



HAL
open science

Hidden Convexity in the l_0 Pseudonorm

Jean-Philippe Chancelier, Michel de Lara

► **To cite this version:**

Jean-Philippe Chancelier, Michel de Lara. Hidden Convexity in the l_0 Pseudonorm. 2019. hal-02146454v1

HAL Id: hal-02146454

<https://hal.science/hal-02146454v1>

Preprint submitted on 4 Jun 2019 (v1), last revised 11 Jun 2021 (v3)

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.

Hidden Convexity in the ℓ_0 Pseudonorm

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

June 4, 2019

Abstract

The so-called ℓ_0 pseudonorm on \mathbb{R}^d counts the number of nonzero components of a vector. It is well-known that the ℓ_0 pseudonorm is not convex, as its Fenchel biconjugate is zero. In this paper, we introduce a suitable conjugacy, induced by a novel coupling, Caprac, having the property of being constant along primal rays, like the ℓ_0 pseudonorm. The coupling Caprac belongs to the class of one-sided linear couplings, that we introduce. We show that they induce conjugacies that share nice properties with the classic Fenchel conjugacy. For the Caprac conjugacy, induced by the coupling Caprac, we prove that the ℓ_0 pseudonorm is equal to its biconjugate: hence, the ℓ_0 pseudonorm is Caprac-convex in the sense of generalized convexity. We also provide expressions for conjugates in terms of two families of dual norms, the $2-k$ -symmetric gauge norms and the k -support norms. As a corollary, we show that the ℓ_0 pseudonorm coincides, on the sphere, with a proper convex lower semicontinuous function — that we characterize, and for which we give explicit formulas in the two dimensional case. This is somewhat surprising as the ℓ_0 pseudonorm is a highly nonconvex function of combinatorial nature.

Key words: ℓ_0 pseudonorm, coupling, Fenchel-Moreau conjugacy, $2-k$ -symmetric gauge norms, k -support norms.

1 Introduction

The *counting function*, also called *cardinality function* or *ℓ_0 pseudonorm*, counts the number of nonzero components of a vector in \mathbb{R}^d . It is related to the rank function defined over matrices [6]. It is well-known that the ℓ_0 pseudonorm is lower semi continuous but is not convex (this can be deduced from the computation of its Fenchel biconjugate, which is zero).

In this paper, we display a suitable conjugacy for which we show that the ℓ_0 pseudonorm is “convex” in the sense of generalized convexity (equal to its biconjugate). As a corollary,

*delara@cermics.enpc.fr

we also show that the ℓ_0 pseudonorm coincides, on the sphere, with a proper convex lower semicontinuous (lsc) function.

The paper is organized as follows. In Sect. 2, we provide background on Fenchel-Moreau conjugacies, then we introduce a novel class of *one-sided linear couplings*, which includes the *constant along primal rays coupling* \dot{c} (*Caprac*). We show that one-sided linear couplings induce conjugacies that share nice properties with the classic Fenchel conjugacy, by giving expressions for conjugate and biconjugate functions. We elucidate the structure of Caprac-convex functions. Then, in Sect. 3, we relate the Caprac conjugate and biconjugate of the ℓ_0 pseudonorm, the characteristic functions of its level sets and the symmetric gauge norms. In particular, we show that the ℓ_0 pseudonorm is Caprac biconjugate (a Caprac-convex function), from which we deduce that it coincides, on the sphere, with a proper convex lsc function defined on the whole space. We establish this result in Sect. 4, where we also provide various expression for the underlying proper convex lsc function. The Appendix A gathers background on J. J. Moreau lower and upper additions, properties of 2- k -symmetric gauge norms and of k -support norms, properties of the ℓ_0 pseudonorm level sets, and technical results on the convex extension of the ℓ_0 pseudonorm.

2 The constant along primal rays coupling (Caprac)

After having recalled background on Fenchel-Moreau conjugacies in §2.1, we introduce *one-sided linear couplings* in §2.2, and finally the *constant along primal rays coupling* \dot{c} (*Caprac*) in §2.3.

2.1 Background on Fenchel-Moreau conjugacies

We review general concepts and notations, then we focus on the special case of the Fenchel conjugacy. We denote $\overline{\mathbb{R}} = [-\infty, +\infty]$. Background on J. J. Moreau lower and upper additions can be found in §A.1.

The general case

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

$$c : \mathbb{X} \times \mathbb{Y} \rightarrow \overline{\mathbb{R}}. \quad (1)$$

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

Definition 1 *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*

$$f^c(y) = \sup_{x \in \mathbb{X}} \left(c(x, y) \dot{+} (-f(x)) \right), \quad \forall y \in \mathbb{Y}. \quad (2)$$

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

$$c' : \mathbb{Y} \times \mathbb{X} \rightarrow \overline{\mathbb{R}}, \quad c'(y, x) = c(x, y), \quad \forall (y, x) \in \mathbb{Y} \times \mathbb{X}. \quad (3)$$

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

$$g^{c'}(x) = \sup_{y \in \mathbb{Y}} \left(c(x, y) \dot{+} (-g(y)) \right), \quad \forall x \in \mathbb{X}. \quad (4)$$

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

$$f^{cc'}(x) = (f^c)^{c'}(x) = \sup_{y \in \mathbb{Y}} \left(c(x, y) \dot{+} (-f^c(y)) \right), \quad \forall x \in \mathbb{X}. \quad (5)$$

For any coupling c ,

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

$$f^{cc'}(x) \leq f(x), \quad \forall x \in \mathbb{X}, \quad (6a)$$

- for any couple of functions $f : \mathbb{X} \rightarrow \overline{\mathbb{R}}$ and $h : \mathbb{X} \rightarrow \overline{\mathbb{R}}$, we have the inequality

$$\sup_{y \in \mathbb{Y}} \left((-f^c(y)) \dot{+} (-h^{-c}(y)) \right) \leq \inf_{x \in \mathbb{X}} \left(f(x) \dot{+} h(x) \right), \quad (6b)$$

where the $(-c)$ -Fenchel-Moreau conjugate is given by

$$h^{-c}(y) = \sup_{x \in \mathbb{X}} \left((-c(x, y)) \dot{+} (-h(x)) \right), \quad \forall y \in \mathbb{Y}, \quad (6c)$$

- for any function $f : \mathbb{X} \rightarrow \overline{\mathbb{R}}$ and subset $X \subset \mathbb{X}$, we have the inequality

$$\sup_{y \in \mathbb{Y}} \left((-f^c(y)) \dot{+} (-\delta_X^{-c}(y)) \right) \leq \inf_{x \in X} \left(f(x) \dot{+} \delta_X(x) \right) = \inf_{x \in X} f(x). \quad (6d)$$

The Fenchel conjugacy

When the sets \mathbb{X} and \mathbb{Y} are vector spaces equipped with a bilinear form $\langle \cdot, \cdot \rangle$, 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 denote¹

$$f^*(y) = \sup_{x \in \mathbb{X}} \left(\langle x, y \rangle \dot{+} (-f(x)) \right), \quad \forall y \in \mathbb{Y}, \quad (7a)$$

$$g^{*'}(x) = \sup_{y \in \mathbb{Y}} \left(\langle x, y \rangle \dot{+} (-g(y)) \right), \quad \forall x \in \mathbb{X} \quad (7b)$$

$$f^{**'}(x) = \sup_{y \in \mathbb{Y}} \left(\langle x, y \rangle \dot{+} (-f^*(y)) \right), \quad \forall x \in \mathbb{X}. \quad (7c)$$

¹In convex analysis, one does not use the notation $^{*'}$, but simply the notation * , as it is often the case that $\mathbb{X} = \mathbb{Y}$ in the Euclidian and Hilbertian cases.

Due to the presence of the coupling $(-c)$ in the Inequality (6b), we also introduce²

$$f^{(-\star)}(y) = \sup_{x \in \mathbb{X}} \left(-\langle x, y \rangle \dot{+} (-f(x)) \right) = f^\star(-y), \quad \forall y \in \mathbb{Y}, \quad (8a)$$

$$g^{(-\star)'}(x) = \sup_{y \in \mathbb{Y}} \left(-\langle x, y \rangle \dot{+} (-g(y)) \right) = g^\star(-x), \quad \forall x \in \mathbb{X} \quad (8b)$$

$$f^{(-\star)(-\star)'}(x) = \sup_{y \in \mathbb{Y}} \left(-\langle x, y \rangle \dot{+} (-f^{(-\star)}(y)) \right) = f^{\star\star'}(x), \quad \forall x \in \mathbb{X}. \quad (8c)$$

When the two vector spaces \mathbb{X} and \mathbb{Y} are *paired* in the sense of convex analysis³, Fenchel conjugates are convex *lower semi continuous (lsc)* functions, and their opposites are concave *upper semi continuous (usc)* functions.

2.2 One-sided linear couplings

Let \mathbb{W} and \mathbb{X} be any two sets and $\theta : \mathbb{W} \rightarrow \mathbb{X}$ be a mapping. We recall the definition [3, p. 214] of the *infimal postcomposition* $(\theta \triangleright h) : \mathbb{X} \rightarrow \overline{\mathbb{R}}$ of a function $h : \mathbb{W} \rightarrow \overline{\mathbb{R}}$:

$$(\theta \triangleright h)(x) = \inf \{h(w) \mid w \in \mathbb{W}, \theta(w) = x\}, \quad \forall x \in \mathbb{X}, \quad (9)$$

with the convention that $\inf \emptyset = +\infty$ (and with the consequence that $\theta : \mathbb{W} \rightarrow \mathbb{X}$ need not be defined on all \mathbb{W} , but only on $\text{dom}(h) = \{w \in \mathbb{W} \mid h(w) < +\infty\}$, the *effective domain* of h). The infimal postcomposition has the following *invariance property*

$$h = f \circ \theta \text{ where } f : \mathbb{X} \rightarrow \overline{\mathbb{R}} \Rightarrow \theta \triangleright h = f \dot{+} \delta_{\theta(\mathbb{W})}, \quad (10)$$

where δ_Z denotes the *characteristic function* of a set Z :

$$\delta_Z(z) = \begin{cases} 0 & \text{if } z \in Z, \\ +\infty & \text{if } z \notin Z. \end{cases} \quad (11)$$

Definition 2 *Let \mathbb{X} and \mathbb{Y} be two vector spaces equipped with a bilinear form $\langle \cdot, \cdot \rangle$. Let \mathbb{W} be a set and $\theta : \mathbb{W} \rightarrow \mathbb{X}$ a mapping. We define the one-sided linear coupling c_θ between \mathbb{W} and \mathbb{Y} by*

$$c_\theta : \mathbb{W} \times \mathbb{Y} \rightarrow \overline{\mathbb{R}}, \quad c_\theta(w, y) = \langle \theta(w), y \rangle, \quad \forall w \in \mathbb{W}, \quad \forall y \in \mathbb{Y}. \quad (12)$$

Notice that, in a one-sided linear coupling, the second set possesses a linear structure (and is even paired with a vector space by means of a bilinear form), whereas the first set is not required to carry any structure.

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

²In convex analysis, one does not use the notations (8), but rather uses $f^\vee(x) = f(-x)$, for all $x \in \mathbb{X}$, and $g^\vee(y) = g(-y)$, for all $y \in \mathbb{Y}$. The connection between both notations is given by $f^{(-\star)} = (f^\vee)^\star = (f^\star)^\vee$.

³That is, \mathbb{X} and \mathbb{Y} are equipped with a bilinear form $\langle \cdot, \cdot \rangle$, and locally convex topologies that are compatible in the sense that the continuous linear forms on \mathbb{X} are the functions $x \in \mathbb{X} \mapsto \langle x, y \rangle$, for all $y \in \mathbb{Y}$, and that the continuous linear forms on \mathbb{Y} are the functions $y \in \mathbb{Y} \mapsto \langle x, y \rangle$, for all $x \in \mathbb{X}$.

Proposition 3 For any function $g : \mathbb{Y} \rightarrow \overline{\mathbb{R}}$, the c'_θ -Fenchel-Moreau conjugate is given by

$$g^{c'_\theta} = g^{*\prime} \circ \theta . \quad (13)$$

For any function $h : \mathbb{W} \rightarrow \overline{\mathbb{R}}$, the c_θ -Fenchel-Moreau conjugate is given by

$$h^{c_\theta} = (\theta \triangleright h)^* , \quad (14)$$

and the c_θ -Fenchel-Moreau biconjugate is given by

$$h^{c_\theta c'_\theta} = (h^{c_\theta})^{*\prime} \circ \theta = (\theta \triangleright h)^{**\prime} \circ \theta . \quad (15)$$

For any subset $W \subset \mathbb{W}$, the $(-c_\theta)$ -Fenchel-Moreau conjugate of the characteristic function of W is given by

$$\delta_W^{-c_\theta} = \sigma_{-\theta(W)} . \quad (16)$$

We recall that, in convex analysis, $\sigma_X : \mathbb{Y} \rightarrow \overline{\mathbb{R}}$ denotes the *support function* of a subset $X \subset \mathbb{X}$:

$$\sigma_X(y) = \sup_{x \in X} \langle x, y \rangle , \quad \forall y \in \mathbb{Y} . \quad (17)$$

Proof. We prove (13). Letting $w \in \mathbb{W}$, we have that

$$\begin{aligned} (g)^{c'_\theta}(w) &= \sup_{y \in \mathbb{Y}} \left(\langle \theta(w), y \rangle \dagger (-g(y)) \right) \quad (\text{by the conjugate formula (2) and the coupling (12)}) \\ &= g^{*\prime}(\theta(w)) . \quad (\text{by the expression (7b) of the Fenchel conjugate}) \end{aligned}$$

We prove (14). Letting $y \in \mathbb{Y}$, we have that

$$\begin{aligned} h^{c_\theta}(y) &= \sup_{w \in \mathbb{X}} \left(\langle \theta(w), y \rangle \dagger (-h(w)) \right) \quad (\text{by the conjugate formula (2) and the coupling (12)}) \\ &= \sup_{x \in \mathbb{X}} \sup_{w \in \mathbb{X}, \theta(w)=x} \left(\langle \theta(w), y \rangle \dagger (-h(w)) \right) \\ &= \sup_{x \in \mathbb{X}} \left(\langle x, y \rangle \dagger \sup_{w \in \mathbb{X}, \theta(w)=x} (-h(w)) \right) \quad (\text{by (47e)}) \\ &= \sup_{x \in \mathbb{X}} \left(\langle x, y \rangle \dagger \left(- \inf_{w \in \mathbb{X}, \theta(w)=x} h(w) \right) \right) \\ &= \sup_{x \in \mathbb{X}} \left(\langle x, y \rangle \dagger \left(- (\theta \triangleright h)(x) \right) \right) \quad (\text{by the infimal postcomposition expression (9)}) \\ &= (\theta \triangleright h)^*(y) . \quad (\text{by the expression (7a) of the Fenchel conjugate}) \end{aligned}$$

We prove (15). Letting $x \in \mathbb{X}$, we have that

$$\begin{aligned} h^{c_\theta c'_\theta}(x) &= (h^{c_\theta})^{c'_\theta}(x) \quad (\text{by the definition (5) of the biconjugate}) \\ &= ((\theta \triangleright h)^*)^{c'_\theta}(x) \quad (\text{by (14)}) \\ &= (\theta \triangleright h)^{**\prime}(\theta(x)) . \quad (\text{by (13)}) \end{aligned}$$

We prove (16):

$$\begin{aligned}
\delta_W^{-c_\theta} &= \delta_W^{c_{(-\theta)}} && \text{(because } -c_\theta = c_{(-\theta)} \text{ by (12))} \\
&= ((-\theta) \triangleright \delta_W)^* && \text{(by (14))} \\
&= \delta_{-\theta(W)}^* && \text{(because } \theta \triangleright \delta_W = \delta_{\theta(W)} \text{ by (9))} \\
&= \sigma_{-\theta(W)}. && \text{(as is well-known in convex analysis)}
\end{aligned}$$

This ends the proof. \square

2.3 Constant along primal rays coupling (Caprac)

Now, we introduce a novel coupling, which is a special case of one-sided linear couplings.

Definition 4 *Let \mathbb{X} and \mathbb{Y} be two vector spaces equipped with a bilinear form $\langle \cdot, \cdot \rangle$, and suppose that \mathbb{X} is equipped with a norm $\|\cdot\|$. We define the coupling $\dot{\zeta}$, or Caprac, between \mathbb{X} and \mathbb{Y} by*

$$\forall y \in \mathbb{Y}, \quad \begin{cases} \dot{\zeta}(x, y) = \frac{\langle x, y \rangle}{\|x\|}, & \forall x \in \mathbb{X} \setminus \{0\}, \\ \dot{\zeta}(0, y) = 0. \end{cases} \quad (18)$$

We stress the point that, in (18), the bilinear form term $\langle x, y \rangle$ and the norm term $\|x\|$ need not be related. Indeed, the bilinear form $\langle \cdot, \cdot \rangle$ is not necessarily a scalar product and the norm $\|\cdot\|$ is not necessarily induced by this latter.

The coupling Caprac has the property of being *constant along primal rays*, hence the acronym Caprac (Constant Along Primal RAYs Coupling). We introduce the unit sphere $S_{\|\cdot\|}$ of the normed space $(\mathbb{X}, \|\cdot\|)$, and the primal *normalization mapping* n as follows:

$$S_{\|\cdot\|} = \{x \in \mathbb{X} \mid \|x\| = 1\} \quad \text{and} \quad n : \mathbb{X} \rightarrow S_{\|\cdot\|} \cup \{0\}, \quad n(x) = \begin{cases} \frac{x}{\|x\|} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases} \quad (19)$$

With these notations, the coupling Caprac in (18) is a special case of one-sided linear coupling c_n , as in (12) with $\theta = n$, the *Fenchel coupling after primal normalization*:

$$\dot{\zeta}(x, y) = c_n(x, y) = \langle n(x), y \rangle, \quad \forall x \in \mathbb{X}, \quad \forall y \in \mathbb{Y}. \quad (20)$$

Here are expressions for the Caprac-conjugates and biconjugates of a function. The following Proposition simply is Proposition 3 in the case where the mapping θ is the normalization mapping n in (19).

Proposition 5 *Let \mathbb{X} and \mathbb{Y} be two vector spaces equipped with a bilinear form $\langle \cdot, \cdot \rangle$, and suppose that \mathbb{X} is equipped with a norm $\|\cdot\|$.*

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

$$g^{\zeta'} = g^* \circ n . \quad (21)$$

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

$$f^{\zeta} = (n \triangleright f)^* , \quad (22)$$

where the infimal postcomposition (9) has the expression

$$(n \triangleright f)(x) = \inf \{ f(x') \mid n(x') = x \} = \begin{cases} \inf_{\lambda > 0} f(\lambda x) & \text{if } x \in S_{\|\cdot\|} \cup \{0\} \\ +\infty & \text{if } x \notin S_{\|\cdot\|} \cup \{0\} \end{cases} \quad (23)$$

and the ζ -Fenchel-Moreau biconjugate is given by

$$f^{\zeta\zeta'} = (f^{\zeta})^* \circ n = (n \triangleright f)^{\star\star'} \circ n . \quad (24)$$

We recall that so-called ζ -convex functions are all functions $f : \mathbb{X} \rightarrow \overline{\mathbb{R}}$ of the form $(g)^{\zeta'}$, for any $g \in \overline{\mathbb{R}}^{\mathbb{Y}}$, or, equivalently, all functions of the form $f^{\zeta\zeta'}$, for any $f \in \overline{\mathbb{R}}^{\mathbb{X}}$, or, equivalently, all functions that are equal to their ζ -biconjugate ($f^{\zeta\zeta'} = f$) [14, 13, 8]. From the expressions (21), (22) and (24), we easily deduce the following result.

Corollary 6 *When \mathbb{X} and \mathbb{Y} are two paired vector spaces, and \mathbb{X} is equipped with a norm $\|\cdot\|$, the ζ -Fenchel-Moreau conjugate f^{ζ} is a convex lsc function on \mathbb{Y} . In addition, using (21), a function is ζ -convex if and only if it is the composition of a convex lsc function on \mathbb{X} with the normalization mapping (19).*

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

3 Caprac conjugates and biconjugates related to the ℓ_0 pseudonorm

From now on, we work on the Euclidian space \mathbb{R}^d (with $d \in \mathbb{N}^*$), equipped with the scalar product $\langle \cdot, \cdot \rangle$ and with the Euclidian norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$. In particular, we consider the *Euclidian unit sphere* and *Euclidian unit ball*

$$S = \{x \in \mathbb{X} \mid \|x\| = 1\} \quad \text{and} \quad B = \{x \in \mathbb{X} \mid \|x\| \leq 1\} , \quad (25)$$

and the (*Euclidian*) *coupling* ζ (Caprac) between \mathbb{R}^d and \mathbb{R}^d by

$$\forall y \in \mathbb{R}^d , \quad \begin{cases} \zeta(x, y) = \frac{\langle x, y \rangle}{\|x\|} = \frac{\langle x, y \rangle}{\sqrt{\langle x, x \rangle}} , & \forall x \in \mathbb{R}^d \setminus \{0\} , \\ \zeta(0, y) = 0 . \end{cases} \quad (26)$$

The so-called ℓ_0 pseudonorm is the function $\ell_0 : \mathbb{R}^d \rightarrow \{0, 1, \dots, d\}$ defined, for any $x \in \mathbb{R}^d$, by

$$\ell_0(x) = |x|_0 = \text{number of nonzero components of } x . \quad (27)$$

The ℓ_0 pseudonorm displays the invariance property

$$\ell_0 \circ n = \ell_0 \quad (28)$$

with respect to the normalization mapping (19). This property will be instrumental to show that the ℓ_0 pseudonorm is a ζ -convex function. For this purpose, we will start by introducing two dual norms.

For any $x \in \mathbb{R}^d$ and $K \subset \{1, \dots, d\}$, we denote by $x_K \in \mathbb{R}^d$ the vector which coincides with x , except for the components outside of K that vanish: x_K is the orthogonal projection of x onto the subspace $\mathbb{R}^K \times \{0\}^{-K} \subset \mathbb{R}^d$. Here, following notation from Game Theory, we have denoted by $-K$ the complementary subset of K in $\{1, \dots, d\}$: $K \cup (-K) = \{1, \dots, d\}$ and $K \cap (-K) = \emptyset$. In what follows, $|K|$ denotes the cardinal of the set K and the notation $\sup_{|K| \leq k}$ is a shorthand for $\sup_{K \subset \{1, \dots, d\}, |K| \leq k}$ (the same holds for $\sup_{|K|=k}$).

Definition 7 For $k \in \{1, \dots, d\}$, we define⁴

$$\|x\|_{(k)}^{\text{sgn}} = \sup_{|K| \leq k} \|x_K\| = \sup_{|K|=k} \|x_K\| , \quad \forall x \in \mathbb{R}^d . \quad (29)$$

Thus defined, $\|\cdot\|_{(k)}^{\text{sgn}}$ is a norm, the 2- k -symmetric gauge norm, or Ky Fan vector norm. Its dual norm (see Definition 12), denoted by $\|\cdot\|_{(k)}^{\text{sn}}$, is called the k -support norm [2]:

$$\|\cdot\|_{(k)}^{\text{sn}} = \left(\|\cdot\|_{(k)}^{\text{sgn}} \right)_\star . \quad (30)$$

The 2- k -symmetric gauge norm is also called the top- $(k, 1)$ norm, as the norm of a vector is obtained with a subvector of size k having the k largest components in module: letting σ be a permutation of $\{1, \dots, d\}$ such that $|x_{\sigma(1)}| \geq |x_{\sigma(2)}| \geq \dots \geq |x_{\sigma(d)}|$, we have that $\|x\|_{(k)}^{\text{sgn}} = \sqrt{\sum_{l=1}^k |x_{\sigma(l)}|^2}$.

The ℓ_0 pseudonorm is used in exact sparse optimization problems of the form $\inf_{|x|_0 \leq k} f(x)$. This is why we introduce the *level sets*

$$\ell_0^{\leq k} = \{x \in \mathbb{R}^d \mid \ell_0(x) \leq k\} , \quad \forall k \in \{0, 1, \dots, d\} , \quad (31a)$$

and the *level curves*

$$\ell_0^=k = \{x \in \mathbb{R}^d \mid \ell_0(x) = k\} , \quad \forall k \in \{0, 1, \dots, d\} . \quad (31b)$$

The ℓ_0 pseudonorm in (27), the characteristic functions $\delta_{\ell_0^{\leq k}}$ of its level sets (31a) and the symmetric gauge norms in (29) are related by the following conjugate formulas. The proof relies on results gathered in the Appendix A.

⁴The property that $\sup_{|K| \leq k} \|x_K\| = \sup_{|K|=k} \|x_K\|$ in (29) comes from the easy observation that $K \subset K' \Rightarrow \|x_K\| \leq \|x_{K'}\|$.

Theorem 8 Let $\dot{\zeta}$ be the Euclidian coupling Caprac (26). Let $k \in \{0, 1, \dots, d\}$. We have that:

$$\delta_{\ell_0^{\leq k}}^{-\dot{\zeta}} = \delta_{\ell_0^{\leq k}}^{\dot{\zeta}} = \|\cdot\|_{(k)}^{\text{sgn}}, \quad (32a)$$

$$\delta_{\ell_0^{\leq k}}^{\dot{\zeta}\dot{\zeta}'} = \delta_{\ell_0^{\leq k}}, \quad (32b)$$

$$\ell_0^{\dot{\zeta}} = \sup_{l=0,1,\dots,d} \left[\|\cdot\|_{(l)}^{\text{sgn}} - l \right], \quad (32c)$$

$$\ell_0^{\dot{\zeta}\dot{\zeta}'} = \ell_0, \quad (32d)$$

with the convention, in (32a) and in (32c), that $\|\cdot\|_{(0)}^{\text{sgn}} = 0$.

Proof. We will use the framework and results of Sect. 2 with $\mathbb{X} = \mathbb{Y} = \mathbb{R}^d$, equipped with the scalar product $\langle \cdot, \cdot \rangle$ and with the Euclidian norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.

- We prove (32a):

$$\begin{aligned} \delta_{\ell_0^{\leq k}}^{-\dot{\zeta}} &= \sigma_{-n(\ell_0^{\leq k})} && \text{(by (16))} \\ &= \sigma_{n(\ell_0^{\leq k})} && \text{(by symmetry of the set } \ell_0^{\leq k} \text{ and of the mapping } n) \\ &= \delta_{\ell_0^{\leq k}}^{\dot{\zeta}} && \text{(by (16))} \\ &= \sigma_{n(\ell_0^{\leq k})} && \text{(by (16))} \\ &= \sigma_{(S \cap \ell_0^{\leq k}) \cup \{0\}} && \text{(by the expression (19) of the normalization mapping } n) \\ &= \sup \{ \sigma_{\ell_0^{\leq k} \cap S}, 0 \} && \text{(as is well-known in convex analysis)} \\ &= \sup \{ \sigma_{\bigcup_{|K| \leq k} S_K}, 0 \} && \text{(as } \ell_0^{\leq k} \cap S = \bigcup_{|K| \leq k} S_K \text{ by (71a))} \\ &= \sup \left\{ \sup_{|K| \leq k} \sigma_{S_K}, 0 \right\} && \text{(as is well-known in convex analysis)} \\ &= \sup \{ \|\cdot\|_{(k)}^{\text{sgn}}, 0 \} && \text{(as } \sup_{|K| \leq k} \sigma_{S_K} = \|\cdot\|_{(k)}^{\text{sgn}} \text{ by (57))} \\ &= \|\cdot\|_{(k)}^{\text{sgn}}. \end{aligned}$$

- Before proving (32b), let us observe that, by definition (19) of the normalization mapping n , we have:

$$0 \in D \subset \mathbb{R}^d \Rightarrow n^{-1}(D) = n^{-1}(\{0\} \cup S_{\|\cdot\|}) \cap D = \{0\} \cup n^{-1}(S_{\|\cdot\|} \cap D). \quad (33)$$

Now, we prove (32b):

$$\begin{aligned}
\delta_{\ell_0^{\leq k}}^{\dot{\zeta}\dot{\zeta}'} &= (\delta_{\ell_0^{\leq k}}^{\dot{\zeta}})^{\star} \circ n && \text{(by the formula (24) for the biconjugate)} \\
&= (\|\cdot\|_{(k)}^{\text{sgn}})^{\star} \circ n && \text{(by (32a))} \\
&= (\sigma_{B_{(k)}^{\text{sn}}})^{\star} \circ n && \text{(by (52), that expresses a norm as a support function)} \\
&= \delta_{B_{(k)}^{\text{sn}}} \circ n && \text{(as } (\sigma_{B_{(k)}^{\text{sn}}})^{\star} = \delta_{B_{(k)}^{\text{sn}}} \text{ since } B_{(k)}^{\text{sn}} \text{ is closed convex)} \\
&= \delta_{n^{-1}(B_{(k)}^{\text{sn}})} && \text{(by the definition (11) of a characteristic function)} \\
&= \delta_{\{0\} \cup n^{-1}(S \cap B_{(k)}^{\text{sn}})} && \text{(by (33) since } 0 \in B_{(k)}^{\text{sn}}) \\
&= \delta_{\{0\} \cup n^{-1}(S \cap \ell_0^{\leq k})} && \text{(as } S \cap B_{(k)}^{\text{sn}} = S \cap \ell_0^{\leq k} \text{ by (71b))} \\
&= \delta_{n^{-1}(\ell_0^{\leq k})} && \text{(by (33) since } 0 \in \ell_0^{\leq k}) \\
&= \delta_{\ell_0^{\leq k}}. && \text{(as } \ell_0 \circ n = \ell_0 \text{ by (28))}
\end{aligned}$$

• We prove (32c):

$$\begin{aligned}
\ell_0^{\dot{\zeta}} &= \left(\inf_{l=0,1,\dots,d} \{ \delta_{\ell_0^{\leq l}}^{\dot{\zeta}} + l \} \right)^{\dot{\zeta}} && \text{(since } \ell_0 = \inf_{l=0,1,\dots,d} \{ \delta_{\ell_0^{\leq l}}^{\dot{\zeta}} + l \} \text{ by using the level curves (31b))} \\
&= \sup_{l=0,1,\dots,d} \{ \delta_{\ell_0^{\leq k}}^{\dot{\zeta}} + (-l) \} && \text{(as conjugacies, being dualities, turn infima into suprema)} \\
&= \sup_{l=0,1,\dots,d} \{ \sigma_{n(\ell_0^{\leq l})} + (-l) \} && \text{(as } \delta_{\ell_0^{\leq k}}^{\dot{\zeta}} = \sigma_{n(\ell_0^{\leq l})} \text{ by (16))} \\
&= \sup \left\{ 0, \sup_{l=1,\dots,d} \{ \sigma_{S \cap \ell_0^{\leq l}} + (-l) \} \right\} && \text{(as } n(\ell_0^{\leq l}) = S \cap \ell_0^{\leq l} \text{ when } l \geq 1 \text{ by (19))} \\
&= \sup \left\{ 0, \sup_{l=1,\dots,d} \{ \sigma_{\overline{S \cap \ell_0^{\leq l}}} + (-l) \} \right\} && \text{(as } \sigma_X = \sigma_{\overline{X}} \text{ for any } X \subset \mathbb{X}) \\
&= \sup \left\{ 0, \sup_{l=1,\dots,d} \{ \sigma_{S \cap \ell_0^{\leq l}} + (-l) \} \right\} && \text{(as } \overline{S \cap \ell_0^{\leq l}} = S \cap \ell_0^{\leq l} \text{ by (71c))} \\
&= \sup \left\{ 0, \sup_{l=1,\dots,d} \{ \sigma_{\cup_{|K| \leq l} S_K} + (-l) \} \right\} && \text{(as } S \cap \ell_0^{\leq l} = \cup_{|K| \leq l} S_K \text{ by (71a))} \\
&= \sup \left\{ 0, \sup_{l=1,\dots,d} \left[\|y\|_{(l)}^{\text{sgn}} - l \right] \right\} && \text{(as } \sup_{|K| \leq l} \sigma_{S_K} = \|\cdot\|_{(l)}^{\text{sgn}} \text{ by (57))} \\
&= \sup_{l=0,1,\dots,d} \left[\|y\|_{(l)}^{\text{sgn}} - l \right]. && \text{(using the convention that } \|\cdot\|_{(0)}^{\text{sgn}} = 0)
\end{aligned}$$

• We prove (32d).

It is easy to check that $\ell_0^{\dot{\zeta}\dot{\zeta}'}(0) = 0 = \ell_0(0)$. Therefore, let $x \in \mathbb{R}^d \setminus \{0\}$ be given and assume that $\ell_0(x) = l \in \{1, \dots, d\}$. We consider the mapping $\phi :]0, +\infty[\rightarrow \mathbb{R}$ defined by

$$\phi(\lambda) = \frac{\langle x, \lambda x \rangle}{\|x\|} + \left(- \sup \left\{ 0, \sup_{j=1,\dots,d} \left[\|\lambda x\|_{(j)}^{\text{sgn}} - j \right] \right\} \right), \quad \forall \lambda > 0, \quad (34)$$

and we will show that $\lim_{\lambda \rightarrow +\infty} \phi(\lambda) = l$. We have

$$\begin{aligned}
\phi(\lambda) &= \lambda \|x\| + \left(- \sup \left\{ 0, \sup_{j=1, \dots, d} \left[\|\lambda x\|_{(j)}^{\text{sgn}} - j \right] \right\} \right) && \text{(by definition (34) of } \phi) \\
&= \lambda \|x\|_{(l)}^{\text{sgn}} + \inf \left\{ 0, - \sup_{j=1, \dots, d} \left[\lambda \|x\|_{(j)}^{\text{sgn}} - j \right] \right\} && \text{(as } \|x\| = \|x\|_{(l)}^{\text{sgn}} \text{ when } \ell_0(x) = l \text{ by (68))} \\
&= \inf \left\{ \lambda \|x\|_{(l)}^{\text{sgn}}, \lambda \|x\|_{(l)}^{\text{sgn}} + \inf_{j=1, \dots, d} \left(- \left[\lambda \|x\|_{(j)}^{\text{sgn}} - j \right] \right) \right\} \\
&= \inf \left\{ \lambda \|x\|_{(l)}^{\text{sgn}}, \inf_{j=1, \dots, d} \left(\lambda (\|x\|_{(l)}^{\text{sgn}} - \|x\|_{(j)}^{\text{sgn}}) + j \right) \right\} \\
&= \inf \left\{ \lambda \|x\|_{(l)}^{\text{sgn}}, \inf_{j=1, \dots, l-1} \left(\lambda (\|x\|_{(l)}^{\text{sgn}} - \|x\|_{(j)}^{\text{sgn}}) + j \right), \inf_{j=l, \dots, d} \left(\lambda (\|x\|_{(l)}^{\text{sgn}} - \|x\|_{(j)}^{\text{sgn}}) + j \right) \right\} \\
&= \inf \left\{ \lambda \|x\|_{(l)}^{\text{sgn}}, \inf_{j=1, \dots, l-1} \left(\lambda (\|x\|_{(l)}^{\text{sgn}} - \|x\|_{(j)}^{\text{sgn}}) + j \right), l \right\}
\end{aligned}$$

as $\|x\|_{(j)}^{\text{sgn}} = \|x\|_{(l)}^{\text{sgn}}$ for $j \geq l$ by (68). Let us show that the two first terms in the infimum go to $+\infty$ when $\lambda \rightarrow +\infty$. The first term $\lambda \|x\|_{(l)}^{\text{sgn}}$ goes to $+\infty$ because $\|x\|_{(l)}^{\text{sgn}} = \|x\| > 0$ by assumption ($x \neq 0$). The second term $\inf_{j=1, \dots, l-1} \left(\lambda (\|x\|_{(l)}^{\text{sgn}} - \|x\|_{(j)}^{\text{sgn}}) + j \right)$ also goes to $+\infty$ because $\ell_0(x) = l$, so that $\|x\| = \|x\|_{(l)}^{\text{sgn}} > \|x\|_{(j)}^{\text{sgn}}$ for $j = 1, \dots, l-1$ by (68). Therefore, $\lim_{\lambda \rightarrow +\infty} \phi(\lambda) = \inf \{ +\infty, +\infty, l \} = l$. This concludes the proof since

$$\begin{aligned}
l &= \lim_{\lambda \rightarrow +\infty} \phi(\lambda) \leq \sup_{y \in \mathbb{R}^d} \left(\frac{\langle x, y \rangle}{\|x\|} \dagger \left(- \sup \left\{ 0, \sup_{j=1, \dots, d} \left[\|y\|_{(j)}^{\text{sgn}} - j \right] \right\} \right) \right) && \text{(by definition (34) of } \phi) \\
&= \sup_{y \in \mathbb{R}^d} \left(\frac{\langle x, y \rangle}{\|x\|} \dagger \left(- \sup_{j=0, 1, \dots, d} \left[\|y\|_{(j)}^{\text{sgn}} - j \right] \right) \right) \\
& && \text{(by the convention that } \|\cdot\|_{(0)}^{\text{sgn}} = 0) \\
&= \sup_{y \in \mathbb{R}^d} \left(\frac{\langle x, y \rangle}{\|x\|} \dagger \left(- \ell_0^{\dot{C}}(y) \right) \right) && \text{(by the formula (32c) for } \ell_0^{\dot{C}}) \\
&= \ell_0^{\dot{C}\dot{C}'}(x) && \text{(by the biconjugate formula (5))} \\
&\leq \ell_0(x) && \text{(by (6a) giving } \ell_0^{\dot{C}\dot{C}'} \leq \ell_0) \\
&= l. && \text{(by assumption)}
\end{aligned}$$

Therefore, we have obtained $l = \ell_0^{\dot{C}\dot{C}'}(x) = \ell_0(x)$.

This ends the proof. \square

We will now present a (rather unexpected) consequence of the just established property that $\ell_0^{\dot{C}\dot{C}'} = \ell_0$.

4 Hidden convexity in the pseudonorm ℓ_0

In this Section, we will show that the ℓ_0 pseudonorm coincides, on the Euclidian sphere S , with a proper⁵ convex lsc function defined on the whole space. This somehow comes as a surprise as the ℓ_0 pseudonorm is a highly nonconvex function of combinatorial nature. Then, we will provide various expression for the underlying proper convex lsc function, and display mathematical expressions and graphical representations in the two-dimensional case.

4.1 The pseudonorm ℓ_0 coincides, on the Euclidian sphere S of \mathbb{R}^d , with a proper convex lsc function \mathcal{L}_0 on \mathbb{R}^d

As a straightforward corollary of Theorem 8, we obtain the following property that the ℓ_0 pseudonorm coincides, on the Euclidian sphere S , with a proper convex lsc function defined on the whole space.

Corollary 9 *The pseudonorm ℓ_0 coincides, on the Euclidian sphere S of \mathbb{R}^d , with a proper convex lsc function on \mathbb{R}^d . In particular, we have that*

$$\ell_0(x) = \mathcal{L}_0(x), \quad \forall x \in S \quad \text{with} \quad \mathcal{L}_0 = \left(\sup_{l=0,1,\dots,d} \left[\|\cdot\|_{(l)}^{\text{sgn}} - l \right] \right)^{\star'}, \quad (35)$$

where the function $\mathcal{L}_0 : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ is proper convex lsc.

As a consequence, the pseudonorm ℓ_0 can be expressed as the composition of the proper convex lsc function \mathcal{L}_0 in (35) with the normalization mapping n in (19):

$$\ell_0(x) = \mathcal{L}_0\left(\frac{x}{\|x\|}\right), \quad \forall x \in \mathbb{R}^d \setminus \{0\}. \quad (36)$$

The proper convex lsc function \mathcal{L}_0 has the property

$$\mathcal{L}_0((x_1, \dots, x_d)) = \mathcal{L}_0((|x_1|, \dots, |x_d|)), \quad \forall (x_1, \dots, x_d) \in \mathbb{R}^d. \quad (37)$$

Proof. First, we prove (35). For $x \in S$, we have

$$\begin{aligned} \ell_0(x) &= \ell_0^{\dot{\mathcal{C}}'}(x) && \text{(by (32d))} \\ &= \sup_{y \in \mathbb{R}^d} \left(\dot{\mathcal{C}}(x, y) \dagger (-\ell_0^{\dot{\mathcal{C}}}(y)) \right) && \text{(by the biconjugate formula (5))} \\ &= \sup_{y \in \mathbb{R}^d} \left(\langle x, y \rangle \dagger (-\ell_0^{\dot{\mathcal{C}}}(y)) \right) && \text{(by (18) with } \|x\| = 1 \text{ since } x \in S) \\ &= \sup_{y \in \mathbb{R}^d} \left(\langle x, y \rangle \dagger \left(- \left(\sup_{l=0,1,\dots,d} \left[\|y\|_{(l)}^{\text{sgn}} - l \right] \right) \right) \right) && \text{(by (32c))} \\ &= \left(\sup_{l=0,1,\dots,d} \left[\|y\|_{(l)}^{\text{sgn}} - l \right] \right)^{\star'}(x) && \text{(by the expression (7a) of the Fenchel conjugate)} \\ &= \mathcal{L}_0(x). \end{aligned}$$

⁵The function $h : \mathbb{W} \rightarrow \overline{\mathbb{R}}$ is said to be *proper* when $-\infty < h(w)$, for all $w \in \mathbb{W}$, and there is at least one $w \in \mathbb{W}$ such that $h(w) < +\infty$.

Second, the equality (36) is an easy consequence of the property that the pseudonorm ℓ_0 is invariant along any open ray of \mathbb{R}^d .

Third, we prove (37). For this purpose, let us take any $\epsilon \in \{-1, 1\}^d$ and consider the symmetry $\tilde{\epsilon}$ of \mathbb{R}^d , defined by

$$\tilde{\epsilon}(x_1, \dots, x_d) = (\epsilon_1 x_1, \dots, \epsilon_d x_d), \quad \forall (x_1, \dots, x_d) \in \mathbb{R}^d. \quad (38)$$

We will show that the proper convex lsc function \mathcal{L}_0 is invariant under the symmetry $\tilde{\epsilon}$ in (38), hence satisfies (37). Indeed, for any $x \in \mathbb{R}^d$, we have

$$\begin{aligned} \mathcal{L}_0(\tilde{\epsilon}x) &= \left(\sup_{l=0,1,\dots,d} \left[\|\cdot\|_{(l)}^{\text{sgn}} - l \right] \right)^{\star'}(\tilde{\epsilon}x) && \text{(by (35))} \\ &= \sup_{y \in \mathbb{R}^d} \left(\langle \tilde{\epsilon}x, y \rangle \dagger \left(- \left(\sup_{l=0,1,\dots,d} \left[\|y\|_{(l)}^{\text{sgn}} - l \right] \right) \right) \right) \\ &\quad \text{(by the expression (4) of the reverse Fenchel conjugate)} \\ &= \sup_{y \in \mathbb{R}^d} \left(\langle x, \tilde{\epsilon}y \rangle \dagger \left(- \left(\sup_{l=0,1,\dots,d} \left[\|y\|_{(l)}^{\text{sgn}} - l \right] \right) \right) \right) && \text{(as easily seen from (38))} \\ &= \sup_{y \in \mathbb{R}^d} \left(\langle x, \tilde{\epsilon}y \rangle \dagger \left(- \left(\sup_{l=0,1,\dots,d} \left[\|\tilde{\epsilon}y\|_{(l)}^{\text{sgn}} - l \right] \right) \right) \right) \end{aligned}$$

as $\|\cdot\|_{(0)}^{\text{sgn}} \equiv 0$ (by convention) and all norms $\|\cdot\|_{(l)}^{\text{sgn}}, l = 1, \dots, d$ are invariant under the symmetry $\tilde{\epsilon}$ as seen from their definition (29)

$$\begin{aligned} &= \left(\sup_{l=0,1,\dots,d} \left[\|\cdot\|_{(l)}^{\text{sgn}} - l \right] \right)^{\star'}(x) && \text{(as } \tilde{\epsilon}^{-1}(\mathbb{R}^d) = \mathbb{R}^d \text{ by (38))} \\ &= \mathcal{L}_0(x). && \text{(by (35))} \end{aligned}$$

This ends the proof. \square

Now, we provide various expression for the underlying proper convex lsc function \mathcal{L}_0 in Corollary 9.

Proposition 10 *The function \mathcal{L}_0 in (35) can also be characterized*

- either by its epigraph⁶

$$\text{epi}\mathcal{L}_0 = \overline{\text{co}} \left(\bigcup_{l=0}^d B_{(l)}^{\text{sn}} \times [l, +\infty[\right), \quad (39)$$

where $\{0\} = B_{(0)}^{\text{sn}}$ (by convention) and $B_{(1)}^{\text{sn}} \subset \dots \subset B_{(l-1)}^{\text{sn}} \subset B_{(l)}^{\text{sn}} \subset \dots \subset B_{(d)}^{\text{sn}} = B$ denote the unit balls associated with the l -support norms defined in (30) for $l = 1, \dots, d$,

⁶For any function $h : \mathbb{W} \rightarrow \overline{\mathbb{R}}$, its epigraph is $\text{epi}h = \{(w, t) \in \mathbb{W} \times \mathbb{R} \mid h(w) \leq t\}$.

- or, equivalently, as the largest proper convex lsc function below the (extended integers valued) function L_0 defined by

$$L_0(x) = \begin{cases} 0 & \text{if } x = 0, \\ l & \text{if } x \in B_{(l)}^{\text{sn}} \setminus B_{(l-1)}^{\text{sn}}, \quad l = 1, \dots, d, \\ +\infty & \text{if } x \notin B, \end{cases} \quad (40)$$

- or also by the expression

$$\mathcal{L}_0(x) = \min_{\substack{x^{(1)} \in \mathbb{R}^d, \dots, x^{(d)} \in \mathbb{R}^d \\ \sum_{l=1}^d \|x^{(l)}\|_{(l)}^{\text{sn}} \leq 1 \\ \sum_{l=1}^d x^{(l)} = x}} \sum_{l=1}^d l \|x^{(l)}\|_{(l)}^{\text{sn}}, \quad \forall x \in \mathbb{R}^d. \quad (41)$$

Proof. First, we prove that the epigraph of \mathcal{L}_0 is given by (39). Indeed, we have that

$$\begin{aligned} \text{epi} \mathcal{L}_0 &= \text{epi} \left(\sup_{l=0,1,\dots,d} \left[\|\cdot\|_{(l)}^{\text{sgn}} - l \right] \right)^{\star'} && \text{(by (35))} \\ &= \overline{\text{co}} \left(\bigcup_{l=0}^d \text{epi} \left[\|\cdot\|_{(l)}^{\text{sgn}} - l \right]^{\star'} \right) && \text{(by [12, Theorem 16.5])} \\ &= \overline{\text{co}} \left(\bigcup_{l=0}^d \text{epi} \left[\sigma_{B_{(l)}^{\text{sn}}} - l \right]^{\star'} \right) && \text{(by (67))} \\ &= \overline{\text{co}} \left(\bigcup_{l=0}^d \text{epi} \left[\delta_{B_{(l)}^{\text{sn}}} + l \right] \right) && \text{(as } \left[\sigma_{B_{(l)}^{\text{sn}}} - l \right]^{\star'} = \delta_{B_{(l)}^{\text{sn}}} + l) \\ &= \overline{\text{co}} \left(\bigcup_{l=0}^d B_{(l)}^{\text{sn}} \times [l, +\infty[\right) \end{aligned}$$

as is easily concluded.

Second, we prove that \mathcal{L}_0 is the largest proper convex lsc function below the function L_0 defined by (40). Indeed, we have that

$$\begin{aligned} \mathcal{L}_0 &= \left(\sup_{l=0,1,\dots,d} \left[\|\cdot\|_{(l)}^{\text{sgn}} - l \right] \right)^{\star'} && \text{(by (35))} \\ &= \left(\sup_{l=0,1,\dots,d} \left[\sigma_{B_{(l)}^{\text{sn}}} - l \right] \right)^{\star'} && \text{(by (67))} \\ &= \left(\sup_{l=0,1,\dots,d} \left[\delta_{B_{(l)}^{\text{sn}}} + l \right]^{\star} \right)^{\star'} && \text{(as } \left[\delta_{B_{(l)}^{\text{sn}}} + l \right]^{\star} = \sigma_{B_{(l)}^{\text{sn}}} - l) \\ &= \left(\left[\inf_{l=0,1,\dots,d} \left[\delta_{B_{(l)}^{\text{sn}}} + l \right]^{\star} \right]^{\star} \right)^{\star'} && \text{(as conjugacies, being dualities, turn infima into suprema)} \\ &= \left(\inf_{l=0,1,\dots,d} \left[\delta_{B_{(l)}^{\text{sn}}} + l \right] \right)^{\star\star'} \\ &= L_0^{\star\star'} \end{aligned}$$

as it is easy to establish that the function $\inf_{l=0,1,\dots,d} [\delta_{B_{(l)}^{\text{sn}}} + l]$ coincides with the function L_0 defined by (40). Indeed, it is deduced from (66) that $\{0\} = B_{(0)}^{\text{sn}} \subset B_{(1)}^{\text{sn}} \subset \dots \subset B_{(l-1)}^{\text{sn}} \subset B_{(l)}^{\text{sn}} \subset \dots \subset B_{(d)}^{\text{sn}} = B$. Finally, from $\mathcal{L}_0 = L_0^{\star\star'}$, we conclude that \mathcal{L}_0 is the largest proper convex lsc function below the function L_0 .

Third, we prove that \mathcal{L}_0 is given by (41). For this purpose, we use a general formula [15, Corollary 2.8.11] for the Fenchel conjugate of the supremum of proper convex functions $f_l : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$, $l = 0, 1, \dots, d$:

$$\bigcap_{l=0,1,\dots,d} \text{dom} f_l \neq \emptyset \Rightarrow \left(\sup_{l=0,1,\dots,d} f_l \right)^{\star} = \min_{\lambda \in \Delta_{d+1}} \left(\sum_{l=0}^d \lambda_l f_l \right)^{\star}, \quad (42)$$

where Δ_{d+1} is the simplex of \mathbb{R}^d . As the functions $f_l = \|\cdot\|_{(l)}^{\text{sgn}} - l$ are proper convex, we obtain

$$\mathcal{L}_0 = \left(\sup_{l=0,1,\dots,d} \left[\|\cdot\|_{(l)}^{\text{sgn}} - l \right] \right)^{\star'} \quad (\text{by (35)})$$

$$= \left(\sup_{l=0,1,\dots,d} \left[\sigma_{B_{(l)}^{\text{sn}}} - l \right] \right)^{\star'} \quad (\text{by (67)})$$

$$= \min_{\lambda \in \Delta_{d+1}} \left(\sum_{l=0}^d \lambda_l \left[\sigma_{B_{(l)}^{\text{sn}}} - l \right] \right)^{\star} \quad (\text{by (42)})$$

$$= \min_{\lambda \in \Delta_{d+1}} \left(\sigma_{\sum_{l=0}^d \lambda_l B_{(l)}^{\text{sn}}} - \sum_{l=0}^d \lambda_l l \right)^{\star}$$

as, for all $l = 0, \dots, d$, $\lambda_l \sigma_{B_{(l)}^{\text{sn}}} = \sigma_{\lambda_l B_{(l)}^{\text{sn}}}$ since $\lambda_l \geq 0$, 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

$$= \min_{\lambda \in \Delta_{d+1}} \left(\delta_{\sum_{l=0}^d \lambda_l B_{(l)}^{\text{sn}}} + \sum_{l=0}^d \lambda_l l \right).$$

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

$$\mathcal{L}_0(x) = \min_{\substack{\lambda \in \Delta_{d+1} \\ x \in \sum_{l=0}^d \lambda_l B_{(l)}^{\text{sn}}}} \sum_{l=0}^d \lambda_l l, \quad (43a)$$

$$= \min_{\substack{z^{(1)} \in B_{(1)}^{\text{sn}}, \dots, z^{(d)} \in B_{(d)}^{\text{sn}} \\ \lambda_1 \geq 0, \dots, \lambda_d \geq 0 \\ \sum_{l=1}^d \lambda_l \leq 1 \\ \sum_{l=1}^d \lambda_l z^{(l)} = x}} \sum_{l=1}^d \lambda_l l \quad (43b)$$

by ignoring $\lambda_0 \geq 0$ since $B_{(0)}^{\text{sn}} = \{0\}$ by convention

$$= \min_{\substack{s^{(1)} \in S_{(1)}^{\text{sn}}, \dots, s^{(d)} \in S_{(d)}^{\text{sn}} \\ \mu_1 \geq 0, \dots, \mu_d \geq 0 \\ \sum_{l=1}^d \mu_l \leq 1 \\ \sum_{l=1}^d \mu_l s^{(l)} = x}} \sum_{l=1}^d \mu_l l \quad (43c)$$

the inequality \leq is obvious as $S_{(l)}^{\text{sn}} \subset B_{(l)}^{\text{sn}}$ for all $l = 1, \dots, d$; the inequality \geq comes from putting, for $l = 1, \dots, d$, $\mu_l = \lambda_l \|z^{(l)}\|_{(l)}^{\text{sn}}$ and observing that i) there exist $s^{(l)} \in S_{(l)}^{\text{sn}}$ such that $\lambda_l z^{(l)} = \mu_l s^{(l)}$ (take any $s^{(l)}$ when $z^{(l)} = 0$ and $s^{(l)} = \frac{z^{(l)}}{\|z^{(l)}\|_{(l)}^{\text{sn}}}$ when $z^{(l)} \neq 0$) ii) $\sum_{l=1}^d \lambda_l l \geq \sum_{l=1}^d \lambda_l \|z^{(l)}\|_{(l)}^{\text{sn}} l = \sum_{l=1}^d \mu_l l$ because $\|z^{(l)}\|_{(l)}^{\text{sn}} \leq 1$

$$= \min_{\substack{x^{(1)} \in \mathbb{R}^d, \dots, x^{(d)} \in \mathbb{R}^d \\ \sum_{l=1}^d \|x^{(l)}\|_{(l)}^{\text{sn}} \leq 1 \\ \sum_{l=1}^d x^{(l)} = x}} \sum_{l=1}^d \|x^{(l)}\|_{(l)}^{\text{sn}} l, \quad (43d)$$

by putting $x^{(l)} = \mu_l s^{(l)}$, for all $l = 1, \dots, d$.

This ends the proof. \square

With Proposition 10, we dispose of expressions that make it possible to obtain more involved formulas for the function \mathcal{L}_0 . For instance, in Proposition 18 we provide characterization of the optimal solutions of the minimization problem (41). With these results, we will now obtain graphical representations and mathematical formulas for the proper convex lsc function \mathcal{L}_0 on \mathbb{R}^2 .

4.2 Graphical representations of the proper convex lsc function \mathcal{L}_0 on \mathbb{R}^2

In dimension $d = 1$, it is easily computed that the function \mathcal{L}_0 in (35) is the absolute value function $|\cdot|$ on the segment $[-1, 1]$ and $+\infty$ outside the segment $[-1, 1]$. The pseudonorm ℓ_0 coincides with \mathcal{L}_0 on the sphere $\{-1, 1\}$ — but also with any convex function taking the value 1 on $\{-1, 1\}$ (the absolute value function $|\cdot|$, the constant function 1, etc.).

In dimension $d = 2$, the function \mathcal{L}_0 in (35) is, by Proposition 10, the largest convex sci function which is below the function which takes the value 0 on the zero $(0, 0)$, the value 1 on the unit lozenge of \mathbb{R}^2 deprived of $(0, 0)$, and the value 2 on the unit disk of \mathbb{R}^2 deprived of the unit lozenge (see Proposition 10). As a consequence, the graph of \mathcal{L}_0 contains segments (in \mathbb{R}^3) that join the zero $(0, 0, 0)$ of the horizontal plane at height $z = 0$ with the unit lozenge of the horizontal plane at height $z = 1$, and this latter with the unit circle of the horizontal plane at height $z = 2$. In Figure 1, we have displayed two views of the topological closure of the graph of \mathcal{L}_0 . As the function \mathcal{L}_0 is not continuous at the four extremal points

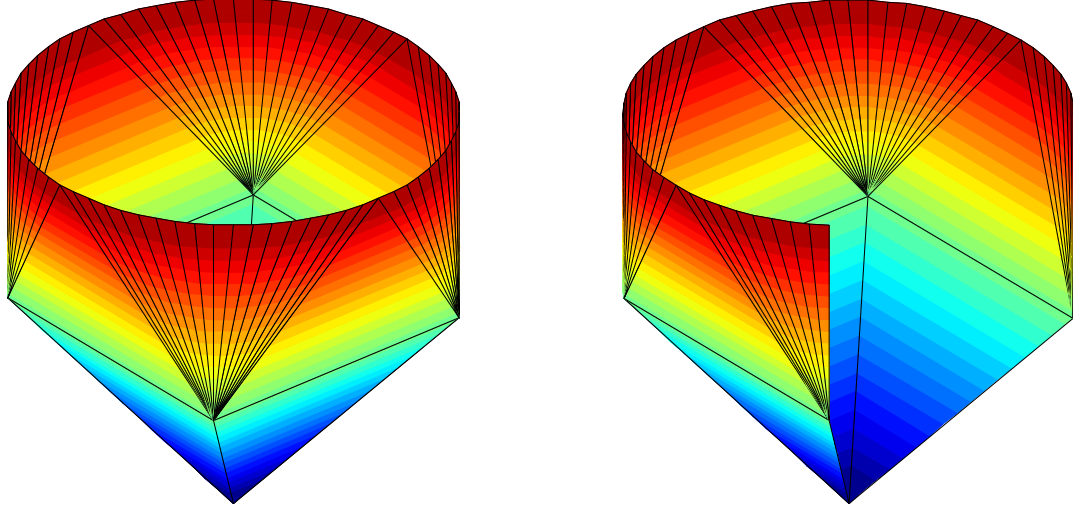


Figure 1: Topological closure of the graph, between heights $z = 0$ and $z = 2$, of the proper convex lsc function \mathcal{L}_0 which coincides, on the sphere S , with the ℓ_0 pseudonorm

— $(0, 1), (1, 0), (0, -1), (-1, 0)$ — of the unit lozenge, it is delicate to depict the graph and easier to do so for its topological closure.

In dimension $d = 2$, the function \mathcal{L}_0 in (35) is given by the following explicit formulas (see also Figure 2).

Proposition 11 *In dimension $d = 2$, the function \mathcal{L}_0 in (35) is given by*

$$\mathcal{L}_0(x_1, x_2) = \begin{cases} +\infty & \text{if } x_1^2 + x_2^2 > 1, & (44a) \\ 1 & \text{if } (x_1, x_2) \in \{(1, 0), (0, 1), (-1, 0), (0, -1)\}, & (44b) \\ 2 & \text{if } x_1^2 + x_2^2 = 1 \text{ and } (x_1, x_2) \notin \{(1, 0), (0, 1), (-1, 0), (0, -1)\} & (44c) \end{cases}$$

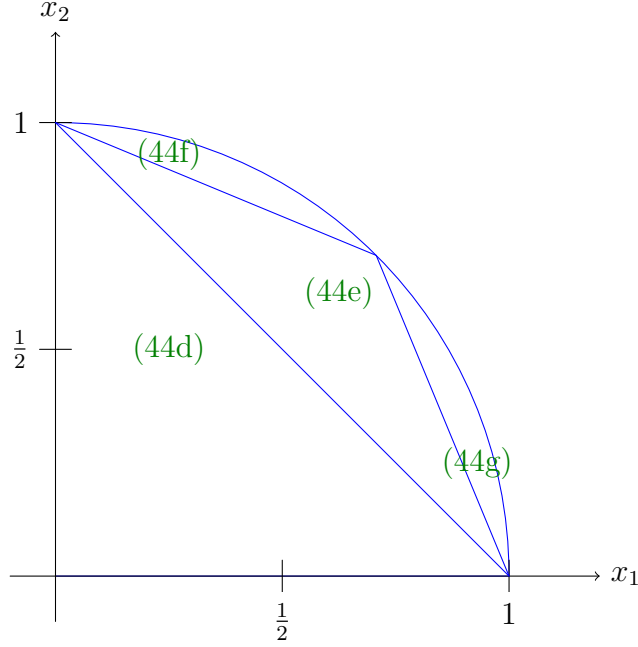


Figure 2: Companion figure for Proposition 11

and, for any (x_1, x_2) such that $x_1^2 + x_2^2 < 1$ by

$$\mathcal{L}_0(x_1, x_2) = \begin{cases} |x_1| + |x_2| & \text{if } |x_1| + |x_2| \leq 1, & (44d) \\ \frac{|x_1| + |x_2| - 2 + \sqrt{2}}{\sqrt{2} - 1} & \text{if } \begin{cases} (\sqrt{2} - 1)|x_1| + |x_2| < 1 < |x_1| + |x_2| \\ \text{or} \\ |x_1| + (\sqrt{2} - 1)|x_2| < 1 < |x_1| + |x_2|, \end{cases} & (44e) \\ \frac{3 - |x_2|}{2} + \frac{x_1^2}{2(1 - |x_2|)} & \text{if } (\sqrt{2} - 1)|x_1| + |x_2| \geq 1 \text{ and } |x_2| > |x_1|, & (44f) \\ \frac{3 - |x_1|}{2} + \frac{x_2^2}{2(1 - |x_1|)} & \text{if } |x_1| + (\sqrt{2} - 1)|x_2| \geq 1 \text{ and } |x_1| > |x_2| & (44g) \end{cases}$$

Proof. By (41) for $d = 2$, we find that

$$\mathcal{L}_0(x) = \min_{(x^{(1)}, x^{(2)}) \in C(x)} \|x^{(1)}\|_{(1)}^{\text{sn}} + 2\|x^{(2)}\|_{(2)}^{\text{sn}}, \quad (45a)$$

where the constraints set is given by

$$C(x) = \left\{ (x^{(1)}, x^{(2)}) \in (\mathbb{R}^2)^2 \mid \|x^{(1)}\|_{(1)}^{\text{sn}} + \|x^{(2)}\|_{(2)}^{\text{sn}} \leq 1, x^{(1)} + x^{(2)} = x \right\}. \quad (45b)$$

If $(x^{(1)}, x^{(2)}) \in C(x)$, we have that

$$\|x\| \leq \|x^{(1)}\|_{(1)}^{\text{sn}} + \|x^{(2)}\|_{(2)}^{\text{sn}} \leq 1, \quad (46a)$$

$$\|x\| = 1 \Rightarrow \|x^{(1)}\|_{(2)}^{\text{sn}} = \|x^{(1)}\|_{(1)}^{\text{sn}} \text{ and } \|x^{(1)}\|_{(1)}^{\text{sn}} + \|x^{(2)}\|_{(2)}^{\text{sn}} = 1, \quad (46b)$$

because

$$\begin{aligned} \|x\| &= \|x\|_{(2)}^{\text{sn}} && \text{(by (65))} \\ &\leq \|x^{(1)}\|_{(2)}^{\text{sn}} + \|x^{(2)}\|_{(2)}^{\text{sn}} && \text{(because } (x^{(1)}, x^{(2)}) \in C(x) \Rightarrow x^{(1)} + x^{(2)} = x) \\ &\leq \|x^{(1)}\|_{(1)}^{\text{sn}} + \|x^{(2)}\|_{(2)}^{\text{sn}} && \text{(because } \|x^{(1)}\|_{(2)}^{\text{sn}} \leq \|x^{(1)}\|_{(1)}^{\text{sn}} \text{ by (65))} \\ &\leq 1. && \text{(because } (x^{(1)}, x^{(2)}) \in C(x) \Rightarrow \|x^{(1)}\|_{(1)}^{\text{sn}} + \|x^{(2)}\|_{(2)}^{\text{sn}} \leq 1) \end{aligned}$$

We are now going to describe the constraints set $C(x)$ in (45b) according to $\|x\|$, then to deduce $\mathcal{L}_0(x)$ from (45a).

1. Suppose that $\|x\| = \sqrt{x_1^2 + x_2^2} > 1$. Then, by (46a), we deduce that $C(x) = \emptyset$ in (45b), hence that $\mathcal{L}_0(x) = +\infty$ by (45a).
2. Suppose that $\|x\| = \sqrt{x_1^2 + x_2^2} = 1$. If $(x^{(1)}, x^{(2)}) \in C(x)$, we obtain by (46b) that

$$\sqrt{|x_1^{(1)}|^2 + |x_2^{(1)}|^2} = \|x^{(1)}\|_{(2)}^{\text{sn}} = \|x^{(1)}\|_{(1)}^{\text{sn}} = |x_1^{(1)}| + |x_2^{(1)}|,$$

from which we deduce that $|x_1^{(1)}| \times |x_2^{(1)}| = 0$. From $x^{(1)} + x^{(2)} = x$ and $\|x^{(1)}\|_{(1)}^{\text{sn}} + \|x^{(2)}\|_{(2)}^{\text{sn}} = 1$, by (46a), we deduce that either $x_1^{(1)} = 0$ and $|x_2^{(1)}| + \sqrt{x_1^2 + (x_2 - x_2^{(1)})^2} = 1$, or $x_2^{(1)} = 0$ and $|x_1^{(1)}| + \sqrt{(x_1 - x_1^{(1)})^2 + x_2^2} = 1$, that is, after calculations, either $x_1^{(1)} = 0$ and $|x_2^{(1)}| = x_2 \times x_2^{(1)}$, or $x_2^{(1)} = 0$ and $|x_1^{(1)}| = x_1 \times x_1^{(1)}$. Therefore, we have the two following subcases.

- (a) If $x \notin \{(1, 0), (0, 1), (-1, 0), (0, -1)\}$, then necessarily $(x^{(1)}, x^{(2)}) = (0, x)$, that is, $C(x) = \{(0, x)\}$. As a consequence, $\mathcal{L}_0(x) = \|0\|_{(1)}^{\text{sn}} + 2\|x\|_{(2)}^{\text{sn}} = 2\|x\| = 2$ by (45a).
- (b) If $x \in \{(1, 0), (0, 1), (-1, 0), (0, -1)\}$, it is easy to check that $(x, 0) \in C(x)$ by (45b). Therefore, $\mathcal{L}_0(x) \leq \|x\|_{(1)}^{\text{sn}} + 2\|0\|_{(2)}^{\text{sn}} = 1$ by (45a). Now, for any $(x^{(1)}, x^{(2)}) \in C(x)$, we have that

$$\|x^{(1)}\|_{(1)}^{\text{sn}} + 2\|x^{(2)}\|_{(2)}^{\text{sn}} \geq \|x^{(1)}\|_{(1)}^{\text{sn}} + \|x^{(2)}\|_{(2)}^{\text{sn}} \geq \|x\| = 1$$

by (46a). To conclude, we obtain that $1 \leq \mathcal{L}_0(x)$ by (45a), hence that $\mathcal{L}_0(x) = 1$.

3. Suppose that $\|x\| = \sqrt{x_1^2 + x_2^2} < 1$. Then, the proof is an application of Proposition 19 in Appendix A, combined with the formula (37).

This ends the proof. □

5 Conclusion

In this paper, we have introduced a novel class of one-sided linear couplings, and have shown that they induce conjugacies that share nice properties with the classic Fenchel conjugacy. Among them, we have distinguished a novel coupling, Caprac, having the property of being constant along primal rays, like the ℓ_0 pseudonorm. For the Caprac conjugacy, induced by the coupling Caprac, we have proved that the ℓ_0 pseudonorm is equal to its biconjugate: hence, the ℓ_0 pseudonorm is Caprac-convex in the sense of generalized convexity. We have also provided expressions for conjugates in terms of two families of dual norms, the $2-k$ -symmetric gauge norms and the k -support norms. Finally, we have shown that the ℓ_0 pseudonorm displays “hidden convexity” as we have proved that it coincides, on the Euclidian sphere, with a proper convex lsc function. This is somewhat surprising as the ℓ_0 pseudonorm is a highly nonconvex function of combinatorial nature.

In a companion paper [4], we apply our results to so-called exact sparse optimization, that is, problems where one looks for solution that have few nonzero components. We provide a systematic way to obtain convex minimization programs (over unit balls of some norms) that are lower bounds for the original exact sparse optimization problem.

Acknowledgements. We want to thank Jean-Baptiste Hiriart-Urruty, Frédéric Meunier, Danil Kadnikov and Pierre Carpentier for discussions on first versions of this work.

A Appendix

A.1 Background on J. J. Moreau lower and upper additions

When we manipulate functions with values in $\overline{\mathbb{R}} = [-\infty, +\infty]$, we adopt the following Moreau *lower addition* or *upper addition*, depending on whether we deal with sup or inf operations. We follow [9]. In the sequel, u , v and w are any elements of $\overline{\mathbb{R}}$.

Moreau lower addition

The Moreau *lower addition* extends the usual addition with

$$(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = -\infty . \quad (47a)$$

With the *lower addition*, $(\overline{\mathbb{R}}, \dot{+})$ is a convex cone, with $\dot{+}$ commutative and associative. The lower addition displays the following properties:

$$u \leq u' , \quad v \leq v' \Rightarrow u \dot{+} v \leq u' \dot{+} v' , \quad (47b)$$

$$(-u) \dot{+} (-v) \leq -(u \dot{+} v) , \quad (47c)$$

$$(-u) \dot{+} u \leq 0 , \quad (47d)$$

$$\sup_{a \in \mathbb{A}} f(a) \dot{+} \sup_{b \in \mathbb{B}} g(b) = \sup_{a \in \mathbb{A}, b \in \mathbb{B}} (f(a) \dot{+} g(b)) , \quad (47e)$$

$$\inf_{a \in \mathbb{A}} f(a) \dot{+} \inf_{b \in \mathbb{B}} g(b) \leq \inf_{a \in \mathbb{A}, b \in \mathbb{B}} (f(a) \dot{+} g(b)) , \quad (47f)$$

$$t < +\infty \Rightarrow \inf_{a \in \mathbb{A}} f(a) \dot{+} t = \inf_{a \in \mathbb{A}} (f(a) \dot{+} t) . \quad (47g)$$

Moreau upper addition

The Moreau *upper addition* extends the usual addition with

$$(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = +\infty . \quad (48a)$$

With the *upper addition*, $(\overline{\mathbb{R}}, \dot{+})$ is a convex cone, with $\dot{+}$ commutative and associative. The upper addition displays the following properties:

$$u \leq u' , \quad v \leq v' \Rightarrow u \dot{+} v \leq u' \dot{+} v' , \quad (48b)$$

$$(-u) \dot{+} (-v) \geq -(u \dot{+} v) , \quad (48c)$$

$$(-u) \dot{+} u \geq 0 , \quad (48d)$$

$$\inf_{a \in \mathbb{A}} f(a) \dot{+} \inf_{b \in \mathbb{B}} g(b) = \inf_{a \in \mathbb{A}, b \in \mathbb{B}} (f(a) \dot{+} g(b)) , \quad (48e)$$

$$\sup_{a \in \mathbb{A}} f(a) \dot{+} \sup_{b \in \mathbb{B}} g(b) \geq \sup_{a \in \mathbb{A}, b \in \mathbb{B}} (f(a) \dot{+} g(b)) , \quad (48f)$$

$$-\infty < t \Rightarrow \sup_{a \in \mathbb{A}} f(a) \dot{+} t = \sup_{a \in \mathbb{A}} (f(a) \dot{+} t) . \quad (48g)$$

Joint properties of the Moreau lower and upper addition

We obviously have that

$$u \dot{+} v \leq u \dot{+} v . \quad (49a)$$

The Moreau lower and upper additions are related by

$$-(u \dot{+} v) = (-u) \dot{+} (-v) , \quad -(u \dot{+} v) = (-u) \dot{+} (-v) . \quad (49b)$$

They satisfy the inequality

$$(u \dot{+} v) \dot{+} w \leq u \dot{+} (v \dot{+} w) . \quad (49c)$$

with

$$(u \dot{+} v) \dot{+} w < u \dot{+} (v \dot{+} w) \iff \begin{cases} u = +\infty \text{ and } w = -\infty , \\ \text{or} \\ u = -\infty \text{ and } w = +\infty \text{ and } -\infty < v < +\infty . \end{cases} \quad (49d)$$

Finally, we have that

$$u \dot{+} (-v) \leq 0 \iff u \leq v \iff 0 \leq v \dot{+} (-u) , \quad (49e)$$

$$u \dot{+} (-v) \leq w \iff u \leq v \dot{+} w \iff u \dot{+} (-w) \leq v , \quad (49f)$$

$$w \leq v \dot{+} (-u) \iff u \dot{+} w \leq v \iff u \leq v \dot{+} (-w) . \quad (49g)$$

A.2 Properties of 2- k -symmetric gauge norms and of k -support norms

Before studying properties of 2- k -symmetric gauge norms and of k -support norms, we recall the notion of dual norm [1].

A.2.1 Dual norm

Let \mathbb{X} and \mathbb{Y} be two vector spaces equipped with a bilinear form $\langle \cdot, \cdot \rangle$, and suppose that \mathbb{X} is equipped with a norm $\|\cdot\|$ with unit ball denoted by

$$B_{\|\cdot\|} = \{x \in \mathbb{X} \mid \|x\| \leq 1\} . \quad (50)$$

Definition 12 *The following expression*

$$\|y\|_{\star} = \sup_{\|x\| \leq 1} \langle x, y \rangle , \quad \forall y \in \mathbb{Y} \quad (51)$$

defines a norm on \mathbb{Y} , called the dual norm $\|\cdot\|_{\star}$.

We have

$$\|\cdot\|_{\star} = \sigma_{B_{\|\cdot\|}} \quad \text{and} \quad \|\cdot\| = \sigma_{B_{\|\cdot\|_{\star}}} , \quad (52)$$

where $B_{\|\cdot\|_{\star}}$, the unit ball of the dual norm, is the polar set $B_{\|\cdot\|}^{\circ}$ of the unit ball $B_{\|\cdot\|}$:

$$B_{\|\cdot\|_{\star}} = \{y \in \mathbb{Y} \mid \|y\|_{\star} \leq 1\} = B_{\|\cdot\|}^{\circ} = \{y \in \mathbb{Y} \mid \langle x, y \rangle \leq 1, \forall x \in B_{\|\cdot\|}\} . \quad (53)$$

A.2.2 Properties of 2- k -symmetric gauge norms

For all $K \subset \{1, \dots, d\}$, we introduce *degenerate* unit “spheres” and “balls” of \mathbb{R}^d equipped with the eucliden norm $\|\cdot\|$ by

$$\mathbb{S}_K = \{x \in \mathbb{R}^d \mid \|x_K\| = 1\} , \quad (54a)$$

$$\mathbb{B}_K = \{x \in \mathbb{R}^d \mid \|x_K\| \leq 1\} , \quad (54b)$$

$$S_K = \{x \in \mathbb{R}^d \mid x_{-K} = 0 \text{ and } \|x_K\| = 1\} , \quad (54c)$$

$$B_K = \{x \in \mathbb{R}^d \mid x_{-K} = 0 \text{ and } \|x_K\| \leq 1\} , \quad (54d)$$

where x_K has been defined right before Definition 7.

In what follows, the notation $\bigcup_{|K|\leq k}$ is a shorthand for $\bigcup_{K\subset\{1,\dots,d\},|K|\leq k}$, $\bigcap_{|K|\leq k}$ for $\bigcap_{K\subset\{1,\dots,d\},|K|\leq k}$, and $\sup_{|K|\leq k}$ for $\sup_{K\subset\{1,\dots,d\},|K|\leq k}$. The same holds true for $\bigcup_{|K|=k}$, $\bigcap_{|K|=k}$ and $\sup_{|K|=k}$.

The 2- k -symmetric gauge norm $\|\cdot\|_{(k)}^{\text{sgn}}$ has been introduced in Definition 7.

Proposition 13 *Let $k \in \{1, \dots, d\}$.*

- *The following inequalities hold true*

$$\sup_{j=1,\dots,d} |x_j| = \|x\|_\infty = \|x\|_{(1)}^{\text{sgn}} \leq \dots \leq \|x\|_{(l)}^{\text{sgn}} \leq \|x\|_{(l+1)}^{\text{sgn}} \leq \dots \leq \|x\|_{(d)}^{\text{sgn}} = \|x\| . \quad (55)$$

- *The two “spheres” in (54a) and (54c) are related by*

$$S_K = S \cap \mathbb{S}_K , \quad \forall K \subset \{1, \dots, d\} . \quad (56)$$

- *The 2- k -symmetric gauge norm $\|\cdot\|_{(k)}^{\text{sgn}}$ satisfies*

$$\|\cdot\|_{(k)}^{\text{sgn}} = \sigma_{\bigcup_{|K|\leq k} B_K} = \sup_{|K|\leq k} \sigma_{B_K} = \sup_{|K|\leq k} \sigma_{S_K} = \sigma_{\bigcup_{|K|\leq k} S_K} , \quad (57)$$

where $|K| \leq k$ can be replaced by $|K| = k$ everywhere.

- *The unit sphere $S_{(k)}^{\text{sgn}}$ and ball $B_{(k)}^{\text{sgn}}$ of \mathbb{R}^d for the 2- k -symmetric gauge norm $\|\cdot\|_{(k)}^{\text{sgn}}$ in Definition 7 satisfy*

$$B_{(k)}^{\text{sgn}} = \left\{ x \in \mathbb{R}^d \mid \|x\|_{(k)}^{\text{sgn}} \leq 1 \right\} = \bigcap_{|K|\leq k} \mathbb{B}_K , \quad (58a)$$

$$S_{(k)}^{\text{sgn}} = \left\{ x \in \mathbb{R}^d \mid \|x\|_{(k)}^{\text{sgn}} = 1 \right\} = B_{(k)}^{\text{sgn}} \cap \left(\bigcup_{|K|\leq k} \mathbb{S}_K \right) , \quad (58b)$$

where $|K| \leq k$ can be replaced by $|K| = k$ everywhere.

- *The units balls $B_{(l)}^{\text{sgn}}$ satisfy the inclusions*

$$B = B_{(d)}^{\text{sgn}} \subset \dots \subset B_{(l+1)}^{\text{sgn}} \subset B_{(l)}^{\text{sgn}} \subset \dots \subset B_{(1)}^{\text{sgn}} . \quad (59)$$

- *The units balls $B_{(l)}^{\text{sgn}}$ satisfy the relation*

$$B_{(1)}^{\text{sgn}} \subset \bigcap_{l=2}^d \sqrt{l} B_{(l)}^{\text{sgn}} . \quad (60)$$

Proof.

• The Inequalities and Equalities (55) easily derive from the very definition (29) of the 2- k -symmetric gauge norm $\|\cdot\|_{(k)}^{\text{sgn}}$.

• We prove Equation (56). Recall that, following notation from Game Theory, we denote by $-K$ the complementary subset of K in $\{1, \dots, d\}$: $K \cup (-K) = \{1, \dots, d\}$ and $K \cap (-K) = \emptyset$. Then, we have that $x = x_K + x_{-K}$, for any $x \in \mathbb{R}^d$, and the decomposition is orthogonal, leading to

$$(\forall x \in \mathbb{R}^d) \quad x = x_K + x_{-K}, \quad x_K \perp x_{-K} \quad \text{and} \quad \|x\|^2 = \|x_K\|^2 + \|x_{-K}\|^2. \quad (61)$$

For $K \subset \{1, \dots, d\}$, we have that

$$\begin{aligned} x \in S \quad \text{and} \quad x \in \mathbb{S}_K &\iff 1 = \|x\|^2 \quad \text{and} \quad 1 = \|x_K\|^2 && \text{(by (54a))} \\ &\iff 1 = \|x\|^2 = \|x_K\|^2 + \|x_{-K}\|^2 \quad \text{and} \quad 1 = \|x_K\|^2 && \text{(by (61))} \\ &\iff \|x_{-K}\| = 0 \quad \text{and} \quad 1 = \|x_K\| && \text{(by (61))} \\ &\iff x \in S_K. && \text{(by (54c))} \end{aligned}$$

• We prove Equation (57). For this purpose, we first establish that

$$\sigma_{B_K}(y) = \|y_K\|, \quad \forall y \in \mathbb{R}^d. \quad (62)$$

Indeed, for $y \in \mathbb{R}^d$, we have

$$\begin{aligned} \sigma_{B_K}(y) &= \sup_{x \in B_K} \langle x, y \rangle && \text{(by definition (17) of a support function)} \\ &= \sup_{x \in B_K} \langle x_K + x_{-K}, y_K + y_{-K} \rangle && \text{(by the decomposition (61))} \\ &= \sup_{x \in B_K} (\langle x_K, y_K \rangle + \langle x_{-K}, y_{-K} \rangle) && \text{(because } x_K \perp y_{-K} \text{ and } x_{-K} \perp y_K \text{ by (61))} \\ &= \sup \{ \langle x_K, y_K \rangle + \langle x_{-K}, y_{-K} \rangle \mid x_{-K} = 0 \text{ and } \|x_K\| \leq 1 \} && \text{(by definition (54d) of } B_K) \\ &= \sup \{ \langle x_K, y_K \rangle \mid \|x_K\| \leq 1 \} \\ &= \|y_K\| \end{aligned}$$

as is well-known for the Euclidian norm $\|\cdot\|$, when restricted to the subspace $\{x \in \mathbb{R}^d \mid x_{-K} = 0\}$ (because it is equal to its dual norm). Then, for all $y \in \mathbb{R}^d$, we have that

$$\begin{aligned} \sigma_{\cup_{|K| \leq k} B_K}(y) &= \sup_{|K| \leq k} \sigma_{B_K}(y) && \text{(as is well-known in convex analysis)} \\ &= \sup_{|K| \leq k} \|y_K\| && \text{(by (62))} \\ &= \|y\|_{(k)}^{\text{sgn}}. && \text{(by definition (29) of } \|\cdot\|_{(k)}^{\text{sgn}}) \end{aligned}$$

Now, by (54c) and (54d), it is straightforward that $\overline{\text{co}}(S_K) = B_K$ and we deduce that

$$\|\cdot\|_{(k)}^{\text{sgn}} = \sigma_{\cup_{|K| \leq k} B_K} = \sup_{|K| \leq k} \sigma_{B_K} = \sup_{|K| \leq k} \sigma_{\overline{\text{co}}(S_K)} = \sup_{|K| \leq k} \sigma_{S_K} = \sigma_{\cup_{|K| \leq k} S_K},$$

giving Equation (57).

If we take over the proof using the property that $\sup_{|K| \leq k} \|y_K\| = \sup_{|K|=k} \|y_K\|$ in (29), we deduce that $|K| \leq k$ can be replaced by $|K| = k$ everywhere.

- We prove Equation (58a):

$$\begin{aligned}
B_{(k)}^{\text{sgn}} &= \left\{ x \in \mathbb{R}^d \mid \|x\|_{(k)}^{\text{sgn}} \leq 1 \right\} && \text{(by definition of the ball } B_{(k)}^{\text{sgn}}) \\
&= \left\{ x \in \mathbb{R}^d \mid \sup_{|K| \leq k} \|x_K\| \leq 1 \right\} && \text{(by definition (29) of } \|\cdot\|_{(k)}^{\text{sgn}}) \\
&= \bigcap_{|K| \leq k} \left\{ x \in \mathbb{R}^d \mid \|x_K\| \leq 1 \right\} \\
&= \bigcap_{|K| \leq k} \mathbb{B}_K. && \text{(by definition (54b) of } \mathbb{B}_K)
\end{aligned}$$

If we take over the proof using the property that $\sup_{|K| \leq k} \|y_K\| = \sup_{|K|=k} \|y_K\|$ in (29), we deduce that $|K| \leq k$ can be replaced by $|K| = k$ everywhere.

- We prove Equation (58b):

$$\begin{aligned}
S_{(k)}^{\text{sgn}} &= \left\{ x \in \mathbb{R}^d \mid \|x\|_{(k)}^{\text{sgn}} = 1 \right\} && \text{(by definition of the sphere } S_{(k)}^{\text{sgn}}) \\
&= \left\{ x \in \mathbb{R}^d \mid \sup_{|K| \leq k} \|x_K\| = 1 \right\} && \text{(by definition (29) of } \|\cdot\|_{(k)}^{\text{sgn}}) \\
&= \left\{ x \in \mathbb{R}^d \mid \sup_{|K| \leq k} \|x_K\| \leq 1 \right\} \\
&\quad \cap \left\{ x \in \mathbb{R}^d \mid \exists K \subset \{1, \dots, d\}, |K| \leq k, \|x_K\| = 1 \right\} \\
&= B_{(k)}^{\text{sgn}} \cap \left(\bigcup_{|K| \leq k} \left\{ x \in \mathbb{R}^d \mid \|x_K\| = 1 \right\} \right) && \text{(by definition of the ball } B_{(k)}^{\text{sgn}}) \\
&= B_{(k)}^{\text{sgn}} \cap \left(\bigcup_{|K| \leq k} \mathbb{S}_K \right). && \text{(by definition (54a) of } \mathbb{S}_K)
\end{aligned}$$

If we take over the proof using the property that $\sup_{|K| \leq k} \|y_K\| = \sup_{|K|=k} \|y_K\|$ in (29), we deduce that $|K| \leq k$ can be replaced by $|K| = k$ everywhere.

- The inclusions (59) directly follow from the Inequalities and Equalities (55).
- The relation (60) easily follows from the property that

$$\|y\|_{(k)}^{\text{sgn}} \leq \sqrt{k} \|y\|_{(1)}^{\text{sgn}}, \quad \forall y \in \mathbb{R}^d, \quad \forall k = 1, \dots, d. \quad (63)$$

Indeed, by definition (29) of $\|\cdot\|_{(k)}^{\text{sgn}}$, for a given $y \in \mathbb{R}^d$, there exists $K \subset \{1, \dots, d\}$ with $|K| = k$ such that $(\|y\|_{(k)}^{\text{sgn}})^2 = \sum_{k \in K} |y_k|^2 \leq \sum_{k \in K} (\|y\|_{(1)}^{\text{sgn}})^2 \leq k (\|y\|_{(1)}^{\text{sgn}})^2$. \square

A.2.3 Properties of k -support norms

The k -support norm $\|\cdot\|_{(k)}^{\text{sn}}$ has been introduced in Definition 7 as the dual norm of the 2 - k -symmetric gauge norm $\|\cdot\|_{(k)}^{\text{sgn}}$.

Proposition 14 *Let $k \in \{1, \dots, d\}$.*

- *The unit balls $B_{(l)}^{\text{sn}}$ satisfy the inclusions*

$$B_{(1)}^{\text{sn}} \subset \dots \subset B_{(l)}^{\text{sn}} \subset B_{(l+1)}^{\text{sn}} \subset \dots \subset B_{(d)}^{\text{sn}} = B. \quad (64)$$

- *The following inequalities hold true*

$$\|x\| = \|x\|_{(d)}^{\text{sn}} \leq \dots \leq \|x\|_{(l+1)}^{\text{sn}} \leq \|x\|_{(l)}^{\text{sn}} \leq \dots \leq \|x\|_{(1)}^{\text{sn}} = \sum_{j=1}^d |x_j|. \quad (65)$$

- *The unit ball $B_{(k)}^{\text{sn}}$ of the k -support norm $\|\cdot\|_{(k)}^{\text{sn}}$ satisfies*

$$B_{(k)}^{\text{sn}} = \{x \in \mathbb{R}^d \mid \|x\|_{(k)}^{\text{sn}} \leq 1\} = \overline{\text{co}}\left(\bigcup_{|K| \leq k} B_K\right) = \overline{\text{co}}\left(\bigcup_{|K| \leq k} S_K\right), \quad (66)$$

where $|K| \leq k$ can be replaced by $|K| = k$ everywhere.

- *For $l = 0, 1, \dots, d$, we have*

$$\|\cdot\|_{(l)}^{\text{sgn}} = \sigma_{B_{(l)}^{\text{sn}}}. \quad (67)$$

Proof.

- The inclusions (64) directly follow from the inclusions (59) and from (53) as $B_{(k)}^{\text{sn}} = (B_{(k)}^{\text{sgn}})^{\odot}$.
- The Inequalities in (65) easily derive from the inclusions (64). The Equalities in (65) are well-known.
- We prove Equation (66). On the one hand, by the first relation in (52), we have that $\|\cdot\|_{(k)}^{\text{sgn}} = \sigma_{B_{(k)}^{\text{sn}}}$. On the other hand, by (57), we have that $\|\cdot\|_{(k)}^{\text{sgn}} = \sigma_{\bigcup_{|K| \leq k} B_K} = \sigma_{\bigcup_{|K| \leq k} S_K}$. Then, as is well-known in convex analysis, we deduce that $\overline{\text{co}}(B_{(k)}^{\text{sn}}) = \overline{\text{co}}(\bigcup_{|K| \leq k} B_K) = \overline{\text{co}}(\bigcup_{|K| \leq k} S_K)$. As the unit ball $B_{(k)}^{\text{sn}}$ is closed and convex, we immediately obtain (66).

If we take over the proof using the property that $\sigma_{\bigcup_{|K| \leq k} B_K} = \sigma_{\bigcup_{|K| \leq k} S_K} = \sigma_{\bigcup_{|K|=k} B_K} = \sigma_{\bigcup_{|K|=k} S_K}$ in (57), we deduce that $|K| \leq k$ can be replaced by $|K| = k$ everywhere.

- We prove Equation (67). By Definition 7, the l -support norm is the dual norm of the 2 - l -symmetric gauge norm. Therefore, the 2 - l -symmetric gauge norm is the dual norm of the l -support norm and (67) follows from (52) for $l = 1, \dots, d$. For $l = 0$, both conventions $\|\cdot\|_{(0)}^{\text{sgn}} = 0$ and $\{0\} = B_{(0)}^{\text{sn}}$ lead to $\|\cdot\|_{(0)}^{\text{sgn}} = 0 = \sigma_0 = \sigma_{B_{(0)}^{\text{sn}}}$.

□

A.3 Properties of the level sets of the ℓ_0 pseudonorm

A connection between the ℓ_0 pseudonorm in (27) and the 2- k -symmetric gauge norm $\|\cdot\|_{(k)}^{\text{sgn}}$ in (29) is given by the (easily proved) following Proposition.

Proposition 15 *Let $k \in \{0, 1, \dots, d\}$. For any $x \in \mathbb{R}^d$, we have*

$$\ell_0(x) = k \iff 0 = \|x\|_{(0)}^{\text{sgn}} \leq \dots \leq \|x\|_{(k-1)}^{\text{sgn}} < \|x\|_{(k)}^{\text{sgn}} = \dots = \|x\|_{(d)}^{\text{sgn}} = \|x\|, \quad (68)$$

from which we deduce the formula

$$\ell_0(x) = \min \left\{ j \in \{0, 1, \dots, d\} \mid \|x\|_{(j)}^{\text{sgn}} = \|x\| \right\}, \quad (69)$$

with the convention that $\|\cdot\|_{(0)}^{\text{sgn}} = 0$.

We prove the following Proposition about the level sets of the ℓ_0 pseudonorm.

Proposition 16 *Let $k \in \{0, 1, \dots, d\}$. The level set $\ell_0^{\leq k}$ in (31a) of the ℓ_0 pseudonorm in (27) satisfies*

$$(\forall x \in \mathbb{R}^d) \quad x \in \ell_0^{\leq k} \iff \ell_0(x) \leq k \iff \|x\|_{(k)}^{\text{sgn}} = \|x\|, \quad (70a)$$

$$(\forall x \in \mathbb{R}^d) \quad x \in \ell_0^{\leq k} \setminus \{0\} \iff 0 < \ell_0(x) \leq k \iff x \neq 0 \text{ and } \frac{x}{\|x\|} \in S \cap S_{(k)}^{\text{sgn}}, \quad (70b)$$

and its intersection with the sphere S has the three following expressions

$$S \cap \ell_0^{\leq k} = \bigcup_{|K| \leq k} S_K = \bigcup_{|K|=k} S_K, \quad (71a)$$

$$S \cap \ell_0^{\leq k} = S \cap B_{(k)}^{\text{sn}}, \quad (71b)$$

$$S \cap \ell_0^{\leq k} = \overline{S \cap \ell_0^{\leq k}}. \quad (71c)$$

Proof.

- The Equivalence (70a) easily follows from (68).
- We prove the Equivalence (70b). Indeed, using Equation (70a) we have that, for $x \in \mathbb{R}^d \setminus \{0\}$:

$$\ell_0(x) \leq k \iff \|x\|_{(k)}^{\text{sgn}} = \|x\| \iff \left\| \frac{x}{\|x\|} \right\|_{(k)}^{\text{sgn}} = 1 \iff \frac{x}{\|x\|} \in S_{(k)}^{\text{sgn}} \iff \frac{x}{\|x\|} \in S \cap S_{(k)}^{\text{sgn}}.$$

- We prove Equation (71a):

$$\begin{aligned}
S \cap \ell_0^{\leq k} &= \left\{ x \in \mathbb{R}^d \mid \|x\| = 1 \text{ and } \ell_0(x) \leq k \right\} && \text{(by definitions (25) of } S \text{ and (31b) of } \ell_0^{\leq k} \text{)} \\
&= \left\{ x \in \mathbb{R}^d \mid \|x\| = 1 \text{ and } \|x\|_{(k)}^{\text{sgn}} = \|x\| \right\} && \text{(by (70a))} \\
&= \left\{ x \in \mathbb{R}^d \mid \|x\| = 1 \text{ and } \|x\|_{(k)}^{\text{sgn}} = 1 \right\} \\
&= S \cap S_{(k)}^{\text{sgn}} && \text{(by definitions (25) and (58b) of the spheres } S \text{ and } S_{(k)}^{\text{sgn}} \text{)} \\
&= S \cap B_{(k)}^{\text{sgn}} \cap \left(\bigcup_{|K| \leq k} \mathbb{S}_K \right) && \text{(by property (58b) of the sphere } S_{(k)}^{\text{sgn}} \text{)} \\
&= S \cap \left(\bigcup_{|K| \leq k} \mathbb{S}_K \right) && \text{(as, by (59), we have that } S \subset B \subset B_{(k)}^{\text{sgn}} \text{)} \\
&= \bigcup_{|K| \leq k} (S \cap \mathbb{S}_K) \\
&= \bigcup_{|K| \leq k} S_K. && \text{(as } S \cap \mathbb{S}_K = S_K \text{ by (56))}
\end{aligned}$$

If we take over the proof where we use $S_{(k)}^{\text{sgn}} = B_{(k)}^{\text{sgn}} \cap \left(\bigcup_{|K|=k} \mathbb{S}_K \right)$ in (58b), we obtain that $S \cap \ell_0^{\leq k} = \bigcup_{|K|=k} S_K$.

- We prove Equation (71b). First, we observe that the level set $\ell_0^{\leq k}$ is closed because, by (70a), it can be expressed as $\ell_0^{\leq k} = \left\{ x \in \mathbb{R}^d \mid \|x\|_{(k)}^{\text{sgn}} = \|x\| \right\}$. This also follows from the well-known property that the pseudonorm ℓ_0 is lower semi continuous. Second, we have

$$\begin{aligned}
S \cap \ell_0^{\leq k} &= S \cap \overline{\text{co}}(S \cap \ell_0^{\leq k}) && \text{(by Lemma 17 since } S \cap \ell_0^{\leq k} \subset S \text{ and is closed)} \\
&= S \cap \overline{\text{co}}\left(\bigcup_{|K| \leq k} S_K \right) && \text{(as } S \cap \ell_0^{\leq k} = \bigcup_{|K| \leq k} S_K \text{ by (71a))} \\
&= S \cap B_{(k)}^{\text{sn}}. && \text{(as } \overline{\text{co}}\left(\bigcup_{|K| \leq k} S_K \right) = B_{(k)}^{\text{sn}} \text{ by (66))}
\end{aligned}$$

- We prove Equation (71c). For this purpose, we first 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. Indeed, as we have seen that $\ell_0^{\leq k}$ is closed, we have $\ell_0^{\leq k} \subset \overline{\ell_0^{\leq k}} \Rightarrow \overline{\ell_0^{\leq k}} \subset \ell_0^{\leq k} = \ell_0^{\leq k}$. 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 $\ell_0(x)$, 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 $S \cap \ell_0^{\leq k} = \overline{S \cap \ell_0^{\leq k}}$. The inclusion $\overline{S \cap \ell_0^{\leq k}} \subset S \cap \ell_0^{\leq k}$, is easy. Indeed, $\overline{\ell_0^{\leq k}} = \ell_0^{\leq k} \Rightarrow \overline{S \cap \ell_0^{\leq k}} \subset \overline{S} \cap \overline{\ell_0^{\leq k}} = S \cap \ell_0^{\leq k}$. To prove the reverse inclusion $S \cap \ell_0^{\leq k} \subset \overline{S \cap \ell_0^{\leq k}}$, we consider $x \in S \cap \ell_0^{\leq k}$. 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 $x \in S$, 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 $x \in S = \{x \in \mathbb{X} \mid \|x\| = 1\}$. 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, $z_n/\|z_n\| \in S$. As a consequence $z_n/\|z_n\| \in S \cap \ell_0^{\leq k}$, and we conclude that $x \in \overline{S \cap \ell_0^{\leq k}}$. Thus, we have proved that $S \cap \ell_0^{\leq k} \subset \overline{S \cap \ell_0^{\leq k}}$.

This ends the proof. \square

Lemma 17 *If A is a subset of the Euclidian sphere S of \mathbb{R}^d , then $A = \text{co}(A) \cap S$. If A is a closed subset of the Euclidian sphere S of \mathbb{R}^d , then $A = \overline{\text{co}(A)} \cap S$.*

Proof. We first prove that $A = \text{co}(A) \cap S$ when $A \subset S$. Since $A \subset \text{co}(A)$ and $A \subset S$, we immediately get that $A \subset \text{co}(A) \cap S$. To prove the reverse inclusion, we first start by proving that $\text{co}(A) \cap S \subset \text{extr}(\text{co}(A))$, the set of extreme points of $\text{co}(A)$.

The proof is by contradiction. Suppose indeed that there exists $x \in \text{co}(A) \cap S$ and $x \notin \text{extr}(\text{co}(A))$. Then, we could find $y \in \text{co}(A)$ and $z \in \text{co}(A)$, distinct from x , and such that $x = \lambda y + (1 - \lambda)z$ for some $\lambda \in (0, 1)$. Notice that necessarily $y \neq z$ (because, else, we would have $x = y = z$ which would contradict $y \neq x$ and $z \neq x$). By assumption $A \subset S$, we deduce that $\text{co}(A) \subset B = \{x \in \mathbb{X} \mid \|x\| \leq 1\}$, the unit ball, and therefore that $\|y\| \leq 1$ and $\|z\| \leq 1$. If y or z were not in S — that is, if either $\|y\| < 1$ or $\|z\| < 1$ — then we would obtain that $\|x\| \leq \lambda\|y\| + (1 - \lambda)\|z\| < 1$ since $\lambda \in (0, 1)$; we would thus arrive at a contradiction since x could not be in S . Thus, both y and z must be in S , and we have a contradiction since no $x \in S$, the Euclidian sphere, can be obtained as a convex combination of $y \in S$ and $z \in S$, with $y \neq z$.

Hence, we have proved by contradiction that $\text{co}(A) \cap S \subset \text{extr}(\text{co}(A))$. We can conclude using the fact that $\text{extr}(\text{co}(A)) \subset A$ (see [5, Exercice 6.4]).

Now, we consider the case where the subset A of the Euclidian sphere S is closed. Using the first part of the proof we have that $A = \text{co}(A) \cap S$. Now, A is closed by assumption and bounded since $A \subset S$. Thus, A is compact and, in a finite dimensional space, we have that $\text{co}(A)$ is compact [12, Th. 17.2], thus closed. We conclude that $A = \text{co}(A) \cap S = \overline{\text{co}(A)} \cap S = \overline{\text{co}(A)} \cap S$, where the last equality comes from [3, Prop. 3.46]. \square

A.4 Additional results on the function \mathcal{L}_0

In Proposition 10, we have provided an expression, for the proper convex lsc function \mathcal{L}_0 in Corollary 9, as the value of the minimization problem (41). Here, we provide a characterization of the optimal solutions of (41).

In the sequel, we will use the following relations regarding faces of unit balls:

$$F_{B_{(l)}^{\text{sgn}}}(0) = B_{(l)}^{\text{sgn}}, \quad \forall l = 1, \dots, d, \quad (72a)$$

$$F_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)}) \subset S_{(l)}^{\text{sgn}}, \quad \text{if } \bar{x}^{(l)} \neq 0, \quad (72b)$$

$$F_{B_{(d)}^{\text{sgn}}}(\bar{x}^{(d)}) = \left\{ \frac{\bar{x}^{(d)}}{\|\bar{x}^{(d)}\|} \right\}, \quad \text{if } \bar{x}^{(d)} \neq 0. \quad (72c)$$

We recall that the *exposed face of the (nonempty) closed convex set* $C \subset \mathbb{R}^d$ at $y \in \mathbb{R}^d$ is [7, p.220]

$$F_C(y) = \{x \in C \mid \langle x, y \rangle = \sigma_C(y)\} = \operatorname{argmax}_{x \in C} \langle x, y \rangle . \quad (73)$$

The boundary of the closed convex set $C \subset \mathbb{R}^d$ is the union of all faces $F_C(y)$, where $y \neq 0$, [7, p.220].

The following Proposition is a characterization of the optimal solutions of (41) in any dimension d .

Proposition 18 *Let $x \in \mathbb{R}^d$ be such that $\|x\| < 1$. The sequence $(\bar{x}^{(1)}, \dots, \bar{x}^{(d)})$ of vectors of \mathbb{R}^d is solution of the minimization problem (41) or, equivalently, of the minimization problem*

$$\min_{\substack{x^{(1)} \in \mathbb{R}^d, \dots, x^{(d)} \in \mathbb{R}^d \\ \sum_{l=1}^d \sigma_{B_{(l)}^{\text{sgn}}}(x^{(l)}) \leq 1 \\ \sum_{l=1}^d x^{(l)} = x}} \sum_{l=1}^d l \sigma_{B_{(l)}^{\text{sgn}}}(x^{(l)}) \quad (74)$$

if and only if

1. either $\sum_{l=1}^d |x_l| = \|x\|_{(1)}^{\text{sn}} \leq 1$ and $(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = (x, 0, \dots, 0)$ (and then, the minimum in (74) or (41) is equal to $\|x\|_{(1)}^{\text{sn}}$),
2. or there exists $\lambda > 0$ such that

$$\bigcap_{l=1}^d (l + \lambda) F_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)}) \neq \emptyset , \quad (75a)$$

$$\sum_{l=1}^d \sigma_{B_{(l)}^{\text{sgn}}}(x^{(l)}) = 1 , \quad (75b)$$

$$\sum_{l=1}^d \bar{x}^{(l)} = x . \quad (75c)$$

Proof. The minimization problems (41) and (74) are the same because $\sigma_{B_{(l)}^{\text{sgn}}}(\cdot) = \|\cdot\|_{(k)}^{\text{sn}}$ since the k -support norm is the dual norm (see Definition 12) of the $2-k$ -symmetric gauge norm (see Definition 7). First, we establish necessary and sufficient Karush-Kuhn-Tucker (KKT) conditions for the optimization problem (74).

The optimization problem (74) is the minimization of the proper convex lsc function

$$f_0(x^{(1)}, \dots, x^{(d)}) = \sum_{l=1}^d l \sigma_{B_{(l)}^{\text{sgn}}}(x^{(l)}) \quad (76a)$$

over a convex domain of $(\mathbb{R}^d)^d$ defined by one scalar inequality constraint, $f_1(x^{(1)}, \dots, x^{(d)}) \leq 0$, represented by the proper convex lsc function

$$f_1(x^{(1)}, \dots, x^{(d)}) = \sum_{l=1}^d \sigma_{B_{(l)}^{\text{sgn}}}(x^{(l)}) - 1, \quad (76b)$$

and d equality constraints, $f_{1+k}(x^{(1)}, \dots, x^{(d)}) = 0$ for $k = 1, \dots, d$, represented by the d affine functions

$$f_{1+k}(x^{(1)}, \dots, x^{(d)}) = \left\langle \sum_{l=1}^d \bar{x}^{(l)} - x, e_k \right\rangle, \quad k = 1, \dots, d, \quad (76c)$$

where e_k is the k -canonical vector of \mathbb{R}^d . It should be noted that all the functions $f_0, f_1, f_2, \dots, f_{1+d}$ are proper and have $(\mathbb{R}^d)^d$ for effective domain.

As $\|x\|_{(d)}^{\text{sn}} = \|x\| < 1$, the sequence $(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = (0, 0, \dots, 0, x)$ strictly satisfies the inequality constraint, that is, $f_1(0, 0, \dots, 0, x) = \|x\| - 1 < 0$ and satisfies also the equality constraints $f_2(0, 0, \dots, 0, x) = \dots = f_{1+d}(0, 0, \dots, 0, x) = 0$.

By the Slater condition, the constraints are qualified. Therefore, the sequence $(\bar{x}^{(1)}, \dots, \bar{x}^{(d)})$ is solution of the convex optimization problem (74) if and only if it satisfies the KKT conditions ([12, Corollary 28.3.1], [10, Example 1''', p. 64], [7, Chapter VII]), that is, there exists $\lambda \geq 0$ and $\mu = (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$ such that

$$0 \in \partial f_0(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) + \lambda \partial f_1(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) + \sum_{k=1}^d \mu_k \partial f_{1+k}(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}), \quad (77a)$$

$$\lambda f_1(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = 0, \quad (77b)$$

$$f_1(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) \leq 0, \quad (77c)$$

$$\forall k = 1, \dots, d, \quad f_{1+k}(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = 0. \quad (77d)$$

Since $\partial \sigma_{B_{(l)}^{\text{sgn}}}(x^{(l)}) = F_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)})$ [11, Corollary 8.25], for $l = 1, \dots, d$, we have, by (76),

$$\partial f_0(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = (F_{B_{(1)}^{\text{sgn}}}(\bar{x}^{(1)}), \dots, lF_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)}), \dots, dF_{B_{(d)}^{\text{sgn}}}(\bar{x}^{(d)})), \quad (78a)$$

$$\partial f_1(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = (F_{B_{(1)}^{\text{sgn}}}(\bar{x}^{(1)}), \dots, F_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)}), \dots, F_{B_{(d)}^{\text{sgn}}}(\bar{x}^{(d)})), \quad (78b)$$

$$\partial f_{1+k}(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = (e_k, \dots, e_k), \quad \forall k = 1, \dots, d. \quad (78c)$$

With these expressions, Equation (77a) is equivalent to $\mu = (\mu_1, \dots, \mu_d) = \sum_{k=1}^d \mu_k e_k \in lF_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)}) + \lambda F_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)})$, for $l = 1, \dots, d$.

We conclude that the sequence $(\bar{x}^{(1)}, \dots, \bar{x}^{(d)})$ of vectors of \mathbb{R}^d is solution of the optimization

problem (74) if and only if there exists $\lambda \geq 0$ such that the following conditions are satisfied

$$\bigcap_{l=1}^d [lF_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)}) + \lambda F_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)})] \neq \emptyset, \quad (79a)$$

$$\lambda \left(\sum_{l=1}^d \sigma_{B_{(l)}^{\text{sgn}}}(x^{(l)}) - 1 \right) = 0, \quad (79b)$$

$$\sum_{l=1}^d \sigma_{B_{(l)}^{\text{sgn}}}(x^{(l)}) \leq 1, \quad (79c)$$

$$\sum_{l=1}^d \bar{x}^{(l)} = x. \quad (79d)$$

Second, we turn to prove Item 1 and Item 2.

1. If $\lambda = 0$ in (79), we obtain

$$\bigcap_{l=1}^d lF_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)}) \neq \emptyset, \quad (80a)$$

$$\sum_{l=1}^d \sigma_{B_{(l)}^{\text{sgn}}}(x^{(l)}) \leq 1, \quad (80b)$$

$$\sum_{l=1}^d \bar{x}^{(l)} = x. \quad (80c)$$

We now show that (80) holds true if and only if $\|x\|_{(1)}^{\text{sn}} \leq 1$ and $(\bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(d)}) = (x, 0, \dots, 0)$.

On the one hand, let $(\bar{x}^{(1)}, \dots, \bar{x}^{(d)})$ be a sequence of vectors of \mathbb{R}^d which satisfies (80). If $(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = (0, \dots, 0)$, then $x = 0$ and we indeed conclude that $\|x\|_{(1)}^{\text{sn}} = \|0\|_{(1)}^{\text{sn}} = 0 \leq 1$ and $(\bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(d)}) = (0, 0, \dots, 0) = (x, 0, \dots, 0)$.

If $(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) \neq (0, \dots, 0)$, then $k = \min \{l \in \{1, \dots, d\} \mid \bar{x}^{(l)} \neq 0\}$ is well defined. By (80a), there exists $y \in \bigcap_{l=1}^d lF_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)})$, where $F_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)}) \subset B_{(l)}^{\text{sgn}}$ for any $l = 1, \dots, d$, by definition (73) of the face. Now, by the inclusion (72b), we have that $F_{B_{(k)}^{\text{sgn}}}(\bar{x}^{(k)}) \subset S_{(k)}^{\text{sgn}}$ since $\bar{x}^{(k)} \neq 0$ by definition of k . Therefore, there exists $y \in B_{(1)}^{\text{sgn}} \cap kS_{(k)}^{\text{sgn}}$, that is, $\|y\|_{(1)}^{\text{sgn}} = \max_{i=1, \dots, d} |y_i| \leq 1$ and $\|y\|_{(k)}^{\text{sgn}} = k$. Hence, by (63) (which follows easily from definition (29) of $\|\cdot\|_{(k)}^{\text{sgn}}$), we get that $k^2 = (\|y\|_{(k)}^{\text{sgn}})^2 \leq k(\|y\|_{(1)}^{\text{sgn}})^2 \leq k$. This gives $k = 1$, hence $\bar{x}^{(1)} \neq 0$ and $\bar{x}^{(l)} = 0$ for all $l = 2, \dots, d$ by definition of k . We conclude that necessarily $(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = (x, 0, \dots, 0)$ by (80c) and $\sigma_{B_{(1)}^{\text{sgn}}}(x) = \|x\|_{(1)}^{\text{sn}} \leq 1$ by (80b).

On the other hand, suppose that $\|x\|_{(1)}^{\text{sn}} \leq 1$ and put $(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = (x, 0, \dots, 0)$. Then, Equations (80b) and (80c) are satisfied. So is (80a) because

$$\begin{aligned} \bigcap_{l=1}^d lF_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)}) &= F_{B_{(1)}^{\text{sgn}}}(x) \cap \left[\bigcap_{l=2}^d lB_{(l)}^{\text{sgn}} \right] && \text{(by (72a))} \\ &= F_{B_{(1)}^{\text{sgn}}}(x) \end{aligned}$$

because $F_{B_{(1)}^{\text{sgn}}}(x) \subset B_{(1)}^{\text{sgn}} \subset \bigcap_{l=2}^d \sqrt{l}B_{(l)}^{\text{sgn}} \subset \bigcap_{l=2}^d lB_{(l)}^{\text{sgn}}$ by (60)
 $\neq \emptyset$.

2. If $\lambda > 0$ in (79), we obtain item 2. Indeed, (79a) is equivalent to (75a) because $lF_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)}) + \lambda F_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)}) = (l + \lambda)F_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)})$ since the face $F_{B_{(l)}^{\text{sgn}}}(\bar{x}^{(l)})$ is convex, and $l > 0, \lambda > 0$.

This ends the proof. \square

Now, we specialize in the two dimensional case $d = 2$. Because the proper convex lsc function \mathcal{L}_0 in Corollary 9 satisfies (37), we restrict the following Proposition to $x = (x_1, x_2) \in \mathbb{R}_+^2$.

Proposition 19 *Let $x = (x_1, x_2) \in \mathbb{R}_+^2$ be such that $x_1^2 + x_2^2 < 1$. The sequence $(\bar{x}^{(1)}, \bar{x}^{(2)})$ of vectors of \mathbb{R}^2 is solution of the optimization problem*

$$\mathcal{L}_0(x) = \min_{\substack{x^{(1)} \in \mathbb{R}^d, x^{(2)} \in \mathbb{R}^2 \\ \|x^{(1)}\|_{(1)}^{\text{sn}} + \|x^{(2)}\|_{(2)}^{\text{sn}} \leq 1 \\ x^{(1)} + x^{(2)} = x}} \|x^{(1)}\|_{(1)}^{\text{sn}} + 2\|x^{(2)}\|_{(2)}^{\text{sn}} \quad (81)$$

if and only if one of the following statements holds true:

1. $x_1 + x_2 \leq 1$, and then $(\bar{x}^{(1)}, \bar{x}^{(2)}) = (x, 0)$, and

$$\mathcal{L}_0((x_1, x_2)) = x_1 + x_2, \quad (82)$$

2. $x_1 > 0, x_1 + (\sqrt{2} - 1)x_2 \geq 1, x_1 > x_2$, and then

$$\begin{aligned} \bar{x}^{(1)} &= \left(\frac{1 - (x_1^2 + x_2^2)}{2(1 - x_1)}, 0 \right), \\ \bar{x}^{(2)} &= \left(\frac{2x_1 - x_1^2 + x_2^2 - 1}{2(1 - x_1)}, x_2 \right), \end{aligned}$$

and

$$\mathcal{L}_0((x_1, x_2)) = \frac{3}{2} - \frac{x_1}{2} + \frac{x_2^2}{2(1 - x_1)}, \quad (84)$$

3. $x_2 > 0$, $x_2 + (\sqrt{2} - 1)x_1 \geq 1$, $x_2 > x_1$, and then

$$\begin{aligned}\bar{x}^{(1)} &= \left(0, \frac{1 - (x_1^2 + x_2^2)}{2(1 - x_2)}\right), \\ \bar{x}^{(2)} &= \left(x_1, \frac{2x_2 - x_2^2 + x_1^2 - 1}{2(1 - x_2)}\right),\end{aligned}$$

and

$$\mathcal{L}_0((x_1, x_2)) = \frac{3}{2} - \frac{x_2}{2} + \frac{x_1^2}{2(1 - x_2)}, \quad (86)$$

4. $x_1 + x_2 > 1$, $(\sqrt{2} - 1)x_1 + x_2 < 1$, $x_1 + (\sqrt{2} - 1)x_2 < 1$, and then

$$\begin{aligned}\bar{x}^{(1)} &= \left(\frac{1 - (\sqrt{2} - 1)x_1 - x_2}{2(\sqrt{2} - 1)}, \frac{1 - x_1 - (\sqrt{2} - 1)x_2}{2(\sqrt{2} - 1)}\right), \\ \bar{x}^{(2)} &= \left(\frac{x_1 + x_2 - 1}{2(\sqrt{2} - 1)}, \frac{x_1 + x_2 - 1}{2(\sqrt{2} - 1)}\right),\end{aligned}$$

and

$$\mathcal{L}_0((x_1, x_2)) = \frac{x_1 + x_2 - 2 + \sqrt{2}}{\sqrt{2} - 1}. \quad (88)$$

Proof. By Proposition 18, the sequence $(\bar{x}^{(1)}, \bar{x}^{(2)})$ of vectors of \mathbb{R}^2 is solution of the optimization problem (81) if and only if

- either $x_1 + x_2 = \|x\|_{(1)}^{\text{sn}} \leq 1$ and $(\bar{x}^{(1)}, \bar{x}^{(2)}) = (x, 0)$, which is equivalent to Item 1,
- or there exists $\lambda > 0$ such that

$$(1 + \lambda)F_{B_{(1)}^{\text{sgn}}}(\bar{x}^{(1)}) \cap (2 + \lambda)F_{B_{(2)}^{\text{sgn}}}(\bar{x}^{(2)}) \neq \emptyset, \quad (89a)$$

$$\|\bar{x}^{(1)}\|_{(1)}^{\text{sn}} + \|\bar{x}^{(2)}\|_{(2)}^{\text{sn}} = 1, \quad (89b)$$

$$\bar{x}^{(1)} + \bar{x}^{(2)} = x. \quad (89c)$$

We are going to prove, in several steps, that $(\bar{x}^{(1)}, \bar{x}^{(2)})$ satisfies (89) for a certain $\lambda > 0$ if and only if it satisfies Item 2, Item 3 or Item 4. For this purpose, we will use the following relations:

$$B_{(1)}^{\text{sgn}} = [-1, 1]^2, \quad (90a)$$

$$F_{B_{(1)}^{\text{sgn}}}(\bar{x}^{(1)}) = \begin{cases} [-1, 1]^2 & \text{if } \ell_0(\bar{x}^{(1)}) = 0, \\ \text{sign}(\bar{x}_1^{(1)}) \times [-1, 1] & \text{if } \ell_0(\bar{x}^{(1)}) = 1 \text{ with } \bar{x}_2^{(1)} = 0, \\ [-1, 1] \times \text{sign}(\bar{x}_2^{(1)}) & \text{if } \ell_0(\bar{x}^{(1)}) = 1 \text{ with } \bar{x}_1^{(1)} = 0, \\ \text{sign}(\bar{x}^{(1)}) & \text{if } \ell_0(\bar{x}^{(1)}) = 2, \end{cases} \quad (90b)$$

where $\text{sign}(\bar{x}^{(1)}) = (\text{sign}(\bar{x}_1^{(1)}), \text{sign}(\bar{x}_2^{(1)}))$ is the vector of \mathbb{R}^2 made of the signs $(-1, 0, 1)$ of the two components.

- Suppose that $(\bar{x}^{(1)}, \bar{x}^{(2)}) = (x, 0)$ satisfies (89) for a certain $\lambda > 0$.

We will show that this is equivalent to $0 < x_1$, $0 < x_2$ and $x_1 + x_2 = 1$, which implies Item 1.

By (72a) for $l = d = 2$, we get that $(2 + \lambda)F_{B_{(2)}^{\text{sgn}}}(0) = (2 + \lambda)B$, where B is the Euclidian unit ball of \mathbb{R}^2 , so that Equation (89) is equivalent to

$$(1 + \lambda)F_{B_{(1)}^{\text{sgn}}}(x) \cap (2 + \lambda)B \neq \emptyset, \quad \|x\|_{(1)}^{\text{sn}} = x_1 + x_2 = 1. \quad (91)$$

By (90b), we distinguish the following subcases that correspond to different expressions for $F_{B_{(1)}^{\text{sgn}}}(x)$.

- If $\ell_0(x) = 0$, then $x_1 = x_2 = 0$. But this contradicts $x_1 + x_2 = 1$ in (91).
- If $\ell_0(x) = 1$ with $x_2 = 0$, then $x = (1, 0)$ because $x_1 + x_2 = 1$ by (91), and $x = (x_1, x_2) \in \mathbb{R}_+^2$ by hypothesis. But this contradicts the assumption that $x_1^2 + x_2^2 < 1$.
- If $\ell_0(x) = 1$ with $x_1 = 0$, we also arrive at a contradiction.
- If $\ell_0(x) = 2$, then $(1 + \lambda)F_{B_{(1)}^{\text{sgn}}}(x) = \{(1 + \lambda, 1 + \lambda)\}$ by (90b).

On the one hand (necessity), we show that necessarily $0 < \lambda \leq \sqrt{2}$. Indeed, (91) implies that $\|((1 + \lambda)\text{sign}(x_1), (1 + \lambda)\text{sign}(x_2))\| \leq 2 + \lambda$, which gives $\sqrt{2}(1 + \lambda) \leq 2 + \lambda$, hence $0 < \lambda \leq \sqrt{2}$.

On the other hand (sufficiency), if we put $(\bar{x}^{(1)}, \bar{x}^{(2)}) = (x, 0)$ where $\|x\|_{(1)}^{\text{sn}} = x_1 + x_2 = 1$ and $\ell_0(x) = 2$, that is, $0 < x_1$, $0 < x_2$, then (91) is satisfied for any $0 < \lambda \leq \sqrt{2}$.

Therefore, we have proven that $(\bar{x}^{(1)}, \bar{x}^{(2)}) = (x, 0)$ satisfies (89) for a certain $\lambda > 0$ if and only if $0 < x_1$, $0 < x_2$ and $x_1 + x_2 = 1$ (condition included in Item 1).

- Suppose that $(\bar{x}^{(1)}, \bar{x}^{(2)}) = (0, x)$ satisfies (89) for a certain $\lambda > 0$.

We will show that this case is impossible.

Indeed, Equation (89b) implies that $\sqrt{x_1^2 + x_2^2} = \|x\| = \|x\|_{(2)}^{\text{sn}} = 1$. But this contradicts the assumption that $x = (x_1, x_2) \in \mathbb{R}_+^2$ is such that $x_1^2 + x_2^2 < 1$.

- Suppose that $\bar{x}^{(1)} \neq 0$ and $\bar{x}^{(2)} \neq 0$ are such that $(\bar{x}^{(1)}, \bar{x}^{(2)})$ satisfies (89) for a certain $\lambda > 0$.

We will show that this is equivalent to Item 2, Item 3 or Item 4.

But, before that, notice that, as $\|\bar{x}^{(1)}\|_{(1)}^{\text{sn}} + \|\bar{x}^{(2)}\|_{(2)}^{\text{sn}} = 1$, by (75b), then

$$\mathcal{L}_0(x) = 1 + \|\bar{x}^{(2)}\|_{(2)}^{\text{sn}} = 2 - \|\bar{x}^{(1)}\|_{(1)}^{\text{sn}}, \quad (92)$$

which will be practical to obtain formulas for $\mathcal{L}_0(x)$.

As $\bar{x}^{(2)} \neq 0$, then $F_{B_{(2)}^{\text{sgn}}}(\bar{x}^{(2)}) = \{\frac{\bar{x}^{(2)}}{\|\bar{x}^{(2)}\|}\}$ by (72c). Therefore, Equation (89) is equivalent to

$$(2 + \lambda) \frac{\bar{x}^{(2)}}{\|\bar{x}^{(2)}\|} \in (1 + \lambda)F_{B_{(1)}^{\text{sgn}}}(\bar{x}^{(1)}), \quad (93a)$$

$$\|\bar{x}^{(1)}\|_{(1)}^{\text{sn}} + \|\bar{x}^{(2)}\|_{(2)}^{\text{sn}} = |\bar{x}_1^{(1)}| + |\bar{x}_2^{(1)}| + \sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2} = 1, \quad (93b)$$

$$\bar{x}^{(1)} + \bar{x}^{(2)} = x. \quad (93c)$$

By (90b), we distinguish the following subcases that correspond to different expressions for $F_{B_{(1)}^{\text{sgn}}}(\bar{x}^{(1)})$.

- As $\bar{x}^{(1)} \neq 0$, we do not consider the case $\ell_0(\bar{x}^{(1)}) = 0$.
- Suppose that $\ell_0(\bar{x}^{(1)}) = 1$ with $\bar{x}_2^{(1)} = 0$. Then, on the one hand, $F_{B_{(1)}^{\text{sgn}}}(\bar{x}^{(1)}) = \text{sign}(\bar{x}_1^{(1)}) \times [-1, 1]$ by (90b), so that Equation (93a) is equivalent to

$$\frac{\bar{x}_1^{(2)}}{\sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2}} = \frac{1 + \lambda}{2 + \lambda} \text{sign}(\bar{x}_1^{(1)}),$$

$$\frac{|\bar{x}_2^{(2)}|}{\sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2}} \leq \frac{1 + \lambda}{2 + \lambda}.$$

On the other hand, $\bar{x}^{(1)} = (\bar{x}_1^{(1)}, 0)$ where $\bar{x}_1^{(1)} \neq 0$, so that Equations (93c)–(93c) are equivalent to

$$|\bar{x}_1^{(1)}| + \sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2} = 1,$$

$$\bar{x}_1^{(1)} + \bar{x}_1^{(2)} = x_1,$$

$$\bar{x}_2^{(2)} = x_2.$$

Therefore, Equation (93) is equivalent to

$$\frac{\bar{x}_1^{(2)}}{1 - |\bar{x}_1^{(1)}|} = \frac{1 + \lambda}{2 + \lambda} \text{sign}(\bar{x}_1^{(1)}), \quad (95a)$$

$$|\bar{x}_2^{(2)}| \leq \text{sign}(\bar{x}_1^{(1)}) \bar{x}_1^{(2)}, \quad (95b)$$

$$|\bar{x}_1^{(1)}| + \sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2} = 1, \quad (95c)$$

$$\bar{x}_1^{(1)} + \bar{x}_1^{(2)} = x_1, \quad (95d)$$

$$\bar{x}_2^{(2)} = x_2, \quad (95e)$$

and we will now show that there exists $\lambda > 0$ such that (95) holds true if and only if Item 2 holds true.

On the one hand (necessity), from (95a), we deduce that $\bar{x}_1^{(2)}$ and $\bar{x}_1^{(1)}$ have the same sign; this common sign must therefore be $\text{sign}(x_1)$, as $\bar{x}_1^{(1)} + \bar{x}_1^{(2)} = x_1$ by (95d); since $x = (x_1, x_2) \in \mathbb{R}_+^2$, we obtain that $x_1 \geq 0$, hence $\bar{x}_1^{(1)} > 0$ and $x_1 > 0$. Therefore, we easily get that $\bar{x}^{(1)} = (\bar{x}_1^{(1)}, 0)$, where $\bar{x}_1^{(1)} > 0$, and that $\bar{x}^{(2)} = (x_1 - \bar{x}_1^{(1)}, x_2)$, by (95d)–(95e), with $x_1 > \bar{x}_1^{(1)}$, since $\bar{x}_1^{(2)} > 0$. Replacing the values in (95c) — where $x_1 > \bar{x}_1^{(1)} > 0$ and $1 > x_1$ since $x_1^2 + x_2^2 < 1$ — we get $\bar{x}_1^{(1)} + \sqrt{(x_1 - \bar{x}_1^{(1)})^2 + x_2^2} = 1$, from which we deduce that $\bar{x}_1^{(1)} = \frac{1 - (x_1^2 + x_2^2)}{2(1 - x_1)}$; we have that $\bar{x}_1^{(1)} > 0$ because $x_1 < 1$; the condition $x_1 > \bar{x}_1^{(1)}$ implies that $x_1 + x_2 > 1$. From (95a), we deduce that $\frac{\bar{x}_1^{(2)}}{1 - |\bar{x}_1^{(1)}|} = \frac{1 + \lambda}{2 + \lambda} \in]1/2, 1[$, hence that

$\frac{x_1 - \bar{x}_1^{(1)}}{1 - \bar{x}_1^{(1)}} < 1$ and $1/2 < \frac{x_1 - \bar{x}_1^{(1)}}{1 - \bar{x}_1^{(1)}}$ by (95b); we are going to detail these two inequalities, one after the other. We have that $\frac{x_1 - \bar{x}_1^{(1)}}{1 - \bar{x}_1^{(1)}} < 1$ because $1 > x_1 > \bar{x}_1^{(1)}$. The condition $1/2 < \frac{x_1 - \bar{x}_1^{(1)}}{1 - \bar{x}_1^{(1)}}$ implies that $0 < x_2 - \sqrt{3}(1 - x_1)$; from (95b), with $\bar{x}_1^{(1)} > 0$ and $\bar{x}_2^{(2)} = x_2 > 0$, we get that $x_2 \leq x_1 - \bar{x}_1^{(1)} = x_1 - \frac{1 - (x_1^2 + x_2^2)}{2(1 - x_1)}$; rearranging terms, we find that this latter inequality is equivalent to $(x_2 - (\sqrt{2} + 1)(1 - x_1))(x_2 + (\sqrt{2} - 1)(1 - x_1)) \geq 0$; as $x_1 < 1$ and $x_2 > 0$, we finally get that $x_2 - (\sqrt{2} + 1)(1 - x_1) \geq 0$. From $x_2 \leq x_1 - \bar{x}_1^{(1)}$ where $\bar{x}_1^{(1)} > 0$, we also deduce that necessarily $x_1 > x_2$.

Finally, Equation (95) implies that $x_1 > 0$, $x_1 + x_2 > 1$, $x_2 - \sqrt{3}(1 - x_1) > 0$, $x_2 - (\sqrt{2} + 1)(1 - x_1) \geq 0$, $x_1 > x_2$, and $\bar{x}^{(1)} = (\frac{1 - (x_1^2 + x_2^2)}{2(1 - x_1)}, 0)$, $\bar{x}^{(2)} = (\frac{2x_1 - x_1^2 + x_2^2 - 1}{2(1 - x_1)}, x_2)$: thus, using the property that

$$x_2 - (\sqrt{2} + 1)(1 - x_1) \geq 0 \text{ and } 1 > x_1 \Rightarrow \begin{cases} x_1 + x_2 \geq \sqrt{2}(1 - x_1) + 1 > 1 \\ x_2 - \sqrt{3}(1 - x_1) > x_2 - (\sqrt{2} + 1)(1 - x_1) \geq 0, \end{cases}$$

we obtain that $x_2 - (\sqrt{2} + 1)(1 - x_1) \geq 0$ and $1 > x_1$; multiplying the first inequality by $\sqrt{2} - 1$, we finally obtain $x_1 + (\sqrt{2} - 1)x_2 \geq 1$ and $1 > x_1$, that is, Item 2.

On the other hand (sufficiency), if we suppose that Item 2 holds it is straightforward to follow all the above computations and to obtain that Equation (95) holds true with $\lambda > 0$ the unique solution to $\frac{\bar{x}_1^{(2)}}{1 - |\bar{x}_1^{(1)}|} = \frac{1 + \lambda}{2 + \lambda} \in]1/2, 1[$.

By (92), we obtain that $\mathcal{L}_0((x_1, x_2)) = 2 - |\bar{x}_1^{(1)}| - |\bar{x}_2^{(1)}| = \frac{3}{2} - \frac{x_1}{2} + \frac{x_2^2}{2(1 - x_1)}$.

- If $\ell_0(\bar{x}^{(1)}) = 1$ with $\bar{x}_1^{(1)} = 0$, we do the same analysis, and we obtain Item 3, and $\mathcal{L}_0((x_1, x_2)) = \frac{3}{2} - \frac{x_2}{2} + \frac{x_1^2}{2(1 - x_2)}$.
- Suppose that $\ell_0(\bar{x}^{(1)}) = 2$. In this case, we have that $F_{B_{(1)}^{\text{sgn}}}(\bar{x}^{(1)}) = \{(\text{sign}(\bar{x}_1^{(1)}), \text{sign}(\bar{x}_2^{(1)}))\}$ by (90b). Therefore, Equation (93) is equivalent to

$$\frac{\bar{x}_1^{(2)}}{\sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2}} = \frac{1 + \lambda}{2 + \lambda} \text{sign}(\bar{x}_1^{(1)}), \quad (96a)$$

$$\frac{\bar{x}_2^{(2)}}{\sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2}} = \frac{1 + \lambda}{2 + \lambda} \text{sign}(\bar{x}_2^{(1)}), \quad (96b)$$

$$|\bar{x}_1^{(1)}| + |\bar{x}_2^{(1)}| + \sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2} = 1, \quad (96c)$$

$$\bar{x}_1^{(1)} + \bar{x}_1^{(2)} = x_1, \quad (96d)$$

$$\bar{x}_2^{(1)} + \bar{x}_2^{(2)} = x_2, \quad (96e)$$

and we will now show that there exists $\lambda > 0$ such that Equation (96) holds true if and only if Item 4 holds true.

On the one hand (necessity), from (96a)–(96b), we deduce that $|\bar{x}_1^{(2)}| = |\bar{x}_2^{(2)}|$ — because $|\text{sign}(\bar{x}_1^{(1)})| = |\text{sign}(\bar{x}_2^{(1)})| = 1$ since $\ell_0(\bar{x}^{(1)}) = 2$ — and that $\text{sign}(\bar{x}^{(1)}) = \text{sign}(\bar{x}^{(2)})$. This common sign must therefore be $\text{sign}(x)$, as $\bar{x}^{(1)} + \bar{x}^{(2)} = x$ by (96d)–(96e). Since $\ell_0(\bar{x}^{(1)}) = 2$ and $x = (x_1, x_2) \in \mathbb{R}_+^2$, we get that $x_1 > 0$ and $x_2 > 0$, so that we put $\bar{x}_1^{(2)} = \bar{x}_2^{(2)} = \beta > 0$. By (96d)–(96e), we get that $\bar{x}_1^{(1)} = x_1 - \beta > 0$ and $\bar{x}_2^{(1)} = x_2 - \beta > 0$; replacing the values in (96c), we obtain that $x_1 - \beta + x_2 - \beta + \sqrt{2}\beta = 1$; this gives $\beta = \frac{x_1 + x_2 - 1}{2 - \sqrt{2}}$. Therefore, $\beta > 0 \iff x_1 + x_2 > 1$, $\beta < x_1 \iff (\sqrt{2} - 1)x_1 + x_2 < 1$ and $\beta < x_2 \iff x_1 + (\sqrt{2} - 1)x_2 < 1$.

Finally, Equation (96) implies that $x_1 + x_2 > 1$, $(\sqrt{2} - 1)x_1 + x_2 < 1$, $x_1 + (\sqrt{2} - 1)x_2 < 1$, and $\bar{x}^{(1)} = \left(\frac{1 - (\sqrt{2} - 1)x_1 - x_2}{2(\sqrt{2} - 1)}, \frac{1 - x_1 - (\sqrt{2} - 1)x_2}{2(\sqrt{2} - 1)}\right)$, $\bar{x}^{(2)} = \left(\frac{x_1 + x_2 - 1}{2(\sqrt{2} - 1)}, \frac{x_1 + x_2 - 1}{2(\sqrt{2} - 1)}\right)$: thus, Item 4 holds true.

On the other hand (sufficiency), if we suppose that Item 4 holds true, it is straightforward to follow all the above computations and to obtain that Equation (96) holds true with $\lambda = \sqrt{2}$.

By (92), we obtain that $\mathcal{L}_0((x_1, x_2)) = 1 + \sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2} = 1 + \frac{x_1 + x_2 - 1}{\sqrt{2} - 1}$.

This ends the proof. □

References

- [1] C. D. Aliprantis and K. C. Border. *Infinite dimensional analysis*. Springer-Verlag, Berlin, second edition, 1999.
- [2] A. Argyriou, R. Foygel, and N. Srebro. Sparse prediction with the k -support norm. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pages 1457–1465, USA, 2012. Curran Associates Inc.
- [3] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer-Verlag, second edition, 2017.
- [4] J.-P. Chancelier and M. De Lara. Lower bound convex programs for exact sparse optimization, 2019. preprint.
- [5] J. Hiriart-Urruty. *Optimisation et analyse convexe*. Presses Universitaires de France, 1998.
- [6] J.-B. Hiriart-Urruty and H. Le. A variational approach of the rank function. *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research*, 21(2):207–240, 2013.

- [7] J. B. Hiriart-Uruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I*. Springer-Verlag, Berlin, 1993.
- [8] J. E. Martínez-Legaz. Generalized convex duality and its economic applications. In S. S. Hadjisavvas N., Komlósi S., editor, *Handbook of Generalized Convexity and Generalized Monotonicity. Nonconvex Optimization and Its Applications*, volume 76, pages 237–292. Springer-Verlag, 2005.
- [9] J. J. Moreau. Inf-convolution, sous-additivité, convexité des fonctions numériques. *J. Math. Pures Appl. (9)*, 49:109–154, 1970.
- [10] R. T. Rockafellar. *Conjugate Duality and Optimization*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1974.
- [11] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer-Verlag, Berlin, 1998.
- [12] T. R. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, N.J., 1970.
- [13] A. Rubinov. *Abstract convexity and global optimization*, volume 44 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 2000.
- [14] I. Singer. *Abstract convex analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons, Inc., New York, 1997.
- [15] C. Zalinescu. *Convex Analysis in General Vector Spaces*. World Scientific, 2002.