Journal of Convex Analysis
Volume 28 (2021), No. 1, 203–236
https://doi.org/10.68381/jca28015

Hidden Convexity in the l_0 Pseudonorm

Jean-Philippe Chancelier, Michel De Lara
CERMICS, École des Ponts, Marne-la-Vallée, France
michel.delara@enpc.fr

Received: June 4, 2019
Accepted: May 15, 2020

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, E-Capra, that has the property of being constant along primal rays like the ℓ_0 pseudonorm. The coupling E-Capra 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 E-Capra conjugacy, induced by the coupling E-Capra, we relate the E-Capra conjugate and biconjugate of the ℓ_0 pseudonorm, the characteristic functions of its level sets and the sequence of so-called top-k norms. In particular, we prove that the ℓ_0 pseudonorm is equal to its biconjugate: hence, the ℓ_0 pseudonorm is E-Capra-convex in the sense of generalized convexity. As a corollary, we show that there exists a proper convex lower semicontinuous function on $\mathbb{R}^d$ such that this function and the ℓ_0 pseudonorm coincide on the Euclidean unit sphere. This hidden convexity property is somewhat surprising as the ℓ_0 pseudonorm is a highly nonconvex function of combinatorial nature. We provide different expressions for this proper convex lower semicontinuous function, and we give explicit formulas in the two-dimensional case.

Keywords: ℓ_0 pseudonorm, coupling, Fenchel-Moreau conjugacy, top-k norms, k-support norms, hidden convexity.

2010 Mathematics Subject Classification: 46N10, 49N15, 46B99, 52A41, 90C46.

1. Introduction

The counting function, also called cardinality function or ℓ_0 pseudonorm, counts the number of nonzero components of a vector in $\mathbb{R}^d$. It is related to the rank function defined over matrices [7]. It is well-known that the ℓ_0 pseudonorm is lower semi continuous (lsc) but is not convex, and that the Fenchel conjugacy fails to provide relevant analysis. Indeed, the Fenchel biconjugate of the characteristic function of the level sets of the ℓ_0 pseudonorm is zero, and the Fenchel biconjugate of the ℓ_0 pseudonorm is also zero.

In this paper, we display a suitable conjugacy for which we prove that the ℓ_0 pseudonorm is "convex" in the sense of generalized convexity, that is, is equal to its biconjugate. As a corollary, we also show that the ℓ_0 pseudonorm function displays hidden convexity in the following sense[1]: the ℓ_0 pseudonorm is equal to the composition of a proper convex lower semicontinuous function on $\mathbb{R}^d$ with the normalization mapping from $\mathbb{R}^d$ to the Euclidean unit sphere.

[1]In [3], the vocable "hidden convexity" refers to *optimization problems* (when an original problem is equivalent to a *convex optimization problem*). Here, the vocable "hidden convexity" refers to *functions* (when a function is the composition of a *convex function* with a mapping).

ISSN 0944-6532 / $ 2.50 © Heldermann Verlag

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 Euclidean constant along primal rays coupling ¢ (E-Capra). 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 also elucidate the structure of E-Capra-convex functions. Then, in Sect. 3, we relate the E-Capra conjugate and biconjugate of the ℓ_0 pseudonorm, the characteristic functions of its level sets and the so-called top-k norms. In particular, we show that the ℓ_0 pseudonorm is E-Capra biconjugate (that is, a E-Capra-convex function). In Sect. 4, we deduce that the ℓ_0 pseudonorm coincides, on the Euclidean unit sphere, with a proper convex lsc function $\mathcal{L}_0$ defined on $\mathbb{R}^d$. We provide various expression for the function $\mathcal{L}_0$. The Appendix A gathers properties of top-k norms and of k-support norms, properties of the ℓ_0 pseudonorm level sets, and technical results on the function $\mathcal{L}_0$.

2. One-sided linear couplings

After having recalled background on Fenchel-Moreau conjugacies in §2.1, we introduce one-sided linear couplings in §2.2.

When we manipulate functions with values in $\overline{\mathbb{R}} = [-\infty, +\infty]$, we adopt the Moreau *lower addition* [10] that extends the usual addition with $(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = -\infty$. Let $\mathbb{W}$ be a set. For any function $h : \mathbb{W} \to \overline{\mathbb{R}}$, its *epigraph* is $\operatorname{epi} h = \{(w,t) \in \mathbb{W} \times \mathbb{R} \mid h(w) \leq t\}$, its *effective domain* is $\operatorname{dom} h = \{w \in \mathbb{W} \mid h(w) < +\infty\}$. A function $h : \mathbb{W} \to \overline{\mathbb{R}}$ is said to be *proper* if it never takes the value $-\infty$ and if $\operatorname{dom} h \neq \emptyset$. When $\mathbb{W}$ is equipped with a topology, the function $h : \mathbb{W} \to \overline{\mathbb{R}}$ is said to be *lower semi continuous (lsc)* if its epigraph is a closed subset of $\mathbb{W} \times \mathbb{R}$.

2.1. Background on Fenchel-Moreau conjugacies

We review concepts and notations related to the Fenchel conjugacy (we refer the reader to [11]), then present how they are extended to general conjugacies [15, 14, 9].

The Fenchel conjugacy

Let $\mathbb{X}$ and $\mathbb{Y}$ be two (real) vector spaces that are *paired* in the following sense [11, p. 13]: there exists a bilinear form $\langle\, ,\rangle : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ 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 $x \in \mathbb{X}$. The classic *Fenchel conjugacy* $\star$ is defined, for any functions $f : \mathbb{X} \to \overline{\mathbb{R}}$ and $g : \mathbb{Y} \to \overline{\mathbb{R}}$, by[2]

$$f^\star(y) = \sup_{x \in \mathbb{X}} \Big(\langle x\, , y\rangle \dot{+} \big(-f(x) \big) \Big)\ , \quad \forall y \in \mathbb{Y}\ , \tag{1a}$$

$$g^{\star'}(x) = \sup_{y \in \mathbb{Y}} \Big(\langle x\, , y\rangle \dot{+} \big(-g(y) \big) \Big)\ , \quad \forall x \in \mathbb{X}\ , \tag{1b}$$

[2]In convex analysis, one does not use the notation $^{\star'}$ in (1b) and (1c), but simply *. We use $^{\star'}$ to be consistent with the notation (3c) and (3d) for general conjugacies.

$$f^{\star\star'}(x) = \sup_{y \in \mathbb{Y}} \Big(\langle x \, , y \rangle \dot{+} \big(- f^{\star}(y) \big) \Big) \, , \ \forall x \in \mathbb{X} \, . \tag{1c}$$

Recall that a function is said to be *convex* if its epigraph is a convex set. Recall that a function is said to be *closed* if it is either lsc and nowhere having the value $-\infty$, or is the constant function $-\infty$ [11, p. 15]. It is proved that the Fenchel conjugacy induces a one-to-one correspondence between the closed convex functions on $\mathbb{X}$ and the closed convex functions on $\mathbb{Y}$ [11, Theorem 5]. Closed convex functions are the two constant functions $-\infty$ and $+\infty$ united with all proper convex lsc functions.[3]

The general case

Let be given two sets $\mathbb{X}$ ("primal"), $\mathbb{Y}$ ("dual"), not necessarily vector spaces, together with a *coupling* function

$$c : \mathbb{X} \times \mathbb{Y} \to \overline{\mathbb{R}} \, . \tag{2}$$

With any coupling, one associates *conjugacies* from the set $\overline{\mathbb{R}}^{\mathbb{X}}$ of functions $\mathbb{X} \to \overline{\mathbb{R}}$ to the set $\overline{\mathbb{R}}^{\mathbb{Y}}$ of functions $\mathbb{Y} \to \overline{\mathbb{R}}$, and from $\overline{\mathbb{R}}^{\mathbb{Y}}$ to $\overline{\mathbb{R}}^{\mathbb{X}}$ as follows.

Definition 2.1. The *c-Fenchel-Moreau conjugate* of a function $f : \mathbb{X} \to \overline{\mathbb{R}}$, with respect to the coupling c, is the function $f^c : \mathbb{Y} \to \overline{\mathbb{R}}$ defined by

$$f^c(y) = \sup_{x \in \mathbb{X}} \Big(c(x,y) \dot{+} \big(- f(x) \big) \Big) \, , \ \forall y \in \mathbb{Y} \, . \tag{3a}$$

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

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

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

$$g^{c'}(x) = \sup_{y \in \mathbb{Y}} \Big(c(x,y) \dot{+} \big(- g(y) \big) \Big) \, , \ \forall x \in \mathbb{X} \, . \tag{3c}$$

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

$$f^{cc'}(x) = \big(f^c \big)^{c'}(x) = \sup_{y \in \mathbb{Y}} \Big(c(x,y) \dot{+} \big(- f^c(y) \big) \Big) \, , \ \forall x \in \mathbb{X} \, . \tag{3d}$$

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

$$f^{cc'}(x) \leq f(x) \, , \ \forall x \in \mathbb{X} \, . \tag{4}$$

With the notion of c-biconjugate, the classic notion of convex function is generalized.

[3] In particular, any closed convex function that takes at least one finite value is necessarily proper convex lsc.

Definition 2.2. A function $f : \mathbb{X} \to \overline{\mathbb{R}}$ is said to be *c-convex* it is equal to its c-biconjugate:

$$f \text{ is } c\text{-convex} \iff f^{cc'} = f \ . \tag{5}$$

In generalized convexity, it is established that c-convex functions are all functions of the form $g^{c'}$, for any $g : \mathbb{Y} \to \overline{\mathbb{R}}$, or, equivalently, all functions of the form $f^{cc'}$,

for any $f : \mathbb{X} \to \overline{\mathbb{R}}$ [9, 14, 15]. As an illustration, the $\star$-convex functions are the closed convex functions since, as recalled above, the Fenchel conjugacy induces a one-to-one correspondence between the closed convex functions on $\mathbb{X}$ and the closed convex functions on $\mathbb{Y}$.

2.2. One-sided linear couplings

Now, we introduce one-sided linear couplings, and we show that they induce conjugacies that share nice properties with the classic Fenchel conjugacy. In what follows, let $\mathbb{X}$ and $\mathbb{Y}$ be two paired vector spaces, and $\mathbb{W}$ be a set.

Definition 2.3. Let $\theta : \mathbb{W} \to \mathbb{X}$ be a mapping. We define the *one-sided linear coupling* c_θ between the set $\mathbb{W}$ and the vector space $\mathbb{Y}$ by[4]

$$c_\theta : \mathbb{W} \times \mathbb{Y} \to \overline{\mathbb{R}} \ , \ \ c_\theta(w, y) = \langle \theta(w) \, , y \rangle \ \ , \ \ \forall (w, y) \in \mathbb{W} \times \mathbb{Y} \ . \tag{6}$$

For any set $W \subset \mathbb{W}$, $\delta_W : \mathbb{W} \to \overline{\mathbb{R}}$ denotes the *characteristic function* of the set W:

$$\delta_W(w) = 0 \text{ if } w \in W \ , \ \ \delta_W(w) = +\infty \text{ if } w \notin W \ . \tag{7}$$

For any subset $X \subset \mathbb{X}$, $\sigma_X : \mathbb{Y} \to \overline{\mathbb{R}}$ denotes the *support function of the subset* X:

$$\sigma_X(y) = \sup_{x \in X} \langle x \, , y \rangle \ \ , \ \ \forall y \in \mathbb{Y} \ . \tag{8}$$

Now, we turn to the c_θ-conjugacy induced by the coupling c_θ. For this purpose, we introduce the notion of conditional infimum.

Definition 2.4. Let $\theta : \mathbb{W} \to \mathbb{X}$ be a mapping and let $h : \mathbb{W} \to \overline{\mathbb{R}}$ be a function. We define the *conditional infimum* (of the function h knowing the mapping θ) as the function $\inf[h \mid \theta] : \mathbb{X} \to \overline{\mathbb{R}}$ given by

$$\big(\inf[h \mid \theta]\big)(x) = \inf\{h(w) \mid w \in \mathbb{W} \ , \ \ \theta(w) = x\} \ \ , \ \ \forall x \in \mathbb{X} \ . \tag{9}$$

If $x \notin \theta(\mathbb{W})$, we get that $\big(\inf[h \mid \theta]\big)(x) = +\infty$ by the convention $\inf \emptyset = +\infty$. Therefore, regarding effective domains, we have the inclusion $\operatorname{dom}\big(\inf[h \mid \theta]\big) \subset \theta(\mathbb{W})$. The notation $\inf[h \mid \theta]$ comes from the analogy with a conditional expectation, and the expression "conditional infimum" is taken from [17]. The conditional infimum is also called epi-composition in [12, p. 27] and infimal postcomposition in [2, p. 214].

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

[4]In a one-sided linear coupling, the second set $\mathbb{Y}$ posesses a linear structure (and is even paired with a vector space by means of a bilinear form), whereas the first set $\mathbb{W}$ is not required to carry any structure.

Proposition 2.5. *For any function $g : \mathbb{Y} \to \overline{\mathbb{R}}$, the c'_θ-Fenchel-Moreau conjugate $g^{c_\theta{}'} : \mathbb{W} \to \overline{\mathbb{R}}$ is given by*

$$g^{c_\theta{}'} = g^{\star'} \circ \theta \; . \tag{10a}$$

For any function $h : \mathbb{W} \to \overline{\mathbb{R}}$, the c_θ-Fenchel-Moreau conjugate $h^{c_\theta} : \mathbb{Y} \to \overline{\mathbb{R}}$ is given by

$$h^{c_\theta} = \big(\inf [h \mid \theta] \big)^\star \; , \tag{10b}$$

and the c_θ-Fenchel-Moreau biconjugate $h^{c_\theta c_\theta{}'} : \mathbb{W} \to \overline{\mathbb{R}}$ is given by

$$h^{c_\theta c_\theta{}'} = \big(h^{c_\theta} \big)^{\star'} \circ \theta = \big(\inf [h \mid \theta] \big)^{\star\star'} \circ \theta \; . \tag{10c}$$

For any subset $W \subset \mathbb{W}$, we have

$$\delta_W^{c_\theta} = \sigma_{\theta(W)} \; . \tag{10d}$$

Proof. We prove (10a). Letting $w \in \mathbb{W}$, we have that

$$\begin{aligned}
\big(g\big)^{c'_\theta}(w) &= \sup_{y \in \mathbb{Y}} \Big(\langle \theta(w) \, , y \rangle \mathbin{\dot{+}} \big(- g(y) \big) \Big) \\
&\qquad \text{(by the conjugate formula (3a) and the coupling (6))} \\
&= g^{\star'}\big(\theta(w)\big) \; . \qquad \text{(by the expression (1b) of the Fenchel conjugate)}
\end{aligned}$$

We prove (10b). Letting $y \in \mathbb{Y}$, we have that

$$\begin{aligned}
h^{c_\theta}(y) &= \sup_{w \in \mathbb{W}} \Big(\langle \theta(w) \, , y \rangle \mathbin{\dot{+}} \big(- h(w) \big) \Big) \\
&\qquad \text{(by the conjugate formula (3a) and the coupling (6))} \\
&= \sup_{x \in \mathbb{X}} \sup_{w \in \mathbb{W}, \theta(w) = x} \Big(\langle \theta(w) \, , y \rangle \mathbin{\dot{+}} \big(- h(w) \big) \Big) \\
&= \sup_{x \in \mathbb{X}} \sup_{w \in \mathbb{W}, \theta(w) = x} \Big(\langle x \, , y \rangle \mathbin{\dot{+}} \big(- h(w) \big) \Big) \\
&= \sup_{x \in \mathbb{X}} \Big(\langle x \, , y \rangle \mathbin{\dot{+}} \sup_{w \in \mathbb{W}, \theta(w) = x} \big(- h(w) \big) \Big) \\
&\text{(since } \textstyle\sup_{w \in W}(r \mathbin{\dot{+}} \varphi(w)) = r \mathbin{\dot{+}} \sup_{w \in W} \varphi(w) \text{ for any } r \in \overline{\mathbb{R}}, \, W \subset \mathbb{W} \text{ and } \varphi : \mathbb{W} \to \overline{\mathbb{R}} \text{ [10])} \\
&= \sup_{x \in \mathbb{X}} \Big(\langle x \, , y \rangle \mathbin{\dot{+}} \big(- \inf_{w \in \mathbb{W}, \theta(w) = x} h(w) \big) \Big) \\
&= \sup_{x \in \mathbb{X}} \Big(\langle x \, , y \rangle \mathbin{\dot{+}} \big(- \big(\inf [h \mid \theta] \big)(x) \big) \Big) \\
&\qquad \text{(by the conditional infimum expression (9))} \\
&= \big(\inf [h \mid \theta] \big)^\star(y) \; . \qquad \text{(by the expression (1a) of the Fenchel conjugate)}
\end{aligned}$$

We prove (10c). Letting $w \in \mathbb{W}$, we have that

$$\begin{aligned}
h^{c_\theta c_\theta{}'}(w) &= \big(h^{c_\theta} \big)^{c'_\theta}(w) \qquad \text{(by the definition (3d) of the biconjugate)} \\
&= \big(\big(\inf [h \mid \theta] \big)^\star \big)^{c'_\theta}(w) \qquad \text{(by (10b))} \\
&= \big(\inf [h \mid \theta] \big)^{\star\star'}\big(\theta(w)\big) \; . \qquad \text{(by (10a))}
\end{aligned}$$

We prove (10d):

$$\begin{aligned}
\delta_W^{c_\theta} &= \big(\inf\left[\delta_W \mid \theta\right]\big)^\star && \text{(by (10b))}\\
&= \delta^\star_{\theta(W)} && \text{(because } \inf\left[\delta_W \mid \theta\right] = \delta_{\theta(W)} \text{ by (9) and (7))}\\
&= \sigma_{\theta(W)}\ . && \text{(by (1a), (7) and (8))}
\end{aligned}$$

This ends the proof. □

Now, we are able to characterize the so-called c_θ-convex functions (see Definition 2.2).

Proposition 2.6. *A function $h : \mathbb{W} \to \overline{\mathbb{R}}$ is c_θ-convex if and only if it is the composition of a closed convex function $f : \mathbb{X} \to \overline{\mathbb{R}}$ with the mapping $\theta : \mathbb{W} \to \mathbb{X}$. More precisely, for any function $h : \mathbb{W} \to \overline{\mathbb{R}}$, we have the equivalences*

$$h \text{ is } c_\theta\text{-convex} \tag{11a}$$

$$\iff h = h^{c_\theta c_\theta{}'} \tag{11b}$$

$$\iff h = \big(h^{c_\theta}\big)^{\star'} \circ \theta \text{ (where } \big(h^{c_\theta}\big)^{\star'} : \mathbb{X} \to \overline{\mathbb{R}} \text{ is a closed convex function)} \tag{11c}$$

$$\iff \text{there exists a closed convex function } f : \mathbb{X} \to \overline{\mathbb{R}} \text{ such that } h = f \circ \theta\ . \tag{11d}$$

Proof. The equivalence between (11a) and (11b) follows from Definition 2.2. The equivalence between (11b) and (11c) follows from (10c); Moreover, the function $\big(h^{c_\theta}\big)^{\star'}$ is closed convex since, as recalled in §2.1, the Fenchel conjugacy induces a one-to-one correspondence between the closed convex functions on $\mathbb{X}$ and the closed convex functions on $\mathbb{Y}$. Obviously, (11c) implies (11d).

Finally, there remains to prove that (11d) implies (11b). If there exists a closed convex function $f : \mathbb{X} \to \overline{\mathbb{R}}$ such that $h = f \circ \theta$, then $\inf\left[h \mid \theta\right] = f \dot{+} \delta_{\theta(\mathbb{W})}$ as easily computed, and therefore

$$h^{c_\theta c_\theta{}'} = \big(\inf\left[h \mid \theta\right]\big)^{\star\star'} \circ \theta = \big(f \dot{+} \delta_{\theta(\mathbb{W})}\big)^{\star\star'} \circ \theta$$

by (10c). Now, as $f \dot{+} \delta_{\theta(\mathbb{W})} \geq f$ by (7), we get that $\big(f \dot{+} \delta_{\theta(\mathbb{W})}\big)^{\star\star'} \geq f^{\star\star'} = f$, where the last equality holds because the function $f : \mathbb{X} \to \overline{\mathbb{R}}$ is closed convex. As a consequence, we obtain that $h^{c_\theta c_\theta{}'} \geq f \circ \theta = h$. Now, by (4), we always have the inequality $h^{c_\theta c_\theta{}'} \leq h$. Thus, we conclude that $h^{c_\theta c_\theta{}'} = h$.

This ends the proof. □

Let us say that *a function $h : \mathbb{W} \to \overline{\mathbb{R}}$ displays hidden convexity* with respect to the mapping $\theta : \mathbb{W} \to \mathbb{X}$ if there exists a closed convex function $f : \mathbb{X} \to \overline{\mathbb{R}}$ such that $h = f \circ \theta$. Then, we have just proved that this notion of hidden convexity for functions (see Footnote 1) coincides with the notion of c_θ-convex functions.

3. The E-Capra conjugacy and the ℓ_0 pseudonorm

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

consider the following *Euclidean unit sphere* $\mathbb{S}$ and *Euclidean unit ball* $\mathbb{B}$:

$$\mathbb{S} = \big\{x \in \mathbb{R}^d \mid \|x\| = 1\big\} \text{ and } \mathbb{B} = \big\{x \in \mathbb{R}^d \mid \|x\| \leq 1\big\} \ . \tag{12}$$

In § 3.1, we introduce the (Euclidean) constant along primal rays coupling ¢ (E-Capra). Then, we recall definitions of the ℓ_0 pseudonorm, and of the top-k and k-support norms in § 3.2. Finally, in § 3.3, we provide expressions for the E-Capra-conjugates and E-Capra-biconjugates of functions related to the ℓ_0 pseudonorm.

3.1. Euclidean Constant along primal rays coupling (E-Capra)

We introduce a novel coupling, which is a special case of one-sided linear coupling.

Definition 3.1. The *E-Capra coupling* ¢ between $\mathbb{R}^d$ and $\mathbb{R}^d$ is defined by

$$\forall y \in \mathbb{R}^d \ , \quad \begin{cases} ¢(x,y) & = \dfrac{\langle x \, , y \rangle}{\|x\|} = \dfrac{\langle x \, , y \rangle}{\sqrt{\langle x \, , x \rangle}} \ , \ \ \forall x \in \mathbb{R}^d \backslash \{0\} \ , \\ ¢(0,y) & = 0. \end{cases} \tag{13}$$

The coupling E-Capra has the property of being *constant along primal rays*, hence the acronym[5] E-Capra (Euclidean Constant Along Primal RAys Coupling). We introduce the primal *normalization mapping* n as follows:

$$n : \mathbb{R}^d \to \mathbb{S} \cup \{0\} \ , \ \ n(x) = \begin{cases} \frac{x}{\|x\|} & \text{if } \ x \neq 0 \ , \\ 0 & \text{if } \ x = 0 \ . \end{cases} \tag{14}$$

With these notations, the coupling E-Capra in (13) is a special case of one-sided linear coupling (see Definition 2.3): ¢ $= c_n$, as in (6) with mapping $\theta = n$, is the *Fenchel coupling after primal normalization*. The following Proposition — that provides expressions for the E-Capra-conjugates and E-Capra-biconjugates of a function — simply is Proposition 2.5 in the case where the mapping θ is the normalization mapping n in (14).

Proposition 3.2. *For any function $g : \mathbb{R}^d \to \overline{\mathbb{R}}$, the ¢′-Fenchel-Moreau conjugate $g^{¢'} : \mathbb{R}^d \to \overline{\mathbb{R}}$ is given by*

$$g^{¢'} = g^\star \circ n \ . \tag{15a}$$

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

$$f^{¢} = \big(\inf [f \mid n] \big)^\star \ , \tag{15b}$$

where the conditional infimum (9) *has the expression*

$$\big(\inf [f \mid n] \big)(x) = \inf \{f(x') \mid n(x') = x\} = \begin{cases} \inf_{\lambda > 0} f(\lambda x) & \text{if } x \in \mathbb{S} \cup \{0\} \ , \\ +\infty & \text{if } x \notin \mathbb{S} \cup \{0\} \ , \end{cases} \tag{15c}$$

[5] In fact, there is large class of couplings that are constant along primal rays. It suffices to replace the Euclidean norm in (13) with any norm. Such couplings are studied in [4, 5]. In this paper, we focus on the constant along primal rays coupling induced by the *Euclidean* norm, hence the acronym E-Capra.

and the ¢-Fenchel-Moreau biconjugate $f^{¢¢'} : \mathbb{R}^d \to \overline{\mathbb{R}}$ *is given by*

$$f^{¢¢'} = \big(f^{¢}\big)^{\star'} \circ n = \big(\inf[f \mid n]\big)^{\star\star'} \circ n \,. \tag{15d}$$

Thanks to Proposition 2.6, we easily deduce the following result.

Proposition 3.3. *A function on* $\mathbb{R}^d$ *is ¢-convex if and only if it is the composition of a closed convex function on* $\mathbb{R}^d$ *with the normalization mapping* (14)*. More precisely, for any function* $h : \mathbb{R}^d \to \overline{\mathbb{R}}$*, we have the equivalences*

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

Now, we turn to analyze the ℓ_0 pseudonorm by means of the E-Capra conjugacy.

3.2. The ℓ_0 pseudonorm, and the top-k and k-support norms

We recall definitions of the so-called ℓ_0 pseudonorm, and of the top-k and k-support norms.

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

$$\ell_0(x) = \big|\, \{j \in \{1, \ldots, d\} \mid x_j \neq 0\} \,\big| \,, \quad \forall x \in \mathbb{R}^d \,, \tag{17}$$

where $|K|$ denotes the cardinal of a subset $K \subset \{1, \ldots, d\}$. The ℓ_0 pseudonorm shares three out of the four axioms of a norm: nonnegativity, positivity except for $x = 0$, subadditivity. The axiom of 1-homogeneity does not hold true; by contrast, the ℓ_0 pseudonorm is 0-homogeneous as $\ell_0(\rho x) = \ell_0(x), \forall \rho \in \mathbb{R}\backslash\{0\}, \forall x \in \mathbb{R}^d$. Thus, the ℓ_0 pseudonorm displays the invariance property

$$\ell_0 \circ n = \ell_0 \tag{18}$$

with respect to the normalization mapping (14). This property will be instrumental to show that the ℓ_0 pseudonorm is a ¢-convex function.

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

$$\ell_0^{\leq k} = \big\{x \in \mathbb{R}^d \mid \ell_0(x) \leq k\big\} \,, \quad \forall k \in \{0, 1, \ldots, d\} \,, \tag{19a}$$

and the *level curves*

$$\ell_0^{=k} = \big\{x \in \mathbb{R}^d \mid \ell_0(x) = k\big\} \,, \quad \forall k \in \{0, 1, \ldots, d\} \,. \tag{19b}$$

For any subset $K \subset \{1, \ldots, d\}$, we denote the subspace of $\mathbb{R}^d$ made of vectors whose components vanish outside of K by[6]

$$\mathcal{R}_K = \mathbb{R}^K \times \{0\}^{-K} = \big\{x \in \mathbb{R}^d \mid x_j = 0 \,, \ \forall j \notin K\big\} \subset \mathbb{R}^d \,, \tag{20}$$

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

where $\mathcal{R}_\emptyset = \{0\}$. For any $x \in \mathbb{R}^d$ and $K \subset \{1, \ldots, d\}$, we denote by $x_K \in \mathbb{R}^d$ the vector which coincides with x, except for the components outside of K that vanish: x_K is the orthogonal projection of x onto the subspace $\mathcal{R}_K$.

The level sets of the ℓ_0 pseudonorm in (19a) are easily related to the subspaces $\mathcal{R}_K$ of $\mathbb{R}^d$, as defined in (20), by[7]

$$\ell_0^{\leq k} = \big\{x \in \mathbb{R}^d \mid \ell_0(x) \leq k\big\} = \bigcup_{|K| \leq k} \mathcal{R}_K \ , \ \ \forall k = 0, 1, \ldots, d \ . \tag{21}$$

The top-k and k-support norms.

Definition 3.4. For $k \in \{1, \ldots, d\}$, we define[8]

$$\|x\|_{(k)}^{\mathrm{tn}} = \sup_{|K| \leq k} \|x_K\| = \sup_{|K| = k} \|x_K\| \ , \ \ \forall x \in \mathbb{R}^d \ . \tag{22}$$

Thus defined, $\|\cdot\|_{(k)}^{\mathrm{tn}}$ is a norm, the so-called *top-k norm*. Its dual norm, as in (39a), denoted by[9] $\|\cdot\|_{(k)}^{\star\mathrm{sn}}$, is called the *k-support norm* [1]:

$$\|\cdot\|_{(k)}^{\star\mathrm{sn}} = \big(\|\cdot\|_{(k)}^{\mathrm{tn}}\big)_\star \ . \tag{23}$$

We follow the terminology of [16], where the top-k norm is also called the top-$(k, 1)$ norm. Indeed, 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 $\{1, \ldots, d\}$ such that $|x_{\sigma(1)}| \geq |x_{\sigma(2)}| \geq \cdots \geq |x_{\sigma(d)}|$, we have that

$$\|x\|_{(k)}^{\mathrm{tn}} = \sqrt{\sum_{l=1}^{k} |x_{\sigma(l)}|^2}.$$

The top-k norm is also known as the 2-*k-symmetric gauge norm*, or *Ky Fan vector norm*.

3.3. E-Capra-conjugates and biconjugates of the ℓ_0 pseudonorm

With the Fenchel conjugacy, we calculate that $\delta^\star_{\ell_0^{\leq k}} = \delta_{\{0\}}$ and $\delta^{\star\star'}_{\ell_0^{\leq k}} = 0$, for all $k = 1, \ldots, d$, and that $\ell_0^\star = \delta_{\{0\}}$ and $\ell_0^{\star\star'} = 0$.

Hence, the Fenchel conjugacy is not suitable to handle the ℓ_0 pseudonorm. We will now see that we obtain more interesting formulas with the E-Capra-conjugacy. Indeed, the ℓ_0 pseudonorm in (17), the characteristic functions $\delta_{\ell_0^{\leq k}}$ of its level sets (21) and the top-k norms in (22) are related by the following conjugate formulas. The proof relies on results gathered in the Appendix A.

[7]The notation $\bigcup_{|K| \leq k}$ is a shorthand for $\bigcup_{K \subset \{1,\ldots,d\}, |K| \leq k}$ (and the same for $\bigcup_{|K|=k}$).

[8]The notation $\sup_{|K| \leq k}$ is a shorthand for $\sup_{K \subset \{1,\ldots,d\}, |K| \leq k}$ (and the same for $\sup_{|K|=k}$). The property that $\sup_{|K| \leq k} \|x_K\| = \sup_{|K|=k} \|x_K\|$ in (22) comes from the easy observation that $K \subset K' \Rightarrow \|x_K\| \leq \|x_{K'}\|$.

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

Theorem 3.5. *Let $\¢$ be the Euclidean coupling E-Capra as defined in* (13)*. Let $k \in \{0, 1, \dots, d\}$. We have that* (*with the convention, in* (24a) *and in* (24c)*, that* $\|\cdot\|^{\mathrm{tn}}_{(0)} = 0$

$$\delta^{-\¢}_{\ell_0^{\leq k}} = \delta^{\¢}_{\ell_0^{\leq k}} = \|\cdot\|^{\mathrm{tn}}_{(k)} \,, \tag{24a}$$

$$\delta^{\¢\¢'}_{\ell_0^{\leq k}} = \delta_{\ell_0^{\leq k}} \,, \tag{24b}$$

$$\ell_0^{\¢} = \sup_{l=0,1,\dots,d} \left[\|\cdot\|^{\mathrm{tn}}_{(l)} - l \right] , \tag{24c}$$

$$\ell_0^{\¢\¢'} = \ell_0 \,. \tag{24d}$$

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 Euclidean norm $\|\cdot\| = \sqrt{\langle \cdot\,, \cdot \rangle}$.

• We prove the first equality in (24a):

$$\begin{aligned}
\delta^{-\¢}_{\ell_0^{\leq k}} &= \sigma_{-n(\ell_0^{\leq k})} && \text{(by (10d) because } -\¢ = c_{-n} \text{ in (6))} \\
&= \sigma_{n(\ell_0^{\leq k})} \\
&\qquad \text{(by symmetry of the set } \ell_0^{\leq k} \text{ in (19a) and of the mapping } n \text{ in (14))} \\
&= \delta^{\¢}_{\ell_0^{\leq k}} \,. && \text{(by (10d))}
\end{aligned}$$

We now turn to prove the second equality in (24a):

$$\begin{aligned}
\delta^{-\¢}_{\ell_0^{\leq k}} &= \sigma_{n(\ell_0^{\leq k})} && \text{(by (10d))} \\
&= \sigma_{(\ell_0^{\leq k} \cap \mathbb{S}) \cup \{0\}} && \text{(by the expression (14) of the normalization mapping } n) \\
&= \sup\{\sigma_{\ell_0^{\leq k} \cap \mathbb{S}}, 0\} \\
&\qquad \text{(as the support function turns a union of sets into a supremum)} \\
&= \sup\{\sigma_{\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S})}, 0\} && \text{(as } \ell_0^{\leq k} \cap \mathbb{S} = \textstyle\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S}) \text{ by (21))} \\
&= \sup\{ \sup_{|K| \leq k} \sigma_{(\mathcal{R}_K \cap \mathbb{S})}, 0\} \\
&\qquad \text{(as the support function turns a union of sets into a supremum)} \\
&= \sup\{\|\cdot\|^{\mathrm{tn}}_{(k)}, 0\} && \text{(as } \sup_{|K| \leq k} \sigma_{(\mathcal{R}_K \cap \mathbb{S})} = \|\cdot\|^{\mathrm{tn}}_{(k)} \text{ by (43))} \\
&= \|\cdot\|^{\mathrm{tn}}_{(k)} \,.
\end{aligned}$$

• Before proving (24b), observe that, by definition (14) of the normalization mapping n, we have:

$$0 \in D \subset \mathbb{R}^d \Rightarrow n^{-1}(D) = n^{-1}\big((\{0\} \cup \mathbb{S}) \cap D\big) = \{0\} \cup n^{-1}(\mathbb{S} \cap D) \,. \tag{25}$$

• We prove (24b):

$$\begin{aligned}
\delta^{\¢\¢'}_{\ell_0^{\leq k}} &= \big(\delta^{\¢}_{\ell_0^{\leq k}}\big)^\star \circ n && \text{(by the formula (15d) for the biconjugate)} \\
&= \big(\|\cdot\|^{\mathrm{tn}}_{(k)}\big)^\star \circ n && \text{(by (24a))} \\
&= \big(\sigma_{\mathbb{B}^{\star \mathrm{sn}}_{(k)}}\big)^\star \circ n
\end{aligned}$$

(by (39b), that expresses a norm as the support function of the unit ball of the dual norm)

$$= \delta_{\mathbb{B}^{\star\mathrm{sn}}_{(k)}} \circ n$$

(as $\left(\sigma_{\mathbb{B}^{\star\mathrm{sn}}_{(k)}}\right)^\star = \delta_{\mathbb{B}^{\star\mathrm{sn}}_{(k)}}$ since $\mathbb{B}^{\star\mathrm{sn}}_{(k)}$ is closed convex [13, Theorem 13.2])

$$= \delta_{n^{-1}(\mathbb{B}^{\star\mathrm{sn}}_{(k)})}$$ (by the definition (7) of a characteristic function)

$$= \delta_{\{0\}\cup n^{-1}(\mathbb{B}^{\star\mathrm{sn}}_{(k)}\cap\mathbb{S})}$$ (by (25) since $0 \in \mathbb{B}^{\star\mathrm{sn}}_{(k)}$)

$$= \delta_{\{0\}\cup n^{-1}(\ell_0^{\leq k}\cap\mathbb{S})}$$ (as $\mathbb{B}^{\star\mathrm{sn}}_{(k)} \cap \mathbb{S} = \ell_0^{\leq k} \cap \mathbb{S}$ by (54a))

$$= \delta_{n^{-1}(\ell_0^{\leq k})}$$ (by (25) since $0 \in \ell_0^{\leq k}$)

$$= \delta_{\ell_0^{\leq k}} \, .$$ (as $\ell_0 \circ n = \ell_0$ by (18))

• We prove (24c):

$$\ell_0^{¢} = \Big(\inf_{l=0,1,\ldots,d} \{\delta_{\ell_0^{=l}} \dot{+} l\} \Big)^{¢}$$

(since $\ell_0 = \inf_{l=0,1,\ldots,d}\{\delta_{\ell_0^{=l}} \dot{+} l\}$ by using the level curves (19b))

$$= \sup_{l=0,1,\ldots,d} \{\delta^{¢}_{\ell_0^{=k}} \dot{+} (-l)\}$$

(as conjugacies, being dualities, turn infima into suprema)

$$= \sup_{l=0,1,\ldots,d} \{\sigma_{n(\ell_0^{=l})} \dot{+} (-l)\}$$ (as $\delta^{¢}_{\ell_0^{=k}} = \sigma_{n(\ell_0^{=l})}$ by (10d))

$$= \sup\Big\{0, \sup_{l=1,\ldots,d} \{\sigma_{\ell_0^{=l}\cap\mathbb{S}} \dot{+} (-l)\}\Big\}$$

(as $\sigma_{\{0\}} = 0$ and $n(\ell_0^{=l}) = \ell_0^{=l} \cap \mathbb{S}$ when $l \geq 1$ by (14))

$$= \sup\Big\{0, \sup_{l=1,\ldots,d} \{\sigma_{\overline{\ell_0^{=l}\cap\mathbb{S}}} \dot{+} (-l)\}\Big\}$$

(as $\sigma_X = \sigma_{\overline{X}}$ for any $X \subset \mathbb{R}^d$ [8, Proposition 2.2.1])

$$= \sup\Big\{0, \sup_{l=1,\ldots,d} \{\sigma_{\ell_0^{\leq l}\cap\mathbb{S}} \dot{+} (-l)\}\Big\}$$ (as $\overline{\ell_0^{=l} \cap \mathbb{S}} = \ell_0^{\leq l} \cap \mathbb{S}$ by (54b))

$$= \sup\Big\{0, \sup_{l=1,\ldots,d} \{\sigma_{\cup_{|K|\leq l}(\mathcal{R}_K\cap\mathbb{S})} \dot{+} (-l)\}\Big\}$$ (as $\ell_0^{\leq l} \cap \mathbb{S} = \cup_{|K|\leq l}(\mathcal{R}_K \cap \mathbb{S})$ by (21))

$$= \sup\Big\{0, \sup_{l=1,\ldots,d} \{\sup_{|K|\leq l} \sigma_{\mathcal{R}_K\cap\mathbb{S}} \dot{+} (-l)\}\Big\}$$

(as the support function turns a union of sets into a supremum)

$$= \sup\Big\{0, \sup_{l=1,\ldots,d} \Big[\|y\|^{\mathrm{tn}}_{(l)} - l\Big]\Big\}$$ (as $\sup_{|K|\leq l} \sigma_{\mathcal{R}_K\cap\mathbb{S}} = \|\cdot\|^{\mathrm{tn}}_{(l)}$ by (43))

$$= \sup_{l=0,1,\ldots,d} \Big[\|y\|^{\mathrm{tn}}_{(l)} - l\Big] \, .$$ (using the convention that $\|\cdot\|^{\mathrm{tn}}_{(0)} = 0$)

• We prove (24d). It is easy to check that $\ell_0^{¢¢'}(0) = 0 = \ell_0(0)$. Therefore, let $x \in \mathbb{R}^d\backslash\{0\}$ be given and assume that $\ell_0(x) = l \in \{1, \ldots, d\}$. We consider the mapping $\phi :]0, +\infty[\to \mathbb{R}$ defined by

$$\phi(\lambda) = \frac{\langle x \, , \lambda x\rangle}{\|x\|} + \Big(- \sup\Big\{0, \sup_{j=1,\ldots,d} \Big[\|\lambda x\|^{\mathrm{tn}}_{(j)} - j\Big]\Big\}\Big) \, , \quad \forall \lambda > 0 \, , \tag{26}$$

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

$$\begin{aligned}
\phi(\lambda) &= \lambda\|x\| + \Big(-\sup\Big\{0, \sup_{j=1,\dots,d}\Big[\|\lambda x\|^{\mathrm{tn}}_{(j)} - j\Big]\Big\}\Big) && \text{(by definition (26) of } \phi\text{)}\\
&= \lambda\|x\|^{\mathrm{tn}}_{(l)} + \inf\Big\{0, -\sup_{j=1,\dots,d}\Big[\lambda\|x\|^{\mathrm{tn}}_{(j)} - j\Big]\Big\} \\
& && \text{(as } \|x\| = \|x\|^{\mathrm{tn}}_{(l)} \text{ when } \ell_0(x) = l \text{ by (53a))}\\
&= \inf\Big\{\lambda\|x\|^{\mathrm{tn}}_{(l)}, \lambda\|x\|^{\mathrm{tn}}_{(l)} + \inf_{j=1,\dots,d}\Big(-\Big[\lambda\|x\|^{\mathrm{tn}}_{(j)} - j\Big]\Big)\Big\}\\
&= \inf\Big\{\lambda\|x\|^{\mathrm{tn}}_{(l)}, \inf_{j=1,\dots,d}\Big(\lambda\big(\|x\|^{\mathrm{tn}}_{(l)} - \|x\|^{\mathrm{tn}}_{(j)}\big) + j\Big)\Big\}\\
&= \inf\Big\{\lambda\|x\|^{\mathrm{tn}}_{(l)}, \inf_{j=1,\dots,l-1}\Big(\lambda\big(\|x\|^{\mathrm{tn}}_{(l)} - \|x\|^{\mathrm{tn}}_{(j)}\big) + j\Big), \inf_{j=l,\dots,d}\Big(\lambda\big(\|x\|^{\mathrm{tn}}_{(l)} - \|x\|^{\mathrm{tn}}_{(j)}\big) + j\Big)\Big\}\\
&= \inf\Big\{\lambda\|x\|^{\mathrm{tn}}_{(l)}, \inf_{j=1,\dots,l-1}\Big(\lambda\big(\|x\|^{\mathrm{tn}}_{(l)} - \|x\|^{\mathrm{tn}}_{(j)}\big) + j\Big), l\Big\}
\end{aligned}$$

as $\|x\|^{\mathrm{tn}}_{(j)} = \|x\|^{\mathrm{tn}}_{(l)}$ for $j \geq l$ by (53a). Let us show that the two first terms in the infimum go to $+\infty$ when $\lambda \to +\infty$. The first term $\lambda\|x\|^{\mathrm{tn}}_{(l)}$ goes to $+\infty$ because $\|x\|^{\mathrm{tn}}_{(l)} = \|x\| > 0$ by assumption ($x \neq 0$). The second term

$$\inf_{j=1,\dots,l-1}\Big(\lambda\big(\|x\|^{\mathrm{tn}}_{(l)} - \|x\|^{\mathrm{tn}}_{(j)}\big) + j\Big)$$

also goes to $+\infty$ because $\ell_0(x) = l$, so that $\|x\| = \|x\|^{\mathrm{tn}}_{(l)} > \|x\|^{\mathrm{tn}}_{(j)}$ for $j = 1, \dots, l-1$ by (53a). Therefore, $\lim_{\lambda\to+\infty}\phi(\lambda) = \inf\{+\infty, +\infty, l\} = l$. This ends the proof since

$$\begin{aligned}
l = \lim_{\lambda\to+\infty}\phi(\lambda) &\leq \sup_{y\in\mathbb{R}^d}\left(\frac{\langle x\,,y\rangle}{\|x\|} \dot{+} \Big(-\sup\Big\{0, \sup_{j=1,\dots,d}\Big[\|y\|^{\mathrm{tn}}_{(j)} - j\Big]\Big\}\Big)\right)\\
& && \text{(by definition (26) of } \phi\text{)}\\
&= \sup_{y\in\mathbb{R}^d}\left(\frac{\langle x\,,y\rangle}{\|x\|} \dot{+} \Big(-\sup_{j=0,1,\dots,d}\Big[\|y\|^{\mathrm{tn}}_{(j)} - j\Big]\Big)\right)\\
& && \text{(by the convention } \|\cdot\|^{\mathrm{tn}}_{(0)} = 0\text{)}\\
&= \sup_{y\in\mathbb{R}^d}\left(\frac{\langle x\,,y\rangle}{\|x\|} \dot{+} \Big(-\ell_0^{\natural}(y)\Big)\right) && \text{(by the formula (24c) for } \ell_0^{\natural}\text{)}\\
&= \ell_0^{\natural\natural'}(x) && \text{(by the biconjugate formula (3d))}\\
&\leq \ell_0(x) && \text{(by (4) giving } \ell_0^{\natural\natural'} \leq \ell_0\text{)}\\
&= l\,. && \text{(by assumption)}
\end{aligned}$$

Therefore, we have obtained $l = \ell_0^{\natural\natural'}(x) = \ell_0(x)$.
This ends the proof. □

In the next Section, we present a (rather unexpected) consequence of the just established property that $\ell_0^{\natural\natural'} = \ell_0$.

4. Hidden convexity in the pseudonorm ℓ_0

In §4.1, we show that there exists a proper convex lsc function on $\mathbb{R}^d$ which takes the same values as the ℓ_0 pseudonorm on the Euclidean unit sphere $\mathbb{S}$. This property of hidden convexity somehow comes as a surprise as the ℓ_0 pseudonorm is a highly nonconvex function of combinatorial nature. Then, we provide various expressions for the underlying proper convex lsc function and, in §4.2, we display mathematical expressions and graphical representations in the two-dimensional case.

4.1. Hidden convexity in the pseudonorm ℓ_0

We introduce the function $\mathcal{L}_0 : \mathbb{R}^d \to \overline{\mathbb{R}}$ defined by

$$\mathcal{L}_0 = \Big(\sup_{l=0,1,\dots,d} \Big[\| \cdot \|_{(l)}^{\mathrm{tn}} - l \Big] \Big)^{\star'} . \tag{27}$$

Theorem 4.1. *The function $\mathcal{L}_0$ in (27) is a proper convex lsc function on $\mathbb{R}^d$. The pseudonorm ℓ_0 coincides, on the Euclidean unit sphere $\mathbb{S}$ of $\mathbb{R}^d$, with the function $\mathcal{L}_0$, that is,*

$$\ell_0(x) = \mathcal{L}_0(x) \ , \ \ \forall x \in \mathbb{S} \, . \tag{28}$$

As a consequence, the pseudonorm ℓ_0 displays hidden convexity, as it can be expressed as the composition of the proper convex lsc function $\mathcal{L}_0$ in (27) with the normalization mapping n in (14):

$$\ell_0(x) = \mathcal{L}_0\Big(\frac{x}{\|x\|} \Big) \ , \ \ \forall x \in \mathbb{R}^d \backslash \{0\} \, . \tag{29}$$

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

$$\mathcal{L}_0\big((x_1, \dots, x_d)\big) = \mathcal{L}_0\big((|x_1|, \dots, |x_d|)\big) \ , \ \ \forall (x_1, \dots, x_d) \in \mathbb{R}^d \, . \tag{30}$$

Proof. First, it is easily seen that the closed convex function $\mathcal{L}_0$ in (27) is proper lsc (see Footnote 3).

Second, we prove (28). For $x \in \mathbb{S}$, we have

$$\begin{aligned}
\ell_0(x) &= \ell_0^{\natural\natural'}(x) && \text{(by (24d))} \\
&= \sup_{y \in \mathbb{R}^d} \Big(\natural(x,y) \dot{+} \big(- \ell_0^{\natural}(y) \big) \Big) && \text{(by the biconjugate formula (3d))} \\
&= \sup_{y \in \mathbb{R}^d} \Big(\langle x \, , y \rangle \dot{+} \big(- \ell_0^{\natural}(y) \big) \Big) && \text{(by (13) with } \|x\| = 1 \text{ since } x \in \mathbb{S}) \\
&= \sup_{y \in \mathbb{R}^d} \Big(\langle x \, , y \rangle \dot{+} \Big(- \big(\sup_{l=0,1,\dots,d} \Big[\|y\|_{(l)}^{\mathrm{tn}} - l \Big] \big) \Big) \Big) && \text{(by (24c))} \\
&= \Big(\sup_{l=0,1,\dots,d} \Big[\|y\|_{(l)}^{\mathrm{tn}} - l \Big] \Big)^{\star'}(x) \\
& && \text{(by the expression (1a) of the Fenchel conjugate)} \\
&= \mathcal{L}_0(x) \, . && \text{(by (27))}
\end{aligned}$$

Third, the equality (29) is an easy consequence of the property (18) that the pseudonorm ℓ_0 is invariant along any open ray of $\mathbb{R}^d$.

Fourth, we prove (30). For this purpose, we take any $\epsilon \in \{-1,1\}^d$ and we consider the symmetry $\tilde{\epsilon}$ of $\mathbb{R}^d$, defined by $\tilde{\epsilon}(x_1,\ldots,x_d) = (\epsilon_1 x_1,\ldots,\epsilon_d x_d)$, for all $(x_1,\ldots,x_d) \in \mathbb{R}^d$. We will show that the proper convex lsc function $\mathcal{L}_0$ is invariant under the symmetry $\tilde{\epsilon}$, hence satisfies (30). Indeed, for any $x \in \mathbb{R}^d$, we have

$$\begin{aligned}
\mathcal{L}_0(\tilde{\epsilon}x) &= \Big(\sup_{l=0,1,\ldots,d} \Big[\|\cdot\|^{\mathrm{tn}}_{(l)} - l \Big] \Big)^{\star'}(\tilde{\epsilon}x) && \text{(by (27))} \\
&= \sup_{y\in\mathbb{R}^d} \Big(\langle \tilde{\epsilon}x \,, y\rangle \dot{+} \Big(- \big(\sup_{l=0,1,\ldots,d} \Big[\|y\|^{\mathrm{tn}}_{(l)} - l \Big] \big) \Big) \Big) \\
& && \text{(by the expression (3c) of the reverse Fenchel conjugate)} \\
&= \sup_{y\in\mathbb{R}^d} \Big(\langle x \,, \tilde{\epsilon}y\rangle \dot{+} \Big(- \big(\sup_{l=0,1,\ldots,d} \Big[\|y\|^{\mathrm{tn}}_{(l)} - l \Big] \big) \Big) \Big) && \text{(as easily seen)} \\
&= \sup_{y\in\mathbb{R}^d} \Big(\langle x \,, \tilde{\epsilon}y\rangle \dot{+} \Big(- \big(\sup_{l=0,1,\ldots,d} \Big[\|\tilde{\epsilon}y\|^{\mathrm{tn}}_{(l)} - l \Big] \big) \Big) \Big)
\end{aligned}$$

as $\|\cdot\|^{\mathrm{tn}}_{(0)} \equiv 0$ (by convention) and all norms $\|\cdot\|^{\mathrm{tn}}_{(l)}, l = 1,\ldots,d$, are invariant under the symmetry $\tilde{\epsilon}$

$$\begin{aligned}
&= \Big(\sup_{l=0,1,\ldots,d} \Big[\|\cdot\|^{\mathrm{tn}}_{(l)} - l \Big] \Big)^{\star'}(x) && \text{(as } \tilde{\epsilon}^{-1}(\mathbb{R}^d) = \mathbb{R}^d\text{)} \\
&= \mathcal{L}_0(x) \,. && \text{(by (27))}
\end{aligned}$$

This ends the proof. □

Now, we provide three expressions for the proper convex lsc function $\mathcal{L}_0$ in (27).

Proposition 4.2. *The proper convex lsc function $\mathcal{L}_0$ in (27) can be characterized*

- *either by its epigraph*

$$\operatorname{epi} \mathcal{L}_0 = \overline{\operatorname{co}}\Big(\bigcup_{l=0}^{d} \mathbb{B}^{\star\mathrm{sn}}_{(l)} \times [l, +\infty[\Big) \,, \tag{31}$$

  where $\mathbb{B}^{\star\mathrm{sn}}_{(0)} = \{0\}$ (by convention) and $\mathbb{B}^{\star\mathrm{sn}}_{(1)} \subset \cdots \subset \mathbb{B}^{\star\mathrm{sn}}_{(l-1)} \subset \mathbb{B}^{\star\mathrm{sn}}_{(l)} \subset \cdots \subset \mathbb{B}^{\star\mathrm{sn}}_{(d)} = \mathbb{B}$ denote the unit balls associated with the l-support norms defined in (23) for $l = 1,\ldots,d$,

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

$$L_0(x) = \begin{cases} 0 & \text{if } x = 0, \\ l & \text{if } x \in \mathbb{B}^{\star\mathrm{sn}}_{(l)} \backslash \mathbb{B}^{\star\mathrm{sn}}_{(l-1)} \,, \quad l = 1,\ldots,d, \\ +\infty & \text{if } x \notin \mathbb{B}^{\star\mathrm{sn}}_{(d)} = \mathbb{B}, \end{cases} \tag{32}$$

- *or also by the expression*

$$\mathcal{L}_0(x) = \min_{\substack{x^{(1)}\in\mathbb{R}^d,\ldots,x^{(d)}\in\mathbb{R}^d \\ \sum_{l=1}^d \|x^{(l)}\|^{\star\mathrm{sn}}_{(l)} \le 1 \\ \sum_{l=1}^d x^{(l)} = x}} \sum_{l=1}^{d} l \|x^{(l)}\|^{\star\mathrm{sn}}_{(l)} \,, \quad \forall x \in \mathbb{R}^d \,. \tag{33}$$

Proof.

• First, we prove that the epigraph of $\mathcal{L}_0$ in (27) is given by (31). Indeed, we have

$$\begin{aligned}
\operatorname{epi} \mathcal{L}_0 &= \operatorname{epi}\Big(\sup_{l=0,1,\dots,d} \Big[\|\cdot\|_{(l)}^{\mathrm{tn}} - l \Big] \Big)^{\star'} && \text{(by (27))} \\
&= \overline{\operatorname{co}}\Big(\bigcup_{l=0}^{d} \operatorname{epi} \Big[\|\cdot\|_{(l)}^{\mathrm{tn}} - l \Big]^{\star'} \Big) && \text{(by [13, Theorem 16.5])} \\
&= \overline{\operatorname{co}}\Big(\bigcup_{l=0}^{d} \operatorname{epi} \Big[\sigma_{\mathbb{B}_{(l)}^{\star \mathrm{sn}}} - l \Big]^{\star'} \Big) && \text{(by (52))} \\
&= \overline{\operatorname{co}}\Big(\bigcup_{l=0}^{d} \operatorname{epi} \Big[\delta_{\mathbb{B}_{(l)}^{\star \mathrm{sn}}} + l \Big] \Big) && \text{(as } \Big[\sigma_{\mathbb{B}_{(l)}^{\star \mathrm{sn}}} - l \Big]^{\star'} = \delta_{\mathbb{B}_{(l)}^{\star \mathrm{sn}}} + l \text{)} \\
&= \overline{\operatorname{co}}\Big(\bigcup_{l=0}^{d} \mathbb{B}_{(l)}^{\star \mathrm{sn}} \times [l, +\infty[\Big) \, . && \text{(as is easily concluded)}
\end{aligned}$$

• Second, we prove that the function $\mathcal{L}_0$ in (27) is the largest proper convex lsc function below the function L_0 defined by (32). Indeed, we have that

$$\begin{aligned}
\mathcal{L}_0 &= \Big(\sup_{l=0,1,\dots,d} \Big[\|\cdot\|_{(l)}^{\mathrm{tn}} - l \Big] \Big)^{\star'} && \text{(by (27))} \\
&= \Big(\sup_{l=0,1,\dots,d} \Big[\sigma_{\mathbb{B}_{(l)}^{\star \mathrm{sn}}} - l \Big] \Big)^{\star'} && \text{(by (52))} \\
&= \Big(\sup_{l=0,1,\dots,d} \Big[\delta_{\mathbb{B}_{(l)}^{\star \mathrm{sn}}} + l \Big]^{\star} \Big)^{\star'} && \text{(as } \Big[\delta_{\mathbb{B}_{(l)}^{\star \mathrm{sn}}} + l \Big]^{\star} = \sigma_{\mathbb{B}_{(l)}^{\star \mathrm{sn}}} - l \text{)} \\
&= \Big(\Big[\inf_{l=0,1,\dots,d} \Big[\delta_{\mathbb{B}_{(l)}^{\star \mathrm{sn}}} + l \Big] \Big]^{\star} \Big)^{\star'} \\
& && \text{(as conjugacies, being dualities, turn infima into suprema)} \\
&= \Big(\inf_{l=0,1,\dots,d} \Big[\delta_{\mathbb{B}_{(l)}^{\star \mathrm{sn}}} + l \Big] \Big)^{\star\star'} && \text{(by definition (1c) of the Fenchel biconjugate)} \\
&= L_0^{\star\star'}
\end{aligned}$$

as it is easy to establish that the function $\inf_{l=0,1,\dots,d} \Big[\delta_{\mathbb{B}_{(l)}^{\star \mathrm{sn}}} + l \Big]$ coincides with the function L_0 defined by (32). Indeed, it is deduced from (51) that

$$\{0\} = \mathbb{B}_{(0)}^{\star \mathrm{sn}} \subset \mathbb{B}_{(1)}^{\star \mathrm{sn}} \subset \cdots \subset \mathbb{B}_{(l-1)}^{\star \mathrm{sn}} \subset \mathbb{B}_{(l)}^{\star \mathrm{sn}} \subset \cdots \subset \mathbb{B}_{(d)}^{\star \mathrm{sn}} = \mathbb{B}.$$

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 $\mathcal{L}_0$ in (27) is given by (33). For this purpose, we use a general formula [18, Corollary 2.8.11] for the Fenchel conjugate of the supremum of proper convex functions $f_l : \mathbb{R}^d \to \overline{\mathbb{R}}$, $l = 0, 1, \dots, d$:

$$\bigcap_{l=0,1,\dots,d} \operatorname{dom} f_l \neq \emptyset \Rightarrow \big(\sup_{l=0,1,\dots,d} f_l \big)^{\star} = \min_{\lambda \in \Delta_{d+1}} \Big(\sum_{l=0}^{d} \lambda_l f_l \Big)^{\star} , \tag{34}$$

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

$$\begin{aligned}
\mathcal{L}_0 &= \Big(\sup_{l=0,1,\dots,d} \Big[\|\cdot\|_{(l)}^{\mathrm{tn}} - l \Big] \Big)^{\star'} && \text{(by (27))} \\
&= \Big(\sup_{l=0,1,\dots,d} \Big[\sigma_{\mathbb{B}_{(l)}^{\star\mathrm{sn}}} - l \Big] \Big)^{\star'} && \text{(by (52))} \\
&= \min_{\lambda \in \Delta_{d+1}} \Big(\sum_{l=0}^{d} \lambda_l \Big[\sigma_{\mathbb{B}_{(l)}^{\star\mathrm{sn}}} - l \Big] \Big)^{\star} && \text{(by (34))} \\
&= \min_{\lambda \in \Delta_{d+1}} \Big(\sigma_{\sum_{l=0}^{d} \lambda_l \mathbb{B}_{(l)}^{\star\mathrm{sn}}} - \sum_{l=0}^{d} \lambda_l l \Big)^{\star}
\end{aligned}$$

as, for all $l = 0, \dots, d$, $\lambda_l \sigma_{\mathbb{B}_{(l)}^{\star\mathrm{sn}}} = \sigma_{\lambda_l \mathbb{B}_{(l)}^{\star\mathrm{sn}}}$ 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 [13, p. 113]

$$= \min_{\lambda \in \Delta_{d+1}} \Big(\delta_{\sum_{l=0}^{d} \lambda_l \mathbb{B}_{(l)}^{\star\mathrm{sn}}} + \sum_{l=0}^{d} \lambda_l l \Big) \, .$$

(as $\Big[\sigma_C - t\Big]^{\star'} = \delta_C + t$ for any closed convex subset C)

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

$$\mathcal{L}_0(x) = \min_{\substack{\lambda \in \Delta_{d+1} \\ x \in \sum_{l=0}^{d} \lambda_l \mathbb{B}_{(l)}^{\star\mathrm{sn}}}} \sum_{l=0}^{d} \lambda_l l \, , \tag{35a}$$

$$= \min_{\substack{z^{(1)} \in \mathbb{B}_{(1)}^{\star\mathrm{sn}}, \dots, z^{(d)} \in \mathbb{B}_{(d)}^{\star\mathrm{sn}} \\ \lambda_1 \geq 0, \dots, \lambda_d \geq 0 \\ \sum_{l=1}^{d} \lambda_l \leq 1 \\ \sum_{l=1}^{d} \lambda_l z^{(l)} = x}} \sum_{l=1}^{d} \lambda_l l \tag{35b}$$

by ignoring $\lambda_0 \geq 0$ since $\mathbb{B}_{(0)}^{\star\mathrm{sn}} = \{0\}$ by convention

$$= \min_{\substack{s^{(1)} \in \mathbb{S}_{(1)}^{\star\mathrm{sn}}, \dots, s^{(d)} \in \mathbb{S}_{(d)}^{\star\mathrm{sn}} \\ \mu_1 \geq 0, \dots, \mu_d \geq 0 \\ \sum_{l=1}^{d} \mu_l \leq 1 \\ \sum_{l=1}^{d} \mu_l s^{(l)} = x}} \sum_{l=1}^{d} \mu_l l \tag{35c}$$

where $\mathbb{S}_{(l)}^{\star\mathrm{sn}}$ is the unit sphere of the l-support norm $\|\cdot\|_{(l)}^{\star\mathrm{sn}}$, and the inequality $\leq$ is obvious as $\mathbb{S}_{(l)}^{\star\mathrm{sn}} \subset \mathbb{B}_{(l)}^{\star\mathrm{sn}}$ for all $l = 1, \dots, d$; the inequality $\geq$ comes from putting, for $l = 1, \dots, d$, $\mu_l = \lambda_l \|z^{(l)}\|_{(l)}^{\star\mathrm{sn}}$ and observing that

(i) there exist $s^{(l)} \in \mathbb{S}_{(l)}^{\star\mathrm{sn}}$ such that $\lambda_l z^{(l)} = \mu_l s^{(l)}$ (take any $s^{(l)}$ when $z^{(l)} = 0$ and $s^{(l)} = \frac{z^{(l)}}{\|z^{(l)}\|_{(l)}^{\star\mathrm{sn}}}$ when $z^{(l)} \neq 0$)

(ii) $\sum_{l=1}^d \lambda_l l \geq \sum_{l=1}^d \lambda_l \|z^{(l)}\|_{(l)}^{\star \mathrm{sn}} l = \sum_{l=1}^d \mu_l l$ because $\|z^{(l)}\|_{(l)}^{\star \mathrm{sn}} \leq 1$

$$= \min_{\substack{x^{(1)} \in \mathbb{R}^d, \ldots, x^{(d)} \in \mathbb{R}^d \\ \sum_{l=1}^d \|x^{(l)}\|_{(l)}^{\star \mathrm{sn}} \leq 1 \\ \sum_{l=1}^d x^{(l)} = x}} \sum_{l=1}^d \|x^{(l)}\|_{(l)}^{\star \mathrm{sn}} l \,, \tag{35d}$$

by putting $x^{(l)} = \mu_l s^{(l)}$, for all $l = 1, \ldots, d$. This ends the proof. □

With Proposition 4.2, we dispose of expressions that make it possible to obtain more involved formulas for the function $\mathcal{L}_0$ in (27). In particular, 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 function $\mathcal{L}_0$ on $\mathbb{R}^2$

In dimension $d = 1$, it is easily computed that the function $\mathcal{L}_0$ in (27) 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 one-dimensional unit sphere $\{-1, 1\}$ — but also with any convex function taking the value 1 on $\{-1, 1\}$ (the function $|\cdot|$, the constant function 1, etc.).

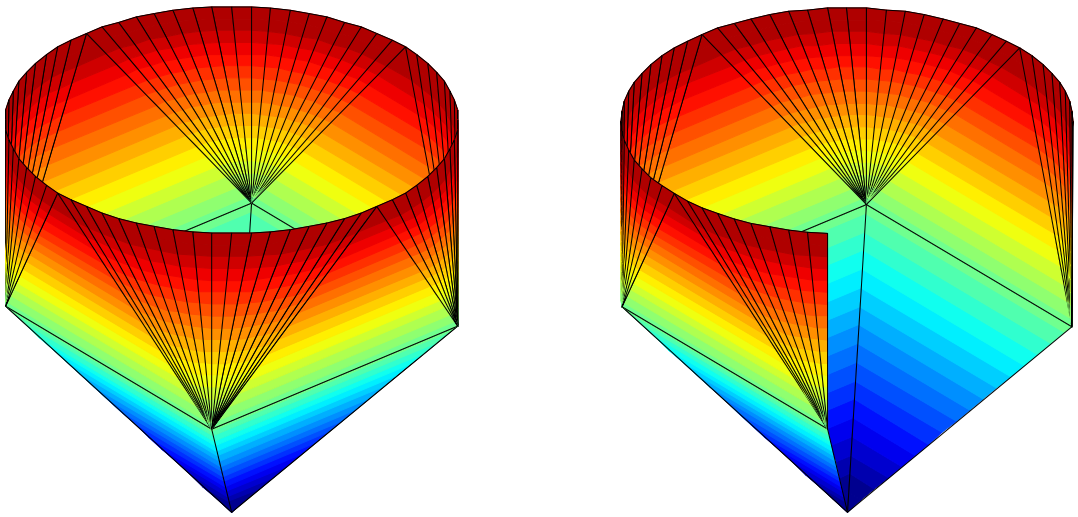

Figure 4.1: Topological closure of the graph, between heights $z = 0$ and $z = 2$, of the proper convex lsc function $\mathcal{L}_0$ which coincides, on the Euclidean unit sphere $\mathbb{S}$, with the ℓ_0 pseudonorm

In dimension $d = 2$, the function $\mathcal{L}_0$ in (27) is, by Proposition 4.2, the largest proper convex lsc 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 4.2). 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 4.1, 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

– $(0,1)$, $(1,0)$, $(0,-1)$, $(-1,0)$ – 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 (27) is given by the following explicit formulas (see also Figure 4.2).

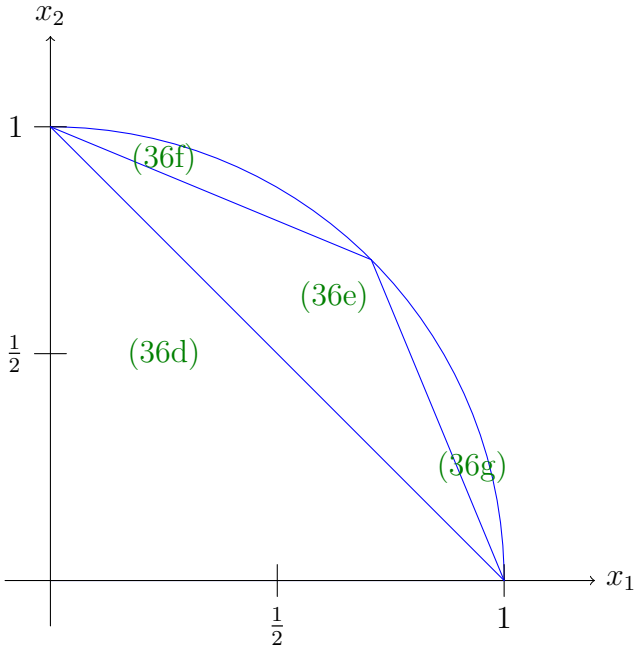

Figure 4.2: Companion figure for Proposition 4.3

Proposition 4.3. *In dimension $d=2$, the function $\mathcal{L}_0$ in (27) is given by*

$$\mathcal{L}_0(x_1,x_2) = \begin{cases} +\infty & \text{if } x_1^2+x_2^2>1\,, \quad (36a)\\ 1 & \text{if } (x_1,x_2)\in\{(1,0),(0,1),(-1,0),(0,-1)\}\,, \quad (36b)\\ 2 & \text{if } x_1^2+x_2^2=1 \text{ and } (x_1,x_2)\notin\{(1,0),(0,1),(-1,0),(0,-1)\}\,, \quad (36c) \end{cases}$$

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

$$\mathcal{L}_0(x_1,x_2) = \begin{cases} |x_1|+|x_2| & \text{if } |x_1|+|x_2|\leq 1\,, \quad (36d)\\ \dfrac{|x_1|+|x_2|-2+\sqrt{2}}{\sqrt{2}-1} & \text{if } \begin{cases}(\sqrt{2}-1)|x_1|+|x_2|<1<|x_1|+|x_2| \\ \text{or} \\ |x_1|+(\sqrt{2}-1)|x_2|<1<|x_1|+|x_2|\,,\end{cases} \quad (36e)\\ \dfrac{3-|x_2|}{2}+\dfrac{x_1^2}{2(1-|x_2|)} & \text{if } (\sqrt{2}-1)|x_1|+|x_2|\geq 1 \text{ and } |x_2|>|x_1|\,, \quad (36f)\\ \dfrac{3-|x_1|}{2}+\dfrac{x_2^2}{2(1-|x_1|)} & \text{if } |x_1|+(\sqrt{2}-1)|x_2|\geq 1 \text{ and } |x_1|>|x_2|\,. \quad (36g) \end{cases}$$

Proof. By (33) for $d=2$, we find that

$$\mathcal{L}_0(x) = \min_{(x^{(1)},x^{(2)})\in C(x)} \|x^{(1)}\|_{(1)}^{\star\mathrm{sn}} + 2\|x^{(2)}\|_{(2)}^{\star\mathrm{sn}}\,, \quad (37a)$$

where the constraints set is given by

$$C(x) = \left\{ \left(x^{(1)}, x^{(2)}\right) \in (\mathbb{R}^2)^2 \;\middle|\; \|x^{(1)}\|^{\star\mathrm{sn}}_{(1)} + \|x^{(2)}\|^{\star\mathrm{sn}}_{(2)} \leq 1 \;,\;\; x^{(1)} + x^{(2)} = x \right\} \;. \quad \text{(37b)}$$

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

$$\|x\| \leq \|x^{(1)}\|^{\star\mathrm{sn}}_{(1)} + \|x^{(2)}\|^{\star\mathrm{sn}}_{(2)} \leq 1 \;, \quad \text{(38a)}$$

$$\|x\| = 1 \Rightarrow \|x^{(1)}\|^{\star\mathrm{sn}}_{(2)} = \|x^{(1)}\|^{\star\mathrm{sn}}_{(1)} \text{ and } \|x^{(1)}\|^{\star\mathrm{sn}}_{(1)} + \|x^{(2)}\|^{\star\mathrm{sn}}_{(2)} = 1 \;, \quad \text{(38b)}$$

$$\begin{aligned}
\text{because } \|x\| &= \|x\|^{\star\mathrm{sn}}_{(2)} && \text{(by (50))} \\
&\leq \|x^{(1)}\|^{\star\mathrm{sn}}_{(2)} + \|x^{(2)}\|^{\star\mathrm{sn}}_{(2)} && \text{(because } (x^{(1)}, x^{(2)}) \in C(x) \Rightarrow x^{(1)} + x^{(2)} = x\text{)} \\
&\leq \|x^{(1)}\|^{\star\mathrm{sn}}_{(1)} + \|x^{(2)}\|^{\star\mathrm{sn}}_{(2)} && \text{(because } \|x^{(1)}\|^{\star\mathrm{sn}}_{(2)} \leq \|x^{(1)}\|^{\star\mathrm{sn}}_{(1)} \text{ by (50))} \\
&\leq 1 \;. && \text{(because } (x^{(1)}, x^{(2)}) \in C(x) \Rightarrow \|x^{(1)}\|^{\star\mathrm{sn}}_{(1)} + \|x^{(2)}\|^{\star\mathrm{sn}}_{(2)} \leq 1\text{)}
\end{aligned}$$

We are now going to describe the constraints set $C(x)$ in (37b) according to $\|x\|$, then to deduce $\mathcal{L}_0(x)$ from (37a).

1. Suppose that $\|x\| = \sqrt{x_1^2 + x_2^2} > 1$. Then, by (38a), we deduce that $C(x) = \emptyset$ in (37b), hence that $\mathcal{L}_0(x) = +\infty$ by (37a).

2. Suppose that $\|x\| = \sqrt{x_1^2 + x_2^2} = 1$. If $(x^{(1)}, x^{(2)}) \in C(x)$, we obtain by (38b) that

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

   from which we deduce that $|x_1^{(1)}| \times |x_2^{(1)}| = 0$. From $x^{(1)} + x^{(2)} = x$ and $\|x^{(1)}\|^{\star\mathrm{sn}}_{(1)} + \|x^{(2)}\|^{\star\mathrm{sn}}_{(2)} = 1$, by (38a), 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 following two 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 by (37a)

$$\mathcal{L}_0(x) = \|0\|^{\star\mathrm{sn}}_{(1)} + 2\|x\|^{\star\mathrm{sn}}_{(2)} = 2\|x\| = 2.$$

   (b) If $x \in \{(1,0), (0,1), (-1,0), (0,-1)\}$, it is easy to check that $(x, 0) \in C(x)$ by (37b). Therefore, $\mathcal{L}_0(x) \leq \|x\|^{\star\mathrm{sn}}_{(1)} + 2\|0\|^{\star\mathrm{sn}}_{(2)} = 1$ by (37a). Now, for any $(x^{(1)}, x^{(2)}) \in C(x)$, we have that

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

   by (38a). To conclude, we obtain that $1 \leq \mathcal{L}_0(x)$ by (37a), 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 A.6 in Appendix A, combined with the formula (30).

This ends the proof. □

5. Conclusion

In this paper, we have introduced a novel class of one-sided linear couplings, and we have shown that they induce conjugacies that share nice properties with the classic Fenchel conjugacy. Among them, we have distinguished a novel coupling, E-Capra, having the property of being constant along primal rays, like the ℓ_0 pseudonorm. For the E-Capra conjugacy, induced by the coupling E-Capra, we have proved that the ℓ_0 pseudonorm is equal to its biconjugate: hence, the ℓ_0 pseudonorm is E-Capra-convex in the sense of generalized convexity. We have also provided expressions for the E-Capra conjugate and biconjugate of the ℓ_0 pseudonorm, and of the characteristic functions of its level sets, in terms of the sequence of so-called top-k norms. Finally, we have shown that the ℓ_0 pseudonorm displays hidden convexity as we have proved that it coincides, on the Euclidean unit sphere, with a proper convex lsc function. This is somewhat surprising as the ℓ_0 pseudonorm is a highly nonconvex function of combinatorial nature.

Acknowledgements. We thank the Editor and an anonymous Reviewer for their comments that contributed to improve the manuscript.

A. Appendix

A.1. Properties of top-k norms and of k-support norms

Before studying properties of top-k norms and of k-support norms, we recall the notion of dual norm. Suppose that $\mathbb{R}^d$ is equipped with a norm $|||\cdot|||$ with unit ball denoted by $\mathbb{B}_{|||\cdot|||} = \big\{x \in \mathbb{R}^d \mid |||x||| \leq 1\big\}$. The expression

$$|||y|||_\star = \sup_{|||x|||\leq 1} \langle x\,, y\rangle \ , \ \forall y \in \mathbb{R}^d \tag{39a}$$

defines a norm on $\mathbb{R}^d$, called the *dual norm* $|||\cdot|||_\star$. We have

$$|||\cdot|||_\star = \sigma_{\mathbb{B}_{|||\cdot|||}} \text{ and } |||\cdot||| = \sigma_{\mathbb{B}_{|||\cdot|||_\star}} \ , \tag{39b}$$

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

$$\begin{aligned}\mathbb{B}_{|||\cdot|||_\star} &= \big\{y \in \mathbb{R}^d \mid |||y|||_\star \leq 1\big\} = \mathbb{B}^\odot_{|||\cdot|||} \\ &= \big\{y \in \mathbb{R}^d \mid \langle x\,, y\rangle \leq 1 \ , \ \forall x \in \mathbb{B}_{|||\cdot|||}\big\} \ .\end{aligned} \tag{39c}$$

A.1.1. Properties of top-k norms

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

$$\mathbb{S}_K = \big\{x \in \mathbb{R}^d \mid \|x_K\| = 1\big\} \ , \tag{40a}$$

$$\mathbb{B}_K = \big\{x \in \mathbb{R}^d \mid \|x_K\| \leq 1\big\} \ , \tag{40b}$$

where x_K has been defined as the orthogonal projection of x onto the subspace $\mathcal{R}_K$ in (20). In what follows, the Euclidean unit sphere $\mathbb{S}$ and ball $\mathbb{B}$ have been defined in (12), and the top-k norm $\|\cdot\|^{\mathrm{tn}}_{(k)}$ has been introduced in Definition 3.4.

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

- *For any $x \in \mathbb{R}^d$, the following equalities and inequalities hold true*

$$\sup_{j=1,\dots,d} |x_j| = \|x\|_\infty = \|x\|^{\mathrm{tn}}_{(1)} \leq \cdots \leq \|x\|^{\mathrm{tn}}_{(l)} \leq \|x\|^{\mathrm{tn}}_{(l+1)} \leq \cdots \leq \|x\|^{\mathrm{tn}}_{(d)} = \|x\| \,. \tag{41}$$

- *We have* $$\mathcal{R}_K \cap \mathbb{S} = \mathbb{S}_K \cap \mathbb{S} \,, \ \forall K \subset \{1, \dots, d\} \,. \tag{42}$$

- *The top-k norm $\|\cdot\|^{\mathrm{tn}}_{(k)}$ satisfies*

$$\|\cdot\|^{\mathrm{tn}}_{(k)} = \sigma_{\cup_{|K|\leq k}(\mathcal{R}_K \cap \mathbb{B})} = \sup_{|K|\leq k} \sigma_{(\mathcal{R}_K \cap \mathbb{B})} = \sup_{|K|\leq k} \sigma_{(\mathcal{R}_K \cap \mathbb{S})} = \sigma_{\cup_{|K|\leq k}(\mathcal{R}_K \cap \mathbb{S})} \,. \tag{43}$$

- *The unit sphere $\mathbb{S}^{\mathrm{tn}}_{(k)}$ and ball $\mathbb{B}^{\mathrm{tn}}_{(k)}$ of $\mathbb{R}^d$ for the top-k norm $\|\cdot\|^{\mathrm{tn}}_{(k)}$ satisfy*

$$\mathbb{B}^{\mathrm{tn}}_{(k)} = \Big\{ x \in \mathbb{R}^d \mid \|x\|^{\mathrm{tn}}_{(k)} \leq 1 \Big\} = \bigcap_{|K|\leq k} \mathbb{B}_K \,, \tag{44a}$$

$$\mathbb{S}^{\mathrm{tn}}_{(k)} = \Big\{ x \in \mathbb{R}^d \mid \|x\|^{\mathrm{tn}}_{(k)} = 1 \Big\} = \mathbb{B}^{\mathrm{tn}}_{(k)} \cap \Big(\bigcup_{|K|\leq k} \mathbb{S}_K \Big) \,. \tag{44b}$$

- *The unit balls $\mathbb{B}^{\mathrm{tn}}_{(l)}$ satisfy the inclusions*

$$\mathbb{B} = \mathbb{B}^{\mathrm{tn}}_{(d)} \subset \cdots \subset \mathbb{B}^{\mathrm{tn}}_{(l+1)} \subset \mathbb{B}^{\mathrm{tn}}_{(l)} \subset \cdots \subset \mathbb{B}^{\mathrm{tn}}_{(1)} \,. \tag{45}$$

- *We have* $$\|y\|^{\mathrm{tn}}_{(k)} \leq \sqrt{k} \|y\|^{\mathrm{tn}}_{(1)} \,, \ \forall y \in \mathbb{R}^d \,, \ \forall k = 1, \dots, d \,. \tag{46}$$

Proof.

• The equalities and inequalities (41) derive from the very definition (22) of the top-k norm $\|\cdot\|^{\mathrm{tn}}_{(k)}$.

• We prove Equation (42). We have that $x = x_K + x_{-K}$, for any $x \in \mathbb{R}^d$, and the decomposition is orthogonal, leading to

$$\big(\forall x \in \mathbb{R}^d\big) \qquad x = x_K + x_{-K} \,, \ x_K \perp x_{-K} \text{ and } \|x\|^2 = \|x_K\|^2 + \|x_{-K}\|^2 \,. \tag{47}$$

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

$$\begin{aligned}
x \in \mathbb{S} \text{ and } x \in \mathbb{S}_K &\iff 1 = \|x\|^2 \text{ and } 1 = \|x_K\|^2 && \text{(by (12) and (40a))} \\
&\iff 1 = \|x\|^2 = \|x_K\|^2 + \|x_{-K}\|^2 \text{ and } 1 = \|x_K\|^2 && \text{(by (47))} \\
&\iff \|x_{-K}\| = 0 \text{ and } 1 = \|x_K\| && \text{(by (47))} \\
&\iff x \in \mathcal{R}_K \cap \mathbb{S} \,. && \text{(by (20) and (12))}
\end{aligned}$$

• We prove Equation (43). For this purpose, we first establish that

$$\sigma_{\mathcal{R}_K \cap \mathbb{B}}(y) = \|y_K\| \,, \ \forall y \in \mathbb{R}^d \,. \tag{48}$$

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

$$\begin{aligned}
\sigma_{\mathcal{R}_K \cap \mathbb{B}}(y) &= \sup_{x \in \mathcal{R}_K \cap \mathbb{B}} \langle x \,, y \rangle && \text{(by definition (8) of a support function)} \\
&= \sup_{x \in \mathcal{R}_K \cap \mathbb{B}} \langle x_K + x_{-K} \,, y_K + y_{-K} \rangle && \text{(by the decomposition (47))}
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{x \in \mathcal{R}_K \cap \mathbb{B}} \big(\langle x_K \,, y_K \rangle + \langle x_{-K} \,, y_{-K} \rangle \big) \\
&\qquad\qquad \text{(because } x_K \perp y_{-K} \text{ and } x_{-K} \perp y_K \text{ by (47))} \\
&= \sup\{ \langle x_K \,, y_K \rangle + \langle x_{-K} \,, y_{-K} \rangle \mid x_{-K} = 0 \text{ and } \|x_K\| \leq 1 \} \\
&\qquad\qquad \text{(by definition of } \mathcal{R}_K \cap \mathbb{B}) \\
&= \sup\{ \langle x_K \,, y_K \rangle \mid \|x_K\| \leq 1 \} \\
&= \|y_K\|
\end{aligned}
$$

as is well-known for the Euclidean norm $\| \cdot \|$, when restricted to the subspace $\mathcal{R}_K$ (because it is equal to its dual norm). Then, for all $y \in \mathbb{R}^d$, we have that

$$
\begin{aligned}
\sigma_{\cup_{|K| \leq k} \mathcal{R}_K \cap \mathbb{B}}(y) &= \sup_{|K| \leq k} \sigma_{\mathcal{R}_K \cap \mathbb{B}}(y) \\
&\qquad \text{(as the support function turns a union of sets into a supremum)} \\
&= \sup_{|K| \leq k} \|y_K\| \qquad \text{(by (48))} \\
&= \|y\|^{\mathrm{tn}}_{(k)} \,. \qquad \text{(by definition (22) of } \| \cdot \|^{\mathrm{tn}}_{(k)})
\end{aligned}
$$

Now, by (12) and (20), it is straightforward that $\overline{\mathrm{co}}(\mathcal{R}_K \cap \mathbb{S}) = \mathcal{R}_K \cap \mathbb{B}$ and

$$
\begin{aligned}
\| \cdot \|^{\mathrm{tn}}_{(k)} = \sigma_{\cup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{B})} &= \sup_{|K| \leq k} \sigma_{(\mathcal{R}_K \cap \mathbb{B})} = \sup_{|K| \leq k} \sigma_{\overline{\mathrm{co}}((\mathcal{R}_K \cap \mathbb{S}))} \\
&= \sup_{|K| \leq k} \sigma_{(\mathcal{R}_K \cap \mathbb{S})} = \sigma_{\cup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S})} \,,
\end{aligned}
$$

giving Equation (43).

- We prove Equation (44a):

$$
\begin{aligned}
\mathbb{B}^{\mathrm{tn}}_{(k)} &= \Big\{ x \in \mathbb{R}^d \mid \|x\|^{\mathrm{tn}}_{(k)} \leq 1 \Big\} \qquad \text{(by definition of the ball } \mathbb{B}^{\mathrm{tn}}_{(k)}) \\
&= \Big\{ x \in \mathbb{R}^d \;\Big|\; \sup_{|K| \leq k} \|x_K\| \leq 1 \Big\} \qquad \text{(by definition (22) of } \| \cdot \|^{\mathrm{tn}}_{(k)}) \\
&= \bigcap_{|K| \leq k} \big\{ x \in \mathbb{R}^d \mid \|x_K\| \leq 1 \big\} = \bigcap_{|K| \leq k} \mathbb{B}_K \,. \qquad \text{(by definition (40b) of } \mathbb{B}_K)
\end{aligned}
$$

- We prove Equation (44b):

$$
\begin{aligned}
\mathbb{S}^{\mathrm{tn}}_{(k)} &= \Big\{ x \in \mathbb{R}^d \mid \|x\|^{\mathrm{tn}}_{(k)} = 1 \Big\} \qquad \text{(by definition of the unit sphere } \mathbb{S}^{\mathrm{tn}}_{(k)}) \\
&= \Big\{ x \in \mathbb{R}^d \;\Big|\; \sup_{|K| \leq k} \|x_K\| = 1 \Big\} \qquad \text{(by definition (22) of } \| \cdot \|^{\mathrm{tn}}_{(k)}) \\
&= \Big\{ x \in \mathbb{R}^d \;\Big|\; \sup_{|K| \leq k} \|x_K\| \leq 1 \Big\} \\
&\qquad \bigcap \big\{ x \in \mathbb{R}^d \mid \exists K \subset \{1, \ldots, d\} \,,\ |K| \leq k \,,\ \|x_K\| = 1 \big\} \\
&= \mathbb{B}^{\mathrm{tn}}_{(k)} \cap \Big(\bigcup_{|K| \leq k} \big\{ x \in \mathbb{R}^d \mid \|x_K\| = 1 \big\} \Big) \qquad \text{(by definition of the ball } \mathbb{B}^{\mathrm{tn}}_{(k)}) \\
&= \mathbb{B}^{\mathrm{tn}}_{(k)} \cap \Big(\bigcup_{|K| \leq k} \mathbb{S}_K \Big) \,. \qquad \text{(by definition (40a) of } \mathbb{S}_K)
\end{aligned}
$$

- The inclusions (45) directly follow from the Equalities and Inequalities (41).
- We prove the Inequality (46). Indeed, by definition (22) of $\|\cdot\|^{\mathrm{tn}}_{(k)}$, for a given $y \in \mathbb{R}^d$, there exists $K \subset \{1, \dots, d\}$ with $|K| \leq k$ such that

$$\big(\|y\|^{\mathrm{tn}}_{(k)}\big)^2 = \sum_{k \in K} |y_k|^2 \leq \sum_{k \in K} \big(\|y\|^{\mathrm{tn}}_{(1)}\big)^2 \leq k\big(\|y\|^{\mathrm{tn}}_{(1)}\big)^2.$$

This ends the proof. □

A.1.2. Properties of k-support norms

The k-support norm $\|\cdot\|^{\star\mathrm{sn}}_{(k)}$ has been introduced in Definition 3.4 as the dual norm of the top-k norm $\|\cdot\|^{\mathrm{tn}}_{(k)}$.

Proposition A.2. *Let $k \in \{1, \dots, d\}$.*

- *The unit balls $\mathbb{B}^{\star\mathrm{sn}}_{(l)}$ satisfy the inclusions*

$$\mathbb{B}^{\star\mathrm{sn}}_{(1)} \subset \cdots \subset \mathbb{B}^{\star\mathrm{sn}}_{(l)} \subset \mathbb{B}^{\star\mathrm{sn}}_{(l+1)} \subset \cdots \subset \mathbb{B}^{\star\mathrm{sn}}_{(d)} = \mathbb{B} \,. \tag{49}$$

- *For any $x \in \mathbb{R}^d$, the following equalities and inequalities hold true*

$$\|x\| = \|x\|^{\star\mathrm{sn}}_{(d)} \leq \cdots \leq \|x\|^{\star\mathrm{sn}}_{(l+1)} \leq \|x\|^{\star\mathrm{sn}}_{(l)} \leq \cdots \leq \|x\|^{\star\mathrm{sn}}_{(1)} = \sum_{j=1}^{d} |x_j| \,. \tag{50}$$

- *The unit ball $\mathbb{B}^{\star\mathrm{sn}}_{(k)}$ of the k-support norm $\|\cdot\|^{\star\mathrm{sn}}_{(k)}$ satisfies*

$$\begin{aligned} \mathbb{B}^{\star\mathrm{sn}}_{(k)} &= \Big\{ x \in \mathbb{R}^d \mid \|x\|^{\star\mathrm{sn}}_{(k)} \leq 1 \Big\} = \overline{\mathrm{co}}\Big(\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{B}) \Big) \\ &= \overline{\mathrm{co}}\Big(\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S}) \Big) \,. \end{aligned} \tag{51}$$

- *For $l = 0, 1, \dots, d$, we have* $\quad \|\cdot\|^{\mathrm{tn}}_{(l)} = \sigma_{\mathbb{B}^{\star\mathrm{sn}}_{(l)}} \,.$ (52)

Proof.

- The inclusions (49) directly follow from the inclusions (45) and from (39c) as $\mathbb{B}^{\star\mathrm{sn}}_{(k)} = \big(\mathbb{B}^{\mathrm{tn}}_{(k)}\big)^{\odot}$.

- The Inequalities in (50) derive from the inclusions (49). The Equalities in (50) are well-known.

- We prove Equation (51). On the one hand, by the first relation in (39b), we have that $\|\cdot\|^{\mathrm{tn}}_{(k)} = \sigma_{\mathbb{B}^{\star\mathrm{sn}}_{(k)}}$. On the other hand, by (43), we have that $\|\cdot\|^{\mathrm{tn}}_{(k)} = \sigma_{\cup_{|K| \leq k}(\mathcal{R}_K \cap \mathbb{B})} = \sigma_{\cup_{|K| \leq k}(\mathcal{R}_K \cap \mathbb{S})}$. Then, as is well-known in convex analysis, we deduce that $\overline{\mathrm{co}}\big(\mathbb{B}^{\star\mathrm{sn}}_{(k)}\big) = \overline{\mathrm{co}}\big(\bigcup_{|K| \leq k}(\mathcal{R}_K \cap \mathbb{B})\big) = \overline{\mathrm{co}}\big(\bigcup_{|K| \leq k}(\mathcal{R}_K \cap \mathbb{S})\big)$. As the unit ball $\mathbb{B}^{\star\mathrm{sn}}_{(k)}$ is closed and convex, we immediately obtain (51).

- We prove Equation (52). By Definition 3.4, the l-support norm is the dual norm of the top-l norm. Therefore, the top-l norm is the dual norm of the l-support norm and (52) follows from (39b) for $l = 1, \dots, d$. For $l = 0$, both conventions $\|\cdot\|^{\mathrm{tn}}_{(0)} = 0$ and $\mathbb{B}^{\star\mathrm{sn}}_{(0)} = \{0\}$ lead to $\|\cdot\|^{\mathrm{tn}}_{(0)} = 0 = \sigma_0 = \sigma_{\mathbb{B}^{\star\mathrm{sn}}_{(l)}}$.

This ends the proof. □

A.2. Properties of the level sets of the ℓ_0 pseudonorm

We establish useful connections between the ℓ_0 pseudonorm in (17) and the top-k norm $\|\cdot\|^{\mathrm{tn}}_{(k)}$ in (22).

Proposition A.3. *Let $k \in \{0, 1, \dots, d\}$. We have*

$$\big(\forall x \in \mathbb{R}^d\big) \;\; \ell_0(x) = k \iff 0 \leq \ldots \leq \|x\|^{\mathrm{tn}}_{(k-1)} < \|x\|^{\mathrm{tn}}_{(k)} = \ldots = \|x\|^{\mathrm{tn}}_{(d)} = \|x\| \,, \tag{53a}$$

$$\big(\forall x \in \mathbb{R}^d\big) \;\; x \in \ell_0^{\leq k} \iff \ell_0(x) \leq k \iff \|x\|^{\mathrm{tn}}_{(k)} = \|x\| \,, \tag{53b}$$

$$\big(\forall x \in \mathbb{R}^d\big) \;\; x \in \ell_0^{\leq k} \backslash \{0\} \iff 0 < \ell_0(x) \leq k \iff x \neq 0 \;\; \textit{and} \;\; \frac{x}{\|x\|} \in \mathbb{S} \cap \mathbb{S}^{\mathrm{tn}}_{(k)} \,. \tag{53c}$$

The intersection of the level set $\ell_0^{\leq k}$ in (21) *of the ℓ_0 pseudonorm in* (17) *with the Euclidean unit sphere $\mathbb{S}$ has the two following expressions*

$$\ell_0^{\leq k} \cap \mathbb{S} = \mathbb{B}^{\star \mathrm{sn}}_{(k)} \cap \mathbb{S} \,, \tag{54a}$$

$$\ell_0^{\leq k} \cap \mathbb{S} = \overline{\ell_0^{=k} \cap \mathbb{S}} \,. \tag{54b}$$

Proof.

- The Equivalences (53a) and (53b) are well-known and easy to prove.
- We prove the Equivalence (53c). Using Equation (53b) we have, for $x \in \mathbb{R}^d \backslash \{0\}$:

$$\ell_0(x) \leq k \iff \|x\|^{\mathrm{tn}}_{(k)} = \|x\| \iff \Big\|\frac{x}{\|x\|}\Big\|^{\mathrm{tn}}_{(k)} = 1 \iff \frac{x}{\|x\|} \in \mathbb{S}^{\mathrm{tn}}_{(k)} \iff \frac{x}{\|x\|} \in \mathbb{S} \cap \mathbb{S}^{\mathrm{tn}}_{(k)}.$$

- We prove Equation (54a). First, we observe that the level set $\ell_0^{\leq k}$ is closed because, by (53b), it can be expressed as $\ell_0^{\leq k} = \big\{x \in \mathbb{R}^d \;\big|\; \|x\|^{\mathrm{tn}}_{(k)} = \|x\|\big\}$. This also follows from the well-known property that the pseudonorm ℓ_0 is lower semi continuous. Second, we have

$$\begin{aligned}
\ell_0^{\leq k} \cap \mathbb{S} &= \mathbb{S} \cap \overline{\mathrm{co}}\big(\ell_0^{\leq k} \cap \mathbb{S}\big) && \text{(by Lemma A.4 since } \ell_0^{\leq k} \cap \mathbb{S} \subset \mathbb{S} \text{ and is closed)} \\
&= \mathbb{S} \cap \overline{\mathrm{co}}\Big(\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S})\Big) && \text{(as } \ell_0^{\leq k} \cap \mathbb{S} = \textstyle\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S}) \text{ by (21))} \\
&= \mathbb{B}^{\star \mathrm{sn}}_{(k)} \cap \mathbb{S} \,. && \text{(as } \overline{\mathrm{co}}\big(\textstyle\bigcup_{|K| \leq k} (\mathcal{R}_K \cap \mathbb{S})\big) = \mathbb{B}^{\star \mathrm{sn}}_{(k)} \text{ by (51))}
\end{aligned}$$

- We prove Equation (54b). For this purpose, we first establish the (known) fact that $\overline{\ell_0^{=k}} = \ell_0^{\leq k}$. The inclusion $\overline{\ell_0^{=k}} \subset \ell_0^{\leq k}$ is easy. Indeed, as we have seen that $\ell_0^{\leq k}$ is closed, we have $\ell_0^{=k} \subset \ell_0^{\leq k} \Rightarrow \overline{\ell_0^{=k}} \subset \overline{\ell_0^{\leq k}} = \ell_0^{\leq k}$. There remains to prove the reverse inclusion $\ell_0^{\leq k} \subset \overline{\ell_0^{=k}}$. For this purpose, we consider $x \in \ell_0^{\leq k}$. If $x \in \ell_0^{=k}$, obviously $x \in \overline{\ell_0^{=k}}$. 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 $x^\epsilon \to x$ when $\epsilon \to 0$. This proves that $\ell_0^{\leq k} \subset \overline{\ell_0^{=k}}$.

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

This ends the proof. □

Lemma A.4. *If A is a subset of the Euclidean unit sphere $\mathbb{S}$ of $\mathbb{R}^d$, then $A = \mathrm{co}(A) \cap \mathbb{S}$. If A is a closed subset of the Euclidean unit sphere $\mathbb{S}$ of $\mathbb{R}^d$, then $A = \overline{\mathrm{co}}(A) \cap \mathbb{S}$.*

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

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

Hence, we have proved by contradiction that $\mathrm{co}(A) \cap \mathbb{S} \subset \mathrm{extr}(\mathrm{co}(A))$. We can conclude using the fact that $\mathrm{extr}(\mathrm{co}(A)) \subset A$ (see [6, Exercice 6.4]).

Now, we consider the case where the subset A of the Euclidean unit sphere $\mathbb{S}$ is closed. Using the first part of the proof we have that $A = \mathrm{co}(A) \cap \mathbb{S}$. Now, A is closed by assumption and bounded since $A \subset \mathbb{S}$. Thus, A is compact and, in a finite dimensional space, we have that $\mathrm{co}(A)$ is compact [13, Th. 17.2], thus closed. We conclude that $A = \mathrm{co}(A) \cap \mathbb{S} = \overline{\mathrm{co}(A)} \cap \mathbb{S} = \overline{\mathrm{co}}(A) \cap \mathbb{S}$, where the last equality comes from [2, Prop. 3.46].

This ends the proof. □

A.3. Additional results on the function $\mathcal{L}_0$

In Proposition 4.2, we have provided an expression, for the proper convex lsc function $\mathcal{L}_0$ in Theorem 4.1, as the value of the minimization problem (33). Here, we provide a characterization of the optimal solutions of (33).

We recall that the *exposed face* of the closed convex set $C \subset \mathbb{R}^d$ at $y \in \mathbb{R}^d$ is [8, p.220]

$$F_C(y) = \{x \in C \mid \langle x\,,y\rangle = \sigma_C(y)\} = \arg\max_{x\in C} \langle x\,,y\rangle \ . \tag{55}$$

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

$$F_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(0) = \mathbb{B}^{\mathrm{tn}}_{(l)} \ , \ \ \forall l = 1, \ldots, d \ , \tag{56a}$$

$$F_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(\bar{x}^{(l)}) \subset \mathbb{S}^{\mathrm{tn}}_{(l)} \ , \quad \text{if } \ \bar{x}^{(l)} \neq 0 \ , \tag{56b}$$

$$F_{\mathbb{B}^{\mathrm{tn}}_{(d)}}(\bar{x}^{(d)}) = \left\{ \frac{\bar{x}^{(d)}}{\|\bar{x}^{(d)}\|} \right\} , \quad \text{if } \ \bar{x}^{(d)} \neq 0 \ . \tag{56c}$$

Proposition A.5. *Let $x \in \mathbb{R}^d$ be such that $\|x\| < 1$. The sequence $(\bar{x}^{(1)}, \ldots, \bar{x}^{(d)})$ of vectors of $\mathbb{R}^d$ is solution of the minimization problem* (33) *or, equivalently, of the minimization problem*

$$\min_{\substack{x^{(1)}\in\mathbb{R}^d,\ldots,x^{(d)}\in\mathbb{R}^d \\ \sum_{l=1}^d \sigma_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(x^{(l)})\leq 1 \\ \sum_{l=1}^d x^{(l)}=x}} \sum_{l=1}^d l\sigma_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(x^{(l)}) \tag{57}$$

if and only if

(1) *either $\sum_{l=1}^d |x_l| = \|x\|^{\star\mathrm{sn}}_{(1)} \leq 1$ and $(\bar{x}^{(1)}, \ldots, \bar{x}^{(d)}) = (x, 0, \ldots, 0)$ (and then, the minimum in* (57) *or* (33) *is equal to $\|x\|^{\star\mathrm{sn}}_{(1)}$),*

(2) *or there exists $\lambda > 0$ such that*

$$\bigcap_{l=1}^d (l+\lambda) F_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(\bar{x}^{(l)}) \neq \emptyset \ , \tag{58a}$$

$$\sum_{l=1}^d \sigma_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(x^{(l)}) = 1 \ , \tag{58b}$$

$$\sum_{l=1}^d \bar{x}^{(l)} = x \ . \tag{58c}$$

Proof. The minimization problems (33) and (57) are identical because we have $\sigma_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(\cdot) = \|\cdot\|^{\star\mathrm{sn}}_{(k)}$, since the k-support norm is the dual norm, as in (39a), of the top-k norm (see Definition 3.4). First, we establish necessary and sufficient Karush-Kuhn-Tucker (KKT) conditions for the optimization problem (57). The optimization problem (57) is the minimization of the proper convex lsc function

$$f_0(x^{(1)}, \ldots, x^{(d)}) = \sum_{l=1}^d l\sigma_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(x^{(l)}) \tag{59a}$$

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

$$f_1(x^{(1)}, \ldots, x^{(d)}) = \sum_{l=1}^d \sigma_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(x^{(l)}) - 1 \ , \tag{59b}$$

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

$$f_{1+k}(x^{(1)}, \dots, x^{(d)}) = \left\langle \sum_{l=1}^{d} \bar{x}^{(l)} - x \, , e_k \right\rangle \ , \quad k = 1, \dots, d \ , \tag{59c}$$

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, ..., f_{1+d} are proper and have $\left(\mathbb{R}^d\right)^d$ for effective domain.

As $\|x\|_{(d)}^{\star \mathrm{sn}} = \|x\| < 1$, the sequence $\left(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}\right) = \left(0, 0, \dots, 0, x\right)$ strictly satisfies the inequality constraint, that is, $f_1(0, 0, \dots, 0, x) = \|x\| - 1 < 0$ and satisfies also the equality constraints $f_2(0, 0, \dots, 0, x) = \cdots = f_{1+d}(0, 0, \dots, 0, x) = 0$. By the Slater condition, the constraints are qualified. Therefore, the sequence $\left(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}\right)$ is solution of the convex optimization problem (57) if and only if it satisfies the KKT conditions ([13, Corollary 28.3.1], [11, Example 1‴, p. 64], [8, Chapter VII]), that is, there exists $\lambda \geq 0$ and $\mu = (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$ such that

$$0 \in \partial f_0(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) + \lambda \partial f_1(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) + \sum_{k=1}^{d} \mu_k \partial f_{1+k}(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) \ , \tag{60a}$$

$$\lambda f_1(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = 0 \ , \tag{60b}$$

$$f_1(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) \leq 0 \ , \tag{60c}$$

$$\forall k = 1, \dots, d \ , \quad f_{1+k}(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = 0 \ . \tag{60d}$$

Since $\partial \sigma_{\mathbb{B}_{(l)}^{\mathrm{tn}}}(x^{(l)}) = F_{\mathbb{B}_{(l)}^{\mathrm{tn}}}(\bar{x}^{(l)})$ [12, Corollary 8.25], for $l = 1, \dots, d$, we have, by (59),

$$\partial f_0(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = \left(F_{\mathbb{B}_{(1)}^{\mathrm{tn}}}(\bar{x}^{(1)}), \dots, lF_{\mathbb{B}_{(l)}^{\mathrm{tn}}}(\bar{x}^{(l)}) \dots, dF_{\mathbb{B}_{(d)}^{\mathrm{tn}}}(\bar{x}^{(d)})\right) \ , \tag{61a}$$

$$\partial f_1(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = \left(F_{\mathbb{B}_{(1)}^{\mathrm{tn}}}(\bar{x}^{(1)}), \dots, F_{\mathbb{B}_{(l)}^{\mathrm{tn}}}(\bar{x}^{(l)}) \dots, F_{\mathbb{B}_{(d)}^{\mathrm{tn}}}(\bar{x}^{(d)})\right) \ , \tag{61b}$$

$$\partial f_{1+k}(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}) = (e_k, \dots, e_k) \ , \ \ \forall k = 1, \dots, d \ . \tag{61c}$$

With these expressions, Equation (60a) is, for $l = 1, \dots, d$, equivalent to

$$\mu = (\mu_1, \dots, \mu_d) = \sum_{k=1}^{d} \mu_k e_k \in lF_{\mathbb{B}_{(l)}^{\mathrm{tn}}}(\bar{x}^{(l)}) + \lambda F_{\mathbb{B}_{(l)}^{\mathrm{tn}}}(\bar{x}^{(l)}).$$

We conclude that the sequence $\left(\bar{x}^{(1)}, \dots, \bar{x}^{(d)}\right)$ of vectors of $\mathbb{R}^d$ is solution of the optimization problem (57) if and only if there exists $\lambda \geq 0$ such that the following conditions are satisfied

$$\bigcap_{l=1}^{d} \left[lF_{\mathbb{B}_{(l)}^{\mathrm{tn}}}(\bar{x}^{(l)}) + \lambda F_{\mathbb{B}_{(l)}^{\mathrm{tn}}}(\bar{x}^{(l)}) \right] \neq \emptyset \ , \tag{62a}$$

$$\lambda \Big(\sum_{l=1}^{d} \sigma_{\mathbb{B}_{(l)}^{\mathrm{tn}}}(x^{(l)}) - 1 \Big) = 0 \ , \tag{62b}$$

$$\sum_{l=1}^{d} \sigma_{\mathbb{B}_{(l)}^{\mathrm{tn}}}(x^{(l)}) \leq 1 \ , \tag{62c}$$

$$\sum_{l=1}^{d} \bar{x}^{(l)} = x \ . \tag{62d}$$

Second, we turn to prove Item A.5(1) and Item A.5(2).

(1) If $\lambda = 0$ in (62), we obtain

$$\bigcap_{l=1}^{d} l F_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(\bar{x}^{(l)}) \neq \emptyset \;, \tag{63a}$$

$$\sum_{l=1}^{d} \sigma_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(x^{(l)}) \leq 1 \;, \tag{63b}$$

$$\sum_{l=1}^{d} \bar{x}^{(l)} = x \;. \tag{63c}$$

We now show that (63) holds true if and only if $\|x\|^{\star\mathrm{sn}}_{(1)} \leq 1$ and $(\bar{x}^{(1)}, \bar{x}^{(2)}, \ldots, \bar{x}^{(d)}) = (x, 0, \ldots, 0)$.

On the one hand, let $(\bar{x}^{(1)}, \ldots, \bar{x}^{(d)})$ be a sequence of vectors of $\mathbb{R}^d$ which satisfies (63). If $(\bar{x}^{(1)}, \ldots, \bar{x}^{(d)}) = (0, \ldots, 0)$, then $x = 0$ and we indeed conclude that $\|x\|^{\star\mathrm{sn}}_{(1)} = \|0\|^{\star\mathrm{sn}}_{(1)} = 0 \leq 1$ and $(\bar{x}^{(1)}, \bar{x}^{(2)}, \ldots, \bar{x}^{(d)}) = (0, 0, \ldots, 0) = (x, 0, \ldots, 0)$.

If $(\bar{x}^{(1)}, \ldots, \bar{x}^{(d)}) \neq (0, \ldots, 0)$, then $k = \min\big\{l \in \{1, \ldots, d\} \mid \bar{x}^{(l)} \neq 0\big\}$ is well defined. By (63a), there exists $y \in \bigcap_{l=1}^{d} l F_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(\bar{x}^{(l)})$, where $F_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(\bar{x}^{(l)}) \subset \mathbb{B}^{\mathrm{tn}}_{(l)}$ for any $l = 1, \ldots, d$, by definition (55) of the face. Now, by the inclusion (56b), we have that $F_{\mathbb{B}^{\mathrm{tn}}_{(k)}}(\bar{x}^{(k)}) \subset \mathbb{S}^{\mathrm{tn}}_{(k)}$ since $\bar{x}^{(k)} \neq 0$ by definition of k. Therefore, there exists $y \in \mathbb{B}^{\mathrm{tn}}_{(1)} \cap k\mathbb{S}^{\mathrm{tn}}_{(k)}$, that is, $\|y\|^{\mathrm{tn}}_{(1)} = \max_{i=1,\ldots,d} |y_i| \leq 1$ and $\|y\|^{\mathrm{tn}}_{(k)} = k$. Hence, it easily follows from definition (22) of $\|\cdot\|^{\mathrm{tn}}_{(k)}$ that (see also (46))

$$k^2 = \big(\|y\|^{\mathrm{tn}}_{(k)}\big)^2 \leq k\big(\|y\|^{\mathrm{tn}}_{(1)}\big)^2 \leq k.$$

This gives $k = 1$, hence $\bar{x}^{(1)} \neq 0$ and $\bar{x}^{(l)} = 0$ for all $l = 2, \ldots, d$ by definition of k. We conclude that necessarily $(\bar{x}^{(1)}, \ldots, \bar{x}^{(d)}) = (x, 0, \ldots, 0)$ by (63c) and $\sigma_{\mathbb{B}^{\mathrm{tn}}_{(1)}}(x) = \|x\|^{\star\mathrm{sn}}_{(1)} \leq 1$ by (63b).

On the other hand, suppose that $\|x\|^{\star\mathrm{sn}}_{(1)} \leq 1$ and put $(\bar{x}^{(1)}, \ldots, \bar{x}^{(d)}) = (x, 0, \ldots, 0)$. Then, Equations (63b) and (63c) are satisfied. So is (63a) because

$$\begin{aligned}\bigcap_{l=1}^{d} l F_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(\bar{x}^{(l)}) &= F_{\mathbb{B}^{\mathrm{tn}}_{(1)}}(x) \cap \left[\bigcap_{l=2}^{d} l\mathbb{B}^{\mathrm{tn}}_{(l)}\right] && \text{(by (56a))} \\ &= F_{\mathbb{B}^{\mathrm{tn}}_{(1)}}(x) \neq \emptyset \;,\end{aligned}$$

because $F_{\mathbb{B}^{\mathrm{tn}}_{(1)}}(x) \subset \mathbb{B}^{\mathrm{tn}}_{(1)} \subset \bigcap_{l=2}^{d} \sqrt{l}\mathbb{B}^{\mathrm{tn}}_{(l)} \subset \bigcap_{l=2}^{d} l\mathbb{B}^{\mathrm{tn}}_{(l)}$ by the Inequality (46).

(2) If $\lambda > 0$ in (62), we obtain (2). Indeed, (62a) is equivalent to (58a) because $l F_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(\bar{x}^{(l)}) + \lambda F_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(\bar{x}^{(l)}) = (l + \lambda) F_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(\bar{x}^{(l)})$ since the face $F_{\mathbb{B}^{\mathrm{tn}}_{(l)}}(\bar{x}^{(l)})$ is convex, and $l > 0$, $\lambda > 0$.

This ends the proof. □

Now, we specialize in the two-dimensional case $d = 2$. Because the function $\mathcal{L}_0$ in (27) satisfies (30), we restrict the following Proposition to $x = (x_1, x_2) \in \mathbb{R}^2_+$.

Proposition A.6. *Let $x = (x_1, x_2) \in \mathbb{R}^2_+$ be such that $x_1^2 + x_2^2 < 1$. The sequence $\big(\bar{x}^{(1)}, \bar{x}^{(2)}\big)$ of vectors of $\mathbb{R}^2$ is solution of the optimization problem*

$$\mathcal{L}_0(x) = \min_{\substack{x^{(1)} \in \mathbb{R}^2, x^{(2)} \in \mathbb{R}^2 \\ \|x^{(1)}\|^{\star \mathrm{sn}}_{(1)} + \|x^{(2)}\|^{\star \mathrm{sn}}_{(2)} \leq 1 \\ x^{(1)} + x^{(2)} = x}} \|x^{(1)}\|^{\star \mathrm{sn}}_{(1)} + 2\|x^{(2)}\|^{\star \mathrm{sn}}_{(2)} \tag{64}$$

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

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

$$\mathcal{L}_0\big((x_1, x_2)\big) = x_1 + x_2 \,, \tag{65a}$$

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

$$\bar{x}^{(1)} = \Big(\frac{1 - (x_1^2 + x_2^2)}{2(1 - x_1)}, 0\Big) \,, \quad \bar{x}^{(2)} = \Big(\frac{2x_1 - x_1^2 + x_2^2 - 1}{2(1 - x_1)}, x_2\Big) \,,$$

$$\mathcal{L}_0\big((x_1, x_2)\big) = \frac{3}{2} - \frac{x_1}{2} + \frac{x_2^2}{2(1 - x_1)} \,, \tag{65b}$$

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

$$\bar{x}^{(1)} = \Big(0, \frac{1 - (x_1^2 + x_2^2)}{2(1 - x_2)}\Big) \,, \quad \bar{x}^{(2)} = \Big(x_1, \frac{2x_2 - x_2^2 + x_1^2 - 1}{2(1 - x_2)}\Big) \,,$$

$$\mathcal{L}_0\big((x_1, x_2)\big) = \frac{3}{2} - \frac{x_2}{2} + \frac{x_1^2}{2(1 - x_2)} \,, \tag{65c}$$

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

$$\bar{x}^{(1)} = \Big(\frac{1 - (\sqrt{2} - 1)x_1 - x_2}{2(\sqrt{2} - 1)}, \frac{1 - x_1 - (\sqrt{2} - 1)x_2}{2(\sqrt{2} - 1)}\Big) \,,$$

$$\bar{x}^{(2)} = \Big(\frac{x_1 + x_2 - 1}{2(\sqrt{2} - 1)}, \frac{x_1 + x_2 - 1}{2(\sqrt{2} - 1)}\Big) \,,$$

$$\mathcal{L}_0\big((x_1, x_2)\big) = \frac{x_1 + x_2 - 2 + \sqrt{2}}{\sqrt{2} - 1} \,. \tag{65d}$$

Proof. By Proposition A.5, the sequence $\big(\bar{x}^{(1)}, \bar{x}^{(2)}\big)$ of vectors of $\mathbb{R}^2$ is solution of the optimization problem (64) if and only if

– either $x_1 + x_2 = \|x\|^{\star \mathrm{sn}}_{(1)} \leq 1$ and $\big(\bar{x}^{(1)}, \bar{x}^{(2)}\big) = (x, 0)$, which is equivalent to (1),

– or there exists $\lambda > 0$ such that

$$(1 + \lambda) F_{\mathbb{B}^{\mathrm{tn}}_{(1)}}(\bar{x}^{(1)}) \cap (2 + \lambda) F_{\mathbb{B}^{\mathrm{tn}}_{(2)}}(\bar{x}^{(2)}) \neq \emptyset \,, \tag{66a}$$

$$\|\bar{x}^{(1)}\|^{\star \mathrm{sn}}_{(1)} + \|\bar{x}^{(2)}\|^{\star \mathrm{sn}}_{(2)} = 1 \,, \tag{66b}$$

$$\bar{x}^{(1)} + \bar{x}^{(2)} = x \,. \tag{66c}$$

We are going to prove, in several steps, that $\left(\bar{x}^{(1)}, \bar{x}^{(2)}\right)$ satisfies (66) for a certain $\lambda > 0$ if and only if it satisfies Item A.6(2), Item A.6(3) or Item A.6(4). For this purpose, we will use the relations

$$\mathbb{B}^{\mathrm{tn}}_{(1)} = [-1,1]^2 \,, \tag{67a}$$

$$F_{\mathbb{B}^{\mathrm{tn}}_{(1)}}(\bar{x}^{(1)}) = \begin{cases} [-1,1]^2 & \text{if } \ell_0(\bar{x}^{(1)}) = 0 \,, \\ \mathrm{sign}(\bar{x}^{(1)}_1) \times [-1,1] & \text{if } \ell_0(\bar{x}^{(1)}) = 1 \text{ with } \bar{x}^{(1)}_2 = 0 \,, \\ [-1,1] \times \mathrm{sign}(\bar{x}^{(1)}_2) & \text{if } \ell_0(\bar{x}^{(1)}) = 1 \text{ with } \bar{x}^{(1)}_1 = 0 \,, \\ \mathrm{sign}(\bar{x}^{(1)}) & \text{if } \ell_0(\bar{x}^{(1)}) = 2 \,, \end{cases} \tag{67b}$$

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

• Suppose that $\left(\bar{x}^{(1)}, \bar{x}^{(2)}\right) = (x, 0)$ satisfies (66) 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 A.6(1).

By (56a) for $l = d = 2$, we get that $(2+\lambda) F_{\mathbb{B}^{\mathrm{tn}}_{(2)}}(0) = (2+\lambda)\mathbb{B}$, where $\mathbb{B}$ is the Euclidean unit ball of $\mathbb{R}^2$, so that Equation (66) is equivalent to

$$(1+\lambda) F_{\mathbb{B}^{\mathrm{tn}}_{(1)}}(x) \cap (2+\lambda)\mathbb{B} \neq \emptyset \,, \quad \|x\|^{\star\mathrm{sn}}_{(1)} = x_1 + x_2 = 1 \,. \tag{68}$$

By (67b), we distinguish the following subcases that correspond to different expressions for $F_{\mathbb{B}^{\mathrm{tn}}_{(1)}}(x)$.

- If $\ell_0(x) = 0$, then $x_1 = x_2 = 0$. But this contradicts $x_1 + x_2 = 1$ in (68).
- If $\ell_0(x) = 1$ with $x_2 = 0$, then $x = (1, 0)$ because $x_1 + x_2 = 1$ by (68), 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 $(1+\lambda) F_{\mathbb{B}^{\mathrm{tn}}_{(1)}}(x) = \{(1+\lambda, 1+\lambda)\}$ by (67b).

  On the one hand (necessity), we show that necessarily $0 < \lambda \leq \sqrt{2}$. Indeed, (68) implies that $\|((1+\lambda)\mathrm{sign}(x_1), (1+\lambda)\mathrm{sign}(x_2))\| \leq 2 + \lambda$, which gives $\sqrt{2}(1+\lambda) \leq 2 + \lambda$, hence $0 < \lambda \leq \sqrt{2}$.
  On the other hand (sufficiency), if we put $\left(\bar{x}^{(1)}, \bar{x}^{(2)}\right) = (x, 0)$ where $\|x\|^{\star\mathrm{sn}}_{(1)} = x_1 + x_2 = 1$ and $\ell_0(x) = 2$, that is, $0 < x_1$, $0 < x_2$, then (68) is satisfied for any $0 < \lambda \leq \sqrt{2}$.

Therefore, we have proven that $\left(\bar{x}^{(1)}, \bar{x}^{(2)}\right) = (x, 0)$ satisfies (66) 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 A.6(1)).

• Suppose that $\left(\bar{x}^{(1)}, \bar{x}^{(2)}\right) = (0, x)$ satisfies (66) for a certain $\lambda > 0$. We will show that this case is impossible. Indeed, Equation (66b) implies that $\sqrt{x_1^2 + x_2^2} = \|x\| = \|x\|^{\star\mathrm{sn}}_{(2)} = 1$. 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 $\left(\bar{x}^{(1)}, \bar{x}^{(2)}\right)$ satisfies (66) for a certain $\lambda > 0$. We will show that this is equivalent to (2), (3) or (4). But, before

that, notice that, as $\|\bar{x}^{(1)}\|_{(1)}^{\star\mathrm{sn}} + \|\bar{x}^{(2)}\|_{(2)}^{\star\mathrm{sn}} = 1$, by (58b), then

$$\mathcal{L}_0(x) = 1 + \|\bar{x}^{(2)}\|_{(2)}^{\star\mathrm{sn}} = 2 - \|\bar{x}^{(1)}\|_{(1)}^{\star\mathrm{sn}} \,, \tag{69}$$

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

As $\bar{x}^{(2)} \neq 0$, then $F_{\mathbb{B}_{(2)}^{\mathrm{tn}}}(\bar{x}^{(2)}) = \{\frac{\bar{x}^{(2)}}{\|\bar{x}^{(2)}\|}\}$ by (56c). Therefore, Equation (66) is equivalent to

$$(2+\lambda)\frac{\bar{x}^{(2)}}{\|\bar{x}^{(2)}\|} \in (1+\lambda) F_{\mathbb{B}_{(1)}^{\mathrm{tn}}}(\bar{x}^{(1)}) \,, \tag{70a}$$

$$\|\bar{x}^{(1)}\|_{(1)}^{\star\mathrm{sn}} + \|\bar{x}^{(2)}\|_{(2)}^{\star\mathrm{sn}} = |\bar{x}_1^{(1)}| + |\bar{x}_2^{(1)}| + \sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2} = 1 \,, \tag{70b}$$

$$\bar{x}^{(1)} + \bar{x}^{(2)} = x \,. \tag{70c}$$

By (67b), we distinguish the following four subcases that correspond to different expressions for the face $F_{\mathbb{B}_{(1)}^{\mathrm{tn}}}(\bar{x}^{(1)})$.

Case 1. As $\bar{x}^{(1)} \neq 0$, we do not consider the case $\ell_0(\bar{x}^{(1)}) = 0$.

Case 2. Suppose that $\ell_0(\bar{x}^{(1)}) = 1$ with $\bar{x}_2^{(1)} = 0$. Then, on the one hand, $F_{\mathbb{B}_{(1)}^{\mathrm{tn}}}(\bar{x}^{(1)}) = \mathrm{sign}(\bar{x}_1^{(1)}) \times [-1, 1]$ by (67b), so that Equation (70a) is equivalent to

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

On the other hand, $\bar{x}^{(1)} = (\bar{x}_1^{(1)}, 0)$ where $\bar{x}_1^{(1)} \neq 0$, so that Equations (70) are equivalent to

$$|\bar{x}_1^{(1)}| + \sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2} = 1 \,, \quad \bar{x}_1^{(1)} + \bar{x}_1^{(2)} = x_1 \,, \quad \bar{x}_2^{(2)} = x_2 \,.$$

Therefore, Equation (70) is equivalent to

$$\frac{\bar{x}_1^{(2)}}{1 - |\bar{x}_1^{(1)}|} = \frac{1+\lambda}{2+\lambda}\mathrm{sign}(\bar{x}_1^{(1)}) \,, \tag{72a}$$

$$|\bar{x}_2^{(2)}| \leq \mathrm{sign}(\bar{x}_1^{(1)})\bar{x}_1^{(2)} \,, \tag{72b}$$

$$|\bar{x}_1^{(1)}| + \sqrt{|\bar{x}_1^{(2)}|^2 + |\bar{x}_2^{(2)}|^2} = 1 \,, \tag{72c}$$

$$\bar{x}_1^{(1)} + \bar{x}_1^{(2)} = x_1 \,, \tag{72d}$$

$$\bar{x}_2^{(2)} = x_2 \,, \tag{72e}$$

and we will now show that there exists $\lambda > 0$ such that (72) holds true if and only if Item A.6(2) holds true.

On the one hand (necessity), from (72a), we deduce that $\bar{x}_1^{(2)}$ and $\bar{x}_1^{(1)}$ have the same sign; this common sign must therefore be $\mathrm{sign}(x_1)$, as $\bar{x}_1^{(1)} + \bar{x}_1^{(2)} = x_1$ by (72d); 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 (72d)–(72e), with $x_1 > \bar{x}_1^{(1)}$, since $\bar{x}_1^{(2)} > 0$. Replacing the values in (72c), 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 (72a), 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 (72b); 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 (72b), 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

$$\big(x_2 - (\sqrt{2}+1)(1-x_1)\big)\big(x_2 + (\sqrt{2}-1)(1-x_1)\big) \geq 0;$$

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 (72) 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

$$\bar{x}^{(1)} = \big(\frac{1-(x_1^2+x_2^2)}{2(1-x_1)}, 0\big), \quad \bar{x}^{(2)} = \big(\frac{2x_1 - x_1^2 + x_2^2 - 1}{2(1-x_1)}, x_2\big) :$$

thus, using the property that

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

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 A.6(2).

On the other hand (sufficiency), if we suppose that Item A.6(2) holds, it is straightforward to follow all the above computations and to obtain that Equation (72) 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 (69), we obtain that $\mathcal{L}_0\big((x_1, x_2)\big) = 2 - |\bar{x}_1^{(1)}| - |\bar{x}_2^{(1)}| = \frac{3}{2} - \frac{x_1}{2} + \frac{x_2^2}{2(1-x_1)}$.

Case 3. If $\ell_0(\bar{x}^{(1)}) = 1$ with $\bar{x}_1^{(1)} = 0$, we do the same analysis, and we obtain Item A.6(3) and $\mathcal{L}_0\big((x_1, x_2)\big) = \frac{3}{2} - \frac{x_2}{2} + \frac{x_1^2}{2(1-x_2)}$.

Case 4. Suppose that $\ell_0(\bar{x}^{(1)}) = 2$. In this case, we have by (67b) that $F_{\mathbb{B}^{\mathrm{tn}}_{(1)}}(\bar{x}^{(1)}) = \{\big(\mathrm{sign}(\bar{x}_1^{(1)}), \mathrm{sign}(\bar{x}_2^{(1)})\big)\}$. Therefore, Equation (70) is equivalent to

$$\frac{\bar{x}_1^{(2)}}{\sqrt{\big|\bar{x}_1^{(2)}\big|^2 + \big|\bar{x}_2^{(2)}\big|^2}} = \frac{1+\lambda}{2+\lambda}\,\mathrm{sign}(\bar{x}_1^{(1)})\ , \tag{73a}$$

$$\frac{\bar{x}_2^{(2)}}{\sqrt{\big|\bar{x}_1^{(2)}\big|^2 + \big|\bar{x}_2^{(2)}\big|^2}} = \frac{1+\lambda}{2+\lambda}\mathrm{sign}(\bar{x}_2^{(1)})\ , \tag{73b}$$

$$\big|\bar{x}_1^{(1)}\big| + \big|\bar{x}_2^{(1)}\big| + \sqrt{\big|\bar{x}_1^{(2)}\big|^2 + \big|\bar{x}_2^{(2)}\big|^2} = 1\ , \tag{73c}$$

$$\bar{x}_1^{(1)} + \bar{x}_1^{(2)} = x_1\ , \tag{73d}$$

$$\bar{x}_2^{(1)} + \bar{x}_2^{(2)} = x_2\ , \tag{73e}$$

and we will now show that there exists $\lambda > 0$ such that Equation (73) holds true if and only if Item A.6(4) holds true.

On the one hand (necessity), from (73a)–(73b), we deduce that $|\bar{x}_1^{(2)}| = |\bar{x}_2^{(2)}|$ because $|\mathrm{sign}(\bar{x}_1^{(1)})| = |\mathrm{sign}(\bar{x}_2^{(1)})| = 1$ since $\ell_0(\bar{x}^{(1)}) = 2$. And that $\mathrm{sign}(\bar{x}^{(1)}) = \mathrm{sign}(\bar{x}^{(2)})$. This common sign must therefore be $\mathrm{sign}(x)$, as $\bar{x}^{(1)} + \bar{x}^{(2)} = x$ by (73d)–(73e). 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 (73d)–(73e), we get $\bar{x}_1^{(1)} = x_1 - \beta > 0$ and $\bar{x}_2^{(1)} = x_2 - \beta > 0$; replacing the values in (73c), we obtain $x_1 - \beta + x_2 - \beta + \sqrt{2}\beta = 1$; 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 (73) implies $x_1 + x_2 > 1$, $(\sqrt{2}-1)x_1 + x_2 < 1$, $x_1 + (\sqrt{2}-1)x_2 < 1$, and $\bar{x}^{(1)} = \big(\frac{1-(\sqrt{2}-1)x_1-x_2}{2(\sqrt{2}-1)}, \frac{1-x_1-(\sqrt{2}-1)x_2}{2(\sqrt{2}-1)}\big)$, $\bar{x}^{(2)} = \big(\frac{x_1+x_2-1}{2(\sqrt{2}-1)}, \frac{x_1+x_2-1}{2(\sqrt{2}-1)}\big)$: thus, Item A.6(4) holds true.

On the other hand (sufficiency), if we suppose that Item A.6(4) holds true, it is straightforward to follow all the above computations and to obtain that Equation (73) holds true with $\lambda = \sqrt{2}$.

By (69), we obtain that $\mathcal{L}_0\big((x_1, x_2)\big) = 1 + \sqrt{\big|\bar{x}_1^{(2)}\big|^2 + \big|\bar{x}_2^{(2)}\big|^2} = 1 + \frac{x_1+x_2-1}{\sqrt{2}-1}$.

This ends the proof. □

References

[1] A. Argyriou, R. Foygel, 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, Curran Associates, Red Hook (2012) 1457–1465.

[2] H. H. Bauschke, P. L. Combettes: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., CMS Books in Mathematics, Springer, Berlin (2017).

[3] A. Ben-Tal, M. Teboulle: *Hidden convexity in some nonconvex quadratically constrained quadratic programming*, Math. Program. 72 (1996) 51–63.

[4] J.-P. Chancelier, M. De Lara: *Constant along primal rays conjugacies and the l_0 pseudonorm*, Optimization, preprint hal-02459673, arXiv: 2001.11828 (2020).

[5] J.-P. Chancelier, M. De Lara: *Variational formulations for the l_0 pseudonorm and applications to sparse optimization*, preprint hal-02459688, arXiv: 2002.01314 (2020).

[6] J.-B. Hiriart-Urruty: *Optimisation et Analyse Convexe*, Presses Universitaires de France, Paris (1998).

[7] J.-B. Hiriart-Urruty, H. Le: *A variational approach of the rank function*, TOP 21(2) (2013) 207–240.

[8] J.-B. Hiriart-Ururty, C. Lemaréchal: *Convex Analysis and Minimization Algorithms I*, Springer, Berlin (1993).

[9] J. E. Martínez-Legaz: *Generalized convex duality and its economic applications*, in: *Handbook of Generalized Convexity and Generalized Monotonicity*, N. Hadjisavvas, S. Komlósi, S. S.Schaible (eds.), Nonconvex Optimization and its Applications Vol. 76, Springer, Berlin (2005) 237–292.

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

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

[12] R. T. Rockafellar, R. J.-B. Wets: *Variational Analysis*, Springer, Berlin (1998).

[13] T. R. Rockafellar: *Convex Analysis*, Princeton University Press, Princeton (1970).

[14] A. Rubinov: *Abstract Convexity and Global Optimization*, Nonconvex Optimization and its Applications 44, Kluwer Academic Publishers, Dordrecht (2000).

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

[16] K. Tono, A. Takeda, J.-y. Gotoh: *Efficient DC algorithm for constrained sparse optimization*, arXiv: 1701.08498 (2017).

[17] H. S. Witsenhausen: *On policy independence of conditional expectations*, Information and Control 28(1) (1975) 65–75.

[18] C. Zalinescu: *Convex Analysis in General Vector Spaces*, World Scientific, Singapore (2002).