



Constant Along Primal Rays Conjugacies and the 10 Pseudonorm

Jean-Philippe Chancelier, Michel de Lara

► To cite this version:

Jean-Philippe Chancelier, Michel de Lara. Constant Along Primal Rays Conjugacies and the 10 Pseudonorm. Optimization, 2022, 71 (2), pp.355-386. 10.1080/02331934.2020.1822836 . hal-02459673v2

HAL Id: hal-02459673

<https://hal.science/hal-02459673v2>

Submitted on 1 Jun 2021

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

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

Constant Along Primal Rays Conjugacies and the ℓ_0 Pseudonorm

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

June 1, 2021

Abstract

The so-called ℓ_0 pseudonorm on \mathbb{R}^d counts the number of nonzero components of a vector. For exact sparse optimization problems — with the ℓ_0 pseudonorm standing either as criterion or in the constraints — the Fenchel conjugacy fails to provide relevant analysis. In this paper, we display a class of conjugacies that are suitable for the ℓ_0 pseudonorm. For this purpose, we suppose given a (source) norm on \mathbb{R}^d . With this norm, we define, on the one hand, a sequence of so-called coordinate- k norms and, on the other hand, a coupling between \mathbb{R}^d and itself, called Capra (constant along primal rays). Then, we provide formulas for the CAPRA-conjugate and biconjugate, and for the CAPRA subdifferentials, of functions of the ℓ_0 pseudonorm, in terms of the coordinate- k norms. As an application, we provide a new family of lower bounds for the ℓ_0 pseudonorm, as a fraction between two norms, the denominator being any norm.

Key words: ℓ_0 pseudonorm, Fenchel-Moreau conjugacy, Capra conjugacy, coordinate- k norm.

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 . The ℓ_0 pseudonorm measures the sparsity of a vector, and the literature in sparse optimization that mentions it is plethoric. However, because of its combinatorial nature, the problems of minimizing the ℓ_0 pseudonorm under constraints or of minimizing a criterion under k -sparsity constraint (ℓ_0 pseudonorm less than a given integer k) are usually not tackled as such. Most of the literature in sparse optimization studies surrogate problems where the ℓ_0 pseudonorm either enters a penalization term or is replaced by a regularizing term. We refer the reader to [1] that provides a brief tour of the

*michel.delara@enpc.fr

literature dealing with least squares minimization constrained by k -sparsity, and to [2] for a survey of the rank function of a matrix, that shares many properties with the ℓ_0 pseudonorm.

Conjugacies, and more generally dualities, are a powerful tool to tackle classes of optimization problems. The Fenchel conjugacy plays a central role in analyzing solutions of convex problems (and beyond) [3]. However, it fails to provide relevant analysis for optimization problems involving the ℓ_0 pseudonorm. Indeed, the Fenchel biconjugate of the characteristic function of the level sets of the ℓ_0 pseudonorm is zero, and the Fenchel biconjugate of the ℓ_0 pseudonorm is also zero. The field of generalized convexity goes beyond the Fenchel conjugacy and convex functions and displays conjugacies adapted to analyze classes of functions such as increasing positive homogeneous, difference of convex, quasi-convex, increasing and convex-along-rays. For more details on the theory, and more examples, we refer the reader to the books [4,5] and to the nice introduction paper [6].

To our knowledge, none of the conjugacies in the literature is adapted to the ℓ_0 pseudonorm (the ℓ_0 pseudonorm is convex-along-rays according to the definition in [7] but not in [4], and calculation shows that the ℓ_0 pseudonorm is not convex for the conjugacy in [7]). In this paper, we study the ℓ_0 pseudonorm as such and we display a suitable class of conjugacies. We extend results of [8] beyond the special Euclidian norm setting.

The paper is organized as follows. In Sect. 2, we recall the definition of the ℓ_0 pseudonorm, and we introduce the notion of sequence of norms on \mathbb{R}^d that are (strictly or not) decreasingly graded with respect to the ℓ_0 pseudonorm. In Sect. 3, we introduce a sequence of coordinate- k norms, all generated from any (source) norm on \mathbb{R}^d , and their dual norms. In Sect. 4, we define a so-called CAPRA coupling between \mathbb{R}^d and itself, that depends on any (source) norm on \mathbb{R}^d . Then, we provide formulas for the CAPRA-conjugate and biconjugate, and for the CAPRA subdifferentials, of functions of the ℓ_0 pseudonorm (hence, in particular, of the ℓ_0 pseudonorm itself and of the characteristic functions of its level sets), in terms of the coordinate- k norms. In Sect. 5, as an application, we provide a new family of lower bounds for the ℓ_0 pseudonorm, as a fraction between two norms, the denominator being any norm. The Appendix A gathers background on Fenchel-Moreau conjugacies.

2 The ℓ_0 pseudonorm and its level sets

First, we introduce basic notations regarding the ℓ_0 pseudonorm. Second, we recall the definition of a sequence of norms on \mathbb{R}^d which is (strictly or not) decreasingly graded with respect to the ℓ_0 pseudonorm (as introduced in the companion paper [9]). We use the notation $\llbracket r, s \rrbracket = \{r, r+1, \dots, s-1, s\}$ for two integers $r \leq s$.

The ℓ_0 pseudonorm. For any vector $x \in \mathbb{R}^d$, $\text{supp}(x) = \{j \in \llbracket 1, d \rrbracket \mid x_j \neq 0\} \subset \llbracket 1, d \rrbracket$ is the support of x . The so-called ℓ_0 *pseudonorm* is the function $\ell_0 : \mathbb{R}^d \rightarrow \llbracket 0, d \rrbracket$ defined by

$$\ell_0(x) = |\text{supp}(x)| = \text{number of nonzero components of } x, \quad \forall x \in \mathbb{R}^d, \quad (1)$$

where $|K|$ denotes the cardinal of a subset $K \subset \llbracket 1, d \rrbracket$. 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 to norms, the ℓ_0 pseudonorm is 0-homogeneous:

$$\ell_0(\rho x) = \ell_0(x) , \quad \forall \rho \in \mathbb{R} \setminus \{0\} , \quad \forall x \in \mathbb{R}^d . \quad (2)$$

The level sets of the ℓ_0 pseudonorm. 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

$$\text{the level sets} \quad \ell_0^{\leq k} = \{x \in \mathbb{R}^d \mid \ell_0(x) \leq k\} , \quad \forall k \in \{0, 1, \dots, d\} , \quad (3a)$$

$$\text{and the level curves} \quad \ell_0^=k = \{x \in \mathbb{R}^d \mid \ell_0(x) = k\} , \quad \forall k \in \{0, 1, \dots, d\} . \quad (3b)$$

For any subset $K \subset \llbracket 1, d \rrbracket$, we denote the subspace of \mathbb{R}^d made of vectors whose components vanish outside of K by¹

$$\mathcal{R}_K = \mathbb{R}^K \times \{0\}^{-K} = \{x \in \mathbb{R}^d \mid x_j = 0 , \quad \forall j \notin K\} \subset \mathbb{R}^d , \quad (4)$$

where $\mathcal{R}_\emptyset = \{0\}$. We denote by $\pi_K : \mathbb{R}^d \rightarrow \mathcal{R}_K$ the *orthogonal projection mapping* and, for any vector $x \in \mathbb{R}^d$, by $x_K = \pi_K(x) \in \mathcal{R}_K$ the vector which coincides with x , except for the components outside of K that are zero. It is easily seen that the orthogonal projection mapping π_K is self-dual, giving

$$\langle x_K, y_K \rangle = \langle x_K, y \rangle = \langle \pi_K(x), y \rangle = \langle x, \pi_K(y) \rangle = \langle x, y_K \rangle , \quad \forall x \in \mathbb{R}^d , \quad \forall y \in \mathbb{R}^d . \quad (5)$$

The level sets of the ℓ_0 pseudonorm in (3a) are easily related to the subspaces \mathcal{R}_K of \mathbb{R}^d by²

$$\ell_0^{\leq k} = \{x \in \mathbb{R}^d \mid \ell_0(x) \leq k\} = \bigcup_{|K| \leq k} \mathcal{R}_K , \quad \forall k \in \llbracket 0, d \rrbracket . \quad (6)$$

Decreasingly graded sequence of norms with respect to the ℓ_0 pseudonorm. Now, we introduce the notion of sequences of norms that are, strictly or not, decreasingly graded with respect to the ℓ_0 pseudonorm: in a sense, the monotone sequence detects the number of nonzero components of a vector in \mathbb{R}^d when it becomes stationary. In the following definition, $\{\|\cdot\|_k\}_{k \in \llbracket 1, d \rrbracket}$ denotes any sequence of norms on \mathbb{R}^d .

Definition 2.1 ([9, Definition ??]) *We say that a sequence $\{\|\cdot\|_k\}_{k \in \llbracket 1, d \rrbracket}$ of norms on \mathbb{R}^d is decreasingly graded (resp. strictly decreasingly graded) w.r.t. (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 implication (resp. equivalence), for any $l \in \llbracket 1, d \rrbracket$,*

$$\ell_0(x) = l \implies \|x\|_1 \geq \dots \geq \|x\|_{l-1} \geq \|x\|_l = \dots = \|x\|_d , \quad (7a)$$

$$(\text{ resp. } \ell_0(x) = l \iff \|x\|_1 \geq \dots \geq \|x\|_{l-1} > \|x\|_l = \dots = \|x\|_d .) \quad (7b)$$

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

²The notation $\bigcup_{|K| \leq k}$ is a shorthand for $\bigcup_{K \subset \llbracket 1, d \rrbracket, |K| \leq k}$.

2. The sequence $k \in \llbracket 1, d \rrbracket \mapsto \|x\|_k$ is nonincreasing and we have the implication (resp. equivalence), for any $l \in \llbracket 1, d \rrbracket$,

$$\ell_0(x) \leq l \implies \|x\|_l = \|x\|_d, \quad (7c)$$

$$(\text{ resp. } \ell_0(x) \leq l \iff \|x\|_l = \|x\|_d \quad (\iff \|x\|_l \leq \|x\|_d) .) \quad (7d)$$

3. The sequence $k \in \llbracket 1, d \rrbracket \mapsto \|x\|_k$ is nonincreasing and we have the inequality (resp. equality)

$$\ell_0(x) \geq \min \{ k \in \llbracket 1, d \rrbracket \mid \|x\|_k = \|x\|_d \}, \quad (7e)$$

$$(\text{ resp. } \ell_0(x) = \min \{ k \in \llbracket 1, d \rrbracket \mid \|x\|_k = \|x\|_d \} .) \quad (7f)$$

3 Coordinate- k norms and dual coordinate- k norms

In § 3.1, we provide background on norms. Then, in § 3.2, we introduce coordinate- k norms and dual coordinate- k norms.

3.1 Background on norms

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

$$\mathbb{S} = \{ x \in \mathbb{R}^d \mid \|x\| = 1 \}, \quad \mathbb{B} = \{ x \in \mathbb{R}^d \mid \|x\| \leq 1 \}. \quad (8)$$

Dual norms. We recall that the expression $\|y\|_\star = \sup_{\|x\| \leq 1} \langle x, y \rangle$, $\forall y \in \mathbb{R}^d$, defines a norm on \mathbb{R}^d , called the *dual norm* $\|\cdot\|_\star$. By definition of the dual norm, we have the inequality

$$\langle x, y \rangle \leq \|x\| \times \|y\|_\star, \quad \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d. \quad (9)$$

We denote the unit sphere \mathbb{S}_\star and the unit ball \mathbb{B}_\star of the dual norm $\|\cdot\|_\star$ by

$$\mathbb{S}_\star = \{ y \in \mathbb{R}^d \mid \|y\|_\star = 1 \}, \quad \mathbb{B}_\star = \{ y \in \mathbb{R}^d \mid \|y\|_\star \leq 1 \}. \quad (10)$$

Denoting by σ_S the *support function* of the set $S \subset \mathbb{R}^d$ ($\sigma_S(y) = \sup_{x \in S} \langle x, y \rangle$), we have

$$\|\cdot\| = \sigma_{\mathbb{B}_\star} = \sigma_{\mathbb{S}_\star} \quad \text{and} \quad \|\cdot\|_\star = \sigma_{\mathbb{B}} = \sigma_{\mathbb{S}}, \quad (11)$$

where $\mathbb{B}_\star = \mathbb{B}^\circ = \{ y \in \mathbb{R}^d \mid \langle x, y \rangle \leq 1, \forall x \in \mathbb{B} \}$ is the polar set \mathbb{B}° of the unit ball \mathbb{B} .

Restriction norms.

Definition 3.1 For any norm $\|\cdot\|$ on \mathbb{R}^d and any subset $K \subset \llbracket 1, d \rrbracket$, we define

- the K -restriction norm $\|\cdot\|_K$ on the subspace \mathcal{R}_K of \mathbb{R}^d , as defined in (4), by

$$\|x\|_K = \|x\|, \quad \forall x \in \mathcal{R}_K. \quad (12)$$

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

We have that [9, Equation (13b)]

$$\|y\|_{K, \star} = \sigma_{\mathcal{R}_K \cap \mathbb{B}}(y) = \sigma_{\mathcal{R}_K \cap \mathbb{S}}(y), \quad \forall y \in \mathcal{R}_K. \quad (13)$$

3.2 Coordinate- k and dual coordinate- k norms

Source norm. Let $\|\cdot\|$ be a norm on \mathbb{R}^d , that we will call the *source norm*.

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

Definition 3.2 For $k \in \llbracket 1, d \rrbracket$, we call coordinate- k norm the norm $\|\cdot\|_{(k)}^{\mathcal{R}}$ whose dual norm is the dual coordinate- k norm, denoted by $\|\cdot\|_{(k), \star}^{\mathcal{R}}$, with expression³

$$\|y\|_{(k), \star}^{\mathcal{R}} = \sup_{|K| \leq k} \|y_K\|_{K, \star}, \quad \forall y \in \mathbb{R}^d, \quad (14)$$

where the (K, \star) -norm $\|\cdot\|_{K, \star}$ is given in Definition 3.1.

It is easily verified that $\|\cdot\|_{(k), \star}^{\mathcal{R}}$ indeed is a norm. We adopt the convention $\|\cdot\|_{(0), \star}^{\mathcal{R}} = 0$ (although this is not a norm on \mathbb{R}^d , but a seminorm).

Examples. Table 1 provides examples [9,10]. With this, we define the top (k, q) -norms in the last right column of Table 1. The (p, k) -support norm, in the middle column of Table 1, is defined as the dual norm of the top (k, q) -norm, with $1/p + 1/q = 1$.

To prepare Sect. 4, we provide properties of coordinate- k and dual coordinate- k norms.

³The notation $\sup_{|K| \leq k}$ is a shorthand for $\sup_{K \subset \llbracket 1, d \rrbracket, |K| \leq k}$.

source norm $\ \cdot\ $	$\ \cdot\ _{(k)}^{\mathcal{R}}$	$\ \cdot\ _{(k),\star}^{\mathcal{R}}$
$\ \cdot\ _p$	(p, k) -support norm $\ x\ _{p,k}^{\text{sn}}$	top (k, q) -norm $\ y\ _{k,q}^{\text{tn}}$ $= (\sum_{j=1}^k y_{\nu(j)} ^q)^{1/q}, 1/p + 1/q = 1$
$\ \cdot\ _1$	$(1, k)$ -support norm ℓ_1 -norm $\ x\ _{1,k}^{\text{sn}} = \ x\ _1$	top (k, ∞) -norm ℓ_∞ -norm $\ y\ _{k,\infty}^{\text{tn}} = y_{\nu(1)} = \ y\ _\infty$
$\ \cdot\ _2$	$(2, k)$ -support norm	top $(k, 2)$ -norm $\ y\ _{k,2}^{\text{tn}} = \sqrt{\sum_{j=1}^k y_{\nu(j)} ^2}$
$\ \cdot\ _\infty$	(∞, k) -support norm	top $(k, 1)$ -norm $\ y\ _{k,1}^{\text{tn}} = \sum_{j=1}^k y_{\nu(j)} $

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

Properties of dual coordinate- k norms. We denote the unit sphere $\mathbb{S}_{(k),\star}^{\mathcal{R}}$ and the unit ball $\mathbb{B}_{(k),\star}^{\mathcal{R}}$ of the dual coordinate- k norm $\|\cdot\|_{(k),\star}^{\mathcal{R}}$ in Definition 3.2 by

$$\mathbb{S}_{(k),\star}^{\mathcal{R}} = \{y \in \mathbb{R}^d \mid \|y\|_{(k),\star}^{\mathcal{R}} = 1\}, \quad \mathbb{B}_{(k),\star}^{\mathcal{R}} = \{y \in \mathbb{R}^d \mid \|y\|_{(k),\star}^{\mathcal{R}} \leq 1\}, \quad k \in \llbracket 1, d \rrbracket. \quad (15)$$

Proposition 3.3

- For $k \in \llbracket 1, d \rrbracket$, the dual coordinate- k norm satisfies

$$\|y\|_{(k),\star}^{\mathcal{R}} = \sup_{|K| \leq k} \sigma_{(\mathcal{R}_K \cap \mathbb{S})}(y) = \sigma_{\ell_0^{\leq k} \cap \mathbb{S}}(y) = \sigma_{\ell_0^{=k} \cap \mathbb{S}}(y), \quad \forall y \in \mathbb{R}^d. \quad (16)$$

- We have the equality

$$\|\cdot\|_{\star} = \|\cdot\|_{(d),\star}^{\mathcal{R}}. \quad (17)$$

- The sequence $\left\{ \|\cdot\|_{(j),\star}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of dual coordinate- k norms in Definition 3.2 is nondecreasing, that is, the following inequalities and equality hold true:

$$\|y\|_{(1),\star}^{\mathcal{R}} \leq \dots \leq \|y\|_{(j),\star}^{\mathcal{R}} \leq \|y\|_{(j+1),\star}^{\mathcal{R}} \leq \dots \leq \|y\|_{(d),\star}^{\mathcal{R}} = \|y\|_{\star}, \quad \forall y \in \mathbb{R}^d. \quad (18)$$

- The sequence $\left\{ \mathbb{B}_{(j),\star}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of units balls of the dual coordinate- k norms in Definition 3.2 is nonincreasing, that is, the following equality and inclusions hold true:

$$\mathbb{B}_{\star} = \mathbb{B}_{(d),\star}^{\mathcal{R}} \subset \dots \subset \mathbb{B}_{(j+1),\star}^{\mathcal{R}} \subset \mathbb{B}_{(j),\star}^{\mathcal{R}} \subset \dots \subset \mathbb{B}_{(1),\star}^{\mathcal{R}}. \quad (19)$$

Proof.

- For any $y \in \mathbb{R}^d$, we have

$$\begin{aligned}
\|y\|_{(k),\star}^{\mathcal{R}} &= \sup_{|K| \leq k} \|y_K\|_{K,\star} && \text{(by definition (14) of } \|y\|_{(k),\star}^{\mathcal{R}} \text{)} \\
&= \sup_{|K| \leq k} \sigma_{(\mathcal{R}_K \cap \mathbb{S})}(y_K) && \text{(as } \|y_K\|_{K,\star} = \sigma_{(\mathcal{R}_K \cap \mathbb{S})}(y_K) \text{ by (13))} \\
&= \sup_{|K| \leq k} \sup_{x \in \mathcal{R}_K \cap \mathbb{S}} \langle x, y_K \rangle && \text{(by definition of the support function } \sigma_{(\mathcal{R}_K \cap \mathbb{S})} \text{)} \\
&= \sup_{|K| \leq k} \sup_{x \in \mathcal{R}_K \cap \mathbb{S}} \langle x, y \rangle && \text{(by (5) as } x \in \mathcal{R}_K \text{)} \\
&= \sup_{|K| \leq k} \sigma_{(\mathcal{R}_K \cap \mathbb{S})}(y) && \text{(by definition of the support function } \sigma_{(\mathcal{R}_K \cap \mathbb{S})} \text{)} \\
&= \sigma_{\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S})}(y) && \text{(as the support function turns a union of sets into a supremum)} \\
&= \sigma_{\ell_0^{\leq k} \cap \mathbb{S}}(y). && \text{(as } \ell_0^{\leq k} \cap \mathbb{S} = \bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S}) \text{ by (6))}
\end{aligned}$$

To finish, we will now prove that $\sigma_{\ell_0^{\leq k} \cap \mathbb{S}} = \sigma_{\overline{\ell_0^{\leq k} \cap \mathbb{S}}}$. For this purpose, we show in two steps that $\ell_0^{\leq k} \cap \mathbb{S} = \overline{\ell_0^{\leq k} \cap \mathbb{S}}$.

First, we establish the (known) fact that $\overline{\ell_0^{\leq k}} = \ell_0^{\leq k}$. The inclusion $\overline{\ell_0^{\leq k}} \subset \ell_0^{\leq k}$ is easy because, on the one hand, $\ell_0^{\leq k} \subset \overline{\ell_0^{\leq k}}$ and, on the other hand, the level set $\ell_0^{\leq k}$ in (3a) is closed, as follows from the well-known property that the pseudonorm ℓ_0 is lower semicontinuous. There remains to prove the reverse inclusion $\ell_0^{\leq k} \subset \overline{\ell_0^{\leq k}}$. For this purpose, we consider $x \in \ell_0^{\leq k}$. If $x \in \ell_0^{\leq k}$, obviously $x \in \overline{\ell_0^{\leq k}}$. Therefore, we suppose that $\ell_0(x) = l < k$. By definition of $\ell_0(x)$ in (1), there exists $L \subset \{1, \dots, d\}$ such that $|L| = l < k$ and $x = x_L$. For $\epsilon > 0$, define x^ϵ as coinciding with x except for $k - l$ indices outside L for which the components are $\epsilon > 0$. By construction $\ell_0(x^\epsilon) = k$ and $x^\epsilon \rightarrow x$ when $\epsilon \rightarrow 0$. This proves that $\ell_0^{\leq k} \subset \overline{\ell_0^{\leq k}}$.

Second, we prove that $\overline{\ell_0^{\leq k} \cap \mathbb{S}} = \ell_0^{\leq k} \cap \mathbb{S}$. The inclusion $\overline{\ell_0^{\leq k} \cap \mathbb{S}} \subset \ell_0^{\leq k} \cap \mathbb{S}$, is easy. Indeed, $\overline{\ell_0^{\leq k}} = \ell_0^{\leq k} \implies \overline{\ell_0^{\leq k} \cap \mathbb{S}} \subset \overline{\ell_0^{\leq k}} \cap \overline{\mathbb{S}} = \ell_0^{\leq k} \cap \mathbb{S}$. To prove the reverse inclusion $\ell_0^{\leq k} \cap \mathbb{S} \subset \overline{\ell_0^{\leq k} \cap \mathbb{S}}$, we consider $x \in \ell_0^{\leq k} \cap \mathbb{S}$. As we have just seen that $\ell_0^{\leq k} = \overline{\ell_0^{\leq k}}$, we deduce that $x \in \overline{\ell_0^{\leq k}}$. Therefore, there exists a sequence $\{z_n\}_{n \in \mathbb{N}}$ in $\ell_0^{\leq k}$ such that $z_n \rightarrow x$ when $n \rightarrow +\infty$. Since $x \in \mathbb{S}$, we can always suppose that $z_n \neq 0$, for all $n \in \mathbb{N}$. Therefore $z_n / \|z_n\|$ is well defined and, when $n \rightarrow +\infty$, we have $z_n / \|z_n\| \rightarrow x / \|x\| = x$ since $x \in \mathbb{S} = \{x \in \mathbb{X} \mid \|x\| = 1\}$. Now, on the one hand, $z_n / \|z_n\| \in \ell_0^{\leq k}$, for all $n \in \mathbb{N}$, and, on the other hand, $z_n / \|z_n\| \in \mathbb{S}$. As a consequence $z_n / \|z_n\| \in \ell_0^{\leq k} \cap \mathbb{S}$, and we conclude that $x \in \overline{\ell_0^{\leq k} \cap \mathbb{S}}$. Thus, we have proved that $\ell_0^{\leq k} \cap \mathbb{S} \subset \overline{\ell_0^{\leq k} \cap \mathbb{S}}$.

From $\ell_0^{\leq k} \cap \mathbb{S} = \overline{\ell_0^{\leq k} \cap \mathbb{S}}$, we get that $\sigma_{\ell_0^{\leq k} \cap \mathbb{S}} = \sigma_{\overline{\ell_0^{\leq k} \cap \mathbb{S}}} = \sigma_{\ell_0^{\leq k} \cap \mathbb{S}}$, by [11, Proposition 7.13]. Thus, we have proved all equalities in (16).

- By the equality $\|y\|_{(k),\star}^{\mathcal{R}} = \sigma_{\ell_0^{\leq k} \cap \mathbb{S}}(y)$ in (16), we get that, for all $y \in \mathbb{R}^d$, $\|y\|_{(d),\star}^{\mathcal{R}} = \sigma_{\ell_0^{\leq d} \cap \mathbb{S}}(y) = \sigma_{\mathbb{S}}(y) = \|y\|_{\star}$ since $\ell_0^{\leq d} = \mathbb{R}^d$ and by (11).
- The inequalities in (18) easily derive from the very definition (14) of the dual coordinate- k norms $\|\cdot\|_{(k),\star}^{\mathcal{R}}$. The last equality is just the equality (17).
- The equality and the inclusions in (19) directly follow from the inequalities and the equality between norms in (18).

This ends the proof. \square

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

$$\mathbb{S}_{(k)}^{\mathcal{R}} = \{x \in \mathbb{R}^d \mid \|x\|_{(k)}^{\mathcal{R}} = 1\}, \quad \mathbb{B}_{(k)}^{\mathcal{R}} = \{x \in \mathbb{R}^d \mid \|x\|_{(k)}^{\mathcal{R}} \leq 1\}. \quad (20)$$

We adopt the convention $\mathbb{B}_{(0)}^{\mathcal{R}} = \{0\}$ (although this is not the unit ball of a norm on \mathbb{R}^d).

Proposition 3.4

- For $k \in \llbracket 1, d \rrbracket$, the coordinate- k norm $\|\cdot\|_{(k)}^{\mathcal{R}}$ has unit ball

$$\mathbb{B}_{(k)}^{\mathcal{R}} = \overline{\text{co}}\left(\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S})\right), \quad (21)$$

where $\overline{\text{co}}(S)$ denotes the closed convex hull of a subset $S \subset \mathbb{R}^d$.

- We have the equality

$$\|\cdot\|_{(d)}^{\mathcal{R}} = \|\cdot\|. \quad (22)$$

- The sequence $\left\{\|\cdot\|_{(j)}^{\mathcal{R}}\right\}_{j \in \llbracket 1, d \rrbracket}$ of coordinate- k norms in Definition 3.2 is nonincreasing, that is, the following equality and inequalities hold true:

$$\|x\| = \|x\|_{(d)}^{\mathcal{R}} \leq \cdots \leq \|x\|_{(j+1)}^{\mathcal{R}} \leq \|x\|_{(j)}^{\mathcal{R}} \leq \cdots \leq \|x\|_{(1)}^{\mathcal{R}}, \quad \forall x \in \mathbb{R}^d. \quad (23)$$

- The sequence $\left\{\mathbb{B}_{(j)}^{\mathcal{R}}\right\}_{j \in \llbracket 1, d \rrbracket}$ of units balls of the coordinate- k norms in (21) is nondecreasing, that is, the following inclusions and equality hold true:

$$\mathbb{B}_{(1)}^{\mathcal{R}} \subset \cdots \subset \mathbb{B}_{(j)}^{\mathcal{R}} \subset \mathbb{B}_{(j+1)}^{\mathcal{R}} \subset \cdots \subset \mathbb{B}_{(d)}^{\mathcal{R}} = \mathbb{B}. \quad (24)$$

Proof.

- For any $y \in \mathbb{R}^d$, we have

$$\begin{aligned} \|y\|_{(k), \star}^{\mathcal{R}} &= \sup_{|K| \leq k} \sigma_{(\mathcal{R}_K \cap \mathbb{S})}(y) && \text{(by (16))} \\ &= \sigma_{\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S})}(y) && \text{(as the support function turns a union of sets into a supremum)} \\ &= \sigma_{\overline{\text{co}}(\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S}))}(y) && \text{(by [11, Proposition 7.13])} \end{aligned}$$

and we conclude that $\mathbb{B}_{(k)}^{\mathcal{R}} = \overline{\text{co}}(\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S}))$ by (11). Thus, we have proved (21).

- From the equality (17), we deduce the equality (22) between the dual norms by definition of the dual norm.
- The equality and inequalities between norms in (23) easily derive from the inclusions and equality between unit balls in (24).
- The inclusions and equality between unit balls in (24) directly follow from the inclusions and equality between unit balls in (19) and from $\mathbb{B}_{(j)}^{\mathcal{R}} = (\mathbb{B}_{(j),\star}^{\mathcal{R}})^{\odot}$, the polar set of $\mathbb{B}_{(j),\star}^{\mathcal{R}}$.

This ends the proof. \square

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 $p \in]1, \infty[$.

We now show that the sequences $\left\{ \|\cdot\|_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of coordinate- k norms (in Definition 3.2) are naturally decreasingly graded with respect to the ℓ_0 pseudonorm (as in Definition 2.1). Part of the proof relies upon the forthcoming Lemma 3.6.

Proposition 3.5

1. The nonincreasing sequence $\left\{ \|\cdot\|_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of coordinate- k norms is decreasingly graded with respect to the ℓ_0 pseudonorm, that is, for any $l \in \llbracket 1, d \rrbracket$,

$$\ell_0(x) \leq l \implies \|x\| = \|x\|_{(l)}^{\mathcal{R}}, \quad \forall x \in \mathbb{R}^d. \quad (25a)$$

2. If the normed space $(\mathbb{R}^d, \|\cdot\|)$ is strictly convex, then the nonincreasing sequence $\left\{ \|\cdot\|_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of coordinate- k norms is strictly decreasingly graded with respect to the ℓ_0 pseudonorm, that is, for any $l \in \llbracket 1, d \rrbracket$,

$$\ell_0(x) \leq l \iff \|x\| = \|x\|_{(l)}^{\mathcal{R}}, \quad \forall x \in \mathbb{R}^d. \quad (25b)$$

Proof.

- We prove Item 1. As the sequence $\left\{ \|\cdot\|_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of coordinate- k norms is nonincreasing by (18), it suffices to show that (7c) holds true — that is, that (25a) holds true — to prove that the sequence is decreasingly graded with respect to the ℓ_0 pseudonorm (see Definition 2.1).

Now, for any $x \in \mathbb{R}^d$ and for any $k \in \llbracket 1, d \rrbracket$, we have⁴

⁴In what follows, by “or” we mean the so-called *exclusive or* (exclusive disjunction). Thus, every “or” should be understood as “or $x \neq 0$ and”

$$x \in \ell_0^{\leq k} \iff x = 0 \text{ or } \frac{x}{\|x\|} \in \ell_0^{\leq k}$$

(by 0-homogeneity (2) of the ℓ_0 pseudonorm, and by definition (3a) of $\ell_0^{\leq k}$)

$$\iff x = 0 \text{ or } \frac{x}{\|x\|} \in \ell_0^{\leq k} \cap \mathbb{S} \quad (\text{as } \frac{x}{\|x\|} \in \mathbb{S} \text{ by definition (8) of the unit sphere } \mathbb{S})$$

$$\iff x = 0 \text{ or } \frac{x}{\|x\|} \in \bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S}) \quad (\text{as } \ell_0^{\leq k} = \bigcup_{|K| \leq k} \mathcal{R}_K \text{ by (6)})$$

$$\implies x = 0 \text{ or } \frac{x}{\|x\|} \in \mathbb{B}_{(k)}^{\mathcal{R}} \quad (\text{as } \mathbb{B}_{(k)}^{\mathcal{R}} = \overline{\text{co}}(\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S})) \text{ by (21)})$$

$$\implies x = 0 \text{ or } \left\| \frac{x}{\|x\|} \right\|_{(k)}^{\mathcal{R}} \leq 1 \quad (\text{since } \mathbb{B}_{(k)}^{\mathcal{R}} \text{ is the unit ball of the norm } \|\cdot\|_{(k)}^{\mathcal{R}} \text{ by (20)})$$

$$\implies \|x\|_{(k)}^{\mathcal{R}} \leq \|x\|$$

$$\implies \|x\|_{(k)}^{\mathcal{R}} \leq \|x\| = \|x\|_{(d)}^{\mathcal{R}} \quad (\text{where the last equality comes from (23)})$$

$$\implies \|x\|_{(k)}^{\mathcal{R}} = \|x\|_{(d)}^{\mathcal{R}} \quad (\text{as } \|x\|_{(k)}^{\mathcal{R}} \geq \|x\|_{(d)}^{\mathcal{R}} \text{ by (23)})$$

Therefore, we have obtained (25a).

• We prove Item 2. As the sequence $\left\{ \|\cdot\|_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of coordinate- k norms is nonincreasing by (18), it suffices to show that (7d) holds true — that is, that (25b) holds true — to prove that the sequence is strictly decreasingly graded with respect to the ℓ_0 pseudonorm (see Definition 2.1).

We suppose that the normed space $(\mathbb{R}^d, \|\cdot\|)$ is strictly convex. Then, for any $x \in \mathbb{R}^d$ and for any $k \in \llbracket 1, d \rrbracket$, we have ⁵

$$x \in \ell_0^{\leq k} \iff x = 0 \text{ or } \frac{x}{\|x\|} \in \ell_0^{\leq k}$$

by 0-homogeneity (2) of the ℓ_0 pseudonorm, and by definition (3a) of $\ell_0^{\leq k}$

$$\iff x = 0 \text{ or } \frac{x}{\|x\|} \in \ell_0^{\leq k} \cap \mathbb{S} \quad (\text{as } \frac{x}{\|x\|} \in \mathbb{S} \text{ by definition (8) of the unit sphere } \mathbb{S})$$

$$\iff x = 0 \text{ or } \frac{x}{\|x\|} \in \mathbb{B}_{(k)}^{\mathcal{R}} \cap \mathbb{S}$$

as $\ell_0^{\leq k} \cap \mathbb{S} = \mathbb{B}_{(k)}^{\mathcal{R}} \cap \mathbb{S}$ by (27) since the assumption of Lemma 3.6 is satisfied, that is, the normed space $(\mathbb{R}^d, \|\cdot\|)$ is strictly convex

$$\iff x = 0 \text{ or } \left\| \frac{x}{\|x\|} \right\|_{(k)}^{\mathcal{R}} \leq 1 \quad (\text{since } \mathbb{B}_{(k)}^{\mathcal{R}} \text{ is the unit ball of the norm } \|\cdot\|_{(k)}^{\mathcal{R}} \text{ by (20)})$$

$$\iff \|x\|_{(k)}^{\mathcal{R}} \leq \|x\|$$

$$\iff \|x\|_{(k)}^{\mathcal{R}} \leq \|x\| = \|x\|_{(d)}^{\mathcal{R}} \quad (\text{where the last equality comes from (23)})$$

$$\iff \|x\|_{(k)}^{\mathcal{R}} = \|x\|_{(d)}^{\mathcal{R}} \quad (\text{as } \|x\|_{(k)}^{\mathcal{R}} \geq \|x\|_{(d)}^{\mathcal{R}} \text{ by (23)})$$

⁵See Footnote 4.

Therefore, we have obtained (25b).

This ends the proof. \square

	$\left\{ \left\ \cdot \right\ _{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$	
	graded	strictly graded
$\left\ \cdot \right\ $ is any norm	✓	
$(\mathbb{R}^d, \left\ \cdot \right\)$ is strictly convex		✓

Table 2: Table of results. It reads as follows: to obtain that the sequence $\left\{ \left\| \cdot \right\|_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ be graded (second column), it suffices that $\left\| \cdot \right\|$ be any norm; to obtain that the sequence $\left\{ \left\| \cdot \right\|_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ be strictly graded (third column), it suffices that $(\mathbb{R}^d, \left\| \cdot \right\|)$ be strictly convex.

Table 2 summarizes the results of Proposition 3.5. As an application with any ℓ_p -norm $\left\| \cdot \right\|_p$ for source norm (for $p \in [1, \infty]$), we obtain that the nonincreasing sequence $\left\{ \left\| \cdot \right\|_{p,j}^{\text{sn}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of (p, k) -support norms (see Table 1) is strictly decreasingly graded w.r.t. the ℓ_0 pseudonorm for $p \in]1, \infty[$. This gives, by (7f):

$$\ell_0(x) = \min \left\{ k \in \llbracket 1, d \rrbracket \mid \|x\|_{p,k}^{\text{sn}} = \|x\|_p \right\}, \quad \forall x \in \mathbb{R}^d, \quad \forall p \in]1, \infty[. \quad (26a)$$

We also have that the sequence $\left\{ \left\| \cdot \right\|_{p,j}^{\text{sn}} \right\}_{j \in \llbracket 1, d \rrbracket}$ is decreasingly graded with respect to the ℓ_0 pseudonorm for $p \in [1, \infty]$. Looking at Table 1, the only interesting case is for $p = \infty$, giving, by (7e):

$$\ell_0(x) \geq \min \left\{ k \in \llbracket 1, d \rrbracket \mid \|x\|_{\infty,k}^{\text{sn}} = \|x\|_{\infty} \right\}, \quad \forall x \in \mathbb{R}^d. \quad (26b)$$

Lemma 3.6 *Let $\left\| \cdot \right\|$ be a norm on \mathbb{R}^d . If the normed space $(\mathbb{R}^d, \left\| \cdot \right\|)$ is strictly convex, we have the equality*

$$\ell_0^{\leq k} \cap \mathbb{S} = \mathbb{B}_{(k)}^{\mathcal{R}} \cap \mathbb{S}, \quad \forall k \in \llbracket 0, d \rrbracket, \quad (27)$$

where $\ell_0^{\leq k}$ is the level set in (3a) of the ℓ_0 pseudonorm in (1), where \mathbb{S} is the unit sphere in (8), and where $\mathbb{B}_{(k)}^{\mathcal{R}}$ in (20) is the unit ball of the norm $\left\| \cdot \right\|_{(k)}^{\mathcal{R}}$.

Proof. It is proved in [9, Proposition 14] that, if the unit ball \mathbb{B} is rotund — that is, if the normed space $(\mathbb{R}^d, \left\| \cdot \right\|)$ is strictly convex — and if A is a closed subset of \mathbb{S} , then $A = \overline{\text{co}}(A) \cap \mathbb{S}$.

Now, we turn to the proof. First, we observe that the level set $\ell_0^{\leq k}$ is closed because the pseudonorm ℓ_0 is lower semi continuous. Second, we have

$$\ell_0^{\leq k} \cap \mathbb{S} = \overline{\text{co}}(\ell_0^{\leq k} \cap \mathbb{S}) \cap \mathbb{S}$$

because $\ell_0^{\leq k} \cap \mathbb{S} \subset \mathbb{S}$ and is closed, and because the unit ball \mathbb{B} is rotund

$$\begin{aligned} &= \overline{\text{co}}\left(\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S})\right) \cap \mathbb{S} && \text{(by (6))} \\ &= \mathbb{B}_{(k)}^{\mathcal{R}} \cap \mathbb{S} . && \text{(by (21))} \end{aligned}$$

This ends the proof. \square

4 The Capra-conjugacy and the ℓ_0 pseudonorm

We introduce the coupling CAPRA in §4.1. Then, we provide formulas for CAPRA-conjugates of functions of the ℓ_0 pseudonorm in §4.2, for CAPRA-biconjugates of functions of the ℓ_0 pseudonorm in §4.3, and for CAPRA-subdifferentials of functions of the ℓ_0 pseudonorm in §4.4.

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). As we manipulate functions with values in $\overline{\mathbb{R}} = [-\infty, +\infty]$, we adopt the Moreau *lower* ($\dot{+}$) and *upper* ($\dot{+}$) additions [12], which extend the usual addition (+) with $(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = -\infty$ and $(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = +\infty$. For any subset $W \subset \mathbb{W}$ of a set \mathbb{W} , $\delta_W : \mathbb{W} \rightarrow \overline{\mathbb{R}}$ denotes the *characteristic function* of the set W : $\delta_W(w) = 0$ if $w \in W$, and $\delta_W(w) = +\infty$ if $w \notin W$.

4.1 Constant along primal rays coupling (Capra)

We introduce the coupling CAPRA, which is a special case of one-sided linear coupling, as introduced in [8]. Fenchel-Moreau conjugacies are recalled in Appendix A.

Definition 4.1 *Let $\|\cdot\|$ be a norm on \mathbb{R}^d . We define the constant along primal rays coupling $\dot{\phi}$, or CAPRA, between \mathbb{R}^d and itself by*

$$\forall y \in \mathbb{R}^d, \quad \dot{\phi}(x, y) = \frac{\langle x, y \rangle}{\|x\|}, \quad \forall x \in \mathbb{R}^d \setminus \{0\} \text{ and } \dot{\phi}(0, y) = 0. \quad (28)$$

We stress the point that, in (28), 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:

$$n : \mathbb{R}^d \rightarrow \mathbb{S} \cup \{0\}, \quad n(x) = \frac{x}{\|x\|} \text{ if } x \neq 0 \text{ and } n(0) = 0. \quad (29)$$

With these notations, the coupling CAPRA in (28) is a special case of one-sided linear coupling, the *Fenchel coupling after primal normalization*: $\dot{\phi}(x, y) = \langle n(x), y \rangle$, $\forall x \in \mathbb{R}^d$, $\forall y \in \mathbb{R}^d$. We will see below that the CAPRA-conjugacy, induced by the coupling CAPRA, shares some relations with the Fenchel conjugacy (see Appendix A).

Capra-conjugates and biconjugates. Here are expressions for the CAPRA-conjugates and biconjugates of a function. The following Proposition simply is [8, Proposition 2.5] with the normalization mapping n in (29).

Proposition 4.2 *For any function $g : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$, the ζ' -Fenchel-Moreau conjugate is given by*

$$g^{\zeta'} = g^{\star'} \circ n . \quad (30a)$$

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

$$f^{\zeta} = (\inf [f \mid n])^{\star} , \quad (30b)$$

where the conditional infimum $\inf [f \mid n]$, defined in [8, Definition 2.4], has the expression

$$\inf [f \mid n](x) = \inf \{ f(x') \mid n(x') = x \} = \begin{cases} \inf_{\lambda > 0} f(\lambda x) & \text{if } x \in \mathbb{S} \cup \{0\} , \\ +\infty & \text{if } x \notin \mathbb{S} \cup \{0\} , \end{cases} \quad (30c)$$

and the ζ -Fenchel-Moreau biconjugate is given by

$$f^{\zeta\zeta'} = (f^{\zeta})^{\star'} \circ n = (\inf [f \mid n])^{\star\star'} \circ n . \quad (30d)$$

The ζ -Fenchel-Moreau conjugate f^{ζ} is a closed convex function (see Appendix A).

Capra-convex functions. We recall that so-called ζ -convex functions are all functions of the form $g^{\zeta'}$, for any $g : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$, or, equivalently, all functions of the form $f^{\zeta\zeta'}$, for any $f : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$, or, equivalently, all functions that are equal to their ζ -biconjugate ($f^{\zeta\zeta'} = f$) [4–6]. 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 Appendix A). The following Proposition simply is [8, Proposition 2.6] with the normalization mapping n in (29).

Proposition 4.3 *A function is ζ -convex if and only if it is the composition of a closed convex function on \mathbb{R}^d with the normalization mapping (29). More precisely, for any function $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$, we have the equivalences*

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

For instance, letting $\|\cdot\|$ be any norm on \mathbb{R}^d (not necessarily the Euclidian norm), the function $\|\cdot\|/\|\cdot\|$ (with the value 0 at 0) is ζ -convex.

Capra-subdifferential. Following the definition of the subdifferential of a function with respect to a duality in [13], the *CAPRA-subdifferential* of the function $f : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ at $x \in \mathbb{R}^d$ has the following expressions

$$\partial_{\dot{\zeta}} f(x) = \{y \in \mathbb{R}^d \mid f^{\dot{\zeta}}(y) = \dot{\zeta}(x, y) \dot{+} (-f(x))\} \quad (32a)$$

$$= \{y \in \mathbb{R}^d \mid (\inf [f \mid n])^*(y) = \langle n(x), y \rangle \dot{+} (-f(x))\}, \quad (32b)$$

so that, thanks to the definition (29) of the normalization mapping n , we deduce that

$$\partial_{\dot{\zeta}} f(0) = \{y \in \mathbb{R}^d \mid (\inf [f \mid n])^*(y) = -f(0)\} \quad (32c)$$

$$\partial_{\dot{\zeta}} f(x) = \{y \in \mathbb{R}^d \mid (\inf [f \mid n])^*(y) = \frac{\langle x, y \rangle}{\|x\|} \dot{+} (-f(x))\}, \quad \forall x \in \mathbb{R}^d \setminus \{0\}. \quad (32d)$$

Now, we turn to analyze the ℓ_0 pseudonorm by means of the CAPRA conjugacy.

4.2 Capra-conjugates related to the ℓ_0 pseudonorm

With the Fenchel conjugacy, we calculate that $\delta_{\ell_0^{\leq k}}^* = \delta_{\{0\}}$ for all $k \in \llbracket 1, d \rrbracket$ — where $\delta_{\ell_0^{\leq k}}$ is the characteristic function of the level sets (3a) — and that $\ell_0^* = \delta_{\{0\}}$. Hence, the Fenchel conjugacy is not suitable to handle the ℓ_0 pseudonorm.

By contrast, we will now show that functions of the ℓ_0 pseudonorm in (1) — including the ℓ_0 pseudonorm itself and the characteristic functions $\delta_{\ell_0^{\leq k}}$ of its level sets (3a) — are related to the sequence of dual coordinate- k norms in Definition 3.2 by the following CAPRA-conjugacy formulas.

Proposition 4.4 *Let $\|\cdot\|$ be a norm on \mathbb{R}^d , with associated sequence $\left\{ \|\cdot\|_{(j),\star}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of dual coordinate- k norms in Definition 3.2, and associated CAPRA-coupling $\dot{\zeta}$ in (28).*

For any function $\varphi : \llbracket 0, d \rrbracket \rightarrow \overline{\mathbb{R}}$, we have (with the convention $\|\cdot\|_{(0),\star}^{\mathcal{R}} = 0$)

$$(\varphi \circ \ell_0)^{\dot{\zeta}} = \sup_{j \in \llbracket 0, d \rrbracket} \left[\|\cdot\|_{(j),\star}^{\mathcal{R}} - \varphi(j) \right]. \quad (33)$$

Proof. We prove (33):

$$(\varphi \circ \ell_0)^{\dot{\zeta}} = \left(\inf_{j \in \llbracket 0, d \rrbracket} [\delta_{\ell_0^{\leq j}} \dot{+} \varphi(j)] \right)^{\dot{\zeta}}$$

because $\varphi \circ \ell_0 = \inf_{j \in \llbracket 0, d \rrbracket} [\delta_{\ell_0^{\leq j}} \dot{+} \varphi(j)]$ since $\varphi \circ \ell_0$ takes the values $\varphi(j)$ on the level curves $\ell_0^{\leq j}$ of ℓ_0 in (3b)

$$= \sup_{j \in \llbracket 0, d \rrbracket} [\delta_{\ell_0^{\leq j}} \dot{+} \varphi(j)]^{\dot{\zeta}}$$

as conjugacies, being dualities, turn infima into suprema

$$\begin{aligned}
&= \sup_{j \in \llbracket 0, d \rrbracket} [\delta_{\ell_0^{-j}}^{\dot{\zeta}} + (-\varphi(j))] && \text{(by property of conjugacies)} \\
&= \sup_{j \in \llbracket 0, d \rrbracket} [\sigma_{n(\ell_0^{-j})} + (-\varphi(j))] && \text{(as } \delta_{\ell_0^{-j}}^{\dot{\zeta}} = \sigma_{n(\ell_0^{-j})} \text{ by [8, Proposition 2.5])} \\
&= \sup_{j \in \llbracket 0, d \rrbracket} \left\{ \sup \{0, \sigma_{\ell_0^{-j} \cap \mathbb{S}}\} + (-\varphi(j)) \right\}
\end{aligned}$$

as $n(\ell_0^{-j}) = \{0\} \cup (\ell_0^{-j} \cap \mathbb{S})$ by (29), and as the support function turns a union of sets into a supremum

$$\begin{aligned}
&= \sup_{j \in \llbracket 0, d \rrbracket} \left\{ \sigma_{\ell_0^{-j} \cap \mathbb{S}} + (-\varphi(j)) \right\} && \text{(as } \sigma_{\ell_0^{-j} \cap \mathbb{S}} \geq 0 \text{ since } \ell_0^{-j} \cap \mathbb{S} = -(\ell_0^{-j} \cap \mathbb{S})) \\
&= \sup \left\{ -\varphi(0), \sup_{j \in \llbracket 1, d \rrbracket} \left[\|\cdot\|_{(j), \star}^{\mathcal{R}} - \varphi(j) \right] \right\} && \text{(as } \sigma_{\ell_0^{-j} \cap \mathbb{S}} = \|\cdot\|_{(j), \star}^{\mathcal{R}} \text{ by (16))} \\
&= \sup_{j \in \llbracket 0, d \rrbracket} \left[\|\cdot\|_{(j), \star}^{\mathcal{R}} - \varphi(j) \right]. && \text{(using the convention that } \|\cdot\|_{(0), \star}^{\mathcal{R}} = 0)
\end{aligned}$$

This ends the proof. \square

With φ the identity function on $\llbracket 0, d \rrbracket$, we find the CAPRA-conjugate of the ℓ_0 pseudonorm. With the functions $\varphi = \delta_{\llbracket 0, k \rrbracket}$ (for any $k \in \llbracket 0, d \rrbracket$), we find the CAPRA-conjugates of the characteristic functions $\delta_{\ell_0^{\leq k}}$ of its level sets (3a). The corresponding expressions are given in Table 3.

4.3 Capra-biconjugates related to the ℓ_0 pseudonorm

With the Fenchel conjugacy, we calculate that $\delta_{\ell_0^{\leq k}}^{\star\star'} = 0$, for all $k \in \llbracket 1, d \rrbracket$, and that $\ell_0^{\star\star'} = 0$. Hence, the Fenchel conjugacy is not suitable to handle the ℓ_0 pseudonorm.

By contrast, we will now show that functions of the ℓ_0 pseudonorm in (1) — including the ℓ_0 pseudonorm itself and the characteristic functions $\delta_{\ell_0^{\leq k}}$ of its level sets (3a) — are related to the sequences of coordinate- k norms and dual coordinate- k norms in Definition 3.2 by the following CAPRA-biconjugacy formulas.

Proposition 4.5 *Let $\|\cdot\|$ be a norm on \mathbb{R}^d , with associated sequence $\left\{ \|\cdot\|_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of coordinate- k norms and sequence $\left\{ \|\cdot\|_{\star(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of dual coordinate- k norms, as in Definition 3.2, and with associated CAPRA coupling $\dot{\zeta}$ in (28).*

1. For any function $\varphi : \llbracket 0, d \rrbracket \rightarrow \overline{\mathbb{R}}$, we have

$$(\varphi \circ \ell_0)^{\dot{\zeta}\dot{\zeta}'}(x) = ((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'}\left(\frac{x}{\|x\|}\right), \quad \forall x \in \mathbb{R}^d \setminus \{0\}, \quad (34a)$$

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

$$((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'} = \left(\sup_{j \in \llbracket 0, d \rrbracket} [\llbracket \cdot \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j)] \right)^{\star'}, \quad (34b)$$

and also has the following four expressions as a Fenchel biconjugate

$$= \left(\inf_{j \in \llbracket 0, d \rrbracket} [\delta_{\mathbb{B}_{(j)}^{\mathcal{R}}} \dot{+} \varphi(j)] \right)^{\star\star'}, \quad (34c)$$

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

$$= \left(x \mapsto \inf \{ \varphi(j) \mid x \in \mathbb{B}_{(j)}^{\mathcal{R}}, j \in \llbracket 0, d \rrbracket \} \right)^{\star\star'}, \quad (34d)$$

$$= \left(\inf_{j \in \llbracket 0, d \rrbracket} [\delta_{\mathbb{S}_{(j)}^{\mathcal{R}}} \dot{+} \varphi(j)] \right)^{\star\star'}, \quad (34e)$$

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

$$= \left(x \mapsto \inf \{ \varphi(j) \mid x \in \mathbb{S}_{(j)}^{\mathcal{R}}, j \in \llbracket 0, d \rrbracket \} \right)^{\star\star'}. \quad (34f)$$

2. For any function $\varphi : \llbracket 0, d \rrbracket \rightarrow \mathbb{R}$, that is, with finite values, the function $((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'}$ is proper convex lsc and has the following variational expression (where Δ_{d+1} denotes the simplex of \mathbb{R}^{d+1})

$$((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'}(x) = \min_{\substack{(\lambda_0, \lambda_1, \dots, \lambda_d) \in \Delta_{d+1} \\ x \in \sum_{j=1}^d \lambda_j \mathbb{B}_{(j)}^{\mathcal{R}}}} \sum_{j=0}^d \lambda_j \varphi(j), \quad \forall x \in \mathbb{R}^d. \quad (34g)$$

3. For any function $\varphi : \llbracket 0, d \rrbracket \rightarrow \mathbb{R}_+$, that is, with nonnegative finite values, and such that $\varphi(0) = 0$, the function $((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'}$ is proper convex lsc and has the following two

variational expressions⁶

$$((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'}(x) = \min_{\substack{(\lambda_0, \lambda_1, \dots, \lambda_d) \in \Delta_{d+1} \\ x \in \sum_{j=1}^d \lambda_j \mathbb{S}_{(j)}^{\mathcal{R}}}} \sum_{j=1}^d \lambda_j \varphi(j), \quad \forall x \in \mathbb{R}^d, \quad (34h)$$

$$= \min_{\substack{z^{(1)} \in \mathbb{R}^d, \dots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d \|z^{(j)}\|_{(j)}^{\mathcal{R}} \leq 1 \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d \varphi(j) \|z^{(j)}\|_{(j)}^{\mathcal{R}}, \quad \forall x \in \mathbb{R}^d, \quad (34i)$$

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

$$(\varphi \circ \ell_0)^{\dot{\zeta}\dot{\zeta}'}(x) = \frac{1}{\|x\|} \min_{\substack{z^{(1)} \in \mathbb{R}^d, \dots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d \|z^{(j)}\|_{(j)}^{\mathcal{R}} \leq \|x\| \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d \|z^{(j)}\|_{(j)}^{\mathcal{R}} \varphi(j), \quad \forall x \in \mathbb{R}^d \setminus \{0\}. \quad (35)$$

Proof. We first note that $(\varphi \circ \ell_0)^{\dot{\zeta}\dot{\zeta}'} = ((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'} \circ n$, by (30d), and we study $((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'}$.

1. Let $\varphi : \llbracket 0, d \rrbracket \rightarrow \overline{\mathbb{R}}$ be a function. The equality (34a) is a straightforward consequence of the expression (30d) for a CAPRA-biconjugate, and of the fact that $n(x) = \frac{x}{\|x\|}$ when $x \neq 0$ by (29). We have

$$\begin{aligned} ((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'} &= \left(\sup_{j \in \llbracket 0, d \rrbracket} [\|\cdot\|_{(j), \star}^{\mathcal{R}} - \varphi(j)] \right)^{\star'} && \text{(by (33))} \\ &= \left(\sup_{j \in \llbracket 0, d \rrbracket} [\sigma_{\mathbb{B}_{(j)}^{\mathcal{R}}} - \varphi(j)] \right)^{\star'} \end{aligned}$$

by (11) as $\mathbb{B}_{(j)}^{\mathcal{R}}$ is the unit ball of the norm $\|\cdot\|_{(j)}^{\mathcal{R}}$ by (20) and with the convention $\mathbb{B}_{(0)}^{\mathcal{R}} = \{0\}$

$$\begin{aligned} &= \left(\sup_{j \in \llbracket 0, d \rrbracket} [\delta_{\mathbb{B}_{(j)}^{\mathcal{R}}}^{\star} - \varphi(j)] \right)^{\star'} && \text{(because } \delta_{\mathbb{B}_{(j)}^{\mathcal{R}}}^{\star} = \sigma_{\mathbb{B}_{(j)}^{\mathcal{R}}}) \\ &= \left(\sup_{j \in \llbracket 0, d \rrbracket} (\delta_{\mathbb{B}_{(j)}^{\mathcal{R}}} + \varphi(j))^{\star} \right)^{\star'} && \text{(by property of conjugacies)} \\ &= \left(\left(\inf_{j \in \llbracket 0, d \rrbracket} [\delta_{\mathbb{B}_{(j)}^{\mathcal{R}}} + \varphi(j)] \right)^{\star} \right)^{\star'} \end{aligned}$$

as conjugacies, being dualities, turn infima into suprema

$$= \left(\inf_{j \in \llbracket 0, d \rrbracket} [\delta_{\mathbb{B}_{(j)}^{\mathcal{R}}} + \varphi(j)] \right)^{\star\star'}. \quad \text{(by (46c))}$$

⁶In (34g), the sum starts from $j = 0$, whereas in (34h) and in (34i), the sum starts from $j = 1$

Thus, we have obtained (34c) and (34d). Now, if we follow again the above sequence of equalities, we see that, everywhere, we can replace the balls $\mathbb{B}_{(j)}^{\mathcal{R}}$ by the spheres $\mathbb{S}_{(j)}^{\mathcal{R}}$, since $\|\cdot\|_{(j),\star}^{\mathcal{R}} = \sigma_{\mathbb{S}_{(j)}^{\mathcal{R}}} = \delta_{\mathbb{S}_{(j)}^{\mathcal{R}}}^{\star}$. Thus, we obtain (34e) and (34f).

2. Let $\varphi : \llbracket 0, d \rrbracket \rightarrow \mathbb{R}$ be a function. Then the closed convex function $((\varphi \circ \ell_0)^{\dot{\varphi}})^{\star'}$ is proper. Indeed, on the one hand, it is easily seen that the function $(\varphi \circ \ell_0)^{\dot{\varphi}}$ takes finite values, from which we deduce that the function $((\varphi \circ \ell_0)^{\dot{\varphi}})^{\star'}$ never takes the value $-\infty$. On the other hand, by (34a) and by the inequality $(\varphi \circ \ell_0)^{\dot{\varphi}\dot{\varphi}'} \leq \varphi \circ \ell_0$ obtained from (45e), we deduce that the function $((\varphi \circ \ell_0)^{\dot{\varphi}})^{\star'}$ never takes the value $+\infty$ on the unit sphere. Therefore, the $((\varphi \circ \ell_0)^{\dot{\varphi}})^{\star'}$ is proper.

For the remaining expressions for $((\varphi \circ \ell_0)^{\dot{\varphi}})^{\star'}$, we use a formula [14, Corollary 2.8.11] for the Fenchel conjugate of the supremum of proper convex functions $f_j : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$, $j \in \llbracket 0, n \rrbracket$:

$$\bigcap_{j=0,1,\dots,n} \text{dom } f_j \neq \emptyset \implies \left(\sup_{j=0,1,\dots,n} f_j \right)^{\star} = \min_{(\lambda_0, \lambda_1, \dots, \lambda_n) \in \Delta_{n+1}} \left(\sum_{j=0}^n \lambda_j f_j \right)^{\star}, \quad (36)$$

where $\text{dom } f = \{x \in \mathbb{R}^d \mid f(x) < +\infty\}$ is the effective domain (see Appendix A), and where Δ_{n+1} is the simplex of \mathbb{R}^{n+1} . We obtain

$$\begin{aligned} ((\varphi \circ \ell_0)^{\dot{\varphi}})^{\star'} &= \left(\sup_{j \in \llbracket 0, d \rrbracket} \left[\|\cdot\|_{(j),\star}^{\mathcal{R}} - \varphi(j) \right] \right)^{\star'} && \text{(by (33))} \\ &= \left(\sup_{j \in \llbracket 0, d \rrbracket} \left[\sigma_{\mathbb{B}_{(j)}^{\mathcal{R}}} - \varphi(j) \right] \right)^{\star'} \end{aligned}$$

by (11) as $\mathbb{B}_{(j)}^{\mathcal{R}}$ is the unit ball of the norm $\|\cdot\|_{(j)}^{\mathcal{R}}$ by (20) and with $\mathbb{B}_{(0)}^{\mathcal{R}} = \{0\}$

$$= \min_{(\lambda_0, \lambda_1, \dots, \lambda_d) \in \Delta_{d+1}} \left(\sum_{j=0}^d \lambda_j \left[\sigma_{\mathbb{B}_{(j)}^{\mathcal{R}}} - \varphi(j) \right] \right)^{\star'} \quad \text{(by (36))}$$

by [14, Corollary 2.8.11], as the functions $f_j = \sigma_{\mathbb{B}_{(j)}^{\mathcal{R}}} - \varphi(j)$ are proper convex (they even take finite values), for $j \in \llbracket 0, d \rrbracket$

$$= \min_{(\lambda_0, \lambda_1, \dots, \lambda_d) \in \Delta_{d+1}} \left(\sigma_{\sum_{j=0}^d \lambda_j \mathbb{B}_{(j)}^{\mathcal{R}}} - \sum_{j=0}^d \lambda_j \varphi(j) \right)^{\star'}$$

as, for all $j \in \llbracket 1, d \rrbracket$, $\lambda_j \sigma_{\mathbb{B}_{(j)}^{\mathcal{R}}} = \sigma_{\lambda_j \mathbb{B}_{(j)}^{\mathcal{R}}}$ since $\lambda_j \geq 0$, and then using the well-known property that the support function of a Minkowski sum of subsets is the sum of the support functions of the individual subsets [15, p. 226]

$$\begin{aligned}
&= \min_{(\lambda_0, \lambda_1, \dots, \lambda_d) \in \Delta_{d+1}} \left(\sigma_{\sum_{j=1}^d \lambda_j \mathbb{B}_{(j)}^{\mathcal{R}}} - \sum_{j=0}^d \lambda_j \varphi(j) \right)^{\star'} \\
&\quad \text{(thanks to the convention } \mathbb{B}_{(0)}^{\mathcal{R}} = \{0\}) \\
&= \min_{(\lambda_0, \lambda_1, \dots, \lambda_d) \in \Delta_{d+1}} \left(\left(\sigma_{\sum_{j=1}^d \lambda_j \mathbb{B}_{(j)}^{\mathcal{R}}} \right)^{\star'} + \sum_{j=0}^d \lambda_j \varphi(j) \right) \\
&\quad \text{(by property of conjugacies)} \\
&= \min_{(\lambda_0, \lambda_1, \dots, \lambda_d) \in \Delta_{d+1}} \left(\delta_{\sum_{j=1}^d \lambda_j \mathbb{B}_{(j)}^{\mathcal{R}}} + \sum_{j=0}^d \lambda_j \varphi(j) \right) \\
&\quad \text{(because } \sum_{j=1}^d \lambda_j \mathbb{B}_{(j)}^{\mathcal{R}} \text{ is a closed convex set.)}
\end{aligned}$$

Therefore, we deduce that, for all $x \in \mathbb{R}^d$,

$$((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'}(x) = \min_{\substack{(\lambda_0, \lambda_1, \dots, \lambda_d) \in \Delta_{d+1} \\ x \in \sum_{j=1}^d \lambda_j \mathbb{B}_{(j)}^{\mathcal{R}}}} \sum_{j=0}^d \lambda_j \varphi(j), \quad \text{which is (34g).}$$

3. Let $\varphi : \llbracket 0, d \rrbracket \rightarrow \mathbb{R}_+$ be a function such that $\varphi(0) = 0$. Then the closed convex function $((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'}$ is proper, as seen above. We go on with

$$\begin{aligned}
((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'}(x) &= \min_{\substack{(\lambda_0, \lambda_1, \dots, \lambda_d) \in \Delta_{d+1} \\ x \in \sum_{j=1}^d \lambda_j \mathbb{B}_{(j)}^{\mathcal{R}}}} \sum_{j=1}^d \lambda_j \varphi(j) && \text{(because } \varphi(0) = 0) \\
&= \min_{\substack{z^{(1)} \in \mathbb{B}_{(1)}^{\mathcal{R}}, \dots, z^{(d)} \in \mathbb{B}_{(d)}^{\mathcal{R}} \\ \lambda_1 \geq 0, \dots, \lambda_d \geq 0 \\ \sum_{j=1}^d \lambda_j \leq 1 \\ \sum_{j=1}^d \lambda_j z^{(j)} = x}} \sum_{j=1}^d \lambda_j \varphi(j)
\end{aligned}$$

because $(\lambda_0, \lambda_1, \dots, \lambda_d) \in \Delta_{d+1}$ if and only if $\lambda_1 \geq 0, \dots, \lambda_d \geq 0$ and $\sum_{j=1}^d \lambda_j \leq 1$ and $\lambda_0 = 1 - \sum_{j=1}^d \lambda_j$

$$\begin{aligned}
&= \min_{\substack{s^{(1)} \in \mathbb{S}_{(1)}^{\mathcal{R}}, \dots, s^{(d)} \in \mathbb{S}_{(d)}^{\mathcal{R}} \\ \mu_1 \geq 0, \dots, \mu_d \geq 0 \\ \sum_{j=1}^d \mu_j \leq 1 \\ \sum_{j=1}^d \mu_j s^{(j)} = x}} \sum_{j=1}^d \mu_j \varphi(j)
\end{aligned}$$

because, on the one hand, the inequality \leq is obvious as the unit sphere $\mathbb{S}_{(j)}^{\mathcal{R}}$ in (15) is included in the unit ball $\mathbb{B}_{(j)}^{\mathcal{R}}$ for all $j \in \llbracket 1, d \rrbracket$; and, on the other hand, the inequality \geq comes from putting, for $j \in \llbracket 1, d \rrbracket$, $\mu_j = \lambda_j \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}}$ and observing that i) $\sum_{i=1}^d \mu_j = \sum_{i=1}^d \lambda_j \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}} \leq \sum_{i=1}^d \lambda_j \leq 1$ because $\llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}} \leq 1$ as $z^{(j)} \in \mathbb{B}_{(j)}^{\mathcal{R}}$ ii) for all $j \in \llbracket 1, d \rrbracket$, there exists $s^{(j)} \in \mathbb{S}_{(j)}^{\mathcal{R}}$ such that $\lambda_j z^{(j)} = \mu_j s^{(j)}$ (take any $s^{(j)}$ when $z^{(j)} = 0$ because $\mu_j = 0$, and take $s^{(j)} = \frac{z^{(j)}}{\llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}}}$ when $z^{(j)} \neq 0$) iii) $\sum_{j=1}^d \lambda_j \varphi(j) \geq \sum_{j=1}^d \lambda_j \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}} \varphi(j) = \sum_{j=1}^d \mu_j \varphi(j)$ because $1 \geq \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}}$ and $\varphi(j) \geq 0$

$$= \min_{\substack{z^{(1)} \in \mathbb{R}^d, \dots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}} \leq 1 \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d \varphi(j) \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}},$$

by putting $z^{(j)} = \mu_j s^{(j)}$, for all $j \in \llbracket 1, d \rrbracket$. Thus, we have obtained (34h).

Finally, from $(\varphi \circ \ell_0)^{\dot{\zeta}\dot{\zeta}'} = ((\varphi \circ \ell_0)^{\dot{\zeta}})^{\star'} \circ n$, by (30d), we get that

$$(\varphi \circ \ell_0)^{\dot{\zeta}\dot{\zeta}'}(x) = \frac{1}{\llbracket x \rrbracket} \min_{\substack{z^{(1)} \in \mathbb{R}^d, \dots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}} \leq \llbracket x \rrbracket \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d \varphi(j) \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}}, \quad \forall x \in \mathbb{R}^d \setminus \{0\},$$

where we have used that $n(x) = \frac{x}{\llbracket x \rrbracket}$ when $x \neq 0$ by (29). Therefore, we have proved (35).

This ends the proof. \square

Before finishing that part on CAPRA-biconjugates, we provide the following characterization of when the characteristic functions $\delta_{\ell_0^{\leq k}}$ are $\dot{\zeta}$ -convex.

Corollary 4.6 *Let $\llbracket \cdot \rrbracket$ be a norm on \mathbb{R}^d , with associated sequence $\left\{ \llbracket \cdot \rrbracket_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of coordinate- k norms in Definition 3.2 and associated CAPRA coupling $\dot{\zeta}$ in (28).*

The following statements are equivalent.

1. *The sequence $\left\{ \llbracket \cdot \rrbracket_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket}$ of coordinate- k norms is strictly decreasingly graded with respect to the ℓ_0 pseudonorm, as in Definition 2.1.*
2. *For all $k \in \llbracket 0, d \rrbracket$, the characteristic functions $\delta_{\ell_0^{\leq k}}$ are $\dot{\zeta}$ -convex, that is,*

$$\delta_{\ell_0^{\leq k}}^{\dot{\zeta}\dot{\zeta}'} = \delta_{\ell_0^{\leq k}}, \quad k \in \llbracket 0, d \rrbracket. \quad (37)$$

Proof. We start by providing an expression for $\delta_{\ell_0^{\leq k}}^{\dot{\mathcal{C}}\dot{\mathcal{C}}'}$. For any $k \in \llbracket 0, d \rrbracket$, we have

$$\begin{aligned} \delta_{\ell_0^{\leq k}}^{\dot{\mathcal{C}}\dot{\mathcal{C}}'} &= \left(\inf_{j \in \llbracket 0, d \rrbracket} [\delta_{\mathbb{B}_{(j)}^{\mathcal{R}}} \dot{+} \delta_{\llbracket 0, k \rrbracket}(j)] \right)^{\star\star'} \circ n && \text{(by (34c) with the functions } \varphi = \delta_{\llbracket 0, k \rrbracket} \text{)} \\ &= \left(\inf_{j=0,1,\dots,k} \delta_{\mathbb{B}_{(j)}^{\mathcal{R}}} \right)^{\star\star'} \circ n \\ &= (\delta_{\mathbb{B}_{(k)}^{\mathcal{R}}})^{\star\star'} \circ n \end{aligned}$$

by the inclusions $\mathbb{B}_{(1)}^{\mathcal{R}} \subset \dots \subset \mathbb{B}_{(k)}^{\mathcal{R}}$ in (24) and by the convention $\mathbb{B}_{(0)}^{\mathcal{R}} = \{0\}$

$$\begin{aligned} &= \delta_{\mathbb{B}_{(k)}^{\mathcal{R}}} \circ n && \text{(because the unit ball } \mathbb{B}_{(k)}^{\mathcal{R}} \text{ is closed and convex)} \\ &= \delta_{n^{-1}(\mathbb{B}_{(k)}^{\mathcal{R}})} \end{aligned}$$

where, by (29), $n^{-1}(\mathbb{B}_{(k)}^{\mathcal{R}}) = \{0\} \cup \{x \in \mathbb{R}^d \setminus \{0\} \mid \|\| \frac{x}{\|x\|} \|\|_{(k)}^{\mathcal{R}} \leq 1\}$, so that we go on with

$$\begin{aligned} &= \delta_{\{x \in \mathbb{R}^d \mid \|\| x \|\|_{(k)}^{\mathcal{R}} \leq \|x\|\}} \\ &= \delta_{\{x \in \mathbb{R}^d \mid \|\| x \|\|_{(k)}^{\mathcal{R}} = \|x\|\}} && \text{(using the equality and inequalities between norms in (23))} \end{aligned}$$

Therefore, we have

$$\begin{aligned} \forall k \in \llbracket 0, d \rrbracket, \quad \delta_{\ell_0^{\leq k}}^{\dot{\mathcal{C}}\dot{\mathcal{C}}'} &= \delta_{\ell_0^{\leq k}} \\ &\Leftrightarrow \forall k \in \llbracket 0, d \rrbracket, \quad \left(x \in \ell_0^{\leq k} \iff \|\| x \|\|_{(k)}^{\mathcal{R}} = \|x\|, \quad \forall x \in \mathbb{R}^d \right) \\ &\Leftrightarrow (7d) \text{ holds true for the sequence } \left\{ \|\| \cdot \|\|_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket} \\ &\quad \text{(because } x \in \ell_0^{\leq k} \iff \ell_0(x) \leq k \text{ by definition of the level sets in (3a))} \\ &\Leftrightarrow \left\{ \|\| \cdot \|\|_{(j)}^{\mathcal{R}} \right\}_{j \in \llbracket 1, d \rrbracket} \text{ is strictly decreasingly graded w.r.t. the } \ell_0 \text{ pseudonorm} \end{aligned}$$

because this sequence is nonincreasing by (18) (see Definition 2.1).

This ends the proof. □

Notice that, by Item 2 in Proposition 3.5, it suffices that the normed space $(\mathbb{R}^d, \|\| \cdot \|\|)$ be strictly convex to obtain that the characteristic functions $\delta_{\ell_0^{\leq k}}$ are $\dot{\mathcal{C}}$ -convex, for all $k \in \llbracket 0, d \rrbracket$. This is the case when the source norm is the ℓ_p -norm $\|\cdot\|_p$ for $p \in]1, \infty[$.

Determinining sufficient conditions under which the ℓ_0 pseudonorm is $\dot{\mathcal{C}}$ -convex requires additional notions. This question is treated in the companion paper [10].

4.4 Capra-subdifferentials related to the ℓ_0 pseudonorm

With the Fenchel conjugacy, we calculate that $\partial\delta_{\ell_0^{\leq k}}(x) = \{0\}$ for $x \in \ell_0^{\leq k}$ and $k \in \llbracket 1, d \rrbracket$ (when $x \notin \ell_0^{\leq k}$, $\partial\delta_{\ell_0^{\leq k}}(x) = \emptyset$). We also calculate that $\partial\ell_0(0) = \{0\}$ and $\partial\ell_0(x) = \emptyset$, for all $x \in \mathbb{R}^d \setminus \{0\}$ (indeed, this is a consequence of $\ell_0^{\star\star'}(x) = 0 \neq \ell_0(x)$ when $x \in \mathbb{R}^d \setminus \{0\}$). Hence, the Fenchel conjugacy is not suitable to handle the ℓ_0 pseudonorm.

By contrast, we will now show that functions of the ℓ_0 pseudonorm in (1) — including the ℓ_0 pseudonorm itself and the characteristic functions $\delta_{\ell_0^{\leq k}}$ of its level sets (3a) — display CAPRA-subdifferentials, as in (32a), that are related to the sequence of dual coordinate- k norms in Definition 3.2 as follows. 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 [15, p.136]

$$N_C(x) = \{y \in \mathbb{R}^d \mid \langle x' - x, y \rangle \leq 0, \forall x' \in C\}. \quad (38)$$

Proposition 4.7 *Let $\|\cdot\|$ be a norm on \mathbb{R}^d , with associated sequence $\left\{\|\cdot\|_{\star(j)}^{\mathcal{R}}\right\}_{j \in \llbracket 1, d \rrbracket}$ of dual coordinate- k norms, as in Definition 3.2, and associated CAPRA-coupling $\dot{\varsigma}$ in (28).*

Let $\varphi : \llbracket 0, d \rrbracket \rightarrow \overline{\mathbb{R}}$ be a function and $x \in \mathbb{R}^d$ be a vector.

- *The CAPRA-subdifferential, as in (32c), of the function $\varphi \circ \ell_0$ at $x = 0$ is given by*

$$\partial_{\dot{\varsigma}}(\varphi \circ \ell_0)(0) = \bigcap_{j \in \llbracket 1, d \rrbracket} [\varphi(j) \dot{+} (-\varphi(0))] \mathbb{B}_{(j), \star}^{\mathcal{R}}, \quad (39)$$

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

- *The CAPRA-subdifferential, as in (32d), 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 $\partial_{\dot{\varsigma}}(\varphi \circ \ell_0)(x) = \mathbb{R}^d$,*
 - *if $l = \ell_0(x) \geq 1$ and $\varphi(l) = +\infty$ and there exists $j \in \llbracket 0, d \rrbracket$ such that $\varphi(j) \neq +\infty$, then $\partial_{\dot{\varsigma}}(\varphi \circ \ell_0)(x) = \emptyset$,*
 - *if $l = \ell_0(x) \geq 1$ and $-\infty < \varphi(l) < +\infty$, then*

$$y \in \partial_{\dot{\varsigma}}(\varphi \circ \ell_0)(x) \iff \begin{cases} y \in N_{\mathbb{B}_{(l)}^{\mathcal{R}}}\left(\frac{x}{\|x\|_{(l)}^{\mathcal{R}}}\right) & \text{and} \\ l \in \arg \max_{j \in \llbracket 0, d \rrbracket} [\|y\|_{(j), \star}^{\mathcal{R}} - \varphi(j)] \end{cases}. \quad (40)$$

Proof. We have

$$y \in \partial_{\dot{\varsigma}}(\varphi \circ \ell_0)(x) \iff (\varphi \circ \ell_0)^{\dot{\varsigma}}(y) = \dot{\varsigma}(x, y) \dot{+} (-\varphi \circ \ell_0)(x)$$

by definition (32a) of the CAPRA-subdifferential

$$\begin{aligned}
&\Longleftrightarrow \sup_{j \in \llbracket 0, d \rrbracket} \left[\llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j) \right] = \dot{\varsigma}(x, y) \dot{+} (-(\varphi \circ \ell_0)(x)) \\
&\quad \text{(as } (\varphi \circ \ell_0)^{\dot{\varsigma}}(y) = \sup_{j \in \llbracket 0, d \rrbracket} \left[\llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j) \right] \text{ by (33))} \\
&\Longleftrightarrow \left(x = 0 \text{ and } \sup_{j \in \llbracket 0, d \rrbracket} \left[\llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j) \right] = -\varphi(0) \right) \\
&\text{or } \left(x \neq 0 \text{ and } \sup_{j \in \llbracket 0, d \rrbracket} \left[\llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j) \right] = \frac{\langle x, y \rangle}{\llbracket x \rrbracket} - \varphi(\ell_0(x)) \right) \\
&\quad \text{(by definition (28) of } \dot{\varsigma}(x, y))
\end{aligned}$$

Therefore, on the one hand, we obtain that

$$\begin{aligned}
y \in \partial_{\dot{\varsigma}}(\varphi \circ \ell_0)(0) &\Longleftrightarrow \llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j) \leq -\varphi(0), \quad \forall j \in \llbracket 1, d \rrbracket \quad \text{(as } \llbracket y \rrbracket_{(0), \star}^{\mathcal{R}} = 0 \text{ by convention)} \\
&\Longleftrightarrow \llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} \leq \varphi(j) \dot{+} (-\varphi(0)), \quad \forall j \in \llbracket 1, d \rrbracket
\end{aligned}$$

by property of the Moreau upper addition [12]

$$\Longleftrightarrow y \in \bigcap_{j \in \llbracket 1, d \rrbracket} [\varphi(j) \dot{+} (-\varphi(0))] \mathbb{B}_{(j), \star}^{\mathcal{R}},$$

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

On the other hand, when $x \neq 0$, we get

$$y \in \partial_{\dot{\varsigma}}(\varphi \circ \ell_0)(x) \Longleftrightarrow \sup_{j \in \llbracket 0, d \rrbracket} \left[\llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j) \right] = \frac{\langle x, y \rangle}{\llbracket x \rrbracket} - \varphi(\ell_0(x)). \quad (41a)$$

We now establish necessary and sufficient conditions for y to belong to $\partial_{\dot{\varsigma}}(\varphi \circ \ell_0)(x)$ when $x \neq 0$.

We consider $x \in \mathbb{R}^d \setminus \{0\}$, and we denote $L = \text{supp}(x)$ and $l = |L| = \ell_0(x)$. We have

$$y \in \partial_{\dot{\varsigma}}(\varphi \circ \ell_0)(x)$$

$$\Longleftrightarrow \sup_{j \in \llbracket 0, d \rrbracket} \left[\llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j) \right] = \frac{\langle x, y \rangle}{\llbracket x \rrbracket} - \varphi(l) \quad \text{(by (41a) with } \ell_0(x) = l)$$

$$\Longleftrightarrow \llbracket y \rrbracket_{(l), \star}^{\mathcal{R}} - \varphi(l) \leq \sup_{j \in \llbracket 0, d \rrbracket} \left[\llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j) \right] = \frac{\langle x, y \rangle}{\llbracket x \rrbracket} - \varphi(l)$$

$$\Longleftrightarrow \llbracket y_L \rrbracket_{L, \star} - \varphi(l) \leq \llbracket y \rrbracket_{(l), \star}^{\mathcal{R}} - \varphi(l) \leq \sup_{j \in \llbracket 0, d \rrbracket} \left[\llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j) \right] = \frac{\langle x, y \rangle}{\llbracket x \rrbracket} - \varphi(l)$$

as $\llbracket y_L \rrbracket_{L, \star} \leq \llbracket y \rrbracket_{(l), \star}^{\mathcal{R}}$ by expression (16) of the dual coordinate- k norm $\llbracket y \rrbracket_{(l), \star}^{\mathcal{R}}$, and because $l = |L|$

$$\Longleftrightarrow \llbracket y_L \rrbracket_{L, \star} - \varphi(l) \leq \llbracket y \rrbracket_{(l), \star}^{\mathcal{R}} - \varphi(l) \leq \sup_{j \in \llbracket 0, d \rrbracket} \left[\llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j) \right] = \frac{\langle x, y \rangle}{\llbracket x \rrbracket} - \varphi(l) \leq \llbracket y_L \rrbracket_{L, \star} - \varphi(l)$$

$$\text{(as we have } \frac{\langle x, y \rangle}{\llbracket x \rrbracket} = \frac{\langle x_L, y_L \rangle}{\llbracket x_L \rrbracket} \leq \llbracket y_L \rrbracket_{L, \star} \text{ since } x = x_L \text{ and by (9))}$$

$$\Longleftrightarrow \llbracket y_L \rrbracket_{L, \star} - \varphi(l) = \llbracket y \rrbracket_{(l), \star}^{\mathcal{R}} - \varphi(l) = \sup_{j \in \llbracket 0, d \rrbracket} \left[\llbracket y \rrbracket_{(j), \star}^{\mathcal{R}} - \varphi(j) \right] = \frac{\langle x, y \rangle}{\llbracket x \rrbracket} - \varphi(l)$$

as all terms in the inequalities are necessarily equal

$$\iff \begin{cases} \text{either } \varphi(l) = -\infty \\ \text{or } (\varphi(l) = +\infty \text{ and } \varphi(j) = +\infty, \forall j \in \llbracket 0, d \rrbracket) \\ \text{or } \left(-\infty < \varphi(l) < +\infty \text{ and } \right. \\ \quad \left. \lVert y_L \rVert_{L, \star} = \lVert y \rVert_{(l), \star}^{\mathcal{R}} = \frac{\langle x, y \rangle}{\lVert x \rVert} \text{ and } \lVert y \rVert_{(l), \star}^{\mathcal{R}} - \varphi(l) = \sup_{j \in \llbracket 0, d \rrbracket} [\lVert y \rVert_{(j), \star}^{\mathcal{R}} - \varphi(j)] \right) . \end{cases}$$

Let us make a brief insert and notice that

$$\begin{aligned} x = x_L, \ell_0(x) = l = |L| > 1, \langle x, y \rangle &= \lVert x \rVert \times \lVert y \rVert_{(l), \star}^{\mathcal{R}} \\ \implies \ell_0(x) = l = |L| > 1, \langle x_L, y_L \rangle &= \lVert x_L \rVert \times \lVert y \rVert_{(l), \star}^{\mathcal{R}} \\ \implies \ell_0(x) = l = |L| > 1, \lVert x_L \rVert \times \lVert y \rVert_{(l), \star}^{\mathcal{R}} &\leq \lVert x_L \rVert \times \lVert y_L \rVert_{L, \star} \quad (\text{by (9)}) \\ \implies l = |L|, \lVert y \rVert_{(l), \star}^{\mathcal{R}} &\leq \lVert y_L \rVert_{L, \star} \\ \implies \lVert y \rVert_{(l), \star}^{\mathcal{R}} &= \lVert y_L \rVert_{L, \star} \end{aligned}$$

as $\lVert y_L \rVert_{L, \star} \leq \lVert y \rVert_{(l), \star}^{\mathcal{R}}$ by expression (16) of the dual coordinate- k norm $\lVert y \rVert_{(l), \star}^{\mathcal{R}}$, and because $l = |L|$.

Now, let us go back to the equivalences regarding $y \in \partial_{\dot{C}}(\varphi \circ \ell_0)(x)$. Focusing on the case where $-\infty < \varphi(l) < +\infty$, we have

$$\begin{aligned} y \in \partial_{\dot{C}}(\varphi \circ \ell_0)(x) &\Leftrightarrow \lVert y_L \rVert_{L, \star} = \lVert y \rVert_{(l), \star}^{\mathcal{R}} = \frac{\langle x, y \rangle}{\lVert x \rVert} \text{ and } l \in \arg \max_{j \in \llbracket 0, d \rrbracket} [\lVert y \rVert_{(j), \star}^{\mathcal{R}} - \varphi(j)] \\ &\Leftrightarrow \lVert y_L \rVert_{L, \star} = \lVert y \rVert_{(l), \star}^{\mathcal{R}} \text{ and } \langle x, y \rangle = \lVert x \rVert \times \lVert y \rVert_{(l), \star}^{\mathcal{R}} \text{ and } l \in \arg \max_{j \in \llbracket 0, d \rrbracket} [\lVert y \rVert_{(j), \star}^{\mathcal{R}} - \varphi(j)] \\ &\Leftrightarrow \langle x, y \rangle = \lVert x \rVert \times \lVert y \rVert_{(l), \star}^{\mathcal{R}} \text{ and } l \in \arg \max_{j \in \llbracket 0, d \rrbracket} [\lVert y \rVert_{(j), \star}^{\mathcal{R}} - \varphi(j)] \end{aligned}$$

as just established in the insert

$$\begin{aligned} &\Leftrightarrow \langle x, y \rangle = \lVert x \rVert_{(l)}^{\mathcal{R}} \times \lVert y \rVert_{(l), \star}^{\mathcal{R}} \text{ and } l \in \arg \max_{j \in \llbracket 0, d \rrbracket} [\lVert y \rVert_{(j), \star}^{\mathcal{R}} - \varphi(j)] \\ &\quad \quad \quad (\text{as } \ell_0(x) = l \implies \lVert x \rVert = \lVert x \rVert_{(l)}^{\mathcal{R}} \text{ by (25a)}) \\ &\Leftrightarrow y \in N_{\mathbb{B}_{(l)}^{\mathcal{R}}} \left(\frac{x}{\lVert x \rVert_{(l)}^{\mathcal{R}}} \right) \text{ and } l \in \arg \max_{j \in \llbracket 0, d \rrbracket} [\lVert y \rVert_{(j), \star}^{\mathcal{R}} - \varphi(j)] \end{aligned}$$

by the equivalence $\langle x, y \rangle = \lVert x \rVert_{(l)}^{\mathcal{R}} \times \lVert y \rVert_{(l), \star}^{\mathcal{R}} \iff y \in N_{\mathbb{B}_{(l)}^{\mathcal{R}}} \left(\frac{x}{\lVert x \rVert_{(l)}^{\mathcal{R}}} \right)$.

This ends the proof. □

With φ the identity function on $\llbracket 0, d \rrbracket$, we find the CAPRA-subdifferential of the ℓ_0 pseudonorm. With the functions $\varphi = \delta_{\llbracket 0, k \rrbracket}$ (for any $k \in \llbracket 0, d \rrbracket$), we find the CAPRA-subdifferentials of the characteristic functions $\delta_{\ell_0 \leq k}$ of its level sets (3a). The corresponding expressions are given in Table 3.

5 Norm ratio lower bounds for the l_0 pseudonorm

As an application, we provide a new family of lower bounds for the l_0 pseudonorm, as a fraction between two norms, the denominator being any norm.

Proposition 5.1 *Let $\|\cdot\|$ be a norm on \mathbb{R}^d , with associated sequence of dual coordinate- k norms, as in Definition 3.2. For any function $\varphi : \llbracket 0, d \rrbracket \rightarrow [0, +\infty[$, such that $\varphi(j) > \varphi(0) = 0$ for all $j \in \llbracket 1, d \rrbracket$, there exists a norm $\|\cdot\|_{(\varphi)}^{\mathcal{R}}$ characterized*

- either by its dual norm $\|\cdot\|_{(\varphi),\star}^{\mathcal{R}}$, which has unit ball $\bigcap_{j \in \llbracket 1, d \rrbracket} \varphi(j) \mathbb{B}_{(j),\star}^{\mathcal{R}}$, that is,

$$\mathbb{B}_{(\varphi),\star}^{\mathcal{R}} = \bigcap_{j \in \llbracket 1, d \rrbracket} \varphi(j) \mathbb{B}_{(j),\star}^{\mathcal{R}} \quad \text{and} \quad \|\cdot\|_{(\varphi)}^{\mathcal{R}} = \sigma_{\mathbb{B}_{(\varphi),\star}^{\mathcal{R}}} , \quad (42a)$$

$$\text{or, equivalently,} \quad \|y\|_{(\varphi),\star}^{\mathcal{R}} = \sup_{j \in \llbracket 1, d \rrbracket} \frac{\|y\|_{(j),\star}^{\mathcal{R}}}{\varphi(j)} , \quad \forall y \in \mathbb{R}^d , \quad (42b)$$

- or by the inf-convolution

$$\|\cdot\|_{(\varphi)}^{\mathcal{R}} = \bigcap_{j \in \llbracket 1, d \rrbracket} \left(\varphi(j) \|\cdot\|_{(j)}^{\mathcal{R}} \right) , \quad (42c)$$

$$\text{that is,} \quad \|x\|_{(\varphi)}^{\mathcal{R}} = \inf_{\substack{z^{(1)} \in \mathbb{R}^d, \dots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d \varphi(j) \|z^{(j)}\|_{(j)}^{\mathcal{R}} , \quad \forall x \in \mathbb{R}^d . \quad (42d)$$

Then, we have the inequalities

$$\frac{\|x\|_{(\varphi)}^{\mathcal{R}}}{\|x\|} \leq \frac{1}{\|x\|} \min_{\substack{z^{(1)} \in \mathbb{R}^d, \dots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d \|z^{(j)}\|_{(j)}^{\mathcal{R}} \leq \|x\| \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d \varphi(j) \|z^{(j)}\|_{(j)}^{\mathcal{R}} \leq \varphi(\ell_0(x)) , \quad \forall x \in \mathbb{R}^d \setminus \{0\} . \quad (43)$$

Proof.

- It is easily seen that $\sigma_{\mathbb{B}_{(\varphi),\star}^{\mathcal{R}}}$ in (42a) defines a norm, and that, for all $y \in \mathbb{R}^d$,

$$\|y\|_{(\varphi),\star}^{\mathcal{R}} = \inf \left\{ \lambda \geq 0 \mid y \in \lambda \bigcap_{j=1}^d \varphi(j) \mathbb{B}_{(j),\star}^{\mathcal{R}} \right\} = \inf \left\{ \lambda \geq 0 \mid \frac{\|y\|_{(j),\star}^{\mathcal{R}}}{\varphi(j)} \leq \lambda \right\} = \sup_{j \in \llbracket 1, d \rrbracket} \frac{\|y\|_{(j),\star}^{\mathcal{R}}}{\varphi(j)} .$$

- We have

$$\begin{aligned} \|\cdot\|_{(\varphi)}^{\mathcal{R}} &= \sigma_{\mathbb{B}_{(\varphi),\star}^{\mathcal{R}}} && \text{(by (42a))} \\ &= \delta_{\mathbb{B}_{(\varphi),\star}^{\mathcal{R}}}^{\star} && \text{(because } \mathbb{B}_{(\varphi),\star}^{\mathcal{R}} \text{ is closed and convex)} \\ &= \left(\sum_{j \in \llbracket 1, d \rrbracket} \delta_{\varphi(j) \mathbb{B}_{(j),\star}^{\mathcal{R}}} \right)^{\star} \end{aligned}$$

by (42a) and by expressing the characteristic function of an intersection of sets as a sum

$$= \bigcap_{j \in \llbracket 1, d \rrbracket} \delta_{\varphi(j)\mathbb{B}_{(j),*}^{\mathcal{R}}}^{\star}$$

using [11, Proposition 15.3 and (v) in Proposition-15.5] because the intersection $\mathbb{B}_{(\varphi),*}^{\mathcal{R}} = \bigcap_{j=1}^d \varphi(j)\mathbb{B}_{(j),*}^{\mathcal{R}}$ of all the domains of the functions $\delta_{\varphi(j)\mathbb{B}_{(j),*}^{\mathcal{R}}}$ contain a neighborhood of 0 since $\varphi(j) > 0$ for all $j \in \llbracket 1, d \rrbracket$

$$\begin{aligned} &= \bigcap_{j \in \llbracket 1, d \rrbracket} \sigma_{\varphi(j)\mathbb{B}_{(j),*}^{\mathcal{R}}} \quad (\text{as } \delta_{\varphi(j)\mathbb{B}_{(j),*}^{\mathcal{R}}}^{\star} = \sigma_{\varphi(j)\mathbb{B}_{(j),*}^{\mathcal{R}}}, \text{ for all } j \in \llbracket 1, d \rrbracket) \\ &= \bigcap_{j \in \llbracket 1, d \rrbracket} \varphi(j) \llbracket \cdot \rrbracket_{(j)}^{\mathcal{R}} \quad (\text{by (11)}) \end{aligned}$$

- We consider the coupling ζ in (28).

By (35) — because the function $\varphi : \llbracket 0, d \rrbracket \rightarrow [0, +\infty[$ satisfies the assumption in Item 3 of Proposition 4.5 — and by the inequality $(\varphi \circ \ell_0)^{\zeta \zeta'} \leq \varphi \circ \ell_0$ obtained from (45e), we get that

$$\frac{1}{\llbracket x \rrbracket} \min_{\substack{z^{(1)} \in \mathbb{R}^d, \dots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}} \leq \llbracket x \rrbracket \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d j \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}} \leq \varphi(\ell_0(x)), \quad \forall x \in \mathbb{R}^d \setminus \{0\}. \quad (44)$$

Thus, we have obtained the right hand side inequality in (43).

By relaxing one constraint in (44), we immediately get that

$$\inf_{\substack{z^{(1)} \in \mathbb{R}^d, \dots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d \varphi(j) \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}} \leq \min_{\substack{z^{(1)} \in \mathbb{R}^d, \dots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}} \leq \llbracket x \rrbracket \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d \varphi(j) \llbracket z^{(j)} \rrbracket_{(j)}^{\mathcal{R}} \leq \varphi(\ell_0(x)), \quad \forall x \in \mathbb{R}^d.$$

Thus, we have obtained the left hand side inequality in (43), thanks to (42d). \square

For any function $\varphi : \llbracket 0, d \rrbracket \rightarrow [0, +\infty[$, such that $\varphi(j) > \varphi(0) = 0$ for all $j \in \llbracket 1, d \rrbracket$, using Table 1 when the source norm $\llbracket \cdot \rrbracket$ is the ℓ_p -norm $\|\cdot\|_p$, for $p \in [1, \infty]$ and $1/p + 1/q = 1$, we denote $\llbracket \cdot \rrbracket_{(\varphi)}^{\mathcal{R}}$ by $\|\cdot\|_{p,\varphi}^{\text{sn}}$. The calculations show that $\|\cdot\|_{1,\varphi}^{\text{sn}} = \|\cdot\|_1$, and that, when $p \in]1, \infty]$, we also have $\|\cdot\|_{p,\varphi}^{\text{sn}} = \|\cdot\|_1$, whatever $p \in [1, \infty]$, if we suppose that $(\varphi(j))^q \geq j$, for all $j \in \llbracket 1, d \rrbracket$. As a consequence, when $p = 1$, the inequality (43) is trivial. When $p \in]1, \infty]$, if we take the function $\varphi(j) = j^{1/q}$ for all $j \in \llbracket 1, d \rrbracket$, the inequality (43) yields that $\frac{\|x\|_1}{\|x\|_p} \leq (\ell_0(x))^{1/q}$, which is easily obtained directly from the Hölder inequality.

Fenchel conjugacy	CAPRA conjugacy
$\delta_{\ell_0^{\leq k}}^{(-\star)} = \delta_{\ell_0^{\leq k}}^{\star} = \delta_{\{0\}}$	$\delta_{\ell_0^{\leq k}}^{-\dot{\zeta}} = \delta_{\ell_0^{\leq k}}^{\dot{\zeta}} = \ \cdot\ _{(k),\star}^{\mathcal{R}}$
$\delta_{\ell_0^{\leq k}}^{\star\star'} = 0$	$\delta_{\ell_0^{\leq k}}^{\dot{\zeta}\dot{\zeta}'} = \delta_{\{x \in \mathbb{R}^d \mid \ x\ _{(k)}^{\mathcal{R}} = \ x\ \}}$
$\partial \delta_{\ell_0^{\leq k}}(x) = \emptyset$ $\forall x \in \mathbb{R}^d$	$\partial_{\dot{\zeta}} \delta_{\ell_0^{\leq k}}(x) = \begin{cases} \emptyset & \text{if } \ell_0(x) = k+1, \dots, d, \\ N_{\mathbb{B}_{(k)}^{\mathcal{R}}}(\frac{x}{\ x\ _{(k)}^{\mathcal{R}}}) & \text{if } \ell_0(x) = 1, \dots, k, \\ \{0\} & \text{if } \ell_0(x) = 0 \end{cases}$ $\forall x \in \mathbb{R}^d$
$\ell_0^{\star} = \delta_{\{0\}}$	$\ell_0^{\dot{\zeta}} = \sup_{j \in \llbracket 0, d \rrbracket} [\ \cdot\ _{(j),\star}^{\mathcal{R}} - j]$
$\ell_0^{\star\star'} = 0$	$\ell_0^{\dot{\zeta}\dot{\zeta}'}(x) = \frac{1}{\ x\ } \min_{\substack{z^{(1)} \in \mathbb{R}^d, \dots, z^{(d)} \in \mathbb{R}^d \\ \sum_{j=1}^d \ z^{(j)}\ _{(j)}^{\mathcal{R}} \leq \ x\ \\ \sum_{j=1}^d z^{(j)} = x}} \sum_{j=1}^d j \ z^{(j)}\ _{(j)}^{\mathcal{R}}, \quad \forall x \in \mathbb{R}^d \setminus \{0\}$ $\ell_0^{\dot{\zeta}\dot{\zeta}'}(0) = 0$
$\partial \ell_0(0) = \{0\}$	$\partial_{\dot{\zeta}} \ell_0(0) = \bigcap_{j \in \llbracket 1, d \rrbracket} j \mathbb{B}_{(j),\star}^{\mathcal{R}} = \mathbb{B}_{(\text{Id}),\star}^{\mathcal{R}}$
$\partial \ell_0(x) = \emptyset$ $\forall x \in \mathbb{R}^d \setminus \{0\}$	$y \in \partial_{\dot{\zeta}} \ell_0(x) \iff \begin{cases} y \in N_{\mathbb{B}_{(l)}^{\mathcal{R}}}(\frac{x}{\ x\ _{(l)}^{\mathcal{R}}}) \\ \text{and } l \in \arg \max_{j \in \llbracket 0, d \rrbracket} [\ y\ _{(j),\star}^{\mathcal{R}} - j] \end{cases}$ $\forall x \in \mathbb{R}^d \setminus \{0\}, \text{ where } l = \ell_0(x) \geq 1$

Table 3: Comparison of Fenchel and CAPRA-conjugates, biconjugates and subdifferentials of the ℓ_0 pseudonorm in (1), and of the characteristic functions $\delta_{\ell_0^{\leq k}}$ of its level sets (3a), for $k \in \llbracket 0, d \rrbracket$

6 Conclusion

In this paper, we have presented a new family of conjugacies, which depend on a given general source norm, and we have shown that they are suitable for the ℓ_0 pseudonorm. More precisely, given a (source) norm on \mathbb{R}^d , we have defined, on the one hand, a sequence of so-called coordinate- k norms and, on the other hand, a coupling between \mathbb{R}^d and itself, called Capra (constant along primal rays). With this, we have provided formulas for the CAPRA-conjugate and biconjugate, and for the CAPRA subdifferentials, of functions of the ℓ_0 pseudonorm, in terms of the coordinate- k norms. Table 3 provides the results of Proposition 4.4, Proposition 4.5, and Proposition 4.7, in the case of the ℓ_0 pseudonorm and of the characteristic functions $\delta_{\ell_0^{\leq k}}$ of its level sets (3a). It compares them with the Fenchel conjugates and biconjugates. As an application, we have provided a new family of lower bounds for the ℓ_0 pseudonorm, as a fraction between two norms, the denominator being any norm.

In the companion paper [10], we provide sufficient conditions under which the ℓ_0 pseudonorm is ψ -convex. We are currently investigating how the Capra conjugacies could provide algorithms for exact sparse optimization

Acknowledgements. We want to thank Guillaume Obozinski for discussions on first versions of this work, as well as the anonymous Referee and Associate Editor whose comments helped us improve the manuscript.

A Background on Fenchel-Moreau conjugacies

We review general concepts and notations on Fenchel-Moreau conjugacies, then focus on the special case of the Fenchel conjugacy.

The general case. Let \mathbb{X} (“primal”), \mathbb{Y} (“dual”) be two sets and $c : \mathbb{X} \times \mathbb{Y} \rightarrow \overline{\mathbb{R}}$ be a so-called *coupling* function. 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.

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

$$f^c(y) = \sup_{x \in \mathbb{X}} \left(c(x, y) + (-f(x)) \right), \quad \forall y \in \mathbb{Y}. \quad (45a)$$

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

$$c' : \mathbb{Y} \times \mathbb{X} \rightarrow \overline{\mathbb{R}}, \quad c'(y, x) = c(x, y), \quad \forall (y, x) \in \mathbb{Y} \times \mathbb{X}. \quad (45b)$$

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

$$g^c(x) = \sup_{y \in \mathbb{Y}} \left(c(x, y) + (-g(y)) \right), \quad \forall x \in \mathbb{X}. \quad (45c)$$

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

$$f^{cc'}(x) = (f^c)'(x) = \sup_{y \in \mathbb{Y}} \left(c(x, y) \dot{+} (-f^c(y)) \right), \quad \forall x \in \mathbb{X}. \quad (45d)$$

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

$$f^{cc'}(x) \leq f(x), \quad \forall x \in \mathbb{X}. \quad (45e)$$

The Fenchel conjugacy. When the sets \mathbb{X} and \mathbb{Y} are two vector spaces that are *paired* with a bilinear form $\langle \cdot, \cdot \rangle$, in the sense of convex analysis [3, p. 13], the corresponding conjugacy is the classical *Fenchel conjugacy*. For any functions $f : \mathbb{X} \rightarrow \overline{\mathbb{R}}$ and $g : \mathbb{Y} \rightarrow \overline{\mathbb{R}}$, we denote⁷

$$f^*(y) = \sup_{x \in \mathbb{X}} \left(\langle x, y \rangle \dot{+} (-f(x)) \right), \quad \forall y \in \mathbb{Y}, \quad (46a)$$

$$g^*(x) = \sup_{y \in \mathbb{Y}} \left(\langle x, y \rangle \dot{+} (-g(y)) \right), \quad \forall x \in \mathbb{X}, \quad (46b)$$

$$f^{**}(x) = \sup_{y \in \mathbb{Y}} \left(\langle x, y \rangle \dot{+} (-f^*(y)) \right), \quad \forall x \in \mathbb{X}. \quad (46c)$$

For any function $h : \mathbb{W} \rightarrow \overline{\mathbb{R}}$, its *epigraph* is $\text{epih} = \{(w, t) \in \mathbb{W} \times \mathbb{R} \mid h(w) \leq t\}$, its *effective domain* is $\text{dom}h = \{w \in \mathbb{W} \mid h(w) < +\infty\}$. A function $h : \mathbb{W} \rightarrow \overline{\mathbb{R}}$ is said to be *proper* if it never takes the value $-\infty$ and that $\text{dom}h \neq \emptyset$. When \mathbb{W} is equipped with a topology, the function $h : \mathbb{W} \rightarrow \overline{\mathbb{R}}$ 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 $-\infty$ [3, p. 15].

It is proved that 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} [3, Theorem 5]. Here, a function is said to be *convex* if its epigraph is convex. The set of closed convex functions is the set of proper convex functions united with the two constant functions $-\infty$ and $+\infty$.

References

- [1] Nikolova M. Relationship between the optimal solutions of least squares regularized with l0-norm and constrained by k-sparsity. *Applied and Computational Harmonic Analysis*. 2016;41(1):237–265.
- [2] Hiriart-Urruty JB, Le H. A variational approach of the rank function. *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research*. 2013;21(2):207–240.
- [3] Rockafellar RT. *Conjugate duality and optimization*. Society for Industrial and Applied Mathematics; 1974. CBMS-NSF Regional Conference Series in Applied Mathematics.

⁷In 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.

- [4] Singer I. Abstract convex analysis. John Wiley & Sons, Inc., New York; 1997. Canadian Mathematical Society Series of Monographs and Advanced Texts.
- [5] Rubinov A. Abstract convexity and global optimization. (Nonconvex Optimization and its Applications; Vol. 44). Kluwer Academic Publishers, Dordrecht; 2000.
- [6] Martínez-Legaz JE. Generalized convex duality and its economic applications. In: Hadjisavvas N SS Komlósi S, editor. Handbook of generalized convexity and generalized monotonicity. nonconvex optimization and its applications. Vol. 76. sv; 2005. p. 237–292.
- [7] Rubinov AM, Glover BM. Increasing convex-along-rays functions with applications to global optimization. Journal of Optimization Theory and Applications. 1999;102(3):615–642.
- [8] Chancelier JP, De Lara M. Hidden convexity in the l_0 pseudonorm. Journal of Convex Analysis. 2021; 28(1):203–236.
- [9] Chancelier JP, De Lara M. Orthant-Strictly Monotonic Norms, Graded Sequences and Generalized Top- k and k -Support Norms for Sparse Optimization ; 2020. Preprint hal-02456433, arXiv:2001.10441.
- [10] Chancelier JP, De Lara M. Variational formulations for the l_0 pseudonorm and applications to sparse optimization; 2020. Preprint hal-02459688, arXiv:2002.01314.
- [11] Bauschke HH, Combettes PL. Convex analysis and monotone operator theory in Hilbert spaces. 2nd ed. Springer-Verlag; 2017. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC.
- [12] Moreau JJ. Inf-convolution, sous-additivité, convexité des fonctions numériques. J Math Pures Appl (9). 1970;49:109–154.
- [13] Akian M, Gaubert S, Kolokoltsov V. Invertibility of functional Galois connections. Comptes Rendus Mathématique. 2002;335(11):883–888.
- [14] Zălinescu C. Convex analysis in general vector spaces. World Scientific; 2002.
- [15] Hiriart-Urruty JB, Lemaréchal C. Convex analysis and minimization algorithms I. Berlin: sv; 1993.