McMaster University

Advanced Optimization Laboratory

Title:

Geometry of sparsity-inducing norms

Authors:

Jean-Philippe Chancelier, Michel de Lara,
Antoine Deza, and Lionel Pournin

AdvOL-Report No. 2025/1

January 2025, Hamilton, Ontario, Canada

Geometry of Sparsity-Inducing Norms

Jean-Philippe Chancelier*, Michel De Lara*,
Antoine Deza†, Lionel Pournin‡

January 13, 2025

Abstract

Sparse optimization seeks an optimal solution with few nonzero entries. To achieve this, it is common to add to the criterion a penalty term proportional to the ℓ_1-norm, which is recognized as the archetype of sparsity-inducing norms. In this approach, the number of nonzero entries is not controlled a priori. By contrast, in this paper, we focus on finding an optimal solution with at most k nonzero coordinates (or for short, k-sparse vectors), where k is a given sparsity level (or "sparsity budget"). For this purpose, we study the class of generalized k-support norms that arise from a given source norm. When added as a penalty term, we provide conditions under which such generalized k-support norms promote k-sparse solutions. The result follows from an analysis of the exposed faces of closed convex sets generated by k-sparse vectors, and of how primal support identification can be deduced from dual information. Finally, we study some of the geometric properties of the unit balls for the k-support norms and their dual norms when the source norm belongs to the family of ℓ_p-norms.

Keywords: sparsity, ℓ_0 pseudonorm, orthant-monotonicity, top-k norm, k-support norm
2020 Mathematics Subject Classification (MSC2020): 49N15 90C25 52A05 52A21

1 Introduction

In 1996, Tibshirani [27] proposed least-square regression with an ℓ_1-norm penalty to achieve sparsity in least-square optimization. Figure 1 is the replica of [27, Figure 2], which provides insight regarding why corresponding optimal solutions are sparse (we copy the comments of [27, Figure 2] with additional precisions in brackets $[\cdots]$):

*CERMICS, École nationale des ponts et chaussées, IP Paris, France
†McMaster University, Hamilton, Ontario, Canada
‡Université Paris 13, Villetaneuse, France

"The elliptical contours of this function [quadratic criterion] are shown by the full curves in Fig. 2(a); they are centred at the OLS [optimal least-square] estimates; the constraint region [ℓ_1-ball in dimension 2] is the rotated square. The lasso solution is the first place that the contours touch the square, and this will sometimes occur at a corner, corresponding to a zero coefficient. The picture for ridge regression is shown in Fig. 2(b): there are no corners for the contours to hit and hence zero solutions will rarely result."

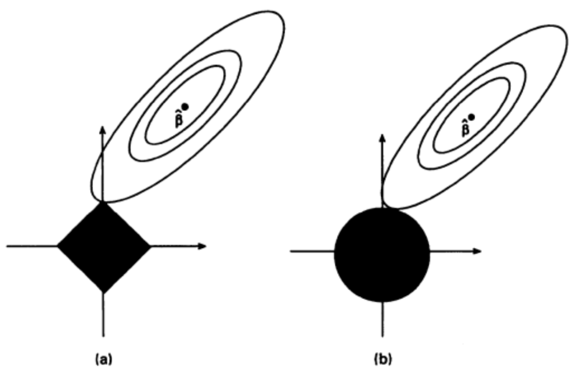

Figure 1: Replica of [27, Figure 2]

Thus, as the kinks of the ℓ_1-ball are located at sparse points, it is common to say that the ℓ_1-norm is *sparsity-inducing*. Figure 2 shows two examples of unit balls with kinks located at sparse points. Both of them arise from norms that are studied in the sequel.

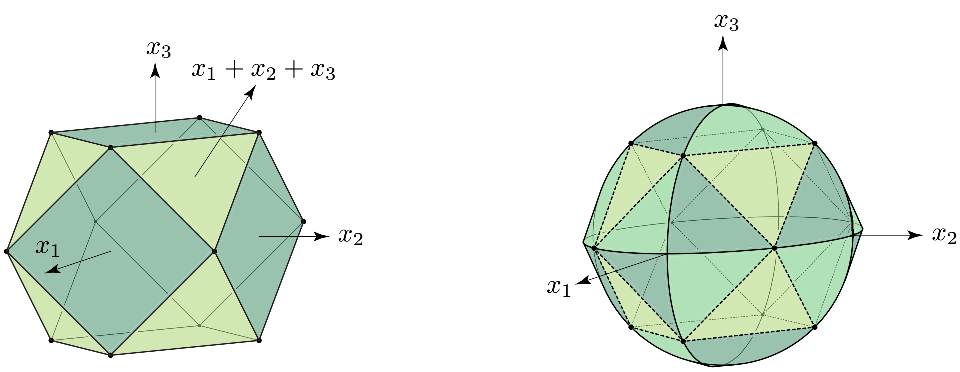

Figure 2: Two examples of unit balls with kinks located at sparse points

A natural question that arises from the comments of [27, Figure 2] is: What could be mathematical conditions for inducing sparsity?

Going beyond least-square regression with an ℓ_1-norm penalty, one can consider optimization problems of the form $\min_{x \in \mathbb{R}^d} \big(f(x) + \gamma \|x\| \big)$, where $f : \mathbb{R}^d \to \mathbb{R}$ is a smooth convex function, $\gamma > 0$ and $\|\cdot\|$ is a norm with unit ball B. This is the approach taken in the papers [2, 18], where the terminology "sparsity-inducing norm" has been introduced. As all functions take finite values — and as $\|\cdot\|$ is the support function $\sigma_{B^\odot}$ of the polar set $B^\odot$ — a solution $x^\sharp$ of the above problem is characterized by the Fermat rule

$$0 \in \nabla f(x^\sharp) + \gamma \partial \sigma_{B^\odot}(x^\sharp) \ ,$$

where $\partial \sigma_{B^\odot}(x^\sharp)$ is the face $F_\perp(B^\odot, x^\sharp)$ of $B^\odot$ exposed by $x^\sharp$. Thus, the optimality condition reads

$$-\nabla f(x^\sharp) \in \gamma F_\perp(B^\odot, x^\sharp)$$

and, by polarity [11, Theorem 5.1], this is equivalent to

$$\frac{x^\sharp}{\|x^\sharp\|} \in F_\perp(B, -\nabla f(x^\sharp)) \ .$$

One could say that the considered norm is *sparsity-inducing* if information about the support of $x^\sharp$ could be obtained from information about the exposed faces $F_\perp(B, -\nabla f(x^\sharp))$ of the unit ball B for that norm. This is the approach that we consider in this paper. More precisely, we analyze the exposed faces of some special convex sets, and in particular of the unit balls of certain norms, and relate them to sparsity.

Our work is related to different trends in the literature. As said above, the terminology "sparsity-inducing norm" has been introduced in the papers [2, 18], which focus on algorithmic issues, whereas we focus on geometric aspects. In particular, we study how the gradient — at a solution, of the original smooth function to be minimized — provides relevant (dual) information about the sparsity of the (primal) solution. Sparsity is also examined for the solutions of undetermined linear systems, and we emphasize the three papers [5, 8, 4]. The paper [5] studies the solutions of an undetermined linear system in the context of compressed sensing; it provides a sufficient property of the sensing matrix, the restricted isometry property, which ensures that the minimal ℓ_1-norm solution coincides with the sparse solution. In [8], it is explained how to design so-called "atomic norms" which promote sparsity, but with respect to a given (compact) atomic set. [8] stresses that it is "the favorable facial structure of the atomic norm ball that makes the atomic norm a suitable convex heuristic to recover simple models" The norms that we present in Sect. 3 are atomic norms where the atomic set is especially designed to provide solutions with an a priori given "sparsity budget" (number of nonzero entries bounded above by a given integer); we focus on the geometric description of the facial structure of the unit balls of these special atomic norms.. As said above, we study in particular how a gradient provides relevant (dual) information about sparsity of the (primal) solution. By contrast, [8] focuses more on measuring the Gaussian width of the tangent cones as a way to to achieve more or less sparsity. The paper [4] also stresses the role of faces in identifying solutions of undetermined linear systems that can be expressed as convex combinations of a small number of atoms. Where [4] studies faces, we focus on *exposed* faces and on how dual information is related to sparsity of a primal solution. Let us mention three more works. The question of decomposing a vector as a convex conical combination of elementary atoms has been studied in [11], with a special role given to the so-called alignment, that is, to normal cones and exposed faces. Our approach intersects that of [11], but with a focus on the classic sparsity along coordinate axis and with the goal to describe the geometry of some unit balls [7]. In [10], the focus is put on stratifying the primal space. Then, an optimal primal-dual pair $(x^\sharp, \nabla f(x^\sharp))$ can provide information on the strata to which $x^\sharp$ belongs. General regularizers (and not only norms) are studied. Convex structured sparsity with norms is the object of [22], but there is no focus on how normal cones and exposed faces relate to sparsity.

To summarize, to the difference with the literature mentioned above, we focus on sparsity-inducing norms that promote solutions within an a priori "sparsity budget" by using dual information.

The paper is organized as follows. In Sect. 2, after providing background on sparsity and on faces of closed convex sets, we state our main result: we characterize the faces of closed convex sets whose extreme points are k-sparse, from which we deduce support identification. In Sect. 3, we characterize the faces of the unit ball of k-support norms, from which we deduce support identification. Then, we provide dual conditions under which the primal optimal solution of a minimization problem, penalized by a k-support norm, is k-sparse. In the cases of orthant-monotonic and orthant-strictly monotonic source norms, we obtain a characterization of the intersection of the k-sparse vectors with the faces of the k-support norm. Sect. 4 deals with the geometric aspects of the face and cone lattices of the unit balls of top-(q,k) norm and (p,k)-support norms, that is, with the ℓ_p as source norms.

2 Face characterization and support identification

In §2.1, we provide background on sparsity and on faces of closed convex sets. In §2.2, we state our main result: we characterize the faces of closed convex sets whose extreme points are k-sparse, from which we deduce support identification. Proofs are given in §2.3.

2.1 Background on sparsity and on faces of closed convex sets

We consider the finite-dimensional real Euclidean vector space $\mathbb{R}^d$ equipped with the scalar product $\langle \mid \rangle$.

Background on sparsity

We use the notation $[\![j, k]\!] = \{j, j+1, \ldots, k-1, k\}$ for any pair of integers such that $j \leq k$. Let d be a natural number. Denoting by $|K|$ the cardinality of a subset $K \subset [\![1, d]\!]$, we define, for any vector x in $\mathbb{R}^d$, the *support* of x by

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

and the ℓ_0 *pseudonorm* of x by the number of nonzero components, that is, by

$$\ell_0(x) = |\mathrm{supp}(x)| \ . \tag{1b}$$

This defines the ℓ_0 *pseudonorm* function $\ell_0 : \mathbb{R}^d \to [\![0, d]\!]$. For any k in $[\![1, d]\!]$, we denote its level sets, made of all the vectors with at most k nonzero coordinates by

$$\ell_0^{\leq k} = \{x \in \mathbb{R}^d : \ell_0(x) \leq k\} \tag{2a}$$

The vectors in $\ell_0^{\leq k}$ will be called *k-sparse vectors*. For any subset K of $[\![1, d]\!]$, we introduce the subspace $\mathcal{R}_K$ of $\mathbb{R}^d$ made of vectors whose components vanish outside of K as

$$\mathcal{R}_K = \left\{x \in \mathbb{R}^d : x_j = 0 \ , \ \forall j \notin K\right\} \tag{2b}$$

with the convention that $\mathcal{R}_\emptyset = \{0\}$. Using $\bigcup_{|K| \leq k}$ as a shorthand for $\bigcup_{K \subset [\![1,d]\!], |K| \leq k}$, we get

$$\ell_0^{\leq k} = \bigcup_{|K| \leq k} \mathcal{R}_K \ . \tag{2c}$$

We denote by $\pi_K : \mathbb{R}^d \to \mathcal{R}_K$ the *orthogonal projection mapping*; for any vector x in $\mathbb{R}^d$, the coordinates of the vector $\pi_K x \in \mathcal{R}_K$ coincide with those of x, except for the ones whose indices range outside of K that are equal to zero. It is easily seen that the orthogonal projection mapping π_K is *self-adjoint* (or *self-dual*), that is,

$$\langle \pi_K x \mid y \rangle = \langle x \mid \pi_K y \rangle = \langle \pi_K x \mid \pi_K y \rangle \ , \ \forall x \in \mathbb{R}^d \ , \ \forall y \in \mathbb{R}^d \ . \tag{3}$$

Background on faces of closed convex sets

For any subset $X \subset \mathbb{R}^d$, the expression

$$\sigma_X(y) = \sup_{x \in X} \langle x \mid y \rangle \ , \ \forall y \in \mathbb{R}^d \tag{4}$$

defines a map $\sigma_X : \mathbb{R}^d \to \overline{\mathbb{R}}$ called the *support function*[1] *of the subset* X. The (negative) *polar set* $X^\odot$ of the subset $X \subset \mathbb{R}^d$ is the closed convex set

$$X^\odot = \left\{y \in \mathbb{R}^d : \langle x, y \rangle \leq 1 \ , \ \forall x \in X\right\} = \{\sigma_X \leq 1\} \ . \tag{5}$$

The face of a nonempty closed convex subset C of $\mathbb{R}^d$ exposed by a dual vector y in $\mathbb{R}^d$ is

$$F_\perp(C, y) = \arg\max_{x \in C} \langle x \mid y \rangle \ , \tag{6}$$

and the normal cone $N(C, x)$ of C at a primal vector $x \in C$ is defined by the conjugacy relation

$$x \in C \text{ and } y \in N(C, x) \iff x \in F_\perp(C, y) \ . \tag{7}$$

[1]Note that the support *function* has nothing to do with the support of a *vector*.

2.2 Convex sets with k-sparse extreme points

As discussed in the introduction, the intuition behind [27, Figure 2] is that the unit ball of a sparsity-inducing norm should have extreme points (vertices) precisely at k-sparse vectors. One way to enforce this is to select a suitable subset of k-sparse vectors, and then take the convex closure. Theorem 1 characterizes the faces of this convex closure.

Theorem 1 (Characterization of faces) *Let $k \in [\![1, d]\!]$ be a natural number and $X \subset \mathbb{R}^d$ be a (primal) nonempty set. We set*

$$X_k = \bigcup_{|K| \leq k} \pi_K(X) \; . \tag{8}$$

Let $y \in \mathbb{R}^d$ be a (dual) vector. We set

$$\mathcal{K}^{\sharp}_{X,k}(y) = \mathop{\arg\max}_{\substack{K \subset [\![1,d]\!] \\ |K| \leq k}} \sigma_{\overline{\mathrm{co}}X}(\pi_K y) \; , \tag{9}$$

which is such that $\emptyset \neq \mathcal{K}^{\sharp}_{X,k}(y) \subset \big\{ K \subset [\![1, d]\!], |K| \leq k \big\} \subset 2^{[\![1,d]\!]}$. Then, we have that the set X_k is made of k-sparse vectors, that is,

$$X_k \subset \ell_0^{\leq k} \; , \tag{10}$$

and the faces of $\overline{\mathrm{co}}X_k$ are related to the faces of $\overline{\mathrm{co}}X$ by

$$X_k \cap F_{\perp}\big(\overline{\mathrm{co}}X_k, y\big) = \Big\{ \pi_{K^\sharp}\big(X \cap F_{\perp}(\overline{\mathrm{co}}X, \pi_{K^\sharp} y)\big) \,:\, K^\sharp \in \mathcal{K}^{\sharp}_{X,k}(y) \Big\} \; , \tag{11}$$

and by

$$F_{\perp}\big(\overline{\mathrm{co}}X_k, y\big) = \overline{\mathrm{co}} \Big\{ \pi_{K^\sharp}\big(X \cap F_{\perp}(\overline{\mathrm{co}}X, \pi_{K^\sharp} y)\big) \,:\, K^\sharp \in \mathcal{K}^{\sharp}_{X,k}(y) \Big\} \; . \tag{12}$$

By construction, the extreme points of $\overline{\mathrm{co}}X_k$ are contained in X_k, while X_k itself is a subset of $\bigcup_{|K| \leq k} \mathcal{R}_K = \ell_0^{\leq k}$ (see Equation (10)), hence is made of k-sparse vectors.

Corollary 2 (Support identification) *Under the assumptions of Theorem 1,*

$$x \in X_k \cap F_{\perp}(\overline{\mathrm{co}}X_k, y) \implies \mathrm{supp}(x) \in \mathcal{K}^{\sharp}_{X,k}(y) \; , \tag{13}$$

$$x \in F_{\perp}(\overline{\mathrm{co}}X_k, y) \implies \mathrm{supp}(x) \subset \bigcup_{K^\sharp \in \mathcal{K}^{\sharp}_{X,k}(y)} K^\sharp \; . \tag{14}$$

2.3 Proofs of Theorem 1 and Corollary 2

We begin the section by proving the following preparatory lemma.

Lemma 3 *Consider a (primal) nonempty subset X of $\mathbb{R}^d$. For any (dual) vector y contained in $\mathbb{R}^d$ and any subset K of $[\![1,d]\!]$, we have that*

$$\mathop{\arg\max}_{x\in\pi_K(X)} \langle x \mid y\rangle = \pi_K\Big(\mathop{\arg\max}_{z\in X} \langle z \mid \pi_K y\rangle\Big) \tag{15}$$

and that

$$\max_{x\in\pi_K(X)} \langle x \mid y\rangle = \sigma_{\overline{\mathrm{co}}X}(\pi_K y)\ . \tag{16}$$

Proof. We prove (15) by establishing two opposite inclusions. On the one hand, the inclusion

$$\mathop{\arg\max}_{x\in\pi_K(X)} \langle x \mid y\rangle \subset \pi_K\Big(\mathop{\arg\max}_{z\in X} \langle z \mid \pi_K y\rangle\Big)$$

holds true as a consequence of the following sequence of equivalences and implications:

$$\begin{aligned} x^\sharp \in \mathop{\arg\max}_{x\in\pi_K(X)} \langle x \mid y\rangle &\iff x^\sharp \in \pi_K(X) \text{ and } \langle x^\sharp \mid y\rangle \geq \langle x \mid y\rangle\ ,\ \ \forall x \in \pi_K(X) \\ &\iff x^\sharp \in \pi_K(X) \text{ and } \langle x^\sharp \mid y\rangle \geq \langle \pi_K z \mid y\rangle\ ,\ \ \forall z \in X \\ &\qquad\qquad \text{(by definition of } \pi_K(X)) \\ &\implies x^\sharp \in \pi_K(X) \text{ and } \langle \pi_K x^\sharp \mid y\rangle \geq \langle \pi_K z \mid y\rangle\ ,\ \ \forall z \in X \end{aligned}$$

because $x^\sharp \in \pi_K(X)$ belongs to the image of the orthogonal projection π_K, hence $x^\sharp = \pi_K x^\sharp$,

$$\begin{aligned} &\implies x^\sharp \in \pi_K(X) \text{ and } \langle x^\sharp \mid \pi_K y\rangle \geq \langle z \mid \pi_K y\rangle\ ,\ \ \forall z \in X \\ &\qquad\qquad \text{(as the orthogonal projection } \pi_K \text{ is self-adjoint, see (3))} \\ &\implies x^\sharp \in \pi_K(X) \text{ and } x^\sharp \in \mathop{\arg\max}_{z\in X} \langle z \mid \pi_K y\rangle \\ &\implies x^\sharp \in \pi_K(X) \text{ and } \pi_K x^\sharp \in \pi_K\Big(\mathop{\arg\max}_{z\in X} \langle z \mid \pi_K y\rangle\Big) \\ &\implies x^\sharp \in \pi_K(X) \text{ and } x^\sharp \in \pi_K\Big(\mathop{\arg\max}_{z\in X} \langle z \mid \pi_K y\rangle\Big) \qquad \text{(as } x^\sharp = \pi_K x^\sharp.) \end{aligned}$$

On the other hand, the inclusion

$$\pi_K\Big(\mathop{\arg\max}_{z\in X} \langle z \mid \pi_K y\rangle\Big) \subset \mathop{\arg\max}_{x\in\pi_K(X)} \langle x \mid y\rangle$$

holds true as a consequence of the following equivalences and implications:

$$\begin{aligned} x^\sharp \in \pi_K\Big(\mathop{\arg\max}_{z\in X} \langle z \mid \pi_K y\rangle\Big) &\iff \exists z^\sharp \in \mathop{\arg\max}_{z\in X} \langle z \mid \pi_K y\rangle \text{ and } x^\sharp = \pi_K z^\sharp \\ &\iff x^\sharp = \pi_K z^\sharp \text{ where } z^\sharp \in X \text{ and } \langle z^\sharp \mid \pi_K y\rangle \geq \langle z \mid \pi_K y\rangle\ ,\ \ \forall z \in X \\ &\iff x^\sharp = \pi_K z^\sharp \text{ where } z^\sharp \in X \text{ and } \langle \pi_K z^\sharp \mid y\rangle \geq \langle \pi_K z \mid y\rangle\ ,\ \ \forall z \in X \\ &\qquad\qquad \text{(as the orthogonal projection } \pi_K \text{ is self-adjoint, see (3))} \\ &\implies x^\sharp \in \pi_K(X) \text{ and } \langle x^\sharp \mid y\rangle \geq \langle x \mid y\rangle\ ,\ \ \forall x \in \pi_K(X) \\ &\implies x^\sharp = \pi_K x^\sharp \in \pi_K(X) \text{ and } \langle x^\sharp \mid y\rangle \geq \langle x \mid y\rangle\ ,\ \ \forall x \in \pi_K(X) \end{aligned}$$

because $x^\sharp \in \pi_K(X)$ belongs to the image of the orthogonal projection π_K, hence $x^\sharp = \pi_K x^\sharp$,

$$\Longrightarrow x^\sharp \in \mathop{\arg\max}_{x \in \pi_K(X)} \langle x \mid y \rangle \qquad \text{(by definition of } \arg\max_{x \in \pi_K(X)} \langle x \mid y \rangle\text{.)}$$

Finally (16) holds true because, as the orthogonal projection π_K is self-adjoint (see (3)),

$$\max_{x \in \pi_K(X)} \langle x \mid y \rangle = \max_{z \in X} \langle \pi_K z \mid y \rangle = \max_{z \in X} \langle z \mid \pi_K y \rangle = \sigma_{\overline{\mathrm{co}} X}(\pi_K y) \ ,$$

by definition (4) and the well-known property $\sigma_X = \sigma_{\overline{\mathrm{co}} X}$ of the support function σ_X. $\square$

2.3.1 Proof of Theorem 1

Proof. Consider a number k in $[\![1, d]\!]$, a (primal) nonempty subset X of $\mathbb{R}^d$, and a (dual) vector y in $\mathbb{R}^d$. First observe that

$$\emptyset \neq \mathcal{K}^\sharp_{X,k}(y) \subset \left\{ K \subset [\![1, d]\!], |K| \leq k \right\} \subset 2^{[\![1,d]\!]}.$$

Second, Equation (10) follows from the definition (8) of X_k as by (2c),

$$X_k = \bigcup_{|K| \leq k} \pi_K(X) \subset \bigcup_{|K| \leq k} \mathcal{R}_K = \ell_0^{\leq k}$$

Third, we prove (11). We have that

$$x^\sharp \in X_k \cap F_\perp(\overline{\mathrm{co}} X_k, y) \iff x^\sharp \in \mathop{\arg\max}_{x \in X_k} \langle x \mid y \rangle$$

as $F_\perp(\overline{\mathrm{co}} X_k, y) = \arg\max_{x \in \overline{\mathrm{co}} X_k} \langle x \mid y \rangle$ and $\max_{x \in X_k} \langle x \mid y \rangle = \max_{x \in \overline{\mathrm{co}} X_k} \langle x \mid y \rangle$,

$$\iff x^\sharp \in X_k \text{ and } \langle x^\sharp \mid y \rangle = \max_{x \in \bigcup_{|K| \leq k} \pi_K(X)} \langle x \mid y \rangle$$

using the definition (8) of X_k and where, in all this proof, the subscript $|K| \leq k$ has to be understood as $K \subset [\![1, d]\!], |K| \leq k$,

$$\begin{aligned} &\iff x^\sharp \in X_k \text{ and } \langle x^\sharp \mid y \rangle = \max_{|K| \leq k} \max_{x \in \pi_K(X)} \langle x \mid y \rangle \\ &\iff x^\sharp \in X_k \text{ and there exists } K^\sharp \subset [\![1, d]\!] \ , \ |K^\sharp| \leq k \text{ such that} \\ &\qquad x^\sharp \in \pi_{K^\sharp}(X) \text{ and } \langle x^\sharp \mid y \rangle = \max_{|K| \leq k} \max_{x \in \pi_K(X)} \langle x \mid y \rangle \end{aligned}$$

because, by definition (8) of X_k, $x^\sharp \in \bigcup_{|K|\leq k} \pi_K(X)$, hence there exists $K^\sharp \subset [\![1,d]\!]$ with $|K^\sharp| \leq k$ such that $x^\sharp \in \pi_{K^\sharp}(X)$,

$$\begin{aligned}
\iff\ & x^\sharp \in X_k \text{ and there exists } K^\sharp \subset [\![1,d]\!]\ ,\ \ |K^\sharp| \leq k \text{ such that} \\
& x^\sharp \in \pi_{K^\sharp}(X) \text{ and } \langle x^\sharp \mid y\rangle = \max_{|K|\leq k} \max_{x\in\pi_K(X)} \langle x \mid y\rangle \\
& \ \text{and } K^\sharp \in \arg\max_{|K|\leq k} \max_{x\in\pi_K(X)} \langle x \mid y\rangle \\
& \hfill \text{(by definition of } \arg\max_{|K|\leq k} \max_{x\in\pi_K(X)}\langle x \mid y\rangle) \\
\iff\ & x^\sharp \in X_k \text{ and there exists } K^\sharp \in \arg\max_{|K|\leq k} \max_{x\in\pi_K(X)} \langle x \mid y\rangle \text{ such that} \\
& x^\sharp \in \pi_{K^\sharp}(X) \text{ and } \langle x^\sharp \mid y\rangle = \max_{x\in\pi_{K^\sharp}(X)} \langle x \mid y\rangle
\end{aligned}$$

because $K^\sharp \in \arg\max_{|K|\leq k} \max_{x\in\pi_K(X)}\langle x \mid y\rangle$ and by definition of $\arg\max_{|K|\leq k} \max_{x\in\pi_K(X)}\langle x \mid y\rangle$,

$$\begin{aligned}
\iff\ & \text{there exists } K^\sharp \in \arg\max_{|K|\leq k} \max_{x\in\pi_K(X)} \langle x \mid y\rangle \\
& \text{such that } x^\sharp \in \arg\max_{x\in\pi_{K^\sharp}(X)} \langle x \mid y\rangle
\end{aligned}$$

because $x^\sharp \in \arg\max_{x\in\pi_{K^\sharp}(X)}\langle x \mid y\rangle$ implies that $x^\sharp \in \pi_{K^\sharp}(X)$, hence that $x^\sharp \in X_k$, by Definition (8) of X_k,

$$\begin{aligned}
\iff\ & \text{there exists } K^\sharp \in \arg\max_{|K|\leq k} \max_{x\in\pi_K(X)} \langle x \mid y\rangle \\
& \text{such that } x^\sharp \in \pi_{K^\sharp}\Big(\arg\max_{z\in X} \langle z \mid \pi_{K^\sharp} y\rangle\Big)\ , \qquad \text{(by (15))} \\
\iff\ & \text{there exists } K^\sharp \in \arg\max_{|K|\leq k} \sigma_{\overline{\mathrm{co}}X}(\pi_K y) \\
& \hfill \text{(as } \max_{x\in\pi_K(X)}\langle x \mid y\rangle = \sigma_{\overline{\mathrm{co}}X}(\pi_K y) \text{ by (16))} \\
& \text{such that } x^\sharp \in \pi_{K^\sharp}\Big(X \cap F_\perp(\overline{\mathrm{co}}X, \pi_{K^\sharp} y)\Big)
\end{aligned}$$

as $\arg\max_{z\in X}\langle z \mid \pi_{K^\sharp} y\rangle = X \cap F_\perp(\overline{\mathrm{co}}X, \pi_{K^\sharp} y)$ by definition (6) of the exposed face $F_\perp(\overline{\mathrm{co}}X, \pi_{K^\sharp} y)$,

$$\begin{aligned}
\iff\ & x^\sharp \in \Big\{\pi_{K^\sharp}\big(X \cap F_\perp(\overline{\mathrm{co}}X, \pi_{K^\sharp} y)\big) : K^\sharp \in \mathcal{K}^\sharp_{X,k}(y)\Big\}\ . \\
& \hfill \text{(by using the notation (9) } \mathcal{K}^\sharp_{X,k}(y) = \arg\max_{|K|\leq k} \sigma_{\overline{\mathrm{co}}X}(\pi_K y))
\end{aligned}$$

Thus, we have proven the equality (11).

We now prove Equation (12) which can be rewritten as

$$F_\perp\big(\overline{\mathrm{co}}X_k, y\big) = \overline{\mathrm{co}}\widehat{F} \ \text{ where }\ \widehat{F} = \big\{\pi_{K^\sharp}\big(X \cap F_\perp(\overline{\mathrm{co}}X, \pi_{K^\sharp} y)\big) : K^\sharp \in \mathcal{K}^\sharp_{X,k}(y)\big\}$$

is the right-hand side term in equality (11), which now writes $X_k \cap F_\perp\big(\overline{\mathrm{co}}X_k, y\big) = \widehat{F}$.

First, we have that $F_\perp\big(\overline{\text{co}}X_k, y\big) \supset X_k \cap F_\perp\big(\overline{\text{co}}X_k, y\big) = \widehat{F}$, where the last equality is Equation (11). An exposed face is closed convex. Thus, $\overline{\text{co}}\widehat{F} \subset F_\perp(\overline{\text{co}}X_k, y)$. Similarly, since $F_\perp(\overline{\text{co}}X_k, y)$ is exposed, an extreme point e of $F_\perp(\overline{\text{co}}X_k, y)$ is also an extreme point of $\overline{\text{co}}X_k$. Thus, it is also contained in X_k. Using $X_k \cap F_\perp\big(\overline{\text{co}}X_k, y\big) = \widehat{F}$ (new form of Equation (11)), we get that $e \in \widehat{F}$. Thus, as $\widehat{F}$ contains all the extreme points of $F_\perp(\overline{\text{co}}X_k, y)$ and a convex set is the convex hull of its extreme points, we obtain that $F_\perp(\overline{\text{co}}X_k, y) \subset \overline{\text{co}}\widehat{F}$. Third, from $\overline{\text{co}}\widehat{F} \subset F_\perp(\overline{\text{co}}X_k, y)$ and $F_\perp(\overline{\text{co}}X_k, y) \subset \overline{\text{co}}\widehat{F}$, we conclude that $F_\perp(\overline{\text{co}}X_k, y) = \overline{\text{co}}\widehat{F}$, finally giving Equation (12). □

2.3.2 Proof of Corollary 2

Proof. The implication (13) is a consequence of (11). Indeed, as

$$\pi_{K^\sharp}\big(X \cap F_\perp(\overline{\text{co}}X, \pi_{K^\sharp}y)\big) \subset \mathcal{R}_{K^\sharp},$$

the support of any point x in $X_k \cap F_\perp(\overline{\text{co}}X_k, y)$ is included in one of the subsets $K^\sharp$ contained in $\mathcal{K}^\sharp_{X,k}(y)$ by (11).

Implication (13) can be interpreted as follows: since $x \in X_k$, it follows from (8) that

$$x \in \bigcup_{|K| \leq k} \mathcal{R}_K$$

and since x belongs to $F_\perp(\overline{\text{co}}X_k, y)$, we can be more precise and obtain from (11) that

$$x \in \pi_{K^\sharp}\big(X \cap F_\perp(\overline{\text{co}}X, \pi_{K^\sharp}y)\big).$$

As a consequence, x belongs to $\mathcal{R}_{K^\sharp}$ or, equivalently, $\text{supp}(x)$ is a subset of $K^\sharp$; the possible supports of x are the $K^\sharp \in \mathcal{K}^\sharp_{X,k}(y)$, determined by the dual vector y by means of (9).

Implication (14) is a direct consequence of (12). Indeed, as any x in $F_\perp(\overline{\text{co}}X_k, y)$ can be expressed as a convex combination of elements of $\mathcal{R}_{K^\sharp}$, with $K^\sharp$ in $\mathcal{K}^\sharp_{X,k}(y)$, the support of x is necessarily a subset of

$$\bigcup_{K^\sharp \in \mathcal{K}^\sharp_{X,k}(y)} K^\sharp$$

as desired. □

3 The case of generalized top-k and k-support norms

In §3.1, we provide background on generalized top-k and k-support dual norms. In §3.2, we apply Theorem 1 with (primal) set the unit ball of a norm, and obtain thus face characterization and support identification with k-support norms. In §3.3, we recall the notion of orthant-monotonic norm and, in this case, we obtain a characterization of the intersection of the k-sparse vectors with the faces of the k-support norm. Finally, in §3.4, we recall the notion of orthant-strictly monotonic norm and, in this case, we obtain a simpler characterization of the intersection of the k-sparse vectors with the faces of the k-support norm.

3.1 Background on generalized top-k and k-support norms

We provide background on generalized top-k and k-support dual norms that are constructed by means of a source norm [6]. In the following, the symbol $\star$ in the superscript indicates that the generalized k-support dual norm $\|\cdot\|^{\top\star}_{\star,(k)}$ is the dual norm of the generalized top-k dual norm $\|y\|^{\top}_{\star,(k)}$ and, thus, is a norm on the primal space. To stress the point, we use x for a primal vector, like in $\|x\|^{\top\star}_{\star,(k)}$, and y for a dual vector, like in $\|y\|^{\top}_{\star,(k)}$.

Definition 4 *Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$, that we call the source norm, with unit ball B. The unit ball $B_\star$ of the dual norm $\|\cdot\|_\star$ is the polar set of B, that is, $B_\star = B^\odot$. For any $k \in [\![1, d]\!]$, and using $\sup_{|K|\leq k}$ as a shorthand for $\sup_{K\subset[\![1,d]\!],|K|\leq k}$, we call*

(*i*) generalized top-k dual norm *the norm [6, Eq. (10)]*

$$\|y\|^{\top}_{\star,(k)} = \underbrace{\sup_{|K|\leq k} \| \overbrace{\pi_K(y)}^{\substack{k\text{-sparse}\\ \text{projection}\\ \text{on } \mathcal{R}_K}} \|_\star}_{\substack{\text{exploring all}\\ k\text{-sparse projections}}} \; , \quad \forall y \in \mathbb{R}^d \; , \tag{17}$$

whose unit ball is

$$B^{\top}_{\star,(k)} = \big\{ y \in \mathbb{R}^d : \|y\|^{\top}_{\star,(k)} \leq 1 \big\} = \bigcap_{|K|\leq k} \underbrace{\pi_K^{-1}(\mathcal{R}_K \cap B_\star)}_{\text{cylinder}} \; , \tag{18}$$

hence is an intersection of cylinders,

(*ii*) generalized k-support dual norm *the corresponding dual norm (of the generalized top-k dual norm) [6, Eq. (11)]*

$$\|\cdot\|^{\top\star}_{\star,(k)} = \big(\|\cdot\|^{\top}_{\star,(k)}\big)_\star \; , \tag{19}$$

whose unit ball is

$$B^{\top\star}_{\star,(k)} = \big\{ x \in \mathbb{R}^d : \|x\|^{\top\star}_{\star,(k)} \leq 1 \big\} = \overline{\mathrm{co}}\big(\bigcup_{|K|\leq k} \pi_K(B) \big) = \overline{\mathrm{co}}\big(\bigcup_{|K|\leq k} \pi_K(S) \big) \; , \tag{20}$$

and unit sphere is denoted by $S^{\top\star}_{(k)}$.

3.2 Face characterization and support identification

Here, we apply the result of Theorem 1 with (primal) set the unit ball of a norm.

Proposition 5 *Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$, that we call the source norm, with unit ball B. For any $k \in [\![1, d]\!]$ and any dual vector $y \in \mathbb{R}^d$, we have that*

$$\big(\bigcup_{|K|\leq k} \pi_K(B) \big) \cap F_\perp\big(B^{\top\star}_{\star,(k)}, y\big) = \Big\{ \pi_{K^\sharp}\big(F_\perp(B, \pi_{K^\sharp} y)\big) : K^\sharp \in \operatorname*{arg\,max}_{|K|\leq k} \|\pi_K y\|_\star \Big\} \; , \tag{21}$$

and the faces of $B^{\top\star}_{\star,(k)}$ *are related to the faces of* B *by*

$$F_\perp(B^{\top\star}_{\star,(k)}, y) = \overline{\mathrm{co}}\Big\{\pi_{K^\sharp}\big(F_\perp(B, \pi_{K^\sharp} y)\big) \,:\, K^\sharp \in \operatorname*{arg\,max}_{|K|\leq k} \|\pi_K y\|_\star\Big\} \,. \tag{22}$$

Proof. The proof results from Theorem 1 with $X = B$, $X_k = \bigcup_{|K|\leq k} \pi_K(X) = \bigcup_{|K|\leq k} \pi_K(B)$ in (8), and

$$\mathcal{K}^\sharp_{X,k}(y) = \operatorname*{arg\,max}_{\substack{K\subset[\![1,d]\!] \\ |K|\leq k}} \sigma_B(\pi_K y) = \operatorname*{arg\,max}_{\substack{K\subset[\![1,d]\!] \\ |K|\leq k}} \sigma_{\overline{\mathrm{co}}X}(\pi_K y) = \operatorname*{arg\,max}_{|K|\leq k} \|\pi_K y\|_\star$$

in (9), as σ_B is the dual norm $\|\cdot\|_\star$. Then, Equation (11) gives Equation (12), and Equation (21) gives Equation (22), where we use the expression (20) of $B^{\top\star}_{\star,(k)} = \overline{\mathrm{co}}\big(\bigcup_{|K|\leq k} \pi_K(B)\big)$. □

We deduce support identification.

Theorem 6 *Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be a smooth convex function, and* $\gamma > 0$*. Let* $\|\cdot\|$ *be a norm on* $\mathbb{R}^d$*. For given sparsity threshold* $k \in [\![1, d]\!]$*, we consider the generalized top-k dual norm* $\|\cdot\|^{\top\star}_{\star,(k)}$ *(see Definition 4). Then, an optimal solution* $x^\sharp$ *of*

$$\min_{x\in\mathbb{R}^d} \big(f(x) + \gamma\|x\|^{\top\star}_{\star,(k)}\big) \tag{23a}$$

has support

$$\mathrm{supp}(x^\sharp) \subset \bigcup_{\substack{K^\sharp \in \operatorname*{arg\,max}_{|K|\leq k} \\ \|\pi_K(-\nabla f(x^\sharp))\|_\star}} K^\sharp \,. \tag{23b}$$

As a consequence, if

$$\operatorname*{arg\,max}_{|K|\leq k} \|\pi_K(-\nabla f(x^\sharp))\|_\star = K^\sharp \quad \textit{is unique} \,, \tag{24a}$$

$$\textit{then} \quad \mathrm{supp}(x^\sharp) \subset K^\sharp \textit{ with } |K^\sharp| \leq k \,, \tag{24b}$$

so that the optimal solution $x^\sharp$ *is k-sparse.*

Proof. We have that

$$\begin{aligned}
& x^\sharp \in \operatorname*{arg\,min}_{x\in\mathbb{R}^d} \big(f(x) + \gamma\|x\|^{\top\star}_{\star,(k)}\big) \\
\iff & x^\sharp \in \operatorname*{arg\,min}_{x\in\mathbb{R}^d} \big(f(x) + \gamma\sigma_{B^{\top}_{\star,(k)}}\big) && \text{(by (18))} \\
\iff & 0 \in \partial\big(f + \gamma\sigma_{B^{\top}_{\star,(k)}}\big)(x^\sharp) && \text{(by the Fermat rule)} \\
\iff & 0 \in \partial f(x^\sharp) + \gamma\partial\sigma_{B^{\top}_{\star,(k)}}(x^\sharp)
\end{aligned}$$

by [3, Corollary 16.48], as both functions are proper convex lsc, and $\mathrm{dom} f = \mathbb{R}^d$,

$$\iff 0 \in \nabla f(x^\sharp) + \gamma \partial \sigma_{B^\top_{\star,(k)}}(x^\sharp) \qquad \text{(by [3, Proposition 17.31 (i)])}$$
$$\iff 0 \in \nabla f(x^\sharp) + \gamma F_\perp(B^\top_{\star,(k)}, x^\sharp)$$

as the subdifferential of a support function is the support function of the corresponding face, see for instance [26, Theorem 1.7.2], [25, Corollary 8.25],

$$\iff -\nabla f(x^\sharp) \in \gamma F_\perp(B^\top_{\star,(k)}, x^\sharp)$$
$$\iff \begin{cases} \text{either } x^\sharp = 0\ , & -\nabla f(0) \in B^\top_{\star,(k)} \\ \text{or } x^\sharp \neq 0\ , & \frac{x^\sharp}{\|x^\sharp\|^{\top\star}_{\star,(k)}} \in F_\perp(B^{\top\star}_{\star,(k)}, -\nabla f(x^\sharp)) \end{cases} \qquad \text{(by polarity [11, Theorem 5.1])}$$
$$\implies \begin{cases} \text{either } x^\sharp = 0\ , & \mathrm{supp}(x^\sharp) = \emptyset \\ \text{or } x^\sharp \neq 0\ , & \mathrm{supp}(x^\sharp) \subset \bigcup_{K^\sharp \in \arg\max_{|K| \leq k} \|\pi_K(-\nabla f(x^\sharp))\|_\star} K^\sharp \end{cases} \qquad \text{(by (14))}$$

Thus, we have proven (23b). Equation (24) follows trivially.

□

Corollary 7 *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a smooth convex function, $\gamma > 0$ and $\|\cdot\|_1$ be the ℓ_1-norm. An optimal solution $x^\sharp$ of*

$$\min_{x \in \mathbb{R}^d} \big(f(x) + \gamma \|x\|_1 \big) \tag{25a}$$

has support

$$\mathrm{supp}(x^\sharp) \subset \arg\max_{j \in [\![1,d]\!]} |\nabla_j f(x^\sharp)|\ . \tag{25b}$$

Proof. If $\|\cdot\|$ is the ℓ_1-norm $\|\cdot\|_1$ on $\mathbb{R}^d$, then the generalized top-k dual norm $\|\cdot\|^{\top\star}_{\star,(k)}$ is also the ℓ_1-norm $\|\cdot\|_1$, for any $k \in [\![1,d]\!]$ (see Table 1). Then, we apply Theorem 6 in the case $k = 1$, and we get (23b), which is exactly (25b) as $\bigcup_{K^\sharp \in \arg\max_{|K| \leq 1} \|\pi_K(-\nabla f(x^\sharp))\|_\star} K^\sharp = \arg\max_{j \in [\![1,d]\!]} |\nabla_j f(x^\sharp)|$. □

3.3 The orthant-monotonic case

The notion of *orthant-monotonic norm*[2] has been introduced in [16, 17] and an equivalent characterization is provided in [7, Item 7 in Proposition 4]. A norm $\|\cdot\|$ is orthant-monotonic if and only if it is *increasing with the coordinate subspaces*, in the sense that $\|x_J\| \leq \|x_K\|$ for any $x \in \mathbb{R}^d$ and any two subsets J and K of $[\![1,d]\!]$ satisfying $J \subset K$. In fact, this is equivalent to $\|x_J\| \leq \|x\|$ for any vector x in $\mathbb{R}^d$ and any subset J of $[\![1,d]\!]$.

[2] A norm is orthant-monotonic if and only if it is monotonic in every orthant, see [16, Lemma 2.12], hence the name.

Proposition 8 *Let $\|\cdot\|$ be a (source) norm on $\mathbb{R}^d$, with unit ball B. For given sparsity threshold $k \in [\![1, d]\!]$, we consider the generalized top-k dual norm $\|\cdot\|^{\top\star}_{\star,(k)}$ (see Definition 4). If the source norm $\|\cdot\|$ is orthant-monotonic, then for any nonzero dual vector $y \in \mathbb{R}^d \setminus \{0\}$, we have that*

$$\ell_0^{\leq k} \cap F_\perp\big(B^{\top\star}_{\star,(k)}, y\big) = \Big\{ \pi_{K^\sharp}\big(F_\perp(B, \pi_{K^\sharp} y)\big) \,:\, K^\sharp \in \mathcal{K}^\sharp_{X,k}(y) \Big\} , \tag{26}$$

and the faces of $B^{\top\star}_{\star,(k)}$ are related to the faces of B by (22).

To the difference of (21), the left-hand side of (26) is exactly the intersection of the level set $\ell_0^{\leq k}$ of the ℓ_0 pseudonorm with the exposed face $F_\perp\big(B^{\top\star}_{\star,(k)}, y\big)$, whereas it was the intersection of a *subset* of the level set $\ell_0^{\leq k}$ of the ℓ_0 pseudonorm with the exposed face $F_\perp\big(B^{\top\star}_{\star,(k)}, y\big)$ in the left-hand side of (21).

Proof. The assumptions of Proposition 5 are satisfied. Thus, the equality (21) holds true. The right-hand sides of (21) and of (26) are identical. By comparing the left-hand side of (21) — namely, $\bigcup_{|K|\leq k} \pi_K(B) \cap F_\perp(B^{\top\star}_{\star,(k)}, y)$ — with the left-hand side of (26) — namely, $\ell_0^{\leq k} \cap F_\perp(B^{\top\star}_{\star,(k)}, y)$ — we conclude that proving (26) amounts to showing that

$$\bigcup_{|K|\leq k} \pi_K(B) \cap F_\perp(B^{\top\star}_{\star,(k)}, y) = \ell_0^{\leq k} \cap F_\perp(B^{\top\star}_{\star,(k)}, y) \; .$$

We prove the equality by two opposite inclusions.

On the one hand, we have that

$$\begin{aligned} \bigcup_{|K|\leq k} \pi_K(B) \cap F_\perp(B^{\top\star}_{\star,(k)}, y) &\subset \bigcup_{|K|\leq k} \mathcal{R}_K \cap F_\perp(B^{\top\star}_{\star,(k)}, y) \\ &\qquad \text{(as } \pi_K(B) \subset \mathcal{R}_K \text{ for all } K\text{, by definition of the orthogonal projection mapping } \pi_K) \\ &= \ell_0^{\leq k} \cap F_\perp(B^{\top\star}_{\star,(k)}, y) \; . \qquad \text{(by (2c))} \end{aligned}$$

On the other hand, we prove the reverse inclusion $\ell_0^{\leq k} \cap F_\perp(B^{\top\star}_{\star,(k)}, y) \subset \bigcup_{|K|\leq k} \pi_K(B) \cap F_\perp(B^{\top\star}_{\star,(k)}, y)$. Indeed, for any nonzero dual vector $y \in \mathbb{R}^d \setminus \{0\}$, we have that

$$\begin{aligned} \ell_0^{\leq k} \cap F_\perp(B^{\top\star}_{\star,(k)}, y) &= \ell_0^{\leq k} \cap S^{\top\star}_{\star,(k)} \cap F_\perp(B^{\top\star}_{\star,(k)}, y) \qquad \text{(because } F_\perp(B^{\top\star}_{\star,(k)}, y) \subset S^{\top\star}_{\star,(k)} \text{ as } y \neq 0) \\ &= \ell_0^{\leq k} \cap S^{\top\star}_{\star,(d)} \cap F_\perp(B^{\top\star}_{\star,(k)}, y) \end{aligned}$$

by [7, Equation (37) in Proposition 20], giving $\ell_0^{\leq k} \cap S^{\top\star}_{\star,(k)} = \ell_0^{\leq k} \cap S^{\top\star}_{\star,(d)}$ using the property that $\|\cdot\|$ is orthant-monotonic,

$$= \ell_0^{\leq k} \cap S \cap F_\perp(B^{\top\star}_{\star,(k)}, y)$$

by [7, Item 2 in Proposition 13], giving $\|\cdot\|^{\top\star}_{\star,(d)} = \|\cdot\|$ using the property that $\|\cdot\|$ is orthant-monotonic,

$$
\begin{aligned}
&= \bigcup_{|K|\leq k} \mathcal{R}_K \cap S \cap F_{\perp}(B^{\top\star}_{\star,(k)}, y) && \text{(as } \ell_0^{\leq k} = \textstyle\bigcup_{|K|\leq k} \mathcal{R}_K \text{ by (2c))} \\
&\subset \bigcup_{|K|\leq k} \pi_K(S) \cap F_{\perp}(B^{\top\star}_{\star,(k)}, y) \\
&\qquad \text{(as } \mathcal{R}_K \cap S \subset \pi_K(S)\text{, by definition of the orthogonal projection mapping } \pi_K) \\
&\subset \bigcup_{|K|\leq k} \pi_K(B) \cap F_{\perp}(B^{\top\star}_{\star,(k)}, y) \; . && \text{(as } \pi_K(S) \subset \pi_K(B))
\end{aligned}
$$

This ends the proof. □

3.4 The orthant-strictly monotonic case

The notion of *orthant-strictly monotonic norm* has been introduced in [7, Definition 5]. A norm $\|\cdot\|$ is orthant-strictly monotonic if and only if, for all x, x' in $\mathbb{R}^d$, we have $\big(\, |x| < |x'| \text{ and } x \circ x' \geq 0 \Rightarrow \|x\| < \|x'\| \, \big)$. An equivalent characterization is provided in [7, Item 3 in Proposition 6]. A norm $\|\cdot\|$ is orthant-strictly monotonic if and only if it is *strictly increasing with the coordinate subspaces*, in the sense that[3], for any $x \in \mathbb{R}^d$ and any $J \subsetneq K \subset [\![1, d]\!]$, we have $x_J \neq x_K \Rightarrow \|x_J\| < \|x_K\|$. An orthant-strictly monotonic norm is orthant-monotonic.

Proposition 9 *Let $\|\cdot\|$ be a (source) norm on $\mathbb{R}^d$, with unit ball B. For given sparsity threshold $k \in [\![1, d]\!]$, we consider the generalized top-k dual norm $\|\cdot\|^{\top\star}_{\star,(k)}$ (see Definition 4). If the source norm $\|\cdot\|$ is orthant-strictly monotonic, then for any nonzero dual vector $y \in \mathbb{R}^d \setminus \{0\}$, we have that*

$$
\ell_0^{\leq k} \cap F_{\perp}\big(B^{\top\star}_{\star,(k)}, y\big) = \Big\{ F_{\perp}(B, \pi_{K^\sharp} y) : \; K^\sharp \in \mathcal{K}^\sharp_{X,k}(y) \Big\} \; , \tag{27}
$$

and the faces of $B^{\top\star}_{\star,(k)}$ are related to the faces of B by

$$
F_{\perp}(B^{\top\star}_{\star,(k)}, y) = \overline{\mathrm{co}} \Big\{ F_{\perp}(B, \pi_{K^\sharp} y) : \; K^\sharp \in \operatorname*{arg\,max}_{|K|\leq k} \|\pi_K y\|_\star \Big\} \; . \tag{28}
$$

To the difference of the right-hand sides of (21), (22), and (26), there is no projection $\pi_{K^\sharp}\big(F_{\perp}(B, \pi_{K^\sharp} y)\big)$, but just $F_{\perp}(B, \pi_{K^\sharp} y)$ in the right-hand sides of (27) and (28).

Proof. An orthant-strictly monotonic norm is orthant-monotonic, the assumptions of Proposition 8 hold true. Thus, Equation (26) holds true and hence, to prove (27), it suffices to show that $\pi_{K^\sharp}\big(F_{\perp}(B, \pi_{K^\sharp} y)\big) = F_{\perp}(B, \pi_{K^\sharp} y)$.

First, notice that, for any $K^\sharp \in \arg\max_{|K|\leq k} \|\pi_K y\|_\star$, we have that $\|\pi_{K^\sharp} y\|_\star > 0$. Indeed, on

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

the contrary, we would have that $\max_{|K|\leq k}\|\pi_K y\|_\star = 0$, hence that $\pi_K y = 0$ for any K with $|K| \leq k$. As $k \geq 1$, this would imply that $y = 0$.

Second, we have that

$$\begin{aligned}
x \in F_\perp(B, \pi_{K^\sharp} y) &\iff \langle x \mid \pi_{K^\sharp} y\rangle = \max_{z\in B} \langle z \mid \pi_{K^\sharp} y\rangle \text{ and } x \in B \\
&\qquad\qquad \text{(by definition (6) of the exposed face } F_\perp(B, \pi_{K^\sharp} y)) \\
&\iff \langle x \mid \pi_{K^\sharp} y\rangle = \|\pi_{K^\sharp} y\|_\star \text{ and } x \in B \quad \text{(by definition of the dual norm } \|\cdot\|_\star) \\
&\iff \langle x \mid \pi_{K^\sharp} y\rangle = \|x\|\|\pi_{K^\sharp} y\|_\star \text{ and } \|x\| = 1 \\
&\iff \langle \pi_{K^\sharp} x \mid \pi_{K^\sharp} y\rangle = \|x\|\|\pi_{K^\sharp} y\|_\star \text{ and } \|x\| = 1 \\
&\qquad\qquad \text{(as the orthogonal projection } \pi_K \text{ is self-adjoint (see (3)))} \\
&\implies \|\pi_{K^\sharp} x\|\|\pi_{K^\sharp} y\|_\star \geq \langle \pi_{K^\sharp} x \mid \pi_{K^\sharp} y\rangle = \|x\|\|\pi_{K^\sharp} y\|_\star \qquad \text{(by polar inequality)} \\
&\implies \|\pi_{K^\sharp} x\| \geq \|x\| \qquad \text{(since } \|\pi_{K^\sharp} y\|_\star > 0) \\
&\implies \|\pi_{K^\sharp} x\| = \|x\|
\end{aligned}$$

because the norm $\|\cdot\|$ is orthant-strictly monotonic, hence orthant-monotonic, hence $\|\pi_{K^\sharp} x\| \leq \|x\|$

$$\implies \pi_{K^\sharp} x = x \ ,$$

because the norm $\|\cdot\|$ is orthant-strictly monotonic hence, if we had $\pi_{K^\sharp} x \neq x$, we would conclude that $\|\pi_{K^\sharp} x\| < \|x\|$.

We conclude that $\pi_{K^\sharp}\big(F_\perp(B, \pi_{K^\sharp} y)\big) = F_\perp(B, \pi_{K^\sharp} y)$. □

4 Geometry of the top-(q,k) and (p,k)-support norms

This section is devoted to the geometric analysis of the face and cone lattices of the unit balls of top-(q,k) norm and (p,k)-support norms. In §4.1, we recall the definition of the ℓ_p-norms, and then of top-(q,k) norm and (p,k)-support norms. In §4.2, we study the case where p is equal to $+\infty$. In §4.3, we study the case where $1 < p < +\infty$. We shall also see in §4.1 that these norms do not depend on k when p is equal to 1.

4.1 The top-(q,k) norm and (p,k)-support norms

For any p in $[1, +\infty[$ and x in $\mathbb{R}^d$, let us recall that the ℓ_p-norm of x is

$$\|x\|_p = \Big(\sum_{i=1}^{d} |x_i|^p\Big)^{\frac{1}{p}} \ ,$$

and that its ℓ_∞-norm is

$$\|x\|_\infty = \max_{i\in[\![1,d]\!]} |x_i| \ .$$

For any p in $[1, +\infty]$, we denote by B_p and S_p the unit ball and the unit sphere for the ℓ_p-norm. When the source norm is the ℓ_p-norm,

- the corresponding generalized k-support dual norm $\left(\|\cdot\|_p\right)^{\top\star}_{\star,(k)}$ is the (p,k)-support norm denoted by $\|\cdot\|^{\top\star}_{p,k}$, with unit ball $B^{\top\star}_{p,k}$ and unit sphere $S^{\top\star}_{p,k}$,

- the corresponding generalized top-k dual norm $\left(\|\cdot\|_p\right)^{\top}_{\star,(k)}$ is the top-(q,k) norm denoted by $\|\cdot\|^{\top}_{q,k}$, where $1/p + 1/q = 1$, with unit ball $B^{\top}_{q,k}$ and unit sphere $S^{\top}_{q,k}$.

For any p and q in $[1, +\infty]$ such that $1/p + 1/q = 1$, we have

$$\|\cdot\|^{\top}_{p,k} = \sigma_{B^{\top\star}_{q,k}} \ , \ \ B^{\top}_{p,k} = \left(B^{\top\star}_{q,k}\right)^{\odot} \text{ and } \ \|\cdot\|^{\top\star}_{q,k} = \sigma_{B^{\top}_{p,k}} \ , \ \ B^{\top\star}_{q,k} = \left(B^{\top}_{p,k}\right)^{\odot} \ . \tag{29}$$

The norms obtained when p varies from 1 to $+\infty$ are summarized in Table 1. The top-$(1,k)$ and top-$(2,k)$ norms arise in various contexts under different names, see [14] and references therein. They are called the *vector k-norm* in [28, Sect. 2], the *largest k-norm* or *CVaR norm* for the ℓ_∞-norm in [15, Sect. 1], the 2-*k-symmetric gauge norm* in [21], and the *Ky Fan vector norm* for the ℓ_2-norm in [22]. Similarly, the $(2,k)$-support norm is referred to as k-support norm in [1]. The (p,k)-support norm for $p \in [1, \infty]$ is defined in [20, Definition 21] where it is showed that the dual norm of the top-(p,k) norm is the (q,k)-support norm, where $1/p + 1/q = 1$. Therefore, the generalized k-support dual norm is the (p,k)-support norm (denoted by $\|\cdot\|^{\top\star}_{p,k}$) when the source norm $\|\cdot\|$ is the ℓ_p-norm $\|\cdot\|_p$.

Let us briefly discuss the cases when p is equal to 1 or to $+\infty$. When p is equal to 1, it follows from the definition that $B^{\top\star}_{p,k}$ is the cross-polytope B_1 independently on k and that its polar $B^{\top}_{q,k}$ coincides with the unit hypercube B_∞. When p is equal to $+\infty$, the balls $B^{\top\star}_{p,k}$ and $B^{\top}_{q,k}$ form two families of polytopes that interpolate between the cross-polytope and the hypercube [9], as illustrated in Figures 3 and 4 when d is equal to 3.

If we apply the result of Proposition 8 to the orthant-monotonic norm ℓ_∞, we obtain a characterization of $\ell_0^{\leq k} \cap F_\perp(B^{\top\star}_{\infty,k}, y)$ in terms of the sets $\pi_K\big(S_\infty \cap F_\perp(B_\infty, \pi_K y)\big)$ for certain subsets K of $[\![1, d]\!]$. If we apply the result of Proposition 9 to the orthant-strictly monotonic norms ℓ_p, where p belongs to $[1, \infty[$, we obtain a characterization of $\ell_0^{\leq k} \cap F_\perp(B^{\top\star}_{p,k}, y)$ in terms of the sets $S_p \cap F_\perp(B_p, \pi_K y)$ for certain subsets K of $[\![1, d]\!]$.

4.2 The case when p is equal to $+\infty$

When the source norm is the ℓ_∞-norm, the corresponding row of Table 1 tells us that we should study the unit balls of top-$(1,k)$ norms and (∞,k)-support norms. This results in families of polytopes whose geometry and combinatorics have been studied in [9]. In this section, we review these families of polytopes. Following the notation of Coxeter, denote by γ_d the d-dimensional hypercube $[-1, 1]^d$ and by β_d the cross-polytope whose vertices are the centers of the facets of γ_d. Note that these two polytopes are related by polarity. It follows from [9, Equations (1.2) and (1.3)] that, for every k in $[\![1, d]\!]$,

$$B^{\top}_{1,k} = \mathrm{co}\bigg(\beta_d \cup \frac{1}{k}\gamma_d\bigg) \text{ and } B^{\top\star}_{\infty,k} = k\beta_d \cap \gamma_d \ . \tag{30}$$

source norm $\|\cdot\|$	$\|\cdot\|^{\top}_{\star,(k)}$, $k \in [\![1,d]\!]$	$\|\cdot\|^{\top\star}_{\star,(k)}$, $k \in [\![1,d]\!]$
$\|\cdot\|_p$	top-(q,k) norm $\|y\|^{\top}_{q,k}$ $\|y\|^{\top}_{q,k} = \left(\sum_{l=1}^{k} \|y_{\nu(l)}\|^q\right)^{\frac{1}{q}}$ $\|y\|^{\top}_{q,1} = \|y\|_\infty$	(p,k)-support norm $\|x\|^{\top\star}_{p,k}$ no analytic expression $\|x\|^{\top\star}_{p,1} = \|x\|_1$
$\|\cdot\|_1$	top-(∞,k) norm ℓ_∞-norm $\|y\|^{\top}_{\infty,k} = \|y\|_\infty$, $\forall k \in [\![1,d]\!]$	(1,k)-support norm ℓ_1-norm $\|x\|^{\top\star}_{1,k} = \|x\|_1$, $\forall k \in [\![1,d]\!]$
$\|\cdot\|_2$	top-(2,k) norm $\|y\|^{\top}_{2,k} = \sqrt{\sum_{l=1}^{k} \|y_{\nu(l)}\|^2}$ $\|y\|^{\top}_{2,1} = \|y\|_\infty$	(2,k)-support norm $\|x\|^{\top\star}_{2,k}$ no analytic expression (computation [1, Prop. 2.1]) $\|x\|^{\top\star}_{2,1} = \|x\|_1$
$\|\cdot\|_\infty$	top-(1,k) norm $\|y\|^{\top}_{1,k} = \sum_{l=1}^{k} \|y_{\nu(l)}\|$ $\|y\|^{\top}_{1,1} = \|y\|_\infty$	(∞,k)-support norm $\|x\|^{\top\star}_{\infty,k} = \max\{\frac{\|x\|_1}{k}, \|x\|_\infty\}$ $\|x\|^{\top\star}_{\infty,1} = \|x\|_1$

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

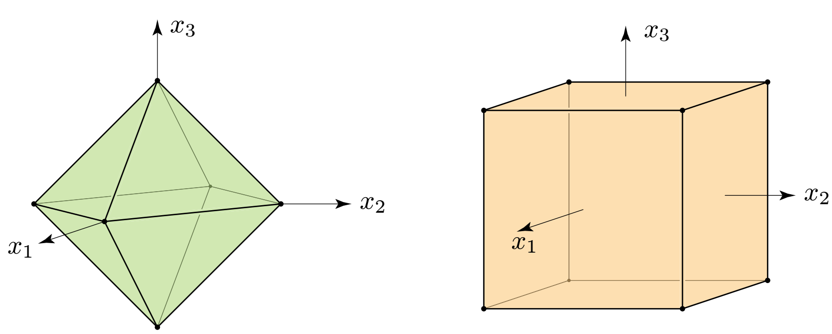

Figure 3: Unit balls $B^{\top\star}_{\infty,1}$ (left) and $B^{\top}_{1,1}$ (right) when $d = 3$

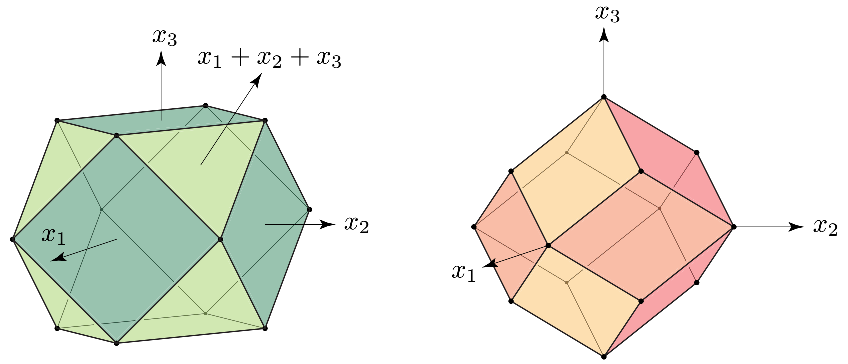

Figure 4: Unit balls $B^{\top\star}_{\infty,2}$ (left) and $B^{\top}_{1,2}$ (right) when $d = 3$

When k is equal to 1, $B^{\top}_{1,k}$ coincides with the hypercube γ_d and $B^{\top\star}_{\infty,k}$ with the cross-polytope β_d. When k is equal to d, the opposite holds: $B^{\top}_{1,k}$ coincides with the cross-polytope β_d and $B^{\top\star}_{\infty,k}$ with the hypercube γ_d. In particular these two families interpolate between the hypercube and the cross-polytope and, as pointed out in [9], $B^{\top}_{1,k}$ and $B^{\top\star}_{\infty,k}$ are related by polarity for all k in $[\![1, d]\!]$ and not just when k is equal to 1 or d. Note that in [9] the parameter k is allowed to take any (possibly non integral) value in the interval $[1, d]$. In dimension 3, these two polytopes are shown on Figure 3 when k is equal to 1 and in Figure 4 when k is equal to 2. Theorem 2.1 from [9] can be rephrased as follows.

Theorem 10 *The facets of $B^{\top}_{1,k}$ are precisely the sets of the form*

$$\mathrm{co}\biggl(F_{\perp}(\beta_d, y) \cup \frac{1}{k} F_{\perp}(\gamma_d, y)\biggr) \tag{31}$$

where y is a vector in $\{-1, 0, 1\}^d$ with exactly k nonzero coordinates.

Observe that (31) is precisely $F_{\perp}(B^{\top}_{1,k}, y)$. As noted in [9], for any vector y in $\{-1, 0, 1\}^d$, the affine hulls of $F_{\perp}(\beta_d, y)$ and $F_{\perp}(\gamma_d, y)$ are orthogonal subspaces of $\mathbb{R}^d$. More precisely, if we denote by k the number of nonzero coordinates of y, hence $k = \ell_0(y)$, then the two polytopes $F_{\perp}(\beta_d, y)$ and $F_{\perp}(\gamma_d, y)/k$ intersect in a single point that belongs to the relative interior of both of these polytopes. We then get, as an immediate consequence, the following description of all the proper faces of $B^{\top}_{1,k}$ from Theorem 10.

Corollary 11 *The proper faces of $B^{\top}_{1,k}$ are precisely the sets of the form*

$$\mathrm{co}\biggl(F \cup \frac{1}{k} G\biggr) \tag{32}$$

where, for some vector y in $\{-1, 0, 1\}^d$ with exactly k nonzero coordinates,

(i) F and G are exposed faces of $F_{\perp}(\beta_d, y)$ and $F_{\perp}(\gamma_d, y)$, respectively,

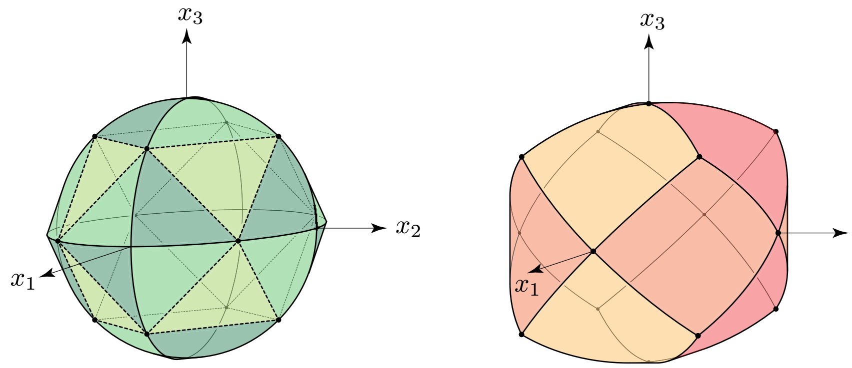

Figure 5: Unit balls $B^{\top\star}_{2,2}$ (left) and $B^{\top}_{2,2}$ (right) when $d = 3$

(ii) F *and* G *are not both empty,*

(iii) F *is equal to* $F_{\perp}(\beta_d, y)$ *if and only if* G *is equal to* $F_{\perp}(\gamma_d, y)$.

Corollary 11 completely describes the face lattice of $B^{\top}_{1,k}$ (which is further enumerated in [9]). Since $B^{\top\star}_{\infty,k}$ is the polar of $B^{\top}_{1,k}$, the normal cones of $B^{\top\star}_{\infty,k}$ are precisely the cones spanned by the faces of $B^{\top}_{1,k}$ and, as a consequence, Corollary 11 also describes the normal fan of $B^{\top\star}_{\infty,k}$. By the duality between the face lattice of a polytope and its normal fan, one then also recovers the face lattice of $B^{\top\star}_{\infty,k}$ from that corollary and the normal fan of $B^{\top}_{1,k}$.

4.3 The case when $1 < p < +\infty$

When the source norm is the ℓ_p-norm where $1 < p < +\infty$, the first row of Table 1 tells us that we should study the unit balls of the top-(q,k) norm, with $1/p + 1/q = 1$, and its dual (p,k)-support norm. Thus, we will describe the exposed faces and the normal cones of $B^{\top\star}_{p,k}$. The exposed faces and normal cones of $B^{\top}_{p,k}$ can then be recovered by polarity. The balls $B^{\top\star}_{2,2}$ and $B^{\top}_{2,2}$ are shown on Figure 5 when d is equal to 3. One can see that $B^{\top}_{2,2}$ is the intersection of three cylinders colored yellow, orange, and red. By duality, $B^{\top\star}_{2,2}$ has eight triangular faces. While these triangular faces are exposed, their edges, shown as dotted lines, are faces of $B^{\top\star}_{2,2}$ that are not exposed.

The face of B_p exposed by a nonzero vector y from $\mathbb{R}^d$ can be recovered from the equality case of Hölder's inequality. Indeed, by this inequality,

$$\big|\langle z \,|\, y\rangle\big| \leq \|y\|_q \tag{33}$$

for any point z in B_p with equality when

$$|z_i|^p = \lambda |y_i|^q \tag{34}$$

for some nonnegative number λ and all integers i satisfying $1 \leq i \leq d$, where

$$\frac{1}{p} + \frac{1}{q} = 1\ . \tag{35}$$

Now assume that z belongs to the face of B_p exposed by y. In that case, $\|z\|_p$ is equal to 1 and (33) must turn into an equality. In particular, there exists a nonnegative number λ satisfying (34) for every i. The value of λ can be recovered by summing (34) over i:

$$\lambda = \frac{1}{\|y\|_q^q} \; . \tag{36}$$

As a consequence, z must be the point such that, when y_i is equal to 0 then so is z_i and when y_i is nonzero, then z_i is the number with the same sign than y_i satisfying

$$|z_i| = \left(\frac{|y_i|}{\|y\|_q}\right)^{q/p} \; . \tag{37}$$

Hence, the face of B_p exposed by y is made of just the point z. From now on, we shall denote z by $v_p(y)$. Note that $v_2(y)$ and y are collinear, more precisely,

$$v_2(y) = \frac{y}{\|y\|_2} \; . \tag{38}$$

We are now able to characterize the exposed faces of $B_{p,k}^{\top\star}$.

Theorem 12 *For any number p satisfying $1 < p < +\infty$, the face of $B_{p,k}^{\top\star}$ exposed by a given nonzero vector y is the convex hull of all the points of the form $v_p(\pi_{K^\sharp} y)$ where*

$$K^\sharp \in \operatorname*{arg\,max}_{|K| \le k} \|\pi_K y\|_1 \; . \tag{39}$$

Proof. Using Equation (28) for the orthant-strictly monotonic ℓ_p-norm, one obtains that

$$F_\perp\left(B_{p,k}^{\top\star}, y\right) = \overline{\mathrm{co}}\Big\{F_\perp(B_p, \pi_{K^\sharp} y) : \; K^\sharp \in \operatorname*{arg\,max}_{|K| \le k} \|\pi_K y\|_p\Big\} \; .$$

Consider any size k subset $K^\sharp$ of $[\![1, d]\!]$ that belongs to $\operatorname{arg\,max}_{|K| \le k} \|\pi_K y\|_p$. As seen in Equation (37), the face of B_p exposed by $\pi_{K^\sharp} y$ is the vertex $v_p(\pi_{K^\sharp} y)$. Hence,

$$F_\perp\left(B_{p,k}^{\top\star}, y\right) = \overline{\mathrm{co}}\Big\{v_p(\pi_{K^\sharp} y) : \; K^\sharp \in \operatorname*{arg\,max}_{|K| \le k} \|\pi_K y\|_p\Big\} \; .$$

Finally, since the ℓ_p-norms are (strictly) monotonic,

$$\operatorname*{arg\,max}_{|K| \le k} \|\pi_K y\|_p = \operatorname*{arg\,max}_{|K| \le k} \|\pi_K y\|_1,$$

which completes the proof. □

Let us first introduce some notation. Consider an integer k in $[\![1, d]\!]$ and a nonzero vector y from $\mathbb{R}^d$. Denote by $m_k(y)$ the largest number such that the set

$$\big\{i \in [\![1, d]\!] : \; |y_i| \ge m_k(y)\big\}$$

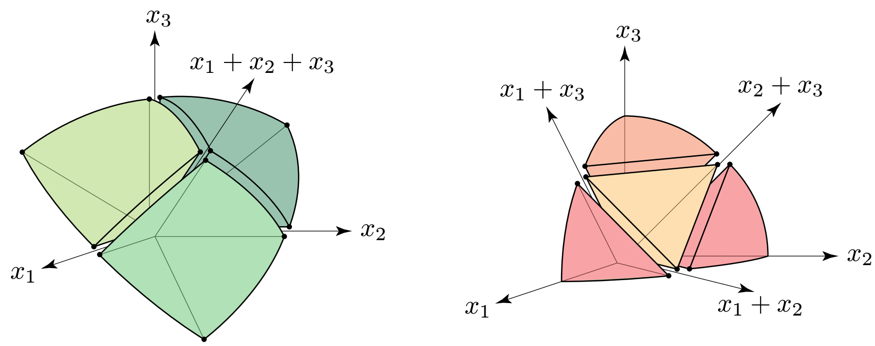

Figure 6: Portions of the normal fan of $B_{2,2}^{\top\star}$ (left) and of $B_{2,2}^{\top}$ (right) when $d = 3$

contains at least k indices. In other words,

$$m_k(y) = \sup\left\{\lambda \geq 0 : \left|\{i \in [\![1, d]\!] : |y_i| \geq \lambda\}\right| \geq k\right\} .$$

We will refer by $L_k(y)$ to the set of the indices i such that $|y_i|$ is greater than $m_k(y)$ and by $\overline{L}_k(y)$ the set of the indices i such that $|y_i|$ is greater than or equal to $m_k(y)$:

$$L_k(y) = \left\{i \in [\![1, d]\!] : |y_i| > m_k(y)\right\} , \tag{40a}$$

$$\overline{L}_k(y) = \left\{i \in [\![1, d]\!] : |y_i| \geq m_k(y)\right\} . \tag{40b}$$

The following statement is an immediate consequence of these definitions.

Proposition 13 *For any integer k in $[\![1, d]\!]$ and any nonzero vector y in $\mathbb{R}^d$,*

$$L_k(y) = \bigcap_{K^\sharp} K^\sharp$$

and

$$\overline{L}_k(y) = \bigcup_{K^\sharp} K^\sharp$$

where the union and the intersection range over the elements $K^\sharp$ of $\arg\max_{|K| \leq k} \|\pi_K y\|_1$.

Using these notations, Theorem 12 allows to recover the description of the normal cones of $B_{p,k}^{\top\star}$ given in [19, Proposition 23]. In our setting, the normal cone of $B_{p,k}^{\top\star}$ at one of its exposed faces F refers to the closure of the set of the vectors y in $\mathbb{R}^d$ such that F is the face of $B_{p,k}^{\top\star}$ exposed by y. The normal fans of $B_{2,2}^{\top\star}$ and $B_{2,2}^{\top}$ are illustrated in Fig. 6 when d is equal to 3. The figure only shows a portion of these fans but both can be reconstructed by symmetry.

Theorem 14 ([19, Proposition 23]) *The normal cones of $B_{p,k}^{\top\star}$ at its exposed faces are the sets of the form*

$$\overline{\text{cone}}\left\{y \in \mathbb{R}^d : \pi_{\overline{L}_k(z)} y = z, \, \overline{L}_k(y) = \overline{L}_k(z)\right\} , \tag{41}$$

where $\overline{\text{cone}}(X)$ denotes the closure of the cone spanned by X and z is a unit vector from $\mathbb{R}^d$ such that z coincides with $\pi_{\overline{L}_k(z)} z$.

Proof. Consider a nonzero vector z in $\mathbb{R}^d$ and denote

$$G = F_\perp\left(B_{p,k}^{\top\star}, z\right).$$

According to Theorem 12,

$$G = \overline{\text{co}}\left\{v_p(\pi_{K^\sharp} z) : K^\sharp \in \underset{|K| \leq k}{\arg\max} \, \|\pi_K z\|_1\right\}. \tag{42}$$

It suffices to determine all the vectors y such that G is the face of $B_{p,k}^{\top\star}$ exposed by y. First observe that a subset $K^\sharp$ of $[\![1, d]\!]$ belongs to $\arg\max_{|K| \leq k} \|\pi_K z\|_1$ if and only if

$$L_k(z) \subset K^\sharp \subset \overline{L}_k(z) \tag{43}$$

and either $m_k(z)$ is equal to 0 or $K^\sharp$ has exactly k elements. In particular by Theorem 12,

$$G = F_\perp\left(B_{p,k}^{\top\star}, \pi_{\overline{L}_k(z)} z\right) , \tag{44}$$

and we can assume, without loss of generality, that z coincides with $\pi_{K^\sharp} z$. We will treat two separate cases depending on whether $m_k(z)$ is equal to 0 or not. First, if $m_k(z)$ is equal to 0, then $\overline{L}_k(z)$ is equal to $[\![1, d]\!]$ and (41) just states that the normal cone of $B_{p,k}^{\top\star}$ at G is the half-line spanned by z. However, in that case, Theorem 12 states that G is equal to $\{v_p(z)\}$. Hence if y is another nonzero vector such that G is the face of $B_{p,k}^{\top\star}$ exposed by y, the points $v_p(z)$ and $v_p(y)$ must coincide, which implies that x and y are multiples of one another by a positive factor and that the normal cone of $B_{p,k}^{\top\star}$ at G is the half-line spanned by z, as desired.

Now assume that $m_k(z)$ is not equal to 0 and consider a nonzero vector y such that G is the face of $B_{p,k}^{\top\star}$ exposed by y. It follows from Theorem 12 that

$$G = \overline{\text{co}}\left\{v_p(\pi_{K^\sharp} y) : K^\sharp \in \underset{|K| \leq k}{\arg\max} \, \|\pi_K y\|_1\right\} . \tag{45}$$

Consider a subset $K^\sharp$ of $[\![1, d]\!]$ such that

$$K^\sharp \in \underset{|K| \leq k}{\arg\max} \, \|\pi_K z\|_1 \ .$$

Since $m_k(z)$ is not equal to 0, $K^\sharp$ contains exactly k elements and $|z_i|$ is nonzero when i belongs to $K^\sharp$. By construction, a coordinate of $v_p(\pi_{K^\sharp} z)$ is nonzero if and only if the corresponding coordinate of $\pi_{K^\sharp} z$ is nonzero. As a consequence, according to (42), $v_p(\pi_{K^\sharp} z)$ is the unique vertex of G contained in $\mathcal{R}_{K^\sharp}$. It then follows from (45) that

$$\underset{|K|\leq k}{\arg\max}\,\|\pi_K y\|_1 = \underset{|K|\leq k}{\arg\max}\,\|\pi_K z\|_1 \ , \tag{46}$$

and that, for every set $K^\sharp$ contained in $\arg\max_{|K|\leq k} \|\pi_K z\|_1$,

$$v_p(\pi_{K^\sharp} y) = v_p(\pi_{K^\sharp} z) \ . \tag{47}$$

According to Proposition 13 and to (46), $\overline{L}_k(z)$ and $\overline{L}_k(y)$ coincide. Now recall that the normal cones of B_p at its proper faces are half lines incident to the origin of $\mathbb{R}^d$. Hence, according to (47), there exists a positive number $\alpha_{K^\sharp}$ such that

$$\pi_{K^\sharp} y = \alpha_{K^\sharp} \pi_{K^\sharp} z \ .$$

It remains to show that the value of $\alpha_{K^\sharp}$ does not depend on $K^\sharp$. Indeed, this will imply that up to a positive multiplicative factor, $\pi_{\overline{L}_k(z)} y$ coincides with z and results in the desired form (41) for the normal cone of $B_{p,k}^{\top\star}$ at G. If $L_k(z)$ is nonempty, this is immediate. Indeed,

$$\alpha_{K^\sharp} = \frac{y_i}{z_i} \ ,$$

for any element i of $L_k(z)$ and as a consequence, $\alpha_{K^\sharp}$ does not depend on $K^\sharp$. Therefore, assume that $L_k(z)$ is empty. In that case, z_i is equal to $m_k(z)$ when i belongs to $\overline{L}_k(z)$, and equal to 0 otherwise. Moreover, the sets $K^\sharp$ are precisely the size k subsets of $\overline{L}_k(z)$. If k is equal to 1, these sets are all the singletons from $\overline{L}_k(z)$ and this implies that $L_k(y)$ is also empty. Therefore y_i is equal to $m_k(y)$ when i belongs to $\overline{L}_k(y)$ which implies that $\alpha_{K^\sharp}$ does not depend on $K^\sharp$.

Finally assume that k is at least 2 and observe that, for any two nondisjoint size k subsets $K^\sharp$ and $\widetilde{K}^\sharp$ of $\overline{L}_k(z)$, the values of $\alpha_{K^\sharp}$ and $\alpha_{\widetilde{K}^\sharp}$ necessarily coincide. As k is at least 2, the graph whose vertices are the size k subsets of $\overline{L}_k(z)$ and whose edges connect two of them when they are nondisjoint is connected, it follows that $\alpha_{K^\sharp}$ does not depend on $K^\sharp$. $\square$

By analogy with the polytopal case, the normal fan of $B_{p,k}^{\top\star}$ refers to the set $\mathcal{N}(B_{p,k}^{\top\star})$ of its normal cones. It is a consequence of Theorem 14 that the normal cones, and therefore the normal fan of $B_{p,k}^{\top\star}$ do not depend on p when $1 < p < +\infty$. Interestingly, B_p has the same property: its normal cones are $\{0\}$ and the half-lines incident to the origin independently on p when $1 < p < +\infty$. We show that $\mathcal{N}(B_{p,k}^{\top\star})$ refines $\mathcal{N}(B_{\infty,k}^{\top\star})$ in the sense of [29, Section 7].

Corollary 15 *Every cone from $\mathcal{N}(B_{p,k}^{\top\star})$ is contained in a cone from $\mathcal{N}(B_{\infty,k}^{\top\star})$.*

Proof. Since the normal cones of $B_{p,k}^{\top\star}$ do not depend on p, it suffices to prove the statement when p is equal to 2. Consider a cone C in $\mathcal{N}(B_{2,k}^{\top\star})$. If C is empty or equal to $\{0\}$, then this is immediate. Assume that C is the normal cone of $B_{2,k}^{\top\star}$ at an exposed face which we will denote by F. According to Theorem 14, there exists a nonzero vector z in $\mathbb{R}^d$ that coincides with $\pi_{\overline{L}_k(z)} z$ such that

$$C = \overline{\mathrm{cone}}\big\{y \in \mathbb{R}^d : \pi_{\overline{L}_k(z)} y = z, \overline{L}_k(y) = \overline{L}_k(z)\big\} \ . \tag{48}$$

Observe that if $m_k(z)$ is equal to 0, then $\overline{L}_k(z)$ is equal to $[\![1, d]\!]$. In that case, C is the half-line spanned by z and it is contained in a normal cone of $B^{\top\star}_{\infty,k}$. Therefore, we shall assume from now on that $m_k(z)$ is positive. Consider a size k set $K^\sharp$ such that $L_k(z) \subset K^\sharp \subset \overline{L}_k(z)$. Since $m_k(z)$ is positive, z_i is nonzero when i belongs to $K^\sharp$. Denote by y the vector such that

$$y_i = \begin{cases} -1 & \text{if } z_i < 0 \text{ and } i \in K^\sharp \ , \\ 1 & \text{if } z_i > 0 \text{ and } i \in K^\sharp \ , \\ 0 & \text{if } i \notin K^\sharp \ . \end{cases} \tag{49}$$

By construction, y has exactly k nonzero coordinates. Denote

$$F = \mathrm{co}\left(F_\perp(\beta_d, y) \cup \frac{1}{k} F_\perp(\gamma_d, y) \right) \ . \tag{50}$$

According to Theorem 10, F is a facet of $B^\top_{1,k}$. We will show that C is contained in the cone spanned by F. By polarity, this cone belongs to $\mathcal{N}(B^{\top\star}_{\infty,k})$ and this will therefore prove the corollary. Consider a vector x such that $\pi_{\overline{L}_k(z)} x$ is equal to z and $\overline{L}_k(x)$ to $\overline{L}_k(z)$. It suffices to show that x belongs to the cone spanned by F. Indeed, since that cone is closed, it follows from (48) that C must be contained in it. Observe that x can be decomposed as

$$x = \pi_{K^\sharp} x + \pi_{[\![1,d]\!] \setminus K^\sharp} x \ . \tag{51}$$

By construction, x_i is equal to $m_k(z) y_i$ when i belongs to $K^\sharp \backslash L_k(z)$. Hence,

$$\pi_{K^\sharp} x = \pi_{L_k(z)} x - m_k(z) \pi_{L_k(z)} y + m_k(z) y \ . \tag{52}$$

Since $\pi_{\overline{L}_k(z)} x$ is equal to z and $L_k(z)$ is a subset of $\overline{L}_k(z)$, the first term in the right-hand side is equal to $\pi_{L_k(z)} z$. As a consequence x can be rewritten into $x' + x''$ where

$$\left\{ \begin{array}{l} x' = \pi_{L_k(z)} z - m_k(z) \pi_{L_k(z)} y \ , \\ x'' = m_k(z) y + \pi_{[\![1,d]\!] \setminus K^\sharp} x \ . \end{array} \right. \tag{53}$$

Observe that x' is contained in the cone spanned by $F_\perp(\beta_d, y)$ and x'' in the cone spanned by $F_\perp(\gamma_d, y)$. Hence, $x' + x''$ is contained in the cone spanned by F, as desired. □

We recall that the *hypersimplex* $\delta_{d,k}$ is the convex hull of the vertices of $[0, 1]^d$ that have exactly k nonzero coordinates. These polytopes appear in algebraic combinatorics [12, 13] and in convex geometry [9, 23, 24]. By extension, we call *hypersimplex* any polytope that coincides up to a bijective affine transformation, with the convex hull of such a subset of vertices of the hypercube. It is observed in [9] that $B^{\top\star}_{\infty,k}$ can be decomposed into a union of hypersimplices with pairwise disjoint interiors and that the proper faces of $B^{\top\star}_{\infty,k}$ are either hypersimplices or isometric copies of $B^{\top\star}_{\infty,k}$, but for another ambient dimension than d. The situation for $B^{\top\star}_{p,k}$ is different as all of its proper faces are hypersimplices.

Corollary 16 *If $1 < p < +\infty$, then all the proper faces of $B^{\top\star}_{p,k}$ are hypersimplices.*

Proof. Consider a proper face F of $B^{\top\star}_{p,k}$. Let us first assume that F is exposed by a vector y. If $m_k(y)$ is equal to 0, then for every subset $K^\sharp$ of $[\![1,d]\!]$ that belongs to $\arg\max_{|K|\leq k} \|\pi_K y\|_p$,

$$\pi_{K^\sharp} y = y \tag{54}$$

It then follows from Theorem 12 that F is the convex hull of a single point and therefore a 0-dimensional hypersimplex. Let us now assume that $m_k(y)$ is positive.

Denote $|\overline{L}_k(y) \backslash L_k(y)|$ by n and let us identify $\mathbb{R}^n$ with the subspace of $\mathbb{R}^d$ made up of the points x such that x_i is equal to 0 when i belongs to either $L_k(y)$ or $[\![1,d]\!] \backslash \overline{L}_k(y)$. Denote by P the orthogonal projection of F on $\mathbb{R}^n$. It follows from Theorem 12 that $P/m_k(y)$ is, up to flipping the signs of the coordinates, the convex hull of the points in $\{0,1\}^d$ with exactly $k - |L_k(y)|$ nonzero coordinates, as desired.

Now assume that F is not an exposed face of $B^{\top\star}_{p,k}$. In that case, F is a proper face of an exposed face of $B^{\top\star}_{p,k}$. Hence by the above, F is proper face of a hypersimplex. As all the proper faces of a hypersimplex are lower-dimensional hypersimplices, this completes the proof. □

5 Conclusion

Our original motivation was to identify a class of norms which, when added as a penalty term, promote sparsity in an optimization problem, *but within a given sparsity budget* k. For this purpose, we have studied in Sect. 2 the exposed faces of closed convex sets generated by k-sparse vectors, hence whose extreme points are k-sparse. We also have deduced support identification from dual information. Thus equipped, we have focused in Sect. 3 on the faces of the unit balls of so-called generalized k-support norms, constructed from k-sparse vectors and from a source norm. In the cases of orthant-monotonic and orthant-strictly monotonic source norms, we have obtained a characterization of the intersection of the k-sparse vectors with the faces of the k-support norm. Theorem 6 makes the link with our original motivation: we have provided dual conditions under which the primal optimal solution of a minimization problem, penalized by a k-support norm, is k-sparse. Going back to the original work of Tibshirani [27], the intuition — behind proposing least-square regression with an ℓ_1-norm penalty to achieve sparsity — is that the kinks of the ℓ_1-unit ball are located at sparse points (see Figure 1). In Sect. 4, we have gone on in that direction by providing geometric descriptions of the face and cone lattices of the unit balls of top-(q,k) norm and (p,k)-support norms. By contrast with the ℓ_1-unit ball, faces intersect in a subtle way, mixing kinks and smoother parts, as illustrated by Figure 5 (left). So, guided by sparsity, we have moved in this paper from optimization to the geometry of unit balls.

Acknowledgements. The third author is partially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant program number RGPIN-2020-06846.

References

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

[2] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Optimization with sparsity-inducing penalties. *Foundations and Trends® in Machine Learning*, 4(1):1–106, January 2012.

[3] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer-Verlag, New York, second edition, 2017.

[4] C. Boyer, A. Chambolle, Y. De Castro, V. Duval, F. de Gournay, and P. Weiss. On representer theorems and convex regularization. *SIAM Journal on Optimization*, 29(2):1260–1281, 2019.

[5] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique*, 346(9):589–592, 2008.

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

[7] J.-P. Chancelier and M. De Lara. Orthant-strictly monotonic norms, generalized top-k and k-support norms and the ℓ_0 pseudonorm. *Journal of Convex Analysis*, 30(3):743–769, 2023.

[8] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, oct 2012.

[9] A. Deza, J.-B. Hiriart-Urruty, and L. Pournin. Polytopal balls arising in optimization. *Contributions to Discrete Mathematics*, 16(3):125–138, 2021.

[10] J. Fadili, J. Malick, and G. Peyré. Sensitivity analysis for mirror-stratifiable convex functions. *SIAM Journal on Optimization*, 28(4):2975–3000, 2018.

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

[12] A. M. Gabriélov, I. M. Gel'fand, and M. V. Losik. Combinatorial computation of characteristic classes. *Functional Analysis and its Applications*, 9:103–115, 1975.

[13] I. M. Gel'fand, M. Goresky, R. D. MacPherson, and V. Serganova. Combinatorial geometries, convex polyhedra and Schubert cells. *Advances in Mathematics*, 63:301–316, 1987.

[14] J.-Y. Gotoh, A. Takeda, and K. Tono. DC formulations and algorithms for sparse optimization problems. *Math. Program.*, 169(1):141–176, may 2018.

[15] J.-Y. Gotoh and S. Uryasev. Two pairs of families of polyhedral norms versus ℓ_p -norms: proximity and applications in optimization. *Math. Program., Ser. A*, 156:391–431, 2016.

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

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

[18] R. Jenatton, J.-Y. Audibert, and F. Bach. Structured variable selection with sparsity-inducing norms. *Journal of Machine Learning Research*, 12(84):2777–2824, 2011.

[19] A. Le Franc, J.-P. Chancelier, and M. De Lara. The capra-subdifferential of the ℓ_0 pseudonorm. *Optimization*, 73(4):1229–1251, 2024.

[20] A. M. McDonald, M. Pontil, and D. Stamos. New perspectives on k-support and cluster norms. *Journal of Machine Learning Research*, 17(155):1–38, 2016.

[21] L. Mirsky. Symmetric gauge functions and unitarily invariant norms. *The Quarterly Journal of Mathematics*, 11(1):50–59, Jan. 1960.

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

[23] L. Pournin. Shallow sections of the hypercube. *Israel Journal of Mathematics*, 255:685–704, 2023.

[24] L. Pournin. Local extrema for hypercube sections. *Journal d'Analyse Mathématique*, 152:557–594, 2024.

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

[26] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*. Cambridge University Press, second edition, 2014.

[27] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[28] B. Wu, C. Ding, D. Sun, and K.-C. Toh. On the Moreau-Yosida regularization of the vector k-norm related functions. *SIAM Journal on Optimization*, 24(2):766–794, 2014.

[29] G. M. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer, 1995.