove (20b) as follows:

$$y \in \partial \mathcal{L}_0^{\varphi}(s) \iff (\mathcal{L}_0^{\varphi})^{\star}(y) = \langle s, y \rangle \dotplus \big(-\mathcal{L}_0^{\varphi}(s)\big)$$

by definition (38a) of the (Rockafellar-Moreau) subdifferential of a function

$$\iff \Big(\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}\Big)^{\star}(y) = \langle s, y \rangle \dotplus \Big(-\Big(\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}\Big)(s)\Big)$$

(by definition (19) of $\mathcal{L}_0^{\varphi} = \big((\varphi \circ \ell_0)^{¢}\big)^{\star'}$)

$$\iff (\varphi \circ \ell_0)^{¢}(y) = \langle s, y \rangle \dotplus \Big(-\Big(\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}\Big)(s)\Big)$$

because the function $(\varphi \circ \ell_0)^{¢}$ is a Fenchel conjugate by [4, Equation (30b)], hence is closed convex, hence is equal to its Fenchel biconjugate $\Big(\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}\Big)^{\star}$

$$\iff (\varphi \circ \ell_0)^{¢}(y) = ¢(s, y) \dotplus \Big(-\Big(\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}\Big)(s)\Big)$$

(by definition (4) of $¢(s, y)$ as $s \in \mathbb{S}$)

$$\iff (\varphi \circ \ell_0)^{¢}(y) = ¢(s, y) \dotplus \Big(-\big((\varphi \circ \ell_0)^{¢¢'}\big)(s)\Big)$$

because $(\varphi \circ \ell_0)^{¢¢'} = \big((\varphi \circ \ell_0)^{¢}\big)^{\star'} \circ n$ by [4, Equation (30d)], and using that $n(s) = s$ since $s \in \mathbb{S}$ by definition (5) of the normalization mapping n

$$\iff (\varphi \circ \ell_0)^{¢}(y) = ¢(s, y) \dotplus \big(-(\varphi \circ \ell_0)(s)\big) \quad (\text{as } (\varphi \circ \ell_0)^{¢¢'} = \varphi \circ \ell_0 \text{ by Theorem 8})$$

$$\iff y \in \partial_{¢}(\varphi \circ \ell_0)(s) \,. \qquad (\text{by definition (13) of the CAPRA-subdifferential})$$

(***d***) The equality (20c) is a consequence of the formula $\varphi \circ \ell_0 = (\varphi \circ \ell_0)^{\cent\cent'} = \big((\varphi \circ \ell_0)^{\cent}\big)^{\star'} \circ n$ given by [4, Equation (30d)]. The equality (20d) is an easy consequence of (20c) and of the definition (5) of the normalization mapping n.

(***e***) As $\mathcal{L}_0^{\varphi} = \big((\varphi \circ \ell_0)^{\cent}\big)^{\star'}$ by definition (19), and as we have that $\big((\varphi \circ \ell_0)^{\cent}\big)^{\star'} = \Big(\sup_{j \in [\![0,d]\!]} \Big[|||\cdot|||_{\star,(j)}^{\mathrm{tn}} - \varphi(j) \Big]\Big)^{\star'}$ by (17), we get (21a).

(***f***) We use Proposition 16 (Appendix B) and especially Equations (35b) and (35c) to obtain (21b). Indeed, we have that

$$\mathcal{L}_0^{\varphi} = \big((\varphi \circ \ell_0)^{\cent}\big)^{\star'} = \Big(\sup_{j \in [\![0,d]\!]} \Big[|||\cdot|||_{\star,(j)}^{\mathrm{tn}} - \varphi(j) \Big]\Big)^{\star'} \qquad \text{(by (21a) proved in Item } (\boldsymbol{e}))$$

$$= \Big(\sup_{j \in [\![0,d]\!]} \big[|||\cdot|||_{(j),\star}^{\mathcal{R}} - \varphi(j) \big]\Big)^{\star'} \tag{23}$$

since $|||\cdot|||_{\star,(j)}^{\mathrm{tn}} = |||\cdot|||_{(j),\star}^{\mathcal{R}}$ as recalled at the beginning of the proof

$$= \Big(\inf_{j \in [\![0,d]\!]} \big[\delta_{\mathbb{B}_{(j)}^{\mathcal{R}}} \dotplus \varphi(j) \big]\Big)^{\star\star'} \qquad \text{(by (35b) and (35c))}$$

$$= \Big(\inf_{j \in [\![0,d]\!]} \big[\delta_{\mathbb{B}_{\star,(j)}^{\star\mathrm{sn}}} \dotplus \varphi(j) \big]\Big)^{\star\star'} ,$$

(as $\mathbb{B}_{(j)}^{\mathcal{R}} = \mathbb{B}_{\star,(j)}^{\star\mathrm{sn}}$ since $|||\cdot|||_{(j)}^{\mathcal{R}} = |||\cdot|||_{\star,(j)}^{\star\mathrm{sn}}$, as recalled at the beginning of the proof)

which gives (21b). The inclusions and equality $\mathbb{B}_{(1)}^{\mathcal{R}} \subset \cdots \subset \mathbb{B}_{(j)}^{\mathcal{R}} \subset \mathbb{B}_{(j+1)}^{\mathcal{R}} \subset \cdots \subset \mathbb{B}_{(d)}^{\mathcal{R}} = \mathbb{B}$ have been established for the generalized coordinate-k norms (see Definition 2) in [4, Equation (24)]. Now, since $|||\cdot|||_{(j)}^{\mathcal{R}} = |||\cdot|||_{\star,(j)}^{\star\mathrm{sn}}$, we get that $\mathbb{B}_{\star,(1)}^{\star\mathrm{sn}} \subset \cdots \subset \mathbb{B}_{\star,(j-1)}^{\star\mathrm{sn}} \subset \mathbb{B}_{\star,(j)}^{\star\mathrm{sn}} \subset \cdots \subset \mathbb{B}_{\star,(d)}^{\star\mathrm{sn}} = \mathbb{B}$.

(***g***) We use Proposition 16 (Appendix B) and especially Equations (35c) and (35e), to obtain (21c). Indeed, we have that

$$\mathcal{L}_0^{\varphi} = \Big(\inf_{j \in [\![0,d]\!]} \big[\delta_{\mathbb{B}_{(j)}^{\mathcal{R}}} \dotplus \varphi(j) \big]\Big)^{\star\star'} \qquad \text{(as seen in Item } (\boldsymbol{f}))$$

$$= \Big(\inf_{j \in [\![0,d]\!]} \big[\delta_{\mathbb{S}_{(j)}^{\mathcal{R}}} \dotplus \varphi(j) \big]\Big)^{\star\star'} \qquad \text{(by (35c) and (35e))}$$

$$= \Big(\inf_{j \in [\![0,d]\!]} \big[\delta_{\mathbb{S}_{\star,(j)}^{\star\mathrm{sn}}} \dotplus \varphi(j) \big]\Big)^{\star\star'} , \qquad \text{(since } \mathbb{S}_{(j)}^{\mathcal{R}} = \mathbb{S}_{\star,(j)}^{\star\mathrm{sn}} \text{ as } |||\cdot|||_{(j)}^{\mathcal{R}} = |||\cdot|||_{\star,(j)}^{\star\mathrm{sn}})$$

which gives (21c).

(***h***) We use the property that, for any $k \in [\![1,d]\!]$, we have $|||\cdot|||_{(k)}^{\mathcal{R}} = |||\cdot|||_{\star,(k)}^{\star\mathrm{sn}}$ and also Proposition 16 (Appendix B) and especially Equations (35g), (35h) and (35i) to obtain (22a), (22b) and (22c).

This ends the proof. $\square$

3.2 Variational formulation for the ℓ_0 pseudonorm

As an application of Proposition 9, we obtain the second main result of this paper, namely a variational formulation for (nonnegative nondecreasing functions of) the ℓ_0 pseudonorm.

Theorem 10 *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$ with associated sequence $\left\{|||\cdot|||^{\mathrm{tn}}_{\star,(j)}\right\}_{j\in[\![1,d]\!]}$ of generalized k-support dual norms as in Definition 3, and with associated* CAPRA *coupling ¢ as in (4). Suppose that both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are orthant-strictly monotonic. Let $\varphi : [\![0,d]\!] \to \mathbb{R}_+$ be a nonnegative nondecreasing function such that $\varphi(0) = 0$. Then, we have the equality*

$$\varphi\big(\ell_0(x)\big) = \frac{1}{|||x|||} \min_{\substack{x^{(1)}\in\mathbb{R}^d,\dots,x^{(d)}\in\mathbb{R}^d \\ \sum_{j=1}^d |||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \leq |||x||| \\ \sum_{j=1}^d x^{(j)}=x}} \sum_{j=1}^d \varphi(j)|||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \; , \quad \forall x \in \mathbb{R}^d \setminus \{0\} \; , \tag{24}$$

where the sequence of generalized k-support dual norms $\left\{|||\cdot|||^{\star\mathrm{sn}}_{\star,(j)}\right\}_{j\in[\![1,d]\!]}$ has been introduced in Definition 3.

When $\ell_0(x) = l \geq 1$, the minimum in (24) is achieved at $(x^{(1)},\dots,x^{(d)}) \in (\mathbb{R}^d)^d$ such that $x^{(j)} = 0$ for $j \neq l$ and $x^{(l)} = x$.

Proof. Equation (24) derives from (20c) where we use the expression (22c) for the function $\mathcal{L}_0^\varphi$ in (19).

Now for the argmin. When $\ell_0(x) = l \geq 1$, we have that $|||x||| = |||x|||^{\mathcal{R}}_{(d)} = \cdots = |||x|||^{\mathcal{R}}_{(l)}$ by [4, Equation (25a)]. Now, for any $k \in [\![1,d]\!]$, we have $|||\cdot|||^{\mathcal{R}}_{(k)} = |||\cdot|||^{\star\mathrm{sn}}_{\star,(k)}$ by (32) in Proposition 14 (Appendix A.3), since the norm $|||\cdot|||$ is orthant-strictly monotonic, hence is orthant-monotonic. As a consequence, we have that $|||x||| = |||x|||^{\star\mathrm{sn}}_{\star,(d)} = \cdots = |||x|||^{\star\mathrm{sn}}_{\star,(l)}$. Therefore, the vectors $x^{(1)} \in \mathbb{R}^d$, …, $x^{(d)} \in \mathbb{R}^d$ defined by $x^{(j)} = 0$ for $j \neq l$ and $x^{(l)} = x$ are admissible for the minimization problem (24). We deduce from (24) 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 10 applies when the norm $|||\cdot|||$ is any of the ℓ_p-norms $||\cdot||_p$ on the space $\mathbb{R}^d$, for $p \in]1,\infty[$, and Equation (24) then gives (see the notations in the second column of Table 1): $\forall x \in \mathbb{R}^d \setminus \{0\}$, $\forall p \in]1,\infty[$,

$$(\varphi \circ \ell_0)(x) = \frac{1}{||x||_p} \min_{\substack{x^{(1)}\in\mathbb{R}^d,\dots,x^{(d)}\in\mathbb{R}^d \\ \sum_{j=1}^d ||x^{(j)}||^{\mathrm{sn}}_{p,j} \leq ||x||_p \\ \sum_{j=1}^d x^{(j)}=x}} \sum_{j=1}^d \varphi(j)||x^{(j)}||^{\mathrm{sn}}_{p,j} \; . \tag{25}$$

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 $1/p + 1/q = 1$. When $p = 1$, the ℓ_1-norm $|||\cdot||| = ||\cdot||_1$ is orthant-strictly monotonic, but the dual norm $|||\cdot||| = ||\cdot||_\infty$ is not; when $p = \infty$, the ℓ_∞-norm

$|||\cdot||| = ||\cdot||_\infty$ is not orthant-strictly monotonic; hence, in those two extreme cases, we cannot conclude (but we obtain inequalities like in [4, Equation (25a)]).

Finally, with the novel expression (24) for the ℓ_0 pseudonorm, we deduce a possible reformulation of exact sparse optimization problems as follows (the proof is a straightforward application of Theorem 10).

Proposition 11 *Let $C \subset \mathbb{R}^d$ be such that $0 \notin C$ (if we had $0 \in C$, the minimization problem below would obviously be achieved at $x = 0$). 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. Let $\varphi : [\![0, d]\!] \to \mathbb{R}_+$ be a nondecreasing function, such that $\varphi(0) = 0$. Then, we have that*

$$\min_{x \in C} \varphi\big(\ell_0(x)\big) = \min_{\substack{x \in C, x^{(1)} \in \mathbb{R}^d, \ldots, x^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d |||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \leq 1 \\ \sum_{j=1}^d x^{(j)} = \frac{x}{|||x|||}}} \sum_{j=1}^d \varphi(j) |||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} , \tag{26a}$$

$$= \min_{\substack{x \in C, x^{(1)} \in \mathbb{R}^d, \ldots, x^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d |||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \leq |||x||| \\ \sum_{j=1}^d x^{(j)} = x}} \frac{1}{|||x|||} \sum_{j=1}^d \varphi(j) |||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} , \tag{26b}$$

$$= \min_{x \in C} \frac{1}{|||x|||} \underbrace{\min_{\substack{x^{(1)} \in \mathbb{R}^d, \ldots, x^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d |||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)} \leq |||x||| \\ \sum_{j=1}^d x^{(j)} = x}} \sum_{j=1}^d \varphi(j) |||x^{(j)}|||^{\star\mathrm{sn}}_{\star,(j)}}_{\text{convex optimization problem}} . \tag{26c}$$

4 Conclusion

In this paper, we have proven that the ℓ_0 pseudonorm is equal to its CAPRA-biconjugate 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 the convex factorization property, a way to express hidden convexity: 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. Then, we have obtained exact variational formulations for the ℓ_0 pseudonorm, suitable for exact sparse optimization. For this purpose, we have introduced sequences of generalized top-k and k-support dual norms. We now briefly sketch a few perspectives for exact sparse optimization.

The reformulations for exact sparse optimization problems, obtained in Proposition 11, have the nice feature to display partial convexity. However, they make use of as many new (latent) vectors as the underlying dimension d. Thus, the algorithmic implementation may be delicate. However, the variational formulation obtained may suggest approximations of the ℓ_0 pseudonorm, involving generalized k-support dual norms, which, themselves, may lead to new smooth sparsity inducing terms. Finally, we have identified elements of the CAPRA-subdifferential of nondecreasing functions of the ℓ_0 pseudonorm, and we have related this CAPRA-subdifferential with the Rockafellar-Moreau subdifferential of the associated convex lsc function (in the convex factorization property). The identification of such subgradients could inspire "gradient-like" algorithms.

Acknowledgements. We thank Guillaume Obozinski for discussions on first versions of this work, and Jean-Baptiste Hiriart-Urruty for his comments (and for proposing the term "convex factorization"). We are indebted to two Reviewers and to the Editor who, by their questions and comments, helped us improve the manuscript.

A Properties of relevant norms for the ℓ_0 pseudonorm

We provide background on properties of norms that prove relevant for the ℓ_0 pseudonorm. In §A.1, we review notions related to dual norms. We establish properties of orthant-monotonic and orthant-strictly monotonic norms in §A.2, and of coordinate-k and dual coordinate-k norms in §A.3.

A.1 Dual norm, $|||\cdot|||$-duality, normal cone

For any norm $|||\cdot|||$ on $\mathbb{R}^d$, we recall that the following expression

$$|||y|||_\star = \sup_{|||x|||\leq 1} \langle x, y\rangle \ , \ \ \forall y \in \mathbb{R}^d \tag{27}$$

defines a norm on $\mathbb{R}^d$, called the *dual norm* $|||\cdot|||_\star$ (in [20, Section 15], this operation is widened to a polarity operation between closed gauges).

By definition of the dual norm in (27), we have the inequality

$$\langle x, y\rangle \leq |||x||| \times |||y|||_\star \ , \ \ \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \ . \tag{28a}$$

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 (28a), that is,

$$y \parallel_{|||\cdot|||} x \iff \langle x, y\rangle = |||x||| \times |||y|||_\star \ . \tag{28b}$$

The terminology $|||\cdot|||$-dual comes from [11, page 2] (see also the vocable of *dual vector pair* in [7, Equation (1.11)] and of *dual vectors* in [8, p. 283], whereas it is refered as *polar alignment* in [6]). 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 [10, p.136]

$$N_C(x) = \big\{ y \in \mathbb{R}^d \, \big| \, \langle y, x' - x\rangle \leq 0 \ , \ \ \forall x' \in C \big\} \ . \tag{29}$$

Now, an easy computation shows that the notion of $|||\cdot|||$-duality can be rewritten in terms of normal cone $N_{\mathbb{B}}$ as follows:

$$y \parallel_{|||\cdot|||} x \iff y \in N_{\mathbb{B}}\Big(\frac{x}{|||x|||}\Big) \ , \ \ \forall (x,y) \in \big(\mathbb{R}^d \setminus \{0\}\big) \times \mathbb{R}^d \ . \tag{30}$$

A.2 Properties of orthant-strictly monotonic norms

We provide useful properties of orthant-monotonic and orthant-strictly monotonic norms (see Definition 5). We recall that $x_K \in \mathcal{R}_K$ denotes the vector which coincides with x, except for the components outside of K that vanish, and that the subspace $\mathcal{R}_K$ of $\mathbb{R}^d$ has been defined in (7).

Proposition 12 *Let $|||\cdot|||$ be an orthant-monotonic norm on $\mathbb{R}^d$. Then, the dual norm $|||\cdot|||_\star$ is orthant-monotonic, and the norm $|||\cdot|||$ is* increasing with the coordinate subspaces*, in the sense that, for any $x \in \mathbb{R}^d$ and any $J \subset K \subset [\![1,d]\!]$, we have $|||x_J||| \leq |||x_K|||$.*

Proof. Let $|||\cdot|||$ be an orthant-monotonic norm on $\mathbb{R}^d$. Then, by [7, Theorem 2.23], the dual norm $|||\cdot|||_\star$ is also orthant-monotonic and, by [11, Proposition 2.4], we have that $|||u||| \leq |||u+v|||$, for any subset $J \subset [\![1,d]\!]$ and for any vectors $u \in \mathcal{R}_J$ and $v \in \mathcal{R}_{-J}$ (following notation from game theory, we have denoted by $-J$ the complementary subset of $J \subset [\![1,d]\!]$, that is, $J \cup (-J) = [\![1,d]\!]$ and $J \cap (-J) = \emptyset$). We consider $x \in \mathbb{R}^d$ and $J \subset K \subset [\![1,d]\!]$. By setting $u = x_J \in \mathcal{R}_J$ and $v = x_K - x_J$, we get that $v \in \mathcal{R}_{-J}$, hence that $|||x_J||| \leq |||x_K|||$. □

Proposition 13 *Let $|||\cdot|||$ be an orthant-strictly monotonic norm on $\mathbb{R}^d$. Then*

(***a***) *the norm $|||\cdot|||$ is* strictly increasing with the coordinate subspaces *in the sense that, for any $x \in \mathbb{R}^d$ and any $J \subsetneq K \subset [\![1,d]\!]$, we have $x_J \neq x_K \Rightarrow |||x_J||| < |||x_K|||$.*

(***b***) *for any vector $u \in \mathbb{R}^d \setminus \{0\}$, there exists a vector $v \in \mathbb{R}^d \setminus \{0\}$ such that* $\mathrm{supp}(v) = \mathrm{supp}(u)$*, that $u \circ v \geq 0$, and that v is $|||\cdot|||$-dual to u, that is, $\langle u, v\rangle = |||u||| \times |||v|||_\star$.*

Proof.

(***a***) Let $x \in \mathbb{R}^d$ and $J \subsetneq K \subset [\![1,d]\!]$ be such that $x_J \neq x_K$. We will show that $|||x_K||| > |||x_J|||$.

For this purpose, we set $u = x_J$ and $v = x_K - x_J$. Thus, we get that $u \in \mathcal{R}_K$ and $v \in \mathcal{R}_{-K} \setminus \{0\}$ (since $J \subsetneq K$ and $x_J \neq x_K$), that is, $u = u_K$ and $v = v_{-K} \neq 0$. We are going to show that $|||u+v||| > |||u|||$.

On the one hand, by definition of the module of a vector, we easily see that $|w| = |w_K| + |w_{-K}|$, for any vector $w \in \mathbb{R}^d$. Thus, we have $|u+v| = |(u+v)_K| + |(u+v)_{-K}| = |u_K + v_K| + |u_{-K} + v_{-K}| = |u_K + 0| + |0 + v_{-K}| = |u_K| + |v_{-K}| > |u_K| = |u|$ since $|v_{-K}| > 0$ as $v = v_{-K} \neq 0$, and since $u = u_K$. On the other hand, we easily get that $(u+v) \circ u = \big((u+v)_K \circ u_K\big) + \big((u+v)_{-K} \circ u_{-K}\big) = \big(u_K \circ u_K\big) + \big(v_{-K} \circ u_{-K}\big) = \big(u_K \circ u_K\big)$, because $u_{-K} = 0$. Therefore, we get that $(u+v) \circ u = \big(u_K \circ u_K\big) \geq 0$.

From $|u+v| > |u|$ and $(u+v) \circ u \geq 0$, we deduce that $|||u+v||| > |||u|||$ by Definition 5 as the norm $|||\cdot|||$ is orthant-strictly monotonic. Since $u = x_J$ and $v = x_K - x_J$, we conclude that $|||x_K||| > |||x_J|||$.

($\boldsymbol{b}$) Let $u \in \mathbb{R}^d \setminus \{0\}$ be given and let us put $K = \mathrm{supp}(u) \neq \emptyset$. As the norm $|||\cdot|||$ is orthant-strictly monotonic, it is orthant-monotonic; hence, by [11, Proposition 2.4], there exists a vector $v \in \mathbb{R}^d \setminus \{0\}$ such that $\mathrm{supp}(v) \subset \mathrm{supp}(u)$, that $u \circ v \geq 0$ and that v is $|||\cdot|||$-dual to u, as in (28b), that is, $\langle u, v \rangle = |||u||| \times |||v|||_\star$. Thus $J = \mathrm{supp}(v) \subset K = \mathrm{supp}(u)$. We will now show that $J \subsetneq K$ is impossible, hence that $J = K$, thus proving that Item ($\boldsymbol{b}$) holds true with the above vector v.

Writing that $\langle u, v \rangle = |||u||| \times |||v|||_\star$ (using that $u = u_K$ and $v = v_K = v_J$), we obtain

$$|||u||| \times |||v|||_\star = \langle u, v \rangle = \langle u_K, v \rangle = \langle u_K, v_K \rangle = \langle u_K, v_J \rangle = \langle u_J, v_J \rangle = \langle u_J, v \rangle \ ,$$

by obvious properties of the scalar product $\langle \cdot, \cdot \rangle$. As a consequence, we get that $\{u_K, u_J\} \subset \arg\max_{|||x||| \leq |||u|||} \langle x, v \rangle$, by definition (27) of $|||v|||_\star$, because $|||u||| = |||u_K||| \geq |||u_J|||$, by Proposition 12 since $J \subset K$ and the norm $|||\cdot|||$ is orthant-monotonic. But any solution in $\arg\max_{|||x||| \leq |||u|||} \langle x, v \rangle$ belongs to the frontier of the ball of radius $|||u|||$, hence has exactly norm $|||u|||$. Thus, we deduce that $|||u||| = |||u_K||| = |||u_J|||$. If we had $J = \mathrm{supp}(v) \subsetneq K = \mathrm{supp}(u)$, we would have $u_J \neq u_K$, hence $|||u_K||| > |||u_J|||$ by Item ($\boldsymbol{a}$); this would be in contradiction with $|||u_K||| = |||u_J|||$. Therefore, $J = \mathrm{supp}(v) = K = \mathrm{supp}(u)$.

This ends the proof. □

A.3 Properties of coordinate-k and dual coordinate-k norms, and of generalized top-k and k-support dual norms

We establish useful properties of coordinate-k and dual coordinate-k norms (Definition 2), and of generalized top-k and k-support dual norms (Definition 3).

Proposition 14 *Let $|||\cdot|||$ be a source norm on $\mathbb{R}^d$.*

Coordinate-k norms are greater than k-support dual norms, that is,

$$|||x|||^{\mathcal{R}}_{(k)} \geq |||x|||^{\star\mathrm{sn}}_{\star,(k)} \ , \ \ \forall x \in \mathbb{R}^d \ , \ \ \forall k \in [\![1, d]\!] \ , \tag{31a}$$

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

$$|||y|||^{\mathcal{R}}_{(k),\star} \leq |||y|||^{\mathrm{tn}}_{\star,(k)} \ , \ \ \forall y \in \mathbb{R}^d \ , \ \ \forall k \in [\![1, d]\!] \ . \tag{31b}$$

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

$$|||\cdot||| \textit{is orthant-monotonic} \Rightarrow \forall k \in [\![1, d]\!] \quad \begin{cases} |||\cdot|||^{\mathcal{R}}_{(k)} &= |||\cdot|||^{\star\mathrm{sn}}_{\star,(k)} \ , \\ |||\cdot|||^{\mathcal{R}}_{(k),\star} &= |||\cdot|||^{\mathrm{tn}}_{\star,(k)} \ . \end{cases} \tag{32}$$

Proof. It is known that, for any nonempty subset $K \subset [\![1,d]\!]$, we have the inequality $|||\cdot|||_{K,\star} \leq |||\cdot|||_{\star,K}$ (see [11, Proposition 2.2]). From the definition (10) of the generalized top-k dual norm, and the definition (8) of the dual coordinate-k norm, we get that $|||y|||^{\mathcal{R}}_{(k),\star} = \sup_{|K|\leq k} |||y_K|||_{K,\star} \leq \sup_{|K|\leq k} |||y_K|||_{\star,K} = |||y|||^{\mathrm{tn}}_{\star,(k)}$, hence we obtain (31b). By taking the dual norms, we get (31a).

The norms for which the equality $|||\cdot|||_{K,\star} = |||\cdot|||_{\star,K}$ holds true for all nonempty subsets $K \subset [\![1,d]\!]$, are the orthant-monotonic norms ([7, Characterization 2.26], [11, Theorem 3.2]). Therefore, if the norm $|||\cdot|||$ is orthant-monotonic, from the definition (10) of the generalized top-k dual norm, we get that the inequality (31b) becomes an equality. Then, the inequality (31a) also becomes an equality by taking the dual norm as in (27). Thus, we have obtained (32).

This ends the proof. □

Proposition 15 *Let $|||\cdot|||$ be a source norm on $\mathbb{R}^d$. Let $y \in \mathbb{R}^d$ and $l \in [\![1,d]\!]$. If the dual norm $|||\cdot|||_\star$ is orthant-strictly monotonic, we have that*

$$\ell_0(y) = l \implies \begin{cases} |||y|||^{\mathcal{R}}_{(1),\star} < \cdots < |||y|||^{\mathcal{R}}_{(l-1),\star} < |||y|||^{\mathcal{R}}_{(l),\star} = \cdots = |||y|||^{\mathcal{R}}_{(d),\star} = |||y|||_\star \;, \\ |||y|||^{\mathrm{tn}}_{\star,(1)} < \cdots < |||y|||^{\mathrm{tn}}_{\star,(l-1)} < |||y|||^{\mathrm{tn}}_{\star,(l)} = \cdots = |||y|||^{\mathrm{tn}}_{\star,(d)} = |||y|||_\star \;. \end{cases} \tag{33}$$

Proof. We consider $y \in \mathbb{R}^d$. We put $L = \mathrm{supp}(y)$ and we suppose that $\ell_0(y) = |L| = l$.

Since the norm $|||\cdot|||_\star$ is orthant-strictly monotonic, it is orthant-monotonic and so is $|||\cdot|||$ by Proposition 12. By (32) in Proposition 14, we get that $|||\cdot|||^{\mathcal{R}}_{(j)} = |||\cdot|||^{\star\mathrm{sn}}_{\star,(j)}$ and $|||\cdot|||^{\mathcal{R}}_{(j),\star} = |||\cdot|||^{\mathrm{tn}}_{\star,(j)}$, for $j \in [\![0,d]\!]$ (with the convention that these are the null seminorms in the case $j = 0$). Therefore, we can translate all the results, obtained in [4], with coordinate-k and dual coordinate-k norms, into results regarding generalized top-k and k-support dual norms. As an application, by [4, Equation (18)], we get, from $\ell_0(y) = l$, that

$$|||y|||^{\mathrm{tn}}_{\star,(1)} \leq \cdots \leq |||y|||^{\mathrm{tn}}_{\star,(j)} \leq |||y|||^{\mathrm{tn}}_{\star,(j+1)} \leq \cdots \leq |||y|||^{\mathrm{tn}}_{\star,(d)} = |||y|||_\star \;, \quad \forall y \in \mathbb{R}^d \;. \tag{34}$$

We now prove (33) in two steps.

We first show that $|||y|||^{\mathrm{tn}}_{\star,(l)} = \cdots = |||y|||^{\mathrm{tn}}_{\star,(d)} = |||y|||_\star$ (the right hand side of (33)). Since $y = y_L$, by definition of the set $L = \mathrm{supp}(y)$, we have that $|||y|||_\star = |||y_L|||_\star \leq \sup_{|K|\leq l} |||y_K|||_\star = |||y|||^{\mathrm{tn}}_{\star,(l)}$ by the very definition (10) of the generalized top-l dual norm $|||\cdot|||^{\mathrm{tn}}_{\star,(l)}$. By (34), we conclude that $|||y|||^{\mathrm{tn}}_{\star,(l)} = \cdots = |||y|||^{\mathrm{tn}}_{\star,(d)} = |||y|||_\star$.

Second, we show that $|||y|||^{\mathrm{tn}}_{\star,(1)} < \cdots < |||y|||^{\mathrm{tn}}_{\star,(l-1)} < |||y|||^{\mathrm{tn}}_{\star,(l)}$ (the left hand side of (33)). There is nothing to show for $l = 0$. Now, for $l \geq 1$ and for any $k \in [\![0, l-1]\!]$, we have

$$
\begin{aligned}
|||y|||^{\text{tn}}_{\star,(k)} &= \sup_{|K| \leq k} |||y_K|||_\star && \text{(by definition (10) of the generalized top-}k\text{ dual norm)} \\
&= \sup_{|K| \leq k} |||y_{K \cap L}|||_\star && \text{(because } y_L = y \text{ by definition of the set } L = \text{supp}(y)\text{)} \\
&= \sup_{|K'| \leq k, K' \subset L} |||y_{K'}|||_\star && \text{(by setting } K' = K \cap L\text{)} \\
&= \sup_{|K| \leq k, K \subset L} |||y_K|||_\star && \text{(the same but with notation } K \text{ instead of } K'\text{)} \\
&= \sup_{|K| \leq k, K \subsetneq L} |||y_K|||_\star && \text{(because } |K| \leq k \leq l-1 < l = |L| \text{ implies that } K \neq L\text{)} \\
&< \sup_{\substack{|K| \leq k, j \in L \setminus K \\ K \subsetneq L}} |||y_{K \cup \{j\}}|||_\star
\end{aligned}
$$

because the set $L \setminus K$ is nonempty (having cardinality $|L| - |K| = l - |K| \geq k + 1 - |K| \geq 1$), and because, since the norm $|||\cdot|||_\star$ is orthant-strictly monotonic, using Item (***a***) in Proposition 13, we obtain that $|||y_K|||_\star < |||y_{K \cup \{j\}}|||_\star$ as $y_K \neq y_{K \cup \{j\}}$ for at least one $j \in L \setminus K$ since $L = \text{supp}(y)$

$$
\begin{aligned}
&\leq \sup_{|J| \leq k+1, J \subset L} |||y_J|||_\star \\
&\qquad \text{(as all the subsets } K' = K \cup \{j\} \text{ are such that } K' \subset L \text{ and } |K'| = k+1\text{)} \\
&\leq |||y|||^{\text{tn}}_{\star,(k+1)}
\end{aligned}
$$

by definition (10) of the generalized top-$(k+1)$ dual norm (in fact the last inequality is easily shown to be an equality as $y_L = y$). Thus, for any $k \in [\![0, l-1]\!]$, we have established that $|||y|||^{\text{tn}}_{\star,(k)} < |||y|||^{\text{tn}}_{\star,(k+1)}$.

This ends the proof. □

B Proposition 16

We reproduce here [4, Proposition 4.5] in order to simplify the reading of the proof of Proposition 9.

Proposition 16 *([4, Proposition 4.5]) Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$, with associated sequence $\left\{|||\cdot|||^{\mathcal{R}}_{(j)}\right\}_{j \in [\![1,d]\!]}$ of coordinate-k norms and sequence $\left\{|||\cdot|||^{\mathcal{R}}_{(j),\star}\right\}_{j \in [\![1,d]\!]}$ of dual coordinate-k norms, as in Definition 2, and with associated* Capra *coupling ¢ in* (4).

1. *For any function $\varphi : [\![0, d]\!] \to \overline{\mathbb{R}}$, we have*

$$
(\varphi \circ \ell_0)^{¢¢'}(x) = \big((\varphi \circ \ell_0)^{¢}\big)^{\star'}\big(\frac{x}{|||x|||}\big) \ , \ \ \forall x \in \mathbb{R}^d \backslash \{0\} \ , \tag{35a}
$$

where the closed convex function $\big((\varphi\circ\ell_0)^{\cancel{C}}\big)^{\star'}$ *has the following expression as a Fenchel conjugate*

$$\big((\varphi\circ\ell_0)^{\cancel{C}}\big)^{\star'} = \Big(\sup_{j\in[\![0,d]\!]}\big[|\!|\!|\cdot|\!|\!|^{\mathcal{R}}_{(j),\star} - \varphi(j)\big]\Big)^{\star'} , \tag{35b}$$

and also has the following four expressions as a Fenchel biconjugate

$$= \Big(\inf_{j\in[\![0,d]\!]}\big[\delta_{\mathbb{B}^{\mathcal{R}}_{(j)}} \dotplus \varphi(j)\big]\Big)^{\star\star'} , \tag{35c}$$

hence the function $\big((\varphi\circ\ell_0)^{\cancel{C}}\big)^{\star'}$ *is the largest closed convex function below the integer valued function* $\inf_{j\in[\![0,d]\!]}\big[\delta_{\mathbb{B}^{\mathcal{R}}_{(j)}} \dotplus \varphi(j)\big]$, *such that* $x\in\mathbb{B}^{\mathcal{R}}_{(j)}\backslash\mathbb{B}^{\mathcal{R}}_{(j-1)}\mapsto\varphi(j)$ *for* $l\in[\![1,d]\!]$, *and* $x\in\mathbb{B}^{\mathcal{R}}_{(0)}=\{0\}\mapsto\varphi(0)$, *the function being infinite outside* $\mathbb{B}^{\mathcal{R}}_{(d)}=\mathbb{B}$, *that is, with the convention that* $\mathbb{B}^{\mathcal{R}}_{(0)}=\{0\}$ *and that* $\inf\emptyset=+\infty$

$$= \Big(x\mapsto\inf\big\{\varphi(j)\,\big|\,x\in\mathbb{B}^{\mathcal{R}}_{(j)}\ ,\ \ j\in[\![0,d]\!]\big\}\Big)^{\star\star'} , \tag{35d}$$

$$= \Big(\inf_{j\in[\![0,d]\!]}\big[\delta_{\mathbb{S}^{\mathcal{R}}_{(j)}} \dotplus \varphi(j)\big]\Big)^{\star\star'} , \tag{35e}$$

hence the function $\big((\varphi\circ\ell_0)^{\cancel{C}}\big)^{\star'}$ *is the largest closed convex function below the integer valued function* $\inf_{j\in[\![0,d]\!]}\big[\delta_{\mathbb{S}^{\mathcal{R}}_{(j)}} \dotplus \varphi(j)\big]$, *that is, with the convention that* $\mathbb{S}^{\mathcal{R}}_{(0)}=\{0\}$ *and that* $\inf\emptyset=+\infty$

$$= \Big(x\mapsto\inf\big\{\varphi(j)\,\big|\,x\in\mathbb{S}^{\mathcal{R}}_{(j)}\ ,\ \ j\in[\![0,d]\!]\big\}\Big)^{\star\star'} . \tag{35f}$$

2. *For any function* $\varphi:[\![0,d]\!]\to\mathbb{R}$, *that is, with finite values, the function* $\big((\varphi\circ\ell_0)^{\cancel{C}}\big)^{\star'}$ *is proper convex lsc and has the following variational expression (where* Δ_{d+1} *denotes the simplex of* $\mathbb{R}^{d+1}$*)*

$$\big((\varphi\circ\ell_0)^{\cancel{C}}\big)^{\star'}(x) = \min_{\substack{(\lambda_0,\lambda_1,\ldots,\lambda_d)\in\Delta_{d+1}\\ x\in\sum_{j=1}^{d}\lambda_j\mathbb{B}^{\mathcal{R}}_{(j)}}} \sum_{j=0}^{d}\lambda_j\varphi(j)\ ,\ \ \forall x\in\mathbb{R}^d\ . \tag{35g}$$

3. *For any function* $\varphi:[\![0,d]\!]\to\mathbb{R}_+$, *that is, with nonnegative finite values, and such that* $\varphi(0)=0$, *the function* $\big((\varphi\circ\ell_0)^{\cancel{C}}\big)^{\star'}$ *is proper convex lsc and has the following two variational expressions (notice that, in* (35g)*, the sum starts from* $j=0$*, whereas*

in (35h) *and in* (35i)*, the sum starts from* $j = 1$*)*

$$\big((\varphi \circ \ell_0)^{¢}\big)^{\star'}(x) = \min_{\substack{(\lambda_0,\lambda_1,\ldots,\lambda_d)\in\Delta_{d+1} \\ x\in\sum_{j=1}^{d}\lambda_j \mathbb{S}^{\mathcal{R}}_{(j)}}} \sum_{j=1}^{d} \lambda_j \varphi(j) \ , \ \ \forall x \in \mathbb{R}^d \ , \tag{35h}$$

$$= \min_{\substack{z^{(1)}\in\mathbb{R}^d,\ldots,z^{(d)}\in\mathbb{R}^d \\ \sum_{j=1}^{d} \|z^{(j)}\|^{\mathcal{R}}_{(j)} \leq 1 \\ \sum_{j=1}^{d} z^{(j)}=x}} \sum_{j=1}^{d} \varphi(j) \|z^{(j)}\|^{\mathcal{R}}_{(j)} \ , \ \ \forall x \in \mathbb{R}^d \ , \tag{35i}$$

and the function $(\varphi \circ \ell_0)^{¢¢'}$ *has the following variational expression*

$$(\varphi \circ \ell_0)^{¢¢'}(x) = \frac{1}{\|x\|} \min_{\substack{z^{(1)}\in\mathbb{R}^d,\ldots,z^{(d)}\in\mathbb{R}^d \\ \sum_{j=1}^{d} \|z^{(j)}\|^{\mathcal{R}}_{(j)} \leq \|x\| \\ \sum_{j=1}^{d} z^{(j)}=x}} \sum_{j=1}^{d} \|z^{(j)}\|^{\mathcal{R}}_{(j)} \varphi(j) \ , \ \ \forall x \in \mathbb{R}^d \backslash \{0\} \ . \tag{36}$$

C Background on the Fenchel conjugacy on $\mathbb{R}^d$

We review concepts and notations related to the Fenchel conjugacy (we refer the reader to [18]). For any function $h : \mathbb{R}^d \to \overline{\mathbb{R}}$, its *epigraph* is $\mathrm{epi}h = \big\{(w,t) \in \mathbb{R}^d \times \mathbb{R} \,\big|\, h(w) \leq t\big\}$, its *effective domain* is $\mathrm{dom}h = \big\{w \in \mathbb{R}^d \,\big|\, h(w) < +\infty\big\}$. A function $h : \mathbb{R}^d \to \overline{\mathbb{R}}$ is said to be *convex* if its epigraph is a convex set, *proper* if it never takes the value $-\infty$ and that $\mathrm{dom}h \neq \emptyset$, *lower semi continuous (lsc)* if its epigraph is closed, *closed* if it either lsc and nowhere having the value $-\infty$, or is the constant function $-\infty$ [18, p. 15]. Closed convex functions are the two constant functions $-\infty$ and $+\infty$ united with all proper convex lsc functions. In particular, any closed convex function that takes at least one finite value is necessarily proper convex lsc.

For any functions $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^d \to \overline{\mathbb{R}}$, we denote

$$f^{\star}(y) = \sup_{x\in\mathbb{R}^d} \Big(\langle x \,,\, y \rangle \stackrel{\cdot}{+} \big(-f(x)\big) \Big) \ , \ \ \forall y \in \mathbb{R}^d \ , \tag{37a}$$

$$g^{\star'}(x) = \sup_{y\in\mathbb{Y}} \Big(\langle x \,,\, y \rangle \stackrel{\cdot}{+} \big(-g(y)\big) \Big) \ , \ \ \forall x \in \mathbb{R}^d \ , \tag{37b}$$

$$f^{\star\star'}(x) = \sup_{y\in\mathbb{R}^d} \Big(\langle x \,,\, y \rangle \stackrel{\cdot}{+} \big(-f^{\star}(y)\big) \Big) \ , \ \ \forall x \in \mathbb{R}^d \ . \tag{37c}$$

In convex analysis, one does not use the notation $^{\star'}$ in (37b) and $^{\star\star'}$ in (37c), but simply * and **. We use $^{\star'}$ and $^{\star\star'}$ to be consistent with the notation (6b) for general conjugacies.

It is proved that the Fenchel conjugacy (indifferently $f \mapsto f^{\star}$ or $g \mapsto g^{\star'}$) induces a one-to-one correspondence between the closed convex functions on $\mathbb{R}^d$ and themselves [18, Theorem 5].

In [20, p. 214-215] (see also the historical note in [19, p. 343]), the notions of (Moreau) subgradient and of (Rockafellar) subdifferential are defined for a convex function. Following the definition of the subdifferential of a function with respect to a duality in [1], we define the *(Rockafellar-Moreau) subdifferential* $\partial f(x)$ of a function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ at $x \in \mathbb{R}^d$ by

$$\partial f(x) = \left\{ y \in \mathbb{R}^d \,\middle|\, f^\star(y) = \langle x \,, y \rangle \dotplus \left(-f(x) \right) \right\} . \tag{38a}$$

When the function f is proper convex and $x \in \mathrm{dom} f$, we recover the classic definition that

$$\partial f(x) = \left\{ y \in \mathbb{R}^d \,\middle|\, \langle x' - x \,, y \rangle + f(x) \leq f(x') \,, \;\; \forall x' \in \mathrm{dom} f \right\} . \tag{38b}$$

References

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

[2] A. Argyriou, R. Foygel, and N. Srebro. Sparse prediction with the k-support norm. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pages 1457–1465, USA, 2012. Curran Associates Inc.

[3] R. Bhatia. *Matrix Analysis*. Springer-Verlag, New York, 1997.

[4] J.-P. Chancelier and M. De Lara. Constant along primal rays conjugacies and the l_0 pseudonorm. *Optimization*, 0(0):1–32, 2020.

[5] J.-P. Chancelier and M. De Lara. Hidden convexity in the l_0 pseudonorm. *Journal of Convex Analysis*, 28(1):203–236, 2021.

[6] Z. Fan, H. Jeong, Y. Sun, and M. P. Friedlander. Atomic decomposition via polar alignment. *Foundations and Trends® in Optimization*, 3(4):280–366, 2020.

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

[8] D. Gries and J. Stoer. Some results on fields of values of a matrix. *SIAM Journal on Numerical Analysis*, 4(2):283–300, 1967.

[9] J.-B. Hiriart-Urruty and H. Le. A variational approach of the rank function. *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research*, 21(2):207–240, 2013.

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

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

[12] 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.

[13] 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.

[14] L. Mirsky. Symmetric Gauge Functions and Unitarily Invariant Norms. *The Quarterly Journal of Mathematics*, 11(1):50–59, Jan. 1960.

[15] J. J. Moreau. Inf-convolution, sous-additivité, convexité des fonctions numériques. *J. Math. Pures Appl. (9)*, 49:109–154, 1970.

[16] M. Nikolova. Relationship between the optimal solutions of least squares regularized with l0-norm and constrained by k-sparsity. *Applied and Computational Harmonic Analysis*, 41(1):237–265, 2016.

[17] G. Obozinski and F. Bach. A unified perspective on convex structured sparsity: Hierarchical, symmetric, submodular norms and beyond. Preprint, Dec. 2016.

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

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

[20] T. R. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, N.J., 1970.

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

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

[23] K. Tono, A. Takeda, and J.-y. Gotoh. Efficient DC algorithm for constrained sparse optimization. Preprint, Jan. 2017.