Journal of Convex Analysis
Volume 30 (2023), No. 3, 743–769

Orthant-Strictly Monotonic Norms, Generalized Top-k and k-Support Norms and the ℓ_0 Pseudonorm

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

Dedicated to Roger J-B Wets on the occasion of his 85th birthday.

Received: December 30, 2021
Accepted: August 4, 2022

The so-called ℓ_0 pseudonorm on the Euclidean space $\mathbb{R}^d$ counts the number of nonzero components of a vector. We say that a sequence of norms is strictly increasingly graded (with respect to the ℓ_0 pseudonorm) if it is nondecreasing and that the sequence of norms of a vector x becomes stationary exactly at the index $\ell_0(x)$. In this paper, with any (source) norm, we associate sequences of generalized top-k and k-support norms, and we also introduce the new class of orthant-strictly monotonic norms (that encompasses the ℓ_p norms, but for the extreme ones). Then, we show that an orthant-strictly monotonic source norm generates a sequence of generalized top-k norms which is strictly increasingly graded. With this, we provide a systematic way to generate sequences of norms with which the level sets of the ℓ_0 pseudonorm are expressed by means of the difference of two norms. Our results rely on the study of orthant-strictly monotonic norms.

Keywords: ℓ_0 pseudonorm, orthant-strictly monotonic norm, generalized top-k norm, generalized k-support norm, strictly graded sequence of norms.

2010 Mathematics Subject Classification: 15A60, 46N10.

1. Introduction

The *counting function*, also called *cardinality function* or ℓ_0 *pseudonorm*, counts the number of nonzero components of a vector in $\mathbb{R}^d$. The ℓ_0 pseudonorm shares three out of the four axioms of a norm — nonnegativity, positivity except for $x = 0$, subadditivity — but the ℓ_0 pseudonorm is 0-homogeneous (hence the axiom of 1-homogeneity does not hold true). The ℓ_0 pseudonorm is used in sparse optimization, either as criterion or in the constraints, to obtain solutions with few nonzero entries. The ℓ_0 pseudonorm is nonconvex, but it has been established that its level sets can be expressed by means of the difference between two convex functions, more precisely two norms, taken from the nondecreasing sequence of so-called top-k norms (see [11] and references therein). In this paper, we generalize this kind of result to a large class of sequences of norms by introducing three concepts and by relating them to the ℓ_0 pseudonorm.

First, we define sequences of generalized top-k and k-support norms, associated with any (source) norm on $\mathbb{R}^d$. This extends already known concepts of top-k and k-support norms [2, 18]. Second, we introduce a new class of orthant-strictly

ISSN 0944-6532 / \$ 2.50

monotonic norms on $\mathbb{R}^d$. We rely on the notion of orthant-monotonic norm[1] introduced and studied in [12, 13] with further developments in [15]. With such an orthant-strictly monotonic norm, when one component of a vector moves away from zero, the norm of the vector strictly grows. Thus, an orthant-strictly monotonic norm is sensitive to the support of a vector, like the ℓ_0 pseudonorm. We study this class of norms, using the notions of dual vector pair for a norm [12, 13, 15] (refered to as polar alignment in [10]), and of Birkhoff orthogonality [6], and strict Birkhoff orthogonality [21]. Third, we define sequences of norms that are strictly increasingly graded (with respect to the ℓ_0 pseudonorm): the sequence of norms of a vector x is nondecreasing and becomes stationary exactly at the index $\ell_0(x)$. Thus equipped, we show why and how these three concepts prove especially relevant for the ℓ_0 pseudonorm.

This paper has some parts in common with the paper [8]. Indeed, the paper [8] built upon [7, 9] to prove hidden convexity of any nondecreasing function of the ℓ_0 pseudonorm, using conjugacies based on a class of norms that were not considered in [7, 9], the orthant-strictly monotonic norms. This is why, we needed specific results on orthant-strictly monotonic norms, and provided them[2] in [8, Appendix 2]. However, the current paper deals with different issues. Indeed, we focus here on a thorough characterization of orthant and orthant-strictly monotonic norms, and on the properties of derived sequences of norms. The only connection with the ℓ_0 pseudonorm is in the notion of (strictly) increasingly graded norms and how this allows to express the level sets of the ℓ_0 pseudonorm by means of the difference between two norms. This last question was not treated in [8].

The paper is organized as follows. In Sect. 2 we introduce a new class of orthant-strictly monotonic norms on $\mathbb{R}^d$, for which we provide different characterizations. In Sect. 3, we define sequences of generalized top-k and k-support norms, generated from a source norm, and we study their properties, be they general or under orthant-monotonicity. Finally, in Sect. 4 we introduce the notion of sequences of norms that are (strictly) increasingly graded with respect to the ℓ_0 pseudonorm. We show that an orthant-strictly monotonic source norm generates a sequence of generalized top-k norms which is strictly increasingly graded with respect to the ℓ_0 pseudonorm. We also study the sequence of generalized k-support norms. In conclusion, we hint at possible applications in sparse optimization.

2. Orthant-monotonic and orthant-strictly monotonic norms

In §2.1, we recall well-known definitions for norms. In §2.2, we provide new characterizations of orthant-monotonic norms. Then, in §2.3, we introduce the new notion of orthant-strictly monotonic norm, and we provide characterizations, as well as properties, that will prove especially relevant for the ℓ_0 pseudonorm.

[1] It is proved in [12, Lemma 2.12] that a norm is orthant-monotonic if and only if it is monotonic in every orthant, hence the name.

[2] More precisely, [8, Proposition 12] corresponds to Item 7 in Proposition 2.4, [8, Proposition 13] corresponds to Item 3 and Item 4 in Proposition 2.6, [8, Proposition 15] corresponds to the second Item in Proposition 4.5.

2.1. Background on norms

We work on the Euclidean space $\mathbb{R}^d$ (where d is a natural number), equipped with the scalar product $\langle \cdot, \cdot \rangle$ (but not necessarily with the Euclidean norm). Thus, all norms define the same (Borel) topology. We use the notation $[\![j,k]\!] = \{j, j+1, \ldots, k-1, k\}$ for any pair of integers such that $j \leq k$. For any vector $x \in \mathbb{R}^d$, we define its *support* by

$$\mathrm{supp}(x) = \{j \in [\![1,d]\!] \mid x_j \neq 0\} \subset [\![1,d]\!] \ . \tag{1}$$

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

$$\mathbb{S} = \{x \in \mathbb{R}^d \mid |\!|\!|x|\!|\!| = 1\} \ , \tag{2a}$$

$$\mathbb{B} = \{x \in \mathbb{R}^d \mid |\!|\!|x|\!|\!| \leq 1\} \ . \tag{2b}$$

Dual norms

We recall that the following expression

$$|\!|\!|y|\!|\!|_\star = \sup_{|\!|\!|x|\!|\!| \leq 1} \langle x, y \rangle \ , \ \forall y \in \mathbb{R}^d \tag{3}$$

defines a norm on $\mathbb{R}^d$, called the *dual norm* $|\!|\!|\cdot|\!|\!|_\star$ [1, Definition 6.7]. In the sequel, we will occasionally consider the ℓ_p-norms $\|\cdot\|_p$ on the space $\mathbb{R}^d$, defined by $\|x\|_p = \big(\sum_{i=1}^d |x_i|^p\big)^{\frac{1}{p}}$ for $p \in [1, \infty[$, and by $\|x\|_\infty = \sup_{i \in [\![1,d]\!]} |x_i|$. It is well-known that the dual norm of the norm $\|\cdot\|_p$ is the ℓ_q-norm $\|\cdot\|_q$, where q is such that $1/p + 1/q = 1$ (with the extreme cases $q = \infty$ when $p = 1$, and $q = 1$ when $p = \infty$).

We denote the unit sphere and the unit ball of the dual norm $|\!|\!|\cdot|\!|\!|_\star$ by

$$\mathbb{S}_\star = \{y \in \mathbb{R}^d \mid |\!|\!|y|\!|\!|_\star = 1\} \ , \tag{4a}$$

$$\mathbb{B}_\star = \{y \in \mathbb{R}^d \mid |\!|\!|y|\!|\!|_\star \leq 1\} \ . \tag{4b}$$

For any subset $X \subset \mathbb{R}^d$, $\sigma_X : \mathbb{R}^d \to [-\infty, +\infty]$ denotes the *support function of the subset* X:

$$\sigma_X(y) = \sup_{x \in X} \langle x, y \rangle \ , \ \forall y \in \mathbb{R}^d \ . \tag{5}$$

It is easily established that

$$|\!|\!|\cdot|\!|\!| = \sigma_{\mathbb{B}_\star} = \sigma_{\mathbb{S}_\star} \ \text{ and } \ |\!|\!|\cdot|\!|\!|_\star = \sigma_{\mathbb{B}} = \sigma_{\mathbb{S}} \ , \tag{6a}$$

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

$$\mathbb{B}_\star = \mathbb{B}^\odot = \{y \in \mathbb{R}^d \mid \langle x, y \rangle \leq 1 \ , \ \forall x \in \mathbb{B}\} \ . \tag{6b}$$

Since the set $\mathbb{B}$ is closed, convex and contains 0, we have [1, Theorem 5.103]

$$\mathbb{B}^{\odot\odot} = \big(\mathbb{B}^\odot\big)^\odot = \mathbb{B} \ , \tag{6c}$$

hence the *bidual norm* $|\!|\!|\cdot|\!|\!|_{\star\star} = \big(|\!|\!|\cdot|\!|\!|_\star\big)_\star$ is the original norm:

$$|\!|\!|\cdot|\!|\!|_{\star\star} = \big(|\!|\!|\cdot|\!|\!|_\star\big)_\star = |\!|\!|\cdot|\!|\!| \ . \tag{6d}$$

$|\!|\!|\cdot|\!|\!|$-duality

By construction of the dual norm in (3), we have the inequality

$$\langle x, y \rangle \leq |\!|\!|x|\!|\!| \times |\!|\!|y|\!|\!|_\star \ , \ \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \ . \tag{7a}$$

One says that $y \in \mathbb{R}^d$ is $|||\cdot|||$-*dual* to $x \in \mathbb{R}^d$, denoted by $y \parallel_{|||\cdot|||} x$, if equality holds in Inequality (7a), that is,

$$y \parallel_{|||\cdot|||} x \iff \langle x, y \rangle = |||x||| \times |||y|||_\star . \tag{7b}$$

The terminology $|||\cdot|||$-dual comes from [15, page 2] (see also the vocable of *dual vector pair* in [12, Equation (1.11)] and of *dual vectors* in [13, p. 283], whereas it is refered as *polar alignment* in [10]).

We illustrate the $|||\cdot|||$-duality in the case of the ℓ_p-norms $\|\cdot\|_p$, for $p \in [1, \infty]$. The notation $x \circ x' = (x_1 x'_1, \ldots, x_d x'_d)$ is for the Hadamard (entrywise) product, for any x, x' in $\mathbb{R}^d$. For any $x \in \mathbb{R}^d$, we denote by $\text{sign}(x) \in \{-1, 0, 1\}^d$ the vector of $\mathbb{R}^d$ with components the signs $\text{sign}(x_i) \in \{-1, 0, 1\}$ of the entries x_i, for $i \in [\![1, d]\!]$. Let $x \in \mathbb{R}^d \backslash \{0\}$ be a given vector (the case $x = 0$ is trivial). We easily obtain that a vector y is

- ℓ_2-dual to x iff (if and only if) there exists $\lambda \in \mathbb{R}_+ (= [0, \infty[)$ such that $y = \lambda x$;
- ℓ_p-dual to x some $p \in]1, \infty[$ iff there exists $\lambda \in \mathbb{R}_+$ such that we have $y = \lambda \text{sign}(x) \circ \big(|x_i|^{p/q}\big)_{i \in [\![1,d]\!]}$, where q satisfies $1/p + 1/q = 1$;
- ℓ_1-dual to x iff the vectors y and $\|y\|_\infty \text{sign}(x)$ coincide on $\text{supp}(x)$, the support of the vector x as defined in (1);
- ℓ_∞-dual to x iff $y_j = 0$ for all $j \in \arg\max_{i \in [\![1,d]\!]} |x_i|$, and $y \circ x \geq 0$.

Restriction norms

For any subset $K \subset [\![1, d]\!]$, we denote by $\mathbb{R}^K$ the set of functions from K to $\mathbb{R}$ – which can be identified with $\mathbb{R}^{|K|}$, where $|K|$ denotes the cardinality of $K \subset [\![1, d]\!]$ – and we introduce the subspace of $\mathbb{R}^d$ made of vectors whose components vanish outside of K by [3]

$$\mathcal{R}_K = \mathbb{R}^K \times \{0\}^{-K} = \{x \in \mathbb{R}^d \mid x_j = 0 \ , \ \forall j \notin K\} \subset \mathbb{R}^d \ , \tag{8}$$

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

$$\langle x_K, y_K \rangle = \langle x_K, y \rangle = \big\langle \pi_K(x), y \big\rangle = \big\langle x, \pi_K(y) \big\rangle = \langle x, y_K \rangle \ , \tag{9}$$

for all $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$.

Definition 2.1. For any norm $|||\cdot|||$ on $\mathbb{R}^d$ and any subset $K \subset [\![1, d]\!]$, we define three norms on the subspace $\mathcal{R}_K$ of $\mathbb{R}^d$, as defined in (8), as follows.

- The *K-restriction norm* $|||\cdot|||_K$ is the norm on $\mathcal{R}_K$ defined by

$$|||x|||_K = |||x||| \ , \ \forall x \in \mathcal{R}_K \ . \tag{10}$$

- The $(\star, K)$-norm $|||\cdot|||_{\star,K}$ is the norm $\big(|||\cdot|||_\star\big)_K$, given by the restriction to the subspace $\mathcal{R}_K$ of the dual norm $|||\cdot|||_\star$ (first dual, as recalled in definition (3) of a dual norm, then restriction),

[3] Here, following notation from game theory, we have denoted by $-K$ the complementary subset of K in $[\![1, d]\!]$: $K \cup (-K) = [\![1, d]\!]$ and $K \cap (-K) = \emptyset$.

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

It has been established (see [15, Proposition 2.2]) that, for any nonempty subset $K \subset [\![1, d]\!]$, one has the inequality $|||\cdot|||_{K,\star} \leq |||\cdot|||_{\star,K}$. We will discuss the equality case in Proposition 2.4.

2.2. New characterizations of orthant-monotonic norms

We recall the definitions of monotonic and of orthant-monotonic norms before introducing, in the next §2.3, the new notion of orthant-strictly monotonic norms. For any $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, we denote $|x| = (|x_1|, \ldots, |x_d|) \in \mathbb{R}^d$.

Definition 2.2. A norm $|||\cdot|||$ on the space $\mathbb{R}^d$ is called

- *monotonic* [3] if, for all x, x' in $\mathbb{R}^d$, we have $|x| \leq |x'| \Rightarrow |||x||| \leq |||x'|||$, where $|x| \leq |x'|$ means $|x_i| \leq |x_i^{'}|$ for all $i \in [\![1, d]\!]$,
- *orthant-monotonic* [12, 13] if, for all x, x' in $\mathbb{R}^d$, we have $\big(|x| \leq |x'|$ and $x \circ x' \geq 0 \Rightarrow |||x||| \leq |||x'|||\big)$.

We will use the following, easy to prove, properties: any monotonic norm is orthant-monotonic; if a norm is orthant-monotonic, so are its restriction norms in Definition 2.1 (as norms on their respective subspaces). All the ℓ_p-norms $\|\cdot\|_p$, for $p \in [1, \infty]$, are monotonic, hence orthant-monotonic. Recall that a *seminorm* satisfies all the axioms of a norm, but for vectors with null seminorm that are not restricted to the singleton $\{0\}$. The definition of an *orthant-monotonic seminorm* is straightforward, and it is easily proven that the supremum of a family of orthant-monotonic seminorms is an orthant-monotonic seminorm.

We recall the definitions of Birkhoff orthogonality [6], and of strict Birkhoff orthogonality [21].

Definition 2.3. Let $\mathcal{U}$ and $\mathcal{V}$ be two subspaces of $\mathbb{R}^d$. Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$.

- We say that the subspace $\mathcal{U}$ is *Birkhoff orthogonal* [6] to the subspace $\mathcal{V}$, denoted by $\mathcal{U} \perp_{|||\cdot|||} \mathcal{V}$ if $|||u + v||| \geq |||u|||$, for any $u \in \mathcal{U}$ and any $v \in \mathcal{V}$, that is,

$$\mathcal{U} \perp_{|||\cdot|||} \mathcal{V} \iff |||u + v||| \geq |||u||| \ , \ \forall u \in \mathcal{U} \ , \ \forall v \in \mathcal{V} \ . \tag{11}$$

- We say that the subspace $\mathcal{U}$ is *strictly Birkhoff orthogonal* [21] to the subspace $\mathcal{V}$, denoted by $\mathcal{U} \perp^{>}_{|||\cdot|||} \mathcal{V}$ if $|||u + v||| > |||u|||$, for any $u \in \mathcal{U}$ and any $v \in \mathcal{V}\backslash\{0\}$, that is,

$$\mathcal{U} \perp^{>}_{|||\cdot|||} \mathcal{V} \iff |||u + v||| > |||u||| \ , \ \forall u \in \mathcal{U} \ , \ \forall v \in \mathcal{V}\backslash\{0\} \ . \tag{12}$$

Now, we are ready to recall established characterizations of orthant-monotonic norms, and to add new characterizations, namely Item 7 and Item 8 in the following Proposition 2.4.

Proposition 2.4. *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$. The following assertions are equivalent.*

(1) *The norm $|||\cdot|||$ is orthant-monotonic.*

(2) *The norm $|||\cdot|||_\star$ is orthant-monotonic.*

(3) $|||\cdot|||_{K,\star} = |||\cdot|||_{\star,K}$*, for all* $K \subset [\![1,d]\!]$.

(4) $\mathcal{R}_K \perp_{|||\cdot|||} \mathcal{R}_{-K}$*, for all* $K \subset [\![1,d]\!]$.

(5) $\mathcal{R}_K \perp_{|||\cdot|||} \mathcal{R}_{-K}$*, for all* $K \subset [\![1,d]\!]$ *with* $|K| = d-1$.

(6) *For any vector* $u \in \mathbb{R}^d\backslash\{0\}$*, there exists a vector* $v \in \mathbb{R}^d\backslash\{0\}$ *such that* $\mathrm{supp}(v) \subset \mathrm{supp}(u)$*, that* $u \circ v \geq 0$ *and that v is $|||\cdot|||$-dual to u as in* (7b).

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

(8) $\pi_K(\mathbb{B}) = \mathcal{R}_K \cap \mathbb{B}$*, for all* $K \subset [\![1,d]\!]$.

Proof. The equivalence between all statements but the two last ones can be found in [15, Proposition 2.4].

It is easily established that Item 7 is equivalent to Item 4. Indeed, suppose that Item 7 holds true. We consider $x \in \mathbb{R}^d$ and $J \subset K \subset [\![1,d]\!]$. By setting $u = x_J \in \mathcal{R}_J$ and $v = x_K - x_J$, we get that $v \in \mathcal{R}_{-J}$. By Item 7, we have that $|||u||| \leq |||u+v|||$, hence that $|||x_J||| \leq |||x_K|||$. The reverse implication is proved in the same way.

We now show that Item 3 and Item 8 are equivalent. For this purpose, let $|||\cdot|||$ be a norm on $\mathbb{R}^d$ and $K \subset [\![1,d]\!]$, and let us admit for a while that

$$|||y|||_{\star,K} = \sigma_{\pi_K(\mathbb{B})}(y) = \sigma_{\pi_K(\mathbb{S})}(y) \ , \ \ \forall y \in \mathcal{R}_K \ , \tag{13a}$$

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

Therefore, the equality $|||\cdot|||_{\star,K} = |||\cdot|||_{K,\star}$ is equivalent to $\sigma_{\pi_K(\mathbb{B})} = \sigma_{\mathcal{R}_K\cap\mathbb{B}}$, when this last equality is restricted to the subspace $\mathcal{R}_K$. Now, on the one hand, the subset $\pi_K(\mathbb{B})$ of $\mathcal{R}_K$ is convex and closed (in the subspace $\mathcal{R}_K$) as the image of the convex and compact set $\mathbb{B}$ by the linear mapping π_K. On the other hand, the subset $\mathcal{R}_K \cap \mathbb{B}$ of $\mathcal{R}_K$ is convex and closed (in the subspace $\mathcal{R}_K$). Therefore, $|||\cdot|||_{\star,K} = |||\cdot|||_{K,\star}$ if and only if $\pi_K(\mathbb{B}) = \mathcal{R}_K \cap \mathbb{B}$. Thus, we have shown that Item 3 and Item 8 are equivalent. It remains to prove (13a) and (13b).

- We prove (13a). For any $y \in \mathcal{R}_K$, we have

$$\begin{aligned}
|||y|||_{\star,K} &= |||y|||_\star && \text{(using Definition 2.1)}\\
&= \sigma_{\mathbb{B}}(y) && \text{(by (6a))}\\
&= \sup_{x\in\mathbb{B}} \langle x,\ y\rangle && \text{(by definition (5) of the support function } \sigma_{\mathbb{B}})\\
&= \sup_{x\in\mathbb{B}} \langle x,\ \pi_K(y)\rangle && \text{(as } y = \pi_K(y) \text{ because } y \in \mathcal{R}_K)\\
&= \sup_{x\in\mathbb{B}} \langle \pi_K(x),\ y\rangle \\
& && \text{(by the self-duality property (9) of the projection mapping } \pi_K)\\
&= \sup_{x'\in\pi_K(\mathbb{B})} \langle x',\ y\rangle = \sigma_{\pi_K(\mathbb{B})}(y) \\
&. && \text{(by definition (5) of the support function } \sigma_{\pi_K(\mathbb{B})})
\end{aligned}$$

Thus, we have proved that $|||y|||_{\star,K} = \sigma_{\pi_K(\mathbb{B})}(y)$.

It remains to prove that $\sigma_{\pi_K(\mathbb{B})}(y) = \sigma_{\pi_K(\mathbb{S})}(y)$. Now, as the unit ball $\mathbb{B}$ is equal to the convex hull $\mathrm{co}(\mathbb{S})$ of the unit sphere $\mathbb{S}$, we get that $\pi_K(\mathbb{B}) = \pi_K(\mathrm{co}(\mathbb{S}))$. As π_K is a linear mapping, we easily obtain that $\pi_K(\mathrm{co}(\mathbb{S})) = \mathrm{co}(\pi_K(\mathbb{S}))$. Since $\sigma_{\mathrm{co}(\pi_K(\mathbb{S}))} = \sigma_{\pi_K(\mathbb{S})}$ [4, Prop. 7.13], we conclude that $|||y|||_{\star,K} = \sigma_{\pi_K(\mathbb{B})} = \sigma_{\mathrm{co}(\pi_K(\mathbb{S}))} = \sigma_{\pi_K(\mathbb{S})}$ on $\mathcal{R}_K$, that is, equality (13a) holds true.

- We prove (13b).

By (6a), we have the equality $|||\cdot|||_{K,\star} = \sigma_{\mathcal{R}_K \cap \mathbb{B}}$ on $\mathcal{R}_K$, as $\mathcal{R}_K \cap \mathbb{B}$ is easily seen to be the unit ball (in $\mathcal{R}_K$) of the restriction norm $|||\cdot|||_K$ in (10). Therefore, we have proved that $|||y|||_{K,\star} = \sigma_{\mathcal{R}_K \cap \mathbb{B}}(y)$ for any $y \in \mathcal{R}_K$.

Now, we prove that $\sigma_{\mathcal{R}_K \cap \mathbb{B}}(y) = \sigma_{\mathcal{R}_K \cap \mathbb{S}}(y)$ for any $y \in \mathcal{R}_K$. It is easy to check that the unit sphere (in $\mathcal{R}_K$) of the restriction norm $|||\cdot|||_K$ in (10) is $\mathcal{R}_K \cap \mathbb{S}$. Then, using the fact that the convex hull (be it in $\mathcal{R}_K$ or in $\mathbb{R}^d$) of the unit sphere $\mathcal{R}_K \cap \mathbb{S}$ is the unit ball $\mathcal{R}_K \cap \mathbb{B}$, we have that $\mathrm{co}(\mathcal{R}_K \cap \mathbb{S}) = \mathcal{R}_K \cap \mathbb{B}$. As $\sigma_{\mathrm{co}(\mathcal{R}_K \cap \mathbb{S})} = \sigma_{\mathcal{R}_K \cap \mathbb{S}}$ [4, Prop. 7.13], we conclude that $|||\cdot|||_{K,\star} = \sigma_{\mathcal{R}_K \cap \mathbb{B}} = \sigma_{\mathrm{co}(\mathcal{R}_K \cap \mathbb{S})} = \sigma_{\mathcal{R}_K \cap \mathbb{S}}$ on $\mathcal{R}_K$, that is, equality (13b) holds true. □

As an example, we illustrate Item 6 of Proposition 2.4 with the ℓ_1 and ℓ_∞ norms, which both are orthant-monotonic. Let $\mathbb{I} \in \mathbb{R}^d$ denote the vector whose components are all equal to one. For any vector $u \in \mathbb{R}^d$,

- the vector $v = \mathrm{sign}(u)$ satisfies $\mathrm{supp}(v) = \mathrm{supp}(u)$, $u \circ v \geq 0$, and it is $\|\cdot\|_1$-dual to the vector u; this last assertion is obvious for $u = 0$ and, when $u \neq 0$, we have that
$$\langle u,\, v\rangle = \langle u,\, \mathrm{sign}(u)\rangle = \langle |u|,\, \mathbb{I}\rangle = \|u\|_1 = \|u\|_1 \times 1 = \|u\|_1 \|v\|_\infty\ ,$$
- the vector $v = \mathrm{sign}(u) \circ \mathbb{I}_U$, where $U = \arg\max_{i \in [\![1,d]\!]} |u_i|$, satisfies $\mathrm{supp}(v) \subset \mathrm{supp}(u)$, $u \circ v \geq 0$, and it is $\|\cdot\|_\infty$-dual to the vector u, as we have
$$\begin{aligned}\langle u,\, v\rangle &= \langle u,\, \mathrm{sign}(u) \circ \mathbb{I}_U\rangle = \langle |u|_U,\, \mathbb{I}_U\rangle = \big\langle \|u\|_\infty \mathbb{I}_U,\, \mathbb{I}_U\big\rangle \\ &= \|u\|_\infty \|\mathbb{I}_U\|_1 = \|u\|_\infty \|v\|_1\ .\end{aligned}$$

2.3. Orthant-strictly monotonic norms

After these recalls, we introduce two new notions, that are the strict versions of monotonic and orthant-monotonic norms. Then, we provide characterizations that will prove especially relevant for the ℓ_0 pseudonorm.

Definition 2.5. A norm $|||\cdot|||$ on the space $\mathbb{R}^d$ is called

- *strictly monotonic* if, for all x, x' in $\mathbb{R}^d$, we have $|x| < |x'| \Rightarrow |||x||| < |||x'|||$, where $|x| < |x'|$ means that $|x_i| \leq |x_i'|$ for all $i \in [\![1,d]\!]$, and that there exists $j \in [\![1,d]\!]$ such that $|x_j| < |x_j'|$,
- *orthant-strictly monotonic* if, for all x, x' in $\mathbb{R}^d$, we have $\big(|x| < |x'|$ and $x \circ x' \geq 0 \Rightarrow |||x||| < |||x'|||\,\big)$.

We will use the following, easy to prove, properties: any strictly monotonic norm is orthant-strictly monotonic; any orthant-strictly monotonic norm is orthant-monotonic.

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

In contrast with orthant-monotonicity (equivalence between Item 1 and Item 2 of Proposition 2.4), the notion of orthant-strictly monotonicity is not necessarily preserved when taking the dual norm: indeed, the ℓ_1-norm $\|\cdot\|_1$ is orthant-strictly monotonic, whereas its dual norm, the ℓ_∞-norm $\|\cdot\|_\infty$ is orthant-monotonic, but not orthant-strictly monotonic.

Now, we provide characterizations of orthant-strictly monotonic norms.

Proposition 2.6. *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$. The following assertions are equivalent.*

(1) *The norm $|||\cdot|||$ is orthant-strictly monotonic.*

(2) *The family $\{\mathcal{R}_K\}_{K\subset[\![1,d]\!]}$ of subspaces of $\mathbb{R}^d$ is* strictly Birkhoff orthogonal, *in the sense that $\mathcal{R}_K \perp^{>}_{|||\cdot|||} \mathcal{R}_{-K}$, for all $K \subset [\![1,d]\!]$, as in* (12).

(3) *The norm $|||\cdot|||$ is* strictly increasing with the coordinate subspaces, *in the sense that*[4], *for any $x \in \mathbb{R}^d$ and any $J \subsetneq K \subset [\![1,d]\!]$, we have $x_J \neq x_K \Rightarrow |||x_J||| < |||x_K|||$.*

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

Proof. • We prove that Item 1 implies Item 2.

Let $K \subset [\![1,d]\!]$. Let $u \in \mathcal{R}_K$ and $v \in \mathcal{R}_{-K}\backslash\{0\}$, that is, $u = u_K$ and $v = v_{-K} \neq 0$. We want to show that $|||u+v||| > |||u|||$, by the definition (12) of strict Birkhoff orthogonality.

On the one hand, by definition of the module of a vector, we easily see that $|x| = |x_K|+|x_{-K}|$, for any vector $x \in \mathbb{R}^d$. Thus, we have $|u+v| = |(u+v)_K|+|(u+v)_{-K}| = |u_K+v_K|+|u_{-K}+v_{-K}| = |u_K+0|+|0+v_{-K}| = |u_K|+|v_{-K}| > |u_K| = |u|$ since $|v_{-K}| > 0$ as $v = v_{-K} \neq 0$, and since $u = u_K$.

On the other hand, we easily get $(u+v)\circ u = \big((u+v)_K \circ u_K\big)+\big((u+v)_{-K}\circ u_{-K}\big) = \big(u_K \circ u_K\big)+\big(v_{-K}\circ u_{-K}\big) = \big(u_K \circ u_K\big)$, because $u_{-K} = 0$ and $v_K = 0$. Therefore, we get that $(u+v)\circ u = \big(u_K \circ u_K\big) \geq 0$.

From $|u+v| > |u|$ and $(u+v)\circ u \geq 0$, we deduce that $|||u+v||| > |||u|||$ by Definition 2.5 as the norm $|||\cdot|||$ is orthant-strictly monotonic. Thus, (12) is satisfied, hence Item 2 holds true.

• We prove that Item 2 implies Item 3.

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

As $J \subsetneq K \subset [\![1,d]\!]$ and $x_J \neq x_K$, there exists $w \in \mathcal{R}_{-J}$, $w \neq 0$, such that $x_K = x_J + w$. Now, as the family $\{\mathcal{R}_K\}_{K\subset[\![1,d]\!]}$ is strictly Birkhoff orthogonal by assumption (Item 2), we have $\mathcal{R}_J \perp^{>}_{|||\cdot|||} \mathcal{R}_{-J}$. As a consequence, we obtain that $|||x_K||| = |||x_J+w||| > |||x_J|||$.

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

- We prove that Item 3 implies Item 4.

Let $u \in \mathbb{R}^d \backslash \{0\}$ be given and let us put $K = \mathrm{supp}(u) \neq \emptyset$. As the norm $|||\cdot|||$ is orthant-strictly monotonic, it is orthant-monotonic; hence, by Item 6 in Proposition 2.4, there exists a vector $v \in \mathbb{R}^d \backslash \{0\}$ such that $\mathrm{supp}(v) \subset \mathrm{supp}(u)$, that $u \circ v \geq 0$ and that v is $|||\cdot|||$-dual to u, as in (7b), that is, $\langle u, v\rangle = |||u||| \times |||v|||_\star$. Thus $J = \mathrm{supp}(v) \subset K = \mathrm{supp}(u)$. We now show that $J \subsetneq K$ is impossible, hence that $J = K$, thus proving that Item 4 holds true with the above vector v.

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

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

As a consequence, $\{u_K, u_J\} \subset \arg\max_{|||x|||\leq|||u|||} \langle x, v\rangle$, by definition (3) of $|||v|||_\star$, because $|||u||| = |||u_K||| \geq |||u_J|||$, by Item 7 in Proposition 2.4 since $J \subset K$ and the norm $|||\cdot|||$ is orthant-monotonic. But any solution in $\arg\max_{|||x|||\leq|||u|||} \langle x, v\rangle$ belongs to the frontier of the ball of radius $|||u|||$, hence has exactly norm $|||u|||$. Thus, we deduce that $|||u||| = |||u_K||| = |||u_J|||$. If we had $J = \mathrm{supp}(v) \subsetneq K = \mathrm{supp}(u)$, we would have $u_J \neq u_K$, hence $|||u_K||| > |||u_J|||$ by Item 3; this would be in contradiction with $|||u_K||| = |||u_J|||$. Therefore, $J = \mathrm{supp}(v) = K = \mathrm{supp}(u)$.

- We prove that Item 4 implies Item 1.

Let x, x' in $\mathbb{R}^d$ be such that $|x| < |x'|$ and $x \circ x' \geq 0$. We are going to prove that $|||x||| < |||x'|||$.

We suppose that $x \neq 0$ (otherwise the proof is trivial). By Item 4, there exists a vector $w \in \mathbb{R}^d$ such that $\mathrm{supp}(w) = \mathrm{supp}(x)$, $x \circ w \geq 0$ and that $\langle x, w\rangle = |||x||| \times |||w|||_\star$. As $\mathrm{supp}(w) = \mathrm{supp}(x)$ with $x \neq 0$, we have $w \neq 0$, so that we can always suppose that $|||w|||_\star = 1$ (after renormalization), giving $|||x||| = \langle x, w\rangle$.

First, we are going to establish that $i \in \mathrm{supp}(x) \Rightarrow x'_i w_i \geq x_i w_i$. From $|x'| > |x|$, we deduce that $|x'|^2 \geq |x'| \circ |x|$, and, as $x' \circ x \geq 0$, we obtain that $|x'|^2 \geq x' \circ x = |x'| \circ |x| \geq 0$. Hence, we deduce

$$(x' \circ x) \circ (x' \circ w) = |x'|^2 \circ (x \circ w) \geq (x' \circ x) \circ (x \circ w) \ ,$$

as $x \circ w \geq 0$. Moving to components, we get that, for all $i \in [\![1, d]\!]$, $x'_i x_i x'_i w_i \geq x'_i x_i x_i w_i$, so that, on the one hand

$$x'_i x_i > 0 \Rightarrow x'_i w_i \geq x_i w_i \ .$$

On the other hand, as $|x'| > |x|$ and $x \circ x' \geq 0$, we easily get that $x'_i x_i > 0 \iff i \in \mathrm{supp}(x)$. Therefore, we deduce that $i \in \mathrm{supp}(x) \Rightarrow x'_i x_i > 0 \Rightarrow x'_i w_i \geq x_i w_i$.

Second, we show that $|||x||| \leq |||x'|||$. Indeed, we have:

$$\begin{aligned}
|||x'||| &= \sup_{|||w'|||_\star \leq 1} \langle x', w'\rangle && \text{(by (3) as } |||\cdot||| = (|||\cdot|||_\star)_\star) \\
&\geq \langle x', w\rangle && \text{(as } |||w|||_\star = 1) \\
&= \sum_{i \in \mathrm{supp}(w)} x'_i w_i
\end{aligned}$$

$$= \sum_{i \in \mathrm{supp}(x)} x'_i w_i \qquad (\text{as } \mathrm{supp}(w) = \mathrm{supp}(x))$$

$$\geq \sum_{i \in \mathrm{supp}(x)} x_i w_i \qquad (\text{as } i \in \mathrm{supp}(x) \Rightarrow x'_i w_i \geq x_i w_i)$$

$$= \langle x,\, w \rangle = |||x||| \qquad (\text{by property of the vector } w.)$$

Third, we show that $|||x||| < |||x'|||$. There are two cases.

In the first case, there exists $j \in \mathrm{supp}(x)$ such that $0 < |x_j| < |x'_j|$. As a consequence, on the one hand, $0 < |w_j||x_j| < |w_j||x'_j|$, since $w_j \neq 0$ because $j \in \mathrm{supp}(x) = \mathrm{supp}(w)$. On the other hand, $x'_j x_j > 0$ implies $x'_j w_j \geq x_j w_j$, as seen above, and $x_j w_j \geq 0$ because $x \circ w \geq 0$. Thus, we get that $x'_j w_j \geq x_j w_j \geq 0$. As $0 < |x_j| < |x'_j|$, we deduce that $x'_j w_j > x_j w_j$. Returning to the last inequality in the sequence of equalities and inequalities above, we observe that it is now strict, and we conclude that $|||x'||| > |||x|||$.

In the second case, $i \in \mathrm{supp}(x) \Rightarrow 0 < |x_i| = |x'_i|$. As $|x| < |x'|$, we deduce that there exists $j \in \mathrm{supp}(x') \backslash \mathrm{supp}(x)$ such that $0 = |x_j| < |x'_j|$. We define a new vector $\tilde{x}$ by $\tilde{x}_j = 1/2x'_j \neq 0$ and $\tilde{x}_i = x_i$ for $i \neq j$. Putting $I = \mathrm{supp}(x)$, we have $\tilde{x} = x_I + 1/2x'_j e_j = \tilde{x}_I + \tilde{x}_{\{j\}}$, where e_j denotes the j-canonical vector of $\mathbb{R}^d$. On the one hand, from the first case we obtain that $|||\tilde{x}||| < |||x'|||$. On the other hand, we have $|||x||| \leq |||\tilde{x}|||$; indeed, by Proposition 2.4, Item 4 implies that the norm $|||\cdot|||$ is orthant-monotonic, hence that $|||\tilde{x}||| = |||\tilde{x}_I + \tilde{x}_{\{j\}}||| \geq |||\tilde{x}_I||| = |||x|||$. We conclude that $|||x||| \leq |||\tilde{x}||| < |||x'|||$.

This ends the proof. □

As an example, we illustrate Item 4 of Proposition 2.6 with the ℓ_1 (orthant-strictly monotonic) and ℓ_∞ (not orthant-strictly monotonic) norms.

- For any vector $u \in \mathbb{R}^d$, we have seen (right after the proof of Proposition 2.6) that the vector $v = \mathrm{sign}(u)$ is such that $\mathrm{supp}(v) = \mathrm{supp}(u)$, that $u \circ v \geq 0$, and is $\|\cdot\|_1$-dual to the vector u. This is another proof that the norm ℓ_1 is orthant-strictly monotonic.
- By contrast, if the vector $v \neq 0$ is $\|\cdot\|_\infty$-dual to the vector $u = (1, 1/2, 0, \ldots, 0)$, then an easy computation shows that, necessarily, $v = (v_1, 0, 0, \ldots, 0)$ with $v_1 > 0$. As a consequence, this gives $\{1\} = \mathrm{supp}(v) \subsetneq \mathrm{supp}(u) = \{1, 2\}$. This suffices to prove that the norm ℓ_∞ is not orthant-strictly monotonic.

We end this §2.3 with additional properties related to exposed and extreme points of the unit ball $\mathbb{B}$ of an orthant-strictly monotonic norm $|||\cdot|||$. We recall that an element x of a convex set C is called an *exposed point* of C if there exists a support hyperplane H to the convex set C at x such that $H \cap C = \{x\}$. We show in the next proposition that orthant-strictly monotonicity implies that the intersection of the unit sphere $\mathbb{S}$ with the subspaces $\mathcal{R}_{\{i\}}$ in (8), for $i \in [\![1, d]\!]$, is made of exposed points of the unit ball $\mathbb{B}$.

Proposition 2.7. *If the norm $|||\cdot|||$ is orthant-strictly monotonic, then the elements of the renormalized canonical basis of $\mathbb{R}^d$, that is the $e_i/|||e_i|||$ for $i \in [\![1, d]\!]$, are exposed points of the unit ball $\mathbb{B}$.*

Proof. Assume that the norm $|||\cdot|||$ is orthant-strictly monotonic and fix $i \in [\![1,d]\!]$. Then, using item 2 of Proposition 2.6, we have that $|||\overline{e}_i + \sum_{j\in[\![1,d]\!]\setminus\{i\}} \lambda_j \overline{e}_j||| > |||\overline{e}_i|||$, for all $\{\lambda_j\}_{j\in[\![1,d]\!]\setminus\{i\}}$ where not all λ_j's are 0 and where $\overline{e}_j = e_j/|||e_j|||$ for all $j \in [\![1,d]\!]$. This means that the renormalized canonical basis is strongly orthonormal relative to $\overline{e}_i$ in the sense of Birkhoff. Using [19, Theorem 2.6], we obtain that $\overline{e}_i$ is an exposed point of the unit ball $\mathbb{B}$. □

We recall that an *extreme point* x of a convex set C cannot be written as $x = \lambda x' + (1-\lambda)x''$ with $x' \in C$, $x'' \in C$, $x' \neq x$, $x'' \neq x$ and $\lambda \in]0,1[$. The normed space $(\mathbb{R}^d, |||\cdot|||)$ is said to be *strictly convex* if the unit ball $\mathbb{B}$ (of the norm $|||\cdot|||$) is *rotund*, that is, if all points of the unit sphere $\mathbb{S}$ are extreme points of the unit ball $\mathbb{B}$. The normed space $(\mathbb{R}^d, \|\cdot\|_p)$, equipped with the ℓ_p-norm $\|\cdot\|_p$ (for $p \in [1,\infty]$), is strictly convex if and only if $p \in]1,\infty[$.

Proposition 2.8. *If the norm $|||\cdot|||$ is orthant-monotonic and if the normed space $(\mathbb{R}^d, |||\cdot|||)$ is strictly convex, then the norm $|||\cdot|||$ is orthant-strictly monotonic.*

Proof. In [21, Theorem 2.2], we find the following result: if the family $\{\mathcal{R}_K\}_{K\subset[\![1,d]\!]}$ of subspaces of $\mathbb{R}^d$ is Birkhoff orthogonal for a norm $|||\cdot|||$, and if the unit ball for that norm is rotund, then the family $\{\mathcal{R}_K\}_{K\subset[\![1,d]\!]}$ is strictly Birkhoff orthogonal.

Now for the proof of the proposition. If the norm $|||\cdot|||$ is orthant-monotonic, then the family $\{\mathcal{R}_K\}_{K\subset[\![1,d]\!]}$ of subspaces of $\mathbb{R}^d$ is Birkhoff orthogonal by Item 4 in Proposition 2.4. As the unit ball for that norm is rotund, we deduce that the family $\{\mathcal{R}_K\}_{K\subset[\![1,d]\!]}$ is strictly Birkhoff orthogonal. As Item 2 implies Item 1 in Proposition 2.6, we conclude that the norm $|||\cdot|||$ is orthant-strictly monotonic. □

3. Generalized top-k and k-support norms

Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$, that we call the *source norm*. In §3.1, we introduce generalized top-k and k-support norms constructed from the source norm, and we provide various examples. In §3.2, we establish properties valid for any source norm, whereas, in §3.3, we establish properties valid when the source norm is orthant-monotonic, making thus the connection with the previous Sect. 2.

3.1. Definition and examples

We introduce generalized top-k and k-support norms that are constructed from the source norm $|||\cdot|||$.

Definition 3.1. For $k \in [\![1,d]\!]$, we call *generalized top-k norm* (associated with the source norm $|||\cdot|||$) the norm defined by[5]

$$|||x|||^{\mathrm{tn}}_{(k)} = \sup_{|K|\leq k} |||x_K||| \ , \ \ \forall x \in \mathbb{R}^d \ . \tag{14}$$

We call *generalized k-support norm* the dual norm of the generalized top-k norm, denoted by[6] $|||\cdot|||^{\star\mathrm{sn}}_{(k)}$:

$$|||\cdot|||^{\star\mathrm{sn}}_{(k)} = \Big(|||\cdot|||^{\mathrm{tn}}_{(k)}\Big)_\star \ . \tag{15}$$

[5] The notation $\sup_{|K|\leq k}$ is a shorthand for $\sup_{K\subset[\![1,d]\!],|K|\leq k}$.

[6] We use the symbol $\star$ in the superscript to indicate that the generalized k-support norm $|||\cdot|||^{\star\mathrm{sn}}_{(k)}$ is a dual norm. To stress the point, we use the letter x for a primal vector, like in $|||x|||^{\mathrm{tn}}_{(k)}$, and the letter y for a dual vector, like in $|||y|||^{\star\mathrm{sn}}_{(k)}$.

It is easily verified that $|||\cdot|||^{\mathrm{tn}}_{(k)}$ indeed is a norm, for all $k \in [\![1, d]\!]$.

We provide examples of generalized top-k and k-support norms in the case of permutation invariant monotonic source norms and of ℓ_p source norms. Table 3.1 provides a summary.

The case of permutation invariant monotonic source norms

Letting $x \in \mathbb{R}^d$ and ν be a permutation of $[\![1, d]\!]$ such that $|x_{\nu(1)}| \geq |x_{\nu(2)}| \geq \cdots \geq |x_{\nu(d)}|$, we note $x^{\downarrow} = \big(|x_{\nu(1)}|, |x_{\nu(2)}|, \ldots, |x_{\nu(d)}|\big)$. The proof of the following Lemma is easy.

Lemma 3.2. *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$. Then, if the norm $|||\cdot|||$ is permutation invariant and monotonic, we have that $|||x|||^{\mathrm{tn}}_{(k)} = |||x^{\downarrow}_{[\![1,k]\!]}|||$, where $x^{\downarrow}_{[\![1,k]\!]} \in \mathbb{R}^d$ is given by $(x^{\downarrow})_{[\![1,k]\!]}$, for all $x \in \mathbb{R}^d$.*

source norm $\lvert\lvert\lvert\cdot\rvert\rvert\rvert$	$\lvert\lvert\lvert\cdot\rvert\rvert\rvert^{\mathrm{tn}}_{(k)}$	$\lvert\lvert\lvert\cdot\rvert\rvert\rvert^{\star\mathrm{sn}}_{(k)}$
$\lVert\cdot\rVert_p$	top-(p,k) norm $\lVert x\rVert^{\mathrm{tn}}_{p,k}$ $= \big(\sum_{j=1}^{k} \lvert x_{\nu(j)}\rvert^p\big)^{1/p}$	(q,k)-support norm $\lVert y\rVert^{\mathrm{sn}}_{q,k}$ $1/p + 1/q = 1$
$\lVert\cdot\rVert_1$	top-(1,k) norm $\lVert x\rVert^{\mathrm{tn}}_{1,k} = \sum_{l=1}^{k} \lvert x_{\nu(l)}\rvert$	(∞,k)-support norm $\lVert y\rVert^{\mathrm{sn}}_{\infty,k} = \max\{\lVert y\rVert_1/k, \lVert y\rVert_\infty\}$
$\lVert\cdot\rVert_2$	top-(2,k) norm $\lVert x\rVert^{\mathrm{tn}}_{2,k} = \sqrt{\sum_{l=1}^{k} \lvert x_{\nu(l)}\rvert^2}$	(2,k)-support norm $\lVert y\rVert^{\mathrm{sn}}_{2,k}$ no analytic expression (computation in [2, Prop. 2.1])
$\lVert\cdot\rVert_\infty$	top-(∞,k) norm ℓ_∞-norm $\lVert x\rVert^{\mathrm{tn}}_{\infty,k} = \lvert x_{\nu(1)}\rvert = \lVert x\rVert_\infty$	(1,k)-support norm ℓ_1-norm $\lVert y\rVert^{\mathrm{sn}}_{1,k} = \lVert y\rVert_1$

Table 3.1: Examples of generalized top-k and k-support norms generated by the ℓ_p source norms $|||\cdot||| = \|\cdot\|_p$ for $p \in [1, \infty]$; ν is a permutation of $[\![1, d]\!]$ such that $|x_{\nu(1)}| \geq |x_{\nu(2)}| \geq \cdots \geq |x_{\nu(d)}|$

The case of ℓ_p source norms

We start with generalized top-k norms as in (14) (see the second column of Table 3.1). When the norm $|||\cdot|||$ is the Euclidean norm $\|\cdot\|_2$ of $\mathbb{R}^d$, the generalized top-k norm is known under different names: the top-$(k, 2)$ norm in [22, p. 8], (in the published version [11], the "top" terminology no longer appears) or the 2-k-symmetric gauge norm [17] or the Ky Fan vector norm [18]. Indeed, in all these cases, the norm of a vector x is obtained with a subvector of size k having the k largest components in module, because the assumptions of Lemma 3.2 are satisfied. More generally, when the norm $|||\cdot|||$ is the ℓ_p-norm $\|\cdot\|_p$, for $p \in [1, \infty]$, the assumptions of Lemma 3.2 are also satisfied, as ℓ_p-norms are permutation invariant and monotonic. Therefore, we obtain that the corresponding generalized top-k norm $\big(\|\cdot\|_p\big)^{\mathrm{tn}}_{(k)}$ has the expression $\big(\|\cdot\|_p\big)^{\mathrm{tn}}_{(k)}(x) = \sup_{|K| \leq k} \|x_K\|_p = \big\|x^{\downarrow}_{[\![1,k]\!]}\big\|_p$, for all $x \in \mathbb{R}^d$. Thus, we have obtained that the generalized top-k norm associated with the ℓ_p-norm is the norm $\|(\cdot)^{\downarrow}_{[\![1,k]\!]}\|_p$: we call it[7] *top-(p,k) norm* and we denote it by $\|\cdot\|^{\mathrm{tn}}_{p,k}$. Notice that $\|\cdot\|^{\mathrm{tn}}_{\infty,k} = \|\cdot\|_\infty$ for all $k \in [\![1,d]\!]$.

Now, we turn to generalized k-support norms as in (15) (see the third column of Table 3.1). When the norm $|\!|\!|\cdot|\!|\!|$ is the Euclidean norm $\|\cdot\|_2$ of $\mathbb{R}^d$, the generalized k-support norm is the so-called k-support norm [2]. More generally, in [16, Definition 21], the authors define the k-support p-norm or *(p,k)-support norm* for $p \in [1,\infty]$. They show, in [16, Corollary 22], that the dual norm $\Big(\big(\|\cdot\|_p\big)^{\mathrm{tn}}_{(k)}\Big)_\star$ of the above top-(k,p) norm is the (q,k)-support norm, where $1/p + 1/q = 1$. Thus, what we call the generalized k-support norm $\big(\|\cdot\|_p\big)^{\star\mathrm{sn}}_{(k)} = \Big(\big(\|\cdot\|_p\big)^{\mathrm{tn}}_{(k)}\Big)_\star$ associated with the ℓ_p-norm is the (q,k)-support norm, that we denote $\|y\|^{\mathrm{sn}}_{q,k}$. The formula $\|x\|^{\mathrm{sn}}_{\infty,k} = \max\{\|x\|_1/k, \|x\|_\infty\}$ can be found in [5, Exercise IV.1.18, p. 90].

3.2. General properties

We establish properties of generalized top-k and k-support norms, valid for any source norm, that will be useful to prove our results in Sect. 4.

Properties of generalized top-k norms

We denote the unit ball of the generalized top-k norm $|\!|\!|\cdot|\!|\!|^{\mathrm{tn}}_{(k)}$ in Definition 3.1 by

$$\mathbb{B}^{\mathrm{tn}}_{(k)} = \{x \in \mathbb{R}^d \mid |\!|\!|x|\!|\!|^{\mathrm{tn}}_{(k)} \leq 1\} \ , \ \ \forall k \in [\![1,d]\!] \ . \tag{16}$$

Proposition 3.3.

- *For $k \in [\![1,d]\!]$, the generalized top-k norm $|\!|\!|\cdot|\!|\!|^{\mathrm{tn}}_{(k)}$ (in Definition 3.1) has the expression*
$$|\!|\!|x|\!|\!|^{\mathrm{tn}}_{(k)} = \sup_{|K|\leq k} \sigma_{\pi_K(\mathbb{S}_\star)}(x) \ , \ \ \forall x \in \mathbb{R}^d \ , \tag{17}$$
where $\mathbb{S}_\star$ is the unit sphere of the dual norm $|\!|\!|\cdot|\!|\!|_\star$ as in (4a).
- *We have the inequality* $$|\!|\!|x|\!|\!| \leq |\!|\!|x|\!|\!|^{\mathrm{tn}}_{(d)} \ , \ \ \forall x \in \mathbb{R}^d \ . \tag{18}$$
- *The sequence $\big\{|\!|\!|\cdot|\!|\!|^{\mathrm{tn}}_{(j)}\big\}_{j\in[\![1,d]\!]}$ of generalized top-k norms in (14) is nondecreasing, in the sense that the following inequalities hold true*
$$|\!|\!|x|\!|\!|^{\mathrm{tn}}_{(1)} \leq \cdots \leq |\!|\!|x|\!|\!|^{\mathrm{tn}}_{(j)} \leq |\!|\!|x|\!|\!|^{\mathrm{tn}}_{(j+1)} \leq \cdots \leq |\!|\!|x|\!|\!|^{\mathrm{tn}}_{(d)} \ , \ \ \forall x \in \mathbb{R}^d \ . \tag{19}$$
- *The sequence $\big\{\mathbb{B}^{\mathrm{tn}}_{(j)}\big\}_{j\in[\![1,d]\!]}$ of units balls of the generalized top-k norms in (16) is nonincreasing, in the sense that the following inclusions hold true:*
$$\mathbb{B}^{\mathrm{tn}}_{(d)} \subset \cdots \subset \mathbb{B}^{\mathrm{tn}}_{(j+1)} \subset \mathbb{B}^{\mathrm{tn}}_{(j)} \subset \cdots \subset \mathbb{B}^{\mathrm{tn}}_{(1)} \ . \tag{20}$$

[7] We invert the indices in the naming convention of [22, p. 5, p. 8], where top-$(k,1)$ and top-$(k,2)$ were used.

Proof. • For any $x \in \mathbb{R}^d$, we have

$$
\begin{aligned}
|||x|||^{\mathrm{tn}}_{(k)} &= \sup_{|K|\le k} |||x_K||| \quad \text{(by definition (14) of the generalized top-}k\text{ norm } |||\cdot|||^{\mathrm{tn}}_{(k)}) \\
&= \sup_{|K|\le k} \sigma_{\mathbb{S}_\star}(x_K) \quad \text{(by (6a))} \\
&= \sup_{|K|\le k} \sup_{y\in\mathbb{S}_\star} \langle x_K ,\, y\rangle \quad \text{(by definition (5) of the support function } \sigma_{\mathbb{S}_\star}) \\
&= \sup_{|K|\le k} \sup_{y\in\mathbb{S}_\star} \langle x ,\, \pi_K(y)\rangle \\
&\qquad \text{(by the self-duality property (9) of the projection mapping } \pi_K) \\
&= \sup_{|K|\le k} \sup_{y'\in\pi_K(\mathbb{S}_\star)} \langle x ,\, y'\rangle \\
&= \sup_{|K|\le k} \sigma_{\pi_K(\mathbb{S}_\star)}(x) \quad \text{(by definition (5) of the support function } \sigma_{\pi_K(\mathbb{S}_\star)})
\end{aligned}
$$

and we get (17).

• From the very definition (14) of the generalized top-d norm $|||\cdot|||^{\mathrm{tn}}_{(d)}$, we get

$$
|||x|||^{\mathrm{tn}}_{(d)} = \sup_{|K|\le d} |||x_K||| \ge |||x_{[\![1,d]\!]}||| = |||x||| \ , \ \ \forall x \in \mathbb{R}^d \ ,
$$

hence (18).

• The inequalities (19) between norms easily derive from the very definition (14) of the generalized top-k norms $|||\cdot|||^{\mathrm{tn}}_{(k)}$.

• The inclusions (20) between unit balls directly follow from the inequalities (19) between norms.

This ends the proof. □

Properties of generalized k-support norms

We denote the unit ball of the generalized k-support norm $|||\cdot|||^{\star\mathrm{sn}}_{(k)}$ in Definition 3.1 by

$$
\mathbb{B}^{\star\mathrm{sn}}_{(k)} = \{ y \in \mathbb{R}^d \mid |||y|||^{\star\mathrm{sn}}_{(k)} \le 1 \} \ , \ \ \forall k \in [\![1,d]\!] \ . \tag{21}
$$

Proposition 3.4.

- *For $k \in [\![1,d]\!]$, the generalized k-support norm $|||\cdot|||^{\star\mathrm{sn}}_{(k)}$ in Definition 3.1 has unit ball*
$$
\mathbb{B}^{\star\mathrm{sn}}_{(k)} = \overline{\mathrm{co}}\Big(\bigcup_{|K|\le k} \pi_K(\mathbb{S}_\star) \Big) \ , \tag{22}
$$
where $\overline{\mathrm{co}}(S)$ denotes the closed convex hull of a subset $S \subset \mathbb{R}^d$.
- *We have the inequality* $\quad |||y|||^{\star\mathrm{sn}}_{(d)} \le |||y|||_\star \ , \ \ \forall y \in \mathbb{R}^d \ . \qquad (23)$
- *The sequence $\big\{ |||\cdot|||^{\star\mathrm{sn}}_{(j)} \big\}_{j\in[\![1,d]\!]}$ of generalized k-support norms in (15) is nonincreasing, in the sense that the following inequalities hold true*
$$
|||y|||^{\star\mathrm{sn}}_{(d)} \le \cdots \le |||y|||^{\star\mathrm{sn}}_{(j+1)} \le |||y|||^{\star\mathrm{sn}}_{(j)} \le \cdots \le |||y|||^{\star\mathrm{sn}}_{(1)} \ , \ \ \forall y \in \mathbb{R}^d \ . \tag{24}
$$

- *The sequence* $\left\{\mathbb{B}^{\star\mathrm{sn}}_{(j)}\right\}_{j\in[\![1,d]\!]}$ *of units balls of the generalized* k*-support norms in* (21) *is nondecreasing, in the sense that the following inclusions hold true:*

$$\mathbb{B}^{\star\mathrm{sn}}_{(1)} \subset \cdots \subset \mathbb{B}^{\star\mathrm{sn}}_{(j)} \subset \mathbb{B}^{\star\mathrm{sn}}_{(j+1)} \subset \cdots \subset \mathbb{B}^{\star\mathrm{sn}}_{(d)} \; . \tag{25}$$

Proof. • For any $x \in \mathbb{R}^d$, we have

$$\begin{aligned}
|||x|||^{\mathrm{tn}}_{(k)} &= \sup_{|K|\leq k} \sigma_{\pi_K(\mathbb{S}_\star)}(x) && \text{(by (17))}\\
&= \sigma_{\bigcup_{|K|\leq k} \pi_K(\mathbb{S}_\star)}(x) \\
& \qquad\text{(as the support function turns a union of sets into a supremum)}\\
&= \sigma_{\overline{\mathrm{co}}\left(\bigcup_{|K|\leq k} \pi_K(\mathbb{S}_\star)\right)}(x) && \text{(by [4, Prop. 7.13])}
\end{aligned}$$

and we obtain (22) thanks to (6a).

• From the inequality (18) between norms, we deduce the inequality (23) between dual norms, by the definition (3) of a dual norm.

• The inequalities in (24) easily derive from the inclusions (25).

• The inclusions (25) directly follow from the inclusions (20) and from (6b) as $\mathbb{B}^{\star\mathrm{sn}}_{(k)} = \left(\mathbb{B}^{\mathrm{tn}}_{(k)}\right)^{\odot}$, the polar set of $\mathbb{B}^{\mathrm{tn}}_{(k)}$.

This ends the proof. □

3.3. Properties under orthant-monotonicity

We establish properties of generalized top-k and k-support norms, valid when the source norm is orthant-monotonic, that will be useful to prove our results in Sect. 4.

Proposition 3.5.

(1) *Let* $k \in [\![1,d]\!]$*. If the source norm* $|||\cdot|||$ *is orthant-monotonic, then*

- *the generalized top-*k *norm has the expression*

$$|||x|||^{\mathrm{tn}}_{(k)} = \sup_{|K|\leq k} \sigma_{\mathcal{R}_K \cap \mathbb{S}_\star}(x) \; , \;\; \forall x \in \mathbb{R}^d \; , \tag{26}$$

where $\mathbb{S}_\star$ *is the unit sphere of the dual norm* $|||\cdot|||_\star$ *as in* (4a)*,*

- *the unit ball of the* k*-support norm is given by*

$$\mathbb{B}^{\star\mathrm{sn}}_{(k)} = \overline{\mathrm{co}}\Big(\bigcup_{|K|\leq k} (\mathcal{R}_K \cap \mathbb{S}_\star) \Big) \; . \tag{27}$$

(2) *The source norm* $|||\cdot|||$ *is orthant-monotonic if and only if* $|||\cdot||| = |||\cdot|||^{\mathrm{tn}}_{(d)}$ *if and only if* $|||\cdot|||_\star = |||\cdot|||^{\star\mathrm{sn}}_{(d)}$.

(3) *If the source norm* $|||\cdot|||$ *is orthant-monotonic, then the generalized top-*k *norms and the generalized* k*-support norms are orthant-monotonic.*

Proof. (1) We suppose that the source norm $|||\cdot|||$ is orthant-monotonic.
Let $k \in [\![1,d]\!]$.

- We prove (26). For any $x \in \mathbb{R}^d$, we have

$$
\begin{aligned}
|||x|||^{\mathrm{tn}}_{(k)} &= \sup_{|K|\leq k} |||x_K||| && \text{(by definition (14) of the generalized top-}k\text{ norm)} \\
&= \sup_{|K|\leq k} |||x_K|||_{\star\star} && \text{(as any norm is equal to its bidual norm by (6d))} \\
&= \sup_{|K|\leq k} \big(|||\cdot|||_{\star}\big)_{\star,K}(x_K) && \text{(by Definition 2.1 of the the }(\star,K)\text{-norm)} \\
&= \sup_{|K|\leq k} \big(|||\cdot|||_{\star}\big)_{K,\star}(x_K) && \text{(by Item 3 in Proposition 2.4 because, as the norm} \\
& && |||\cdot|||\text{ is orthant-monotonic, so is also the dual norm} \\
& && |||\cdot|||_{\star}\text{ (equivalence between (1) and (2) in Prop. 2.4))} \\
&= \sup_{|K|\leq k} \sigma_{\mathcal{R}_K\cap\mathbb{S}_{\star}}(x_K) && \text{(by (13b) applied to }|||\cdot|||_{\star}\text{ with }x_K\in\mathcal{R}_K) \\
&= \sup_{|K|\leq k} \sigma_{\mathcal{R}_K\cap\mathbb{S}_{\star}}(x)
\end{aligned}
$$

by the self-duality property (9) of the projection mapping π_K, and by definition (8) of the subspace $\mathcal{R}_K$.

- We prove (27). Indeed, by (26), we know that $|||\cdot|||^{\mathrm{tn}}_{(k)} = \sup_{|K|\leq k} \sigma_{\mathcal{R}_K\cap\mathbb{S}_{\star}}$. As we have $\sup_{|K|\leq k} \sigma_{\mathcal{R}_K\cap\mathbb{S}_{\star}} = \sigma_{\bigcup_{|K|\leq k}(\mathcal{R}_K\cap\mathbb{S}_{\star})}$, we have just established that this implies $|||\cdot|||^{\mathrm{tn}}_{(k)} = \sigma_{\cup_{|K|\leq k}(\mathcal{R}_K\cap\mathbb{S}_{\star})}$. On the other hand, by (6a) we have that $|||\cdot|||^{\mathrm{tn}}_{(k)} = \sigma_{\mathbb{B}^{\star\mathrm{sn}}_{(k)}}$ since, by Definition 3.1, the k-support norm is the dual norm of the top-k norm. Then, by [4, Prop. 7.13], we deduce that $\overline{\mathrm{co}}\big(\mathbb{B}^{\star\mathrm{sn}}_{(k)}\big) = \overline{\mathrm{co}}\big(\bigcup_{|K|\leq k}(\mathcal{R}_K\cap\mathbb{S}_{\star})\big)$. As the unit ball $\mathbb{B}^{\star\mathrm{sn}}_{(k)}$ in (21) is closed and convex, we immediately obtain (27).

(2) First, let us observe that, from the very definition (14) of the generalized top-d norm $|||\cdot|||^{\mathrm{tn}}_{(d)}$, and by (18), we have, for all $x \in \mathbb{R}^d$:

$$
|||x|||^{\mathrm{tn}}_{(d)} = |||x||| \iff \sup_{|K|\leq d} |||x_K||| = |||x||| \iff |||x_K||| \leq |||x||| \ , \ \forall K \subset [\![1,d]\!] \ . \tag{28}
$$

Now, we turn to prove Item 2 as two reverse implications.

Suppose that the source norm $|||\cdot|||$ is orthant-monotonic, and let us prove that $|||x|||^{\mathrm{tn}}_{(d)} = |||x|||$. By Item 7 in Proposition 2.4, we get that $|||x_K||| \leq |||x|||$, for all $K \subset [\![1,d]\!]$, hence $|||x|||^{\mathrm{tn}}_{(d)} = |||x|||$, for all $x \in \mathbb{R}^d$ by the just proven equivalence (28).

Suppose that $|||x|||^{\mathrm{tn}}_{(d)} = |||x|||$ and let us prove that the source norm $|||\cdot|||$ is orthant-monotonic. By (28), we have that $|||x_J||| \leq |||x|||$, for all $x \in \mathbb{R}^d$ and all $J \subset [\![1,d]\!]$. This gives, in particular, $|||x_{K\cap J}||| = |||(x_K)_J||| \leq |||x_K|||$; if $J \subset K$, we deduce that $|||x_J||| \leq |||x_K|||$. Thus, Item 7 in Proposition 2.4 holds true, and we obtain that the source norm $|||\cdot|||$ is orthant-monotonic.

We end the proof by taking the dual norms, as in (3), of both sides of the equality $|||\cdot||| = |||\cdot|||^{\mathrm{tn}}_{(d)}$, yielding $|||\cdot|||_{\star} = |||\cdot|||^{\star\mathrm{sn}}_{(d)}$ by (15).

(3) The generalized top-k norm in (14) is the supremum of the subfamily, when $|K| \leq k$, of the seminorms $|||\pi_K(\cdot)|||_K$. As already mentioned, the definition of orthant-monotonic norms can be extended to seminorms. With this extension, it is easily seen that the seminorms $|||\pi_K(\cdot)|||_K$ are orthant-monotonic as soon as the

source norm $|||\cdot|||$ is orthant-monotonic. Therefore, if the source norm $|||\cdot|||$ is orthant-monotonic, so is the supremum in (14), thanks to the property claimed right after the Definition 2.2: the supremum of a family of orthant-monotonic seminorms is an orthant-monotonic seminorm. Thus, we have established that the generalized top-k norm in (14) is orthant-monotonic. We deduce that its dual norm, the generalized k-support norm $|||\cdot|||^{\star\mathrm{sn}}_{(k)}$ in (15), is orthant-monotonic. Indeed, the dual norm of an orthant-monotonic norm $|||\cdot|||$ is orthant-monotonic, as proved in [12, Theorem 2.23] (equivalence between Item 1 and Item 2 in Proposition 2.4).

This ends the proof. □

4. The ℓ_0 pseudonorm, orthant-monotonicity and generalized top-k and k-support norms

In §4.1, we introduce basic notation regarding the ℓ_0 pseudonorm. In §4.2, we introduce the notions of (strictly) increasingly or decreasingly graded sequences of norms, and we display conditions for generalized top-k norms or generalized k-support norms to be graded sequences.

4.1. Level sets of the ℓ_0 pseudonorm

The so-called ℓ_0 *pseudonorm* is the function $\ell_0 : \mathbb{R}^d \to [\![0, d]\!]$ defined, for any $x \in \mathbb{R}^d$, by

$$\ell_0(x) = |\mathrm{supp}(x)| = \text{number of nonzero components of } x \ . \tag{29}$$

The ℓ_0 pseudonorm shares three out of the four axioms of a norm: nonnegativity, positivity except for $x = 0$, subadditivity. The axiom of 1-homogeneity does not hold true; by contrast, the ℓ_0 pseudonorm is 0-homogeneous:

$$\ell_0(\rho x) = \ell_0(x) \ , \ \ \forall \rho \in \mathbb{R}\backslash\{0\} \ , \ \ \forall x \in \mathbb{R}^d \ . \tag{30}$$

We introduce the *level sets*

$$\ell_0^{\leq k} = \{x \in \mathbb{R}^d \mid \ell_0(x) \leq k\} \ , \ \ \forall k \in [\![0, d]\!] \ . \tag{31}$$

The level sets of the ℓ_0 pseudonorm in (31) are easily related to the subspaces $\mathcal{R}_K$ of $\mathbb{R}^d$, as defined in (8), by

$$\ell_0^{\leq k} = \{x \in \mathbb{R}^d \mid \ell_0(x) \leq k\} = \bigcup_{|K| \leq k} \mathcal{R}_K \ , \ \ \forall k \in [\![0, d]\!] \ , \tag{32}$$

where the notation $\bigcup_{|K|\leq k}$ is a shorthand for $\bigcup_{K\subset[\![1,d]\!],|K|\leq k}$.

If the source norm $|||\cdot|||$ is orthant-monotonic, the expression (27) of the unit ball of the k-support norm can be written with the level sets of the ℓ_0 pseudonorm as

$$\mathbb{B}^{\star\mathrm{sn}}_{(k)} = \overline{\mathrm{co}}\Big(\bigcup_{|K|\leq k} (\mathcal{R}_K \cap \mathbb{S}_\star) \Big) = \overline{\mathrm{co}}\big(\ell_0^{\leq k} \cap \mathbb{S}_\star \big) \ . \tag{33}$$

This formula is reminiscent of (and generalizes) [2, Equation (2)], which was established for the Euclidean source norm. With an additional assumption, we obtain a refinement. The proof of the following Proposition 4.1 relies on Lemma 4.2 and its Corollary 4.3.

Proposition 4.1. *If the source norm $|||\cdot|||$ is orthant-monotonic and if the normed space $\big(\mathbb{R}^d, |||\cdot|||_\star\big)$ is strictly convex, then we have*

$$\ell_0^{\leq k} \cap \mathbb{S}_\star = \mathbb{B}^{\star\mathrm{sn}}_{(k)} \cap \mathbb{S}_\star \; , \;\; \forall k \in [\![0,d]\!] \; , \tag{34}$$

where $\ell_0^{\leq k}$ is the level set in (31) *of the ℓ_0 pseudonorm in* (29)*, where $\mathbb{S}_\star$ in* (4a) *is the unit sphere of the dual norm $|||\cdot|||_\star$, and where $\mathbb{B}^{\star\mathrm{sn}}_{(k)}$ in* (21) *is the unit ball of the generalized k-support norm $|||\cdot|||^{\star\mathrm{sn}}_{(k)}$.*

Proof. First, let us observe that the level set $\ell_0^{\leq k}$ in (31) is closed because the pseudonorm ℓ_0 is lower semi continuous. Then, we get

$$\begin{aligned}
\ell_0^{\leq k} \cap \mathbb{S}_\star &= \overline{\mathrm{co}}\big(\ell_0^{\leq k} \cap \mathbb{S}_\star\big) \cap \mathbb{S}_\star && \text{(by Corollary 4.3 because } \ell_0^{\leq k} \cap \mathbb{S}_\star \subset \mathbb{S}_\star \text{ and is} \\
& && \text{closed, and because the unit ball } \mathbb{B}_\star \text{ is rotund)} \\
&= \overline{\mathrm{co}}\Big(\bigcup_{|K|\leq k} (\mathcal{R}_K \cap \mathbb{S}_\star)\Big) \cap \mathbb{S}_\star && \text{(as } \ell_0^{\leq k} = \textstyle\bigcup_{|K|\leq k} \mathcal{R}_K \text{ by (32))} \\
&= \mathbb{B}^{\star\mathrm{sn}}_{(k)} \cap \mathbb{S}_\star
\end{aligned}$$

as $\overline{\mathrm{co}}\big(\bigcup_{|K|\leq k}(\mathcal{R}_K \cap \mathbb{S}_\star)\big) = \mathbb{B}^{\star\mathrm{sn}}_{(k)}$ by (27) because the source norm $|||\cdot|||$ is orthant-monotonic.

This ends the proof. □

The result of Proposition 4.1 applies to the ℓ_p-norm $\|\cdot\|_p$ for $p \in]1,\infty[$.

Lemma 4.2. *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$. Let $\widetilde{\mathbb{S}}$ be a subset of* $\mathrm{extr}(\mathbb{B}) \subset \mathbb{S}$*, the set of extreme points of* $\mathbb{B}$*. If A is a subset of $\widetilde{\mathbb{S}}$, then $A = \mathrm{co}(A) \cap \widetilde{\mathbb{S}}$. If A is a closed subset of $\widetilde{\mathbb{S}}$, then $A = \overline{\mathrm{co}}(A) \cap \widetilde{\mathbb{S}}$.*

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

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

Hence, we have proved by contradiction that $\mathrm{co}(A) \cap \widetilde{\mathbb{S}} \subset \mathrm{extr}\big(\mathrm{co}(A)\big)$. We can conclude using the fact that $\mathrm{extr}\big(\mathrm{co}(A)\big) \subset A$, because the convex closure operation cannot generate new extreme points, as proved in [14, Exercice 6.4].

Now, we consider the case where the subset A of $\widetilde{\mathbb{S}}$ is closed. Using the first part of the proof we have that $A = \mathrm{co}(A) \cap \widetilde{\mathbb{S}}$. Now, A is closed by assumption and bounded since $A \subset \widetilde{\mathbb{S}} \subset \mathbb{S}$. Thus, A is a compact subset of $\mathbb{R}^d$ and, in a finite dimensional space, we get that $\mathrm{co}(A)$ is compact [20, Theorem 17.2], thus closed. We conclude that $A = \mathrm{co}(A) \cap \widetilde{\mathbb{S}} = \overline{\mathrm{co}(A)} \cap \widetilde{\mathbb{S}} = \overline{\mathrm{co}}(A) \cap \widetilde{\mathbb{S}}$, where the last equality comes from [4, Proposition 3.46].

This ends the proof. □

If the unit ball $\mathbb{B}$ is rotund, we then have that $\mathbb{S} = \mathrm{extr}(\mathbb{B})$, and we can apply Lemma 4.2 with $\widetilde{\mathbb{S}} = \mathbb{S}$ to obtain the following corollary.

Corollary 4.3. *Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$. Suppose that the unit ball of the norm $|||\cdot|||$ is rotund. If A is a subset of the unit sphere $\mathbb{S}$, then $A = \mathrm{co}(A) \cap \mathbb{S}$. If A is a closed subset of $\mathbb{S}$, then $A = \overline{\mathrm{co}}(A) \cap \mathbb{S}$.*

4.2. Graded sequences of norms

In [9], we introduced the notions of (strictly) decreasingly graded sequences of norms. In §4.2.1, we define (strictly) increasingly graded sequences of norms. In §4.2.2, we display conditions for generalized top-k norms to be (strictly) increasingly graded sequences. In §4.2.3, we display conditions for generalized k-support norms to be (strictly) decreasingly graded sequences. In §4.2.4, we express the level sets of the ℓ_0 pseudonorm in (31) by means of the difference between two norms.

4.2.1. Definitions of graded sequences of norms

In a sense, a graded sequence of norms is a monotone sequence that detects the number of nonzero components of a vector in $\mathbb{R}^d$ when the sequence becomes stationary.

Definition 4.4. We say that a sequence $\{|||\cdot|||_k\}_{k\in[\![1,d]\!]}$ of norms on $\mathbb{R}^d$ is *increasingly graded* (resp. *strictly increasingly graded*) w.r.t. (with respect to) the ℓ_0 pseudonorm if, for any $x \in \mathbb{R}^d$, one of the three following equivalent statements holds true.

(1) We have the implication (resp. equivalence), for any $l \in [\![1,d]\!]$,

$$\ell_0(x) = l \implies |||x|||_1 \leq \cdots \leq |||x|||_{l-1} \leq |||x|||_l = \cdots = |||x|||_d \;, \tag{35a}$$

$$(\text{resp.} \quad \ell_0(x) = l \iff |||x|||_1 \leq \cdots \leq |||x|||_{l-1} < |||x|||_l = \cdots = |||x|||_d.) \tag{35b}$$

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

$$\ell_0(x) \leq l \implies |||x|||_l = |||x|||_d \;, \tag{35c}$$

$$(\text{ resp.} \quad \ell_0(x) \leq l \iff |||x|||_l = |||x|||_d \quad \big(\iff |||x|||_l \leq |||x|||_d\big) \;. \;) \tag{35d}$$

(3) The sequence $k \in [\![1,d]\!] \mapsto |||x|||_k$ is nondecreasing and we have the inequality (resp. equality)

$$\ell_0(x) \geq \min\{k \in [\![1,d]\!] \mid |||x|||_k = |||x|||_d\} \ , \tag{35e}$$

$$(\text{ resp. } \quad \ell_0(x) = \min\{k \in [\![1,d]\!] \mid |||x|||_k = |||x|||_d\} \ . \ ) \tag{35f}$$

These definitions of (strictly) increasingly graded mimic the ones of (strictly) decreasingly graded in [9, Definition 1] (replace $\leq$ in (35a) by $\geq$, replace $\leq$ and $<$ in (35b) by $\geq$ and $>$, replace nondecreasing by nonincreasing in the two last items).

The property of orthant-strict monotonicity for norms, as introduced in Definition 2.5, proves especially relevant for the ℓ_0 pseudonorm and sequences of generalized top-k norms, as the following Propositions 4.5 and 4.7 reveal.

4.2.2. Sufficient conditions for increasingly graded sequence of generalized top-k norms

We show that, when the source norm is orthant-(strictly) monotonic, the sequence of induced generalized top-k norms is (strictly) increasingly graded.

Proposition 4.5.

- *If the norm $|||\cdot|||$ is orthant-monotonic, then the nondecreasing sequence $\{|||\cdot|||^{\mathrm{tn}}_{(j)}\}_{j\in[\![1,d]\!]}$ of generalized top-k norms in* (14) *is increasingly graded with respect to the ℓ_0 pseudonorm, that is,*

$$\ell_0(x) \leq l \Rightarrow |||x|||^{\mathrm{tn}}_{(l)} = |||x|||^{\mathrm{tn}}_{(d)} \ , \ \ \forall x \in \mathbb{R}^d \ , \ \ \forall l \in [\![0,d]\!] \ .$$

- *If the norm $|||\cdot|||$ is orthant-strictly monotonic, then the nondecreasing sequence $\{|||\cdot|||^{\mathrm{tn}}_{(j)}\}_{j\in[\![1,d]\!]}$ of generalized top-k norms in* (14) *is strictly increasingly graded with respect to the ℓ_0 pseudonorm, that is,*

$$\ell_0(x) \leq l \iff |||x|||^{\mathrm{tn}}_{(l)} = |||x|||^{\mathrm{tn}}_{(d)} \ , \ \ \forall x \in \mathbb{R}^d \ , \ \ \forall l \in [\![0,d]\!] \ .$$

Proof. • We suppose that the norm $|||\cdot|||$ is orthant-monotonic. As the sequence $\{|||\cdot|||^{\mathrm{tn}}_{(j)}\}_{j\in[\![1,d]\!]}$ of generalized top-k norms in (14) is nondecreasing by the inequalities (19), it suffices to show (35c) – that is, $\ell_0(x) \leq l \Rightarrow |||x|||^{\mathrm{tn}}_{(d)} = |||x|||^{\mathrm{tn}}_{(l)}$ – to prove that the sequence is increasingly graded with respect to the ℓ_0 pseudonorm.

For this purpose, we consider $x \in \mathbb{R}^d$, we put $L = \mathrm{supp}(x)$ and we suppose that $\ell_0(x) = |L| \leq l$. We now show that $|||x|||^{\mathrm{tn}}_{(d)} = |||x|||^{\mathrm{tn}}_{(l)}$. Since $x = x_L$, we have $|||x||| = |||x_L||| = |||x_L|||_L \leq |||x|||^{\mathrm{tn}}_{(l)}$, by the very definition (14) of the generalized top-l norm $|||\cdot|||^{\mathrm{tn}}_{(l)}$. On the one hand, we have just obtained that $|||x||| \leq |||x|||^{\mathrm{tn}}_{(l)}$. On the other hand, we have that $|||x|||^{\mathrm{tn}}_{(l)} \leq |||x|||^{\mathrm{tn}}_{(l+1)} \leq \cdots \leq |||x|||^{\mathrm{tn}}_{(d)} = |||x|||$ by the inequalities (19) and the last equality comes from Item 2 in Proposition 3.5 since the norm $|||\cdot|||$ is orthant-monotonic. This implies that $|||x||| = |||x|||^{\mathrm{tn}}_{(d)} = \cdots = |||x|||^{\mathrm{tn}}_{(l)}$, so that $|||x|||^{\mathrm{tn}}_{(k)}$ is stationary for $k \geq l$.

• We suppose that the norm $|||\cdot|||$ is orthant-strictly monotonic. To prove that the equivalence (35b) holds true for the sequence $\{|||\cdot|||^{\mathrm{tn}}_{(j)}\}_{j\in[\![1,d]\!]}$, it is easily seen that it suffices to show that

$$\ell_0(x) = l \Rightarrow |||x|||^{\mathrm{tn}}_{(1)} < \cdots < |||x|||^{\mathrm{tn}}_{(l-1)} < |||x|||^{\mathrm{tn}}_{(l)} = |||x|||^{\mathrm{tn}}_{(l+1)} = \cdots = |||x|||^{\mathrm{tn}}_{(d)} \; , \;\; \forall x \in \mathbb{R}^d \; . \tag{36}$$

We consider $x \in \mathbb{R}^d$. We put $L = \mathrm{supp}(x)$ and we suppose that $\ell_0(x) = |L| = l$. As the norm $|||\cdot|||$ is orthant-strictly monotonic, it is orthant-monotonic, so that the equalities $|||x|||^{\mathrm{tn}}_{(l)} = |||x|||^{\mathrm{tn}}_{(l+1)} = \cdots = |||x|||^{\mathrm{tn}}_{(d)}$ above hold true (as just established in the first part of the proof).

Therefore, it only remains to prove that $|||x|||^{\mathrm{tn}}_{(1)} < \cdots < |||x|||^{\mathrm{tn}}_{(l-1)} < |||x|||^{\mathrm{tn}}_{(l)}$.

There is nothing to show for $l = 0$. Now, for $l \geq 1$ and for any $k \in [\![0, l-1]\!]$, we have

$$\begin{aligned}
|||x|||^{\mathrm{tn}}_{(k)} &= \sup_{|K| \leq k} |||x_K||| && \text{(by definition (14) of the generalized top-}k\text{ norm)} \\
&= \sup_{|K| \leq k} |||x_{K \cap L}||| && (x_L = x \text{ by definition of the set } L = \mathrm{supp}(x)) \\
&= \sup_{|K'| \leq k, K' \subset L} |||x_{K'}||| && \text{(by setting } K' = K \cap L) \\
&= \sup_{|K| \leq k, K \subset L} |||y_K|||_\star && \text{(the same but with } K \text{ instead of } K') \\
&= \sup_{|K| \leq k, K \subsetneq L} |||x_K||| && (|K| \leq k \leq l-1 < l = |L| \text{ implies } K \neq L) \\
&< \sup_{\substack{|K| \leq k, j \in L \setminus K \\ K \subsetneq L}} |||x_{K \cup \{j\}}|||
\end{aligned}$$

(because the set $L \setminus K$ is nonempty (having cardinality $|L| - |K| = l - |K| \geq k + 1 - |K| \geq 1$), and because, since the norm $|||\cdot|||$ is orthant-strictly monotonic, using Item 3 in Proposition 2.6, we obtain that $|||x_K||| < |||x_{K \cup \{j\}}|||$ as $x_K \neq x_{K \cup \{j\}}$ for at least one $j \in L \setminus K$ since $L = \mathrm{supp}(x)$)

$$\begin{aligned}
&\leq \sup_{|J| \leq k+1, J \subset L} |||x_J||| && \text{(as all subsets } K' = K \cup \{j\} \text{ satisfy } K' \subset L \text{ and } |K'| = k+1) \\
&\leq |||x|||^{\mathrm{tn}}_{(k+1)}
\end{aligned}$$

by definition (14) of the generalized top-$k+1$ norm (in fact the last inequality is easily shown to be an equality as $x_L = x$). Thus, for any $k \in [\![0, l-1]\!]$, we have established that $|||x|||^{\mathrm{tn}}_{(k)} < |||x|||^{\mathrm{tn}}_{(k+1)}$.

This ends the proof. □

We show that, when the source norm is orthant-strictly monotonic, it is equivalent either that the sequence of induced generalized top-k norms be strictly increasingly graded or that the dual norm $|||\cdot|||_\star$ be orthant-strictly monotonic.

Proposition 4.6. *The following statements are equivalent.*

(1) *The dual norm $|||\cdot|||_\star$ is orthant-strictly monotonic and the sequence $\{|||\cdot|||^{\mathrm{tn}}_{(j)}\}_{j \in [\![1,d]\!]}$ of generalized top-k norms in* (14) *is strictly increasingly graded with respect to the ℓ_0 pseudonorm.*

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

Proof. • Suppose that Item 1 is satisfied and let us show that Item 2 holds true. For this, it suffices to prove that the norm $|||\cdot|||$ is orthant-strictly monotonic. To prove that the norm $|||\cdot|||$ is orthant-strictly monotonic, we will show that Item 3 in Proposition 2.6 holds true for $|||\cdot|||$. For this purpose, we consider $x \in \mathbb{R}^d$ and $J \subsetneq K \subset [\![1, d]\!]$ such that $x_J \neq x_K$. By definition of the ℓ_0 pseudonorm in (29), we have $j = \ell_0(x_J) < k = \ell_0(x_K)$.

On the one hand, as the dual norm $|||\cdot|||_\star$ is orthant-strictly monotonic, it is orthant-monotonic, so that the norm $|||\cdot|||$ is also orthant-monotonic, as proved in [12, Theorem 2.23] (equivalence between Item 1 and Item 2 in Proposition 2.4). As a consequence, so are the norms in the sequence $\{|||\cdot|||^{\mathrm{tn}}_{(j)}\}_{j\in[\![1,d]\!]}$ by Item 3 in Proposition 3.5, and we get that $|||x_J|||^{\mathrm{tn}}_{(k-1)} \leq |||x_K|||^{\mathrm{tn}}_{(k-1)}$, in particular, by the equivalence between Item 1 and Item 7 in Proposition 2.4.

On the other hand, since, by assumption, the sequence $\{|||\cdot|||^{\mathrm{tn}}_{(j)}\}_{j\in[\![1,d]\!]}$ of generalized top-k norms is strictly increasingly graded with respect to the ℓ_0 pseudonorm, we have by (35b) that, on the one hand,

$$|||x_J|||^{\mathrm{tn}}_{(1)} \leq \cdots \leq |||x_J|||^{\mathrm{tn}}_{(j-1)} < |||x_J|||^{\mathrm{tn}}_{(j)} = \cdots = |||x_J|||^{\mathrm{tn}}_{(d)} = |||x_J|||,$$

because $j = \ell_0(x_J)$, and, on the other hand,

$$|||x|||^{\mathrm{tn}}_{(1)} \leq \cdots \leq |||x_K|||^{\mathrm{tn}}_{(k-1)} < |||x_K|||^{\mathrm{tn}}_{(k)} = \cdots = |||x_K|||^{\mathrm{tn}}_{(d)} = |||x_K|||,$$

because $k = \ell_0(x_K)$. Since $j < k$, we deduce that

$$|||x_J||| = |||x_J|||^{\mathrm{tn}}_{(j)} = |||x_J|||^{\mathrm{tn}}_{(k-1)} \leq |||x_K|||^{\mathrm{tn}}_{(k-1)} < |||x_K|||^{\mathrm{tn}}_{(k)} = |||x_K|||\ ,$$

and therefore that $|||x_J||| < |||x_K|||$. Thus, Item 3 in Proposition 2.6 holds true for $|||\cdot|||$, so that the norm $|||\cdot|||$ is orthant-strictly monotonic. Hence, we have shown that Item 2 is satisfied.

• Suppose that Item 2 is satisfied and let us show that Item 1 holds true. Since the norm $|||\cdot|||$ is orthant-strictly monotonic, it has been proved in Proposition 4.5 that the sequence $\{|||\cdot|||^{\mathrm{tn}}_{(j)}\}_{j\in[\![1,d]\!]}$ is strictly increasingly graded with respect to the ℓ_0 pseudonorm. Hence, Item 1 holds true.

This ends the proof. □

4.2.3. Sufficient conditions for decreasingly graded sequence of generalized k-support norms

There is an asymetry in that the property of orthant-strict monotonicity for norms does not proves especially relevant for the ℓ_0 pseudonorm and sequences of k-support norms. Indeed, consider the source norm $|||\cdot||| = \|\cdot\|_1$, that is, the ℓ_1 norm which is orthant-strict monotonic. By Table 3.1 (third column), we know that the k-support norms are the norms $|||\cdot|||^{\star\mathrm{sn}}_{(k)} = \|\cdot\|^{\mathrm{sn}}_{\infty,k} = \max\{\|\cdot\|_1/k, \|\cdot\|_\infty\}$, for $k \in [\![1, d]\!]$. Now, the nonincreasing sequence $\{|||\cdot|||^{\star\mathrm{sn}}_{(j)}\}_{j\in[\![1,d]\!]}$ of norms is not strictly decreasingly graded with respect to the ℓ_0 pseudonorm when $d \geq 2$. Indeed, for any $\varepsilon \in]0, 1[$, the vector $y = \big(\varepsilon/(d-1), \ldots, \varepsilon/(d-1), 1\big)$ satisfies

$$\ell_0(y) = d \ \text{ and } \ |||y|||^{\star\mathrm{sn}}_{(1)} > |||y|||^{\star\mathrm{sn}}_{(2)} = \cdots = |||y|||^{\star\mathrm{sn}}_{(d)}$$

because $|||y|||^{\star\mathrm{sn}}_{(k)} = \max\{\|y\|_1/k, \|y\|_\infty\} = \max\{(\varepsilon+1)/k, 1\}$, for $k \in [\![1,d]\!]$, so that $\varepsilon + 1 = |||y|||^{\star\mathrm{sn}}_{(1)} > |||y|||^{\star\mathrm{sn}}_{(2)} = \cdots = |||y|||^{\star\mathrm{sn}}_{(d)} = 1$. However, we establish the following result.

Proposition 4.7.

(1) *If the source norm $|||\cdot|||$ is orthant-monotonic, then the nonincreasing sequence $\{|||\cdot|||^{\star\mathrm{sn}}_{(j)}\}_{j\in[\![1,d]\!]}$ of generalized k-support norms in* (15) *is decreasingly graded with respect to the ℓ_0 pseudonorm, that is,*

$$\ell_0(y) \leq l \Rightarrow |||y|||^{\star\mathrm{sn}}_{(l)} = |||y|||^{\star\mathrm{sn}}_{(d)} \ , \ \ \forall y \in \mathbb{R}^d \ , \ \ \forall l \in [\![0,d]\!] \ . \tag{37}$$

(2) *If the source norm $|||\cdot|||$ is orthant-monotonic, and if the normed space $(\mathbb{R}^d, |||\cdot|||_\star)$ is strictly convex, then the nonincreasing sequence $\{|||\cdot|||^{\star\mathrm{sn}}_{(j)}\}_{j\in[\![1,d]\!]}$ of generalized k-support norms in* (15) *is strictly decreasingly graded with respect to the ℓ_0 pseudonorm, that is,*

$$\ell_0(y) \leq l \iff |||y|||^{\star\mathrm{sn}}_{(l)} = |||y|||^{\star\mathrm{sn}}_{(d)} \ , \ \ \forall y \in \mathbb{R}^d \ , \ \ \forall l \in [\![0,d]\!] \ . \tag{38}$$

Proof. A direct proof would use [9, Proposition 6] with $|||\cdot|||_\star$ as source norm, and the property that $|||\cdot|||^{\star\mathrm{sn}}_{(j)}$ coincides with the coordinate-k norm [9, Definition 3] induced by $|||\cdot|||_\star$ when the norm $|||\cdot|||$ is orthant-monotonic. We give a self-contained proof for the sake of completeness.

- We suppose that the source norm $|||\cdot|||$ is orthant-monotonic.

For any $y \in \mathbb{R}^d$ and for any $k \in [\![1,d]\!]$, we have [8]

$$\begin{aligned}
y \in \ell_0^{\leq k} &\Leftrightarrow y = 0 \text{ or } \frac{y}{|||y|||_\star} \in \ell_0^{\leq k} \\
&\qquad \text{(by 0-homogeneity (30) of the } \ell_0 \text{ pseudonorm, and by definition (31) of } \ell_0^{\leq k}) \\
&\Leftrightarrow y = 0 \text{ or } \frac{y}{|||y|||_\star} \in \ell_0^{\leq k} \cap \mathbb{S}_\star \qquad \text{(as } \tfrac{y}{|||y|||_\star} \in \mathbb{S}_\star) \\
&\Leftrightarrow y = 0 \text{ or } \frac{y}{|||y|||_\star} \in \bigcup_{|K|\leq k} (\mathcal{R}_K \cap \mathbb{S}_\star) \qquad \text{(as } \ell_0^{\leq k} = \textstyle\bigcup_{|K|\leq k} \mathcal{R}_K \text{ by (32))} \\
&\Rightarrow y = 0 \text{ or } \frac{y}{|||y|||_\star} \in \overline{\mathrm{co}}\Big(\bigcup_{|K|\leq k} (\mathcal{R}_K \cap \mathbb{S}_\star)\Big) \\
&\qquad \text{(as } S \subset \overline{\mathrm{co}}(S) \text{ for any subset } S \text{ of } \mathbb{R}^d) \\
&\Rightarrow y = 0 \text{ or } \frac{y}{|||y|||_\star} \in \mathbb{B}^{\star\mathrm{sn}}_{(k)} \qquad \text{(as } \overline{\mathrm{co}}\big(\textstyle\bigcup_{|K|\leq k}(\mathcal{R}_K \cap \mathbb{S}_\star)\big) = \mathbb{B}^{\star\mathrm{sn}}_{(k)} \text{ by (27)} \\
&\qquad \text{because the source norm } |||\cdot||| \text{ is orthant-monotonic)} \\
&\Rightarrow y = 0 \text{ or } \Big|\Big|\Big|\frac{y}{|||y|||_\star}\Big|\Big|\Big|^{\star\mathrm{sn}}_{(k)} \leq 1 \qquad \text{(by definition (21) of the unit ball } \mathbb{B}^{\star\mathrm{sn}}_{(k)}) \\
&\Rightarrow |||y|||^{\star\mathrm{sn}}_{(k)} \leq |||y|||_\star = |||y|||^{\star\mathrm{sn}}_{(d)} \quad \text{(where the last equality comes from Item 2} \\
&\qquad \text{in Proposition 3.5 since the norm } |||\cdot||| \text{ is orthant-monotonic)}
\end{aligned}$$

[8] In what follows, by "or", we mean the so-called *exclusive or* (exclusive disjunction). Thus, every "or" should be understood as "or $y \neq 0$ and".

$$\Rightarrow |||y|||_{(k)}^{\star\mathrm{sn}} = |||y|||_{(d)}^{\star\mathrm{sn}} \,. \qquad (\text{as } |||y|||_{(k)}^{\star\mathrm{sn}} \geq |||y|||_{(d)}^{\star\mathrm{sn}} \text{ by (24)})$$

Therefore, we have obtained (37). As the sequence $\big\{|||\cdot|||_{(j)}^{\star\mathrm{sn}}\big\}_{j\in[\![1,d]\!]}$ of generalized k-support norms is nonincreasing by (24), we conclude that it is decreasingly graded with respect to the ℓ_0 pseudonorm (see the comments after Definition 4.4).

• We suppose that the source norm $|||\cdot|||$ is orthant-monotonic and that the normed space $\big(\mathbb{R}^d, |||\cdot|||_\star\big)$ is strictly convex.

For any $y \in \mathbb{R}^d$ and for any $k \in [\![1,d]\!]$, we have[9]

$$y \in \ell_0^{\leq k} \Leftrightarrow y = 0 \text{ or } \frac{y}{|||y|||_\star} \in \ell_0^{\leq k}$$

$$\text{(by 0-homogeneity (30) of the } \ell_0 \text{ pseudonorm, and by definition (31) of } \ell_0^{\leq k})$$

$$\Leftrightarrow y = 0 \text{ or } \frac{y}{|||y|||_\star} \in \ell_0^{\leq k} \cap \mathbb{S}_\star \qquad (\text{as } \tfrac{y}{|||y|||_\star} \in \mathbb{S}_\star)$$

$$\Leftrightarrow y = 0 \text{ or } \frac{y}{|||y|||_\star} \in \mathbb{B}_{(k)}^{\star\mathrm{sn}} \cap \mathbb{S}_\star$$

(by (34) since the assumptions of Proposition 4.1 – namely, the source norm $|||\cdot|||$ is orthant-monotonic and the normed space $\big(\mathbb{R}^d, |||\cdot|||_\star\big)$ is strictly convex – are satisfied)

$$\Leftrightarrow y = 0 \text{ or } \frac{y}{|||y|||_\star} \in \mathbb{B}_{(k)}^{\star\mathrm{sn}} \qquad (\text{as } \tfrac{y}{|||y|||_\star} \in \mathbb{S}_\star)$$

$$\Leftrightarrow y = 0 \text{ or } \Big|\Big|\Big|\frac{y}{|||y|||_\star}\Big|\Big|\Big|_{(k)}^{\star\mathrm{sn}} \leq 1 \qquad (\text{by definition (21) of the unit ball } \mathbb{B}_{(k)}^{\star\mathrm{sn}})$$

$$\Leftrightarrow |||y|||_{(k)}^{\star\mathrm{sn}} \leq |||y|||_\star = |||y|||_{(d)}^{\star\mathrm{sn}} \quad \text{(where the last equality comes from Item 2 in Proposition 3.5 since the norm } |||\cdot||| \text{ is orthant-monotonic)}$$

$$\Leftrightarrow |||y|||_{(k)}^{\star\mathrm{sn}} = |||y|||_{(d)}^{\star\mathrm{sn}} \,. \qquad (\text{as } |||y|||_{(k)}^{\star\mathrm{sn}} \geq |||y|||_{(d)}^{\star\mathrm{sn}} \text{ by (24)})$$

Therefore, we have obtained (38). As the sequence $\big\{|||\cdot|||_{(j)}^{\star\mathrm{sn}}\big\}_{j\in[\![1,d]\!]}$ of generalized k-support norms is nonincreasing by (24), we conclude that it is strictly decreasingly graded with respect to the ℓ_0 pseudonorm (see the comments after Definition 4.4).

This ends the proof. □

4.2.4. Expressing the ℓ_0 pseudonorm by means of the difference between two norms

Propositions 4.5 and 4.7 open the way for so-called "difference of convex" (DC) optimization methods [11] to achieve sparsity.

Indeed, if the source norm $|||\cdot|||$ is orthant-strictly monotonic, the level sets of the ℓ_0 pseudonorm in (31) can be expressed by means of the difference between two norms (one being a generalized top-k norm), as follows,

$$\ell_0^{\leq k} = \{x \in \mathbb{R}^d \mid |||x||| = |||x|||_{(k)}^{\mathrm{tn}}\} = \{x \in \mathbb{R}^d \mid |||x||| \leq |||x|||_{(k)}^{\mathrm{tn}}\} \,, \ \ \forall k \in [\![0,d]\!] \,, \tag{39a}$$

[9] See Footnote 8.

and the ℓ_0 pseudonorm has the expression (see (35f))

$$\ell_0(x) = \min \left\{ k \in [\![1,d]\!] \mid |||x|||^{\mathrm{tn}}_{(k)} = |||x||| \right\} \ , \ \ \forall x \in \mathbb{R}^d \ . \tag{39b}$$

As the ℓ_p-norm $\|\cdot\|_p$ and its dual norm are orthant-strictly monotonic for $p \in]1,\infty[$, the formulas above hold true with the top-(p,k) norm $|||\cdot|||^{\mathrm{tn}}_{(k)} = \|\cdot\|^{\mathrm{tn}}_{p,k}$ (see second column of Table 3.1).

If the source norm $|||\cdot|||$ is orthant-monotonic and the normed space $\left(\mathbb{R}^d, |||\cdot|||_\star\right)$ is strictly convex, the level sets of the ℓ_0 pseudonorm in (31) can be expressed by means of the difference between two norms (one being a generalized k-support norm), as follows,

$$\ell_0^{\leq k} = \{y \in \mathbb{R}^d \mid |||y|||^{\star\mathrm{sn}}_{(k)} = |||y|||_\star\} = \{y \in \mathbb{R}^d \mid |||y|||^{\star\mathrm{sn}}_{(k)} \leq |||y|||_\star\} \ , \ \ \forall k \in [\![0,d]\!] \ , \tag{40a}$$

and the ℓ_0 pseudonorm has the expression (see (35f))

$$\ell_0(y) = \min \left\{ k \in [\![1,d]\!] \mid |||y|||^{\star\mathrm{sn}}_{(k)} = |||y|||_\star \right\} \ , \ \ \forall y \in \mathbb{R}^d \ . \tag{40b}$$

As the ℓ_p-norm $\|\cdot\|_p$ is orthant-monotonic and the normed space $\left(\mathbb{R}^d, \|\cdot\|_q\right)$ is strictly convex, when $p \in]1,\infty[$ and $1/p + 1/q = 1$, the formulas above hold true with the (q,k)-support norm $|||\cdot|||^{\star\mathrm{sn}}_{(k)} = \|y\|^{\mathrm{sn}}_{q,k}$ for $q \in]1,\infty[$ (see Table 3.1).

5. Conclusion

In sparse optimization problems, one looks for solution that have few nonzero components, that is, sparsity is exactly measured by the ℓ_0 pseudonorm. However, the mathematical expression of the ℓ_0 pseudonorm, taking integer values, makes it difficult to handle it in optimization problems. To overcome this difficulty, one can try to replace the embarrassing ℓ_0 pseudonorm by nicer terms, like norms. In this paper, we contribute to this program by bringing up three new concepts for norms, and show how they prove especially relevant for the ℓ_0 pseudonorm.

First, we have introduced a new class of orthant-strictly monotonic norms, inspired from orthant-monotonic norms. With such a norm, when one component of a vector moves away from zero, the norm of the vector strictly grows. Thus, an orthant-strictly monotonic norm is sensitive to the support of a vector, like the ℓ_0 pseudonorm. We have provided different characterizations of orthant-strictly monotonic norms (and added a new characterization of orthant-monotonic norms). Second, we have extended already known concepts of top-k and k-support norms to sequences of generalized top-k and k-support norms, generated from any source norm (and not only from the ℓ_p norms), and have studied their properties. Third, we have introduced the notion of sequences of norms that are strictly increasingly graded with respect to the ℓ_0 pseudonorm. A graded sequence detects the number of nonzero components of a vector when the sequence becomes stationary.

With these three notions, we have proved that, when the source norm is orthant-strictly monotonic, the sequence of induced generalized top-k norms is strictly increasingly graded. We have also shown that, when the source norm is orthant-monotonic and that the normed space $\mathbb{R}^d$ is strictly convex when equipped with the dual norm, the sequence of induced generalized k-support norms is strictly decreasingly graded.

	$\left\{ \vert\vert\vert\cdot\vert\vert\vert^{\mathrm{tn}}_{(j)} \right\}_{j\in[\![1,d]\!]}$ increasingly			$\left\{ \vert\vert\vert\cdot\vert\vert\vert^{\star\mathrm{sn}}_{(j)} \right\}_{j\in[\![1,d]\!]}$ decreasingly			
	graded		strictly graded	graded		strictly graded	
$\vert\vert\vert\cdot\vert\vert\vert$ is orthant-monotonic	✓			✓		✓	
$\vert\vert\vert\cdot\vert\vert\vert$ is orthant-strictly monotonic			✓				
$\vert\vert\vert\cdot\vert\vert\vert_\star$ is orthant-monotonic		✓			✓		✓
$(\mathbb{R}^d, \vert\vert\vert\cdot\vert\vert\vert_\star)$ is strictly convex						✓	✓

Table 5.1: Table of results. It reads by columns as follows: to obtain that $\left\{ |||\cdot|||^{\mathrm{tn}}_{(j)} \right\}_{j\in[\![1,d]\!]}$ is increasingly strictly graded (column 4), it suffices that $|||\cdot|||$ be orthant-strictly monotonic (the only checkmark ✓in column 4); to obtain that $\left\{ |||\cdot|||^{\mathrm{tn}}_{(j)} \right\}_{j\in[\![1,d]\!]}$ is increasingly graded (columns 2 and 3), it suffices that either $|||\cdot|||$ be orthant-monotonic (the only checkmark ✓in column 2) or $|||\cdot|||_\star$ be orthant-monotonic (the only checkmark ✓in column 3); to obtain that $\left\{ |||\cdot|||^{\star\mathrm{sn}}_{(j)} \right\}_{j\in[\![1,d]\!]}$ is decreasingly strictly graded (columns 7 and 8), it suffices either that $|||\cdot|||$ be orthant-monotonic and that $(\mathbb{R}^d, |||\cdot|||_\star)$ be strictly convex (two checkmarks ✓in column 7) or that $|||\cdot|||_\star$ be orthant-monotonic and that $(\mathbb{R}^d, |||\cdot|||_\star)$ be strictly convex (two checkmarks ✓in column 8)

These results – summarized in Table 5.1 – open the way for so-called "difference of convex" (DC) optimization methods to achieve sparsity. Indeed, the level sets of the ℓ_0 pseudonorm can be expressed by means of the difference between norms, taken from an increasingly or decreasingly graded sequence of norms. And we provide a way to generate such sequences from a class of source norms that encompasses the ℓ_p norms (but for the extreme ones).

To complete the possible applications, we add that, in another paper [8], we show that, with orthant-strictly monotonic norms, we can define conjugacies for which the ℓ_0 pseudonorm is equal to its biconjugate.

Acknowledgements. We thank Jean-Baptiste Hiriart-Urruty for his comments on first versions of this work.

References

[1] C. D. Aliprantis, K. C. Border: *Infinite Dimensional Analysis*, 2nd ed., Springer, Berlin (1999).

[2] A. Argyriou, R. Foygel, N. Srebro: *Sparse prediction with the k-support norm*, in: Proc. 25th Int. Conf. on Neural Information Processing Systems, Vol. 1, Lake Tahoe, Curran Associates, Red Hook (2012) 1457–1465.

[3] F. L. Bauer, J. Stoer, C. Witzgall: *Absolute and monotonic norms*, Numer. Math. 3 (1961) 257–264.

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

[5] R. Bhatia: *Matrix Analysis*, Graduate Texts in Mathematics 169, Springer, New York (1997).

[6] G. Birkhoff: *Orthogonality in linear metric spaces*, Duke Math. J. 1/2 (1935) 169–172.

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

[8] J.-P. Chancelier, M. De Lara: *Capra-convexity, convex factorization and variational formulations for the l_0 pseudonorm*, Set-Valued Var. Analysis 30 (2022) 597–619.

[9] J.-P. Chancelier, M. De Lara: *Constant along primal rays conjugacies and the l_0 pseudonorm*, Optimization 71/2 (2022) 355–386.

[10] Z. Fan, H. Jeong, Y. Sun, M. P. Friedlander: *Atomic decomposition via polar alignment*, Found. Trends Optimization 3/4 (2020) 280–366.

[11] J.-y. Gotoh, A. Takeda, K. Tono: *Efficient DC algorithm for constrained sparse optimization*, Math. Program. 169/1 (2018) 141–176.

[12] D. Gries: *Characterization of certain classes of norms*, Numer. Math. 10 (1967) 30–41.

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

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

[15] E. Marques de Sà, M.-J. Sodupe: *Characterizations of *orthant-monotonic norms*, Linear Algebra Appl. 193 (1993) 1–9.

[16] A. M. McDonald, M. Pontil, D. Stamos: *New perspectives on k-support and cluster norms*, J. Machine Learning Res. 17/155 (2016) 1–38.

[17] L. Mirsky: *Symmetric gauge functions and unitarily invariant norms*, Quart. J. Math. 11/1 (1960) 50–59.

[18] G. Obozinski, F. Bach: *A unified perspective on convex structured sparsity: Hierarchical, symmetric, submodular norms and beyond*, preprint (2016).

[19] K. Paul, D. Sain, K. Jha: *On strong orthogonality and strictly convex normed linear spaces*, J. Inequal. Appl. (2013), art. no. 242, 7 pp.

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

[21] D. Sain, K. Paul, K. Jha: *Strictly convex space: Strong orthogonality and conjugate diameters*, J. Convex Analysis 22 (2015) 1215–1225.

[22] K. Tono, A. Takeda, J.-y. Gotoh: *Efficient DC algorithm for constrained sparse optimization*, preprint (2017), published in revised form as [11].