Lower Bound Convex Programs for Exact Sparse Optimization

Jean-Philippe Chancelier and Michel De Lara,
CERMICS, École des Ponts ParisTech

February 11, 2019

Abstract

In exact sparse optimization problems on $\mathbb{R}^d$ (also known as sparsity constrained problems), one looks for solution that have few nonzero components. In this paper, we consider problems where sparsity is exactly measured either by the nonconvex l_0 pseudonorm (and not by substitute penalty terms) or by the belonging of the solution to a finite union of subsets. Due to the combinatorial nature of the sparsity constraint, such problems do not generally display convexity properties, even if the criterion to minimize is convex. In the most common approach to tackle them, one replaces the sparsity constraint by a convex penalty term, supposed to induce sparsity. Thus doing, one loses the original exact sparse optimization problem, but gains convexity. However, by doing so, it is not clear that one obtains a lower bound of the original exact sparse optimization problem. In this paper, we propose another approach, where we lose convexity but where we gain at keeping the original exact sparse optimization formulation, by displaying lower bound convex minimization programs. For this purpose, we introduce suitable conjugacies, induced by a novel class of one-sided linear couplings. Thus equipped, we present a systematic way to design norms and lower bound convex minimization programs over their unit ball. The family of norms that we display encompasses most of the sparsity inducing norms used in machine learning. Therefore, our approach provides foundation and interpretation for their use.

Key words: sparse optimization, l_0 pseudonorm, sparsity inducing norm, machine learning, Fenchel-Moreau conjugacy.

1 Introduction

In exact sparse optimization problems on $\mathbb{R}^d$ (also known as sparsity constrained problems), one looks for solution that have few nonzero components. The *counting function*, also called *cardinality function* or l_0 *pseudonorm*, counts the number of nonzero components of a vector in $\mathbb{R}^d$. It is well-known that the l_0 pseudonorm is lower semi continuous but is not convex.

As a consequence, a minimization problem under the constraint that the l_0 pseudonorm is less than a given integer is not convex in general. Then, it is common practice to *replace* the nonconvex sparsity constraint *by substitute (convex) penalty terms*, supposed to induce sparsity. By doing so, on the one hand, one gains convexity and benefits of duality tools with the Fenchel conjugacy. However, on the other hand, it is not clear that one obtains a lower bound of the original exact sparse optimization problem.

In this paper, we consider exact sparse optimization problems, that is, problems with combinatorial sparsity constraint. More precisely, we focus on problems where *sparsity is exactly measured* either by the nonconvex l_0 pseudonorm (and not by substitute penalty terms) or by the belonging of the solution to a finite union of given subsets. Our main contribution is to provide a systematic way to design norms, and associated convex programs that are lower bounds for the original exact sparse optimization problem.

The paper is organized as follows. In Sect. 2, we recall the definition and properties of so-called one-sided linear couplings, introduced in the companion paper [4], and we show how to use them to obtain concave maximization/convex minimization problems that are lower bounds of a given optimization problem. In Sect. 3, we consider minimization problems under the constraint that the l_0 pseudonorm is less than a given integer. To provide a lower bound, we make use of a suitable conjugacy (not the Fenchel one) induced by the so-called coupling Caprac, introduced in [4]. We obtain a concave maximization program as lower bound and, under a mild assumption, it coincides with a convex minimization program on the unit ball of the so-called k-support norm. In Sect. 4, we consider generalized exact sparse optimization problems. These are minimization problems under the constraint that the solution belongs to a finite union of given subsets. We present a systematic way to design norms and lower bound convex minimization programs over their unit ball.

2 One-sided linear couplings and lower bound convex programs

In §2.1, we recall the definition and properties of *one-sided linear couplings*. In §2.2, we show how to use them to obtain concave maximization/convex minimization problems that are lower bounds of a given optimization problem.

2.1 One-sided linear couplings and conjugacies

The material here is mostly taken from [4]. Basic recalls and notations used in analysis can be found in §A.2.

2.1.1 Background on couplings and conjugacies

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

The general case

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

$$c : \mathbb{X} \times \mathbb{Y} \to \overline{\mathbb{R}} \,. \tag{1}$$

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

Definition 1 *The* c-Fenchel-Moreau conjugate *of a function* $f : \mathbb{X} \to \overline{\mathbb{R}}$*, with respect to the coupling* c*, is the function* $f^c : \mathbb{Y} \to \overline{\mathbb{R}}$ *defined by*

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

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

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

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

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

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

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

For any coupling c,

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

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

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

$$\sup_{y \in \mathbb{Y}} \Big(\big(-f^c(y) \big) \mathbin{\dot{+}} \big(-h^{-c}(y) \big) \Big) \leq \inf_{x \in \mathbb{X}} \Big(f(x) \mathbin{\dot{+}} h(x) \Big) \,, \tag{6b}$$

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

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

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

$$\sup_{y \in \mathbb{Y}} \Big(\big(-f^c(y) \big) \mathbin{\dot{+}} \big(-\delta_X^{-c}(y) \big) \Big) \leq \inf_{x \in \mathbb{X}} \Big(f(x) \mathbin{\dot{+}} \delta_X(x) \Big) = \inf_{x \in X} f(x) \,. \tag{6d}$$

The Fenchel conjugacy

When the sets $\mathbb{X}$ and $\mathbb{Y}$ are vector spaces equipped with a bilinear form $\langle\ ,\ \rangle$, the corresponding conjugacy is the classical *Fenchel conjugacy*. For any functions $f : \mathbb{X} \to \overline{\mathbb{R}}$ and $g : \mathbb{Y} \to \overline{\mathbb{R}}$, we denote

$$f^\star(y) = \sup_{x\in\mathbb{X}} \Big(\langle x\,,y\rangle \dot{+} \big(-f(x)\big)\Big)\ ,\ \ \forall y \in \mathbb{Y}\ , \tag{7a}$$

$$g^{\star'}(x) = \sup_{y\in\mathbb{Y}} \Big(\langle x\,,y\rangle \dot{+} \big(-g(y)\big)\Big)\ ,\ \ \forall x \in \mathbb{X} \tag{7b}$$

$$f^{\star\star'}(x) = \sup_{y\in\mathbb{Y}} \Big(\langle x\,,y\rangle \dot{+} \big(-f^\star(y)\big)\Big)\ ,\ \ \forall x \in \mathbb{X}\ . \tag{7c}$$

Due to the presence of the coupling $(-c)$ in the Inequality (6b), we also introduce[1]

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

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

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

When the two vector spaces $\mathbb{X}$ and $\mathbb{Y}$ are *paired* in the sense of convex analysis[2], Fenchel conjugates are convex *lower semi continuous (lsc)* functions, and their opposites are concave *upper semi continuous (usc)* functions.

2.1.2 One-sided linear couplings and conjugacies

In the companion paper [4], we have introduced and studied a novel family of couplings defined as follows.

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

$$\big(\theta \rhd h\big)(x) = \inf\big\{h(w)\ \big|\ w \in \mathbb{W}\ ,\ \ \theta(w) = x\big\}\ ,\ \ \forall x \in \mathbb{X}\ , \tag{9}$$

with the convention that $\inf \emptyset = +\infty$ (and with the consequence that $\theta : \mathbb{W} \to \mathbb{X}$ need not be defined on all $\mathbb{W}$, but only on the effective domain $\mathrm{dom} h = \{w \in \mathbb{W}\ |\ h(w) < +\infty\}$ of the function $h : \mathbb{W} \to \overline{\mathbb{R}}$).

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

$$c_\theta : \mathbb{W} \times \mathbb{Y} \to \overline{\mathbb{R}}\ ,\ \ c_\theta(w,y) = \langle \theta(w)\,,y\rangle\ \ ,\ \ \forall w \in \mathbb{W}\ ,\ \ \forall y \in \mathbb{Y}\ . \tag{10}$$

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

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

Here are expressions for the c_θ-conjugates in function of the Fenchel conjugate. The proof of Proposition 3 can be found in [4].

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

$$g^{c'_\theta} = g^\star \circ \theta \ . \tag{11a}$$

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

$$h^{c_\theta} = \big(\theta \rhd h\big)^\star \ , \tag{11b}$$

and the c_θ-Fenchel-Moreau biconjugate is given by

$$h^{c_\theta c_\theta{}'} = \big(h^{c_\theta}\big)^\star \circ \theta = \big(\theta \rhd h\big)^{\star\star'} \circ \theta \ . \tag{11c}$$

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

$$\delta_W^{-c_\theta} = \sigma_{-\theta(W)} \ . \tag{11d}$$

2.2 Lower bound convex programs

To illustrate how we can obtain lower bounds with one-sided linear couplings, we start with general problems of the form

$$\inf_{w \in W} h(w) \ , \tag{12}$$

where $h : \mathbb{W} \to \overline{\mathbb{R}}$ and $W \subset \mathbb{W}$ (we can always replace the subset W by $\mathrm{dom} h \cap W$).

Proposition 4 *Let $\mathbb{X}$ and $\mathbb{Y}$ be two vector spaces equipped with a bilinear form $\langle \, , \rangle$. Let $\mathbb{W}$ be a set. For any function $h : \mathbb{W} \to \overline{\mathbb{R}}$, nonempty set $W \subset \mathbb{W}$ and mapping $\theta : W \to \mathbb{X}$, we have the following lower bound*

$$\sup_{y \in \mathbb{Y}} \Big(\big(- \big(\theta \rhd h\big)^\star(y) \big) \mathbin{\dot{+}} \big(- \sigma_{-\theta(W)}(y) \big) \Big) \leq \inf_{w \in W} h(w) \ . \tag{13a}$$

Proof. As $\inf_{w \in W} h(w) = \inf_{w \in \mathbb{W}} \big(h(w) \mathbin{\dot{+}} \delta_W \big)$, it suffices to use the Inequality (6b), with the c_θ-Fenchel-Moreau conjugate of the function h given by (11b) and the $(-c_\theta)$-Fenchel-Moreau conjugate of the characteristic function δ_W given by (11d). □

When $\mathbb{X}$ and $\mathbb{Y}$ are two paired vector spaces, the dual problem to the left hand side of (13a) consists in the maximization of a usc concave function.

When $\mathbb{X}$ and $\mathbb{Y}$ for a dual system (see §A.2.2), recall that a set $X \subset \mathbb{X}$ is said to be *weakly bounded* if $\sup_{x \in X} \langle x \, , y \rangle < +\infty$ for all $y \in \mathbb{Y}$ (see (52) in Definition 17). Now, when the primal and dual spaces in Proposition 4 are a Hilbert space, we provide conditions for the lower bound, to the left of (13a), to display an alternative primal expression as the minimization of a lsc convex function on a weakly bounded and closed convex set.

Corollary 5 *Let $\mathbb{X} = \mathbb{Y}$ be a Hilbert space. Let $\mathbb{W}$ be a set. Let $h : \mathbb{W} \to \overline{\mathbb{R}}$ be a function, $W \subset \mathbb{W}$ be a nonempty set and $\theta : W \to \mathbb{X}$ be a mapping. If*

1. *the set $-\theta(W)$ is weakly bounded, that is, the barrier cone of $-\theta(W)$ in (51) is the full space, namely $\mathrm{bar}\big(-\theta(W)\big) = \mathbb{Y}$,*
2. *the convex lsc function $\big(\theta \rhd h\big)^\star : \mathbb{Y} \to \overline{\mathbb{R}}$ is proper,*

then the lower bound, to the left of (13a), has the alternative primal expression

$$\sup_{y\in\mathbb{Y}} \Big(\big(-\big(\theta \rhd h\big)^\star(y)\big) \dot{+} \big(-\sigma_{-\theta(W)}(y)\big)\Big) = \min_{x\in\overline{\mathrm{co}}(-\theta(W))} \big(\theta \rhd h\big)^{\star\star'}(x) \leq \inf_{w\in W} h(w)\ , \quad (13b)$$

where the primal problem to the left consists in the minimization of a lsc convex function on a weakly bounded and closed convex set.

Proof. We consider the Inequality (13a). On the one hand, the convex lsc function $\big(n \rhd f\big)^\star$ is proper by assumption. On the other hand, the support function $\sigma_{-\theta(W)}$ is convex lsc, and also proper. Indeed, $-\infty < \sigma_{-\theta(W)}$ since $\theta(W) \neq \emptyset$ by assumption, and $\mathrm{dom}\sigma_{-\theta(W)} = \mathbb{Y}$ since $\mathrm{bar}\big(-\theta(W)\big) = \mathbb{Y}$ by assumption. As a consequence, the support function $\sigma_{-\theta(W)}$ has for effective domain the full space $\mathbb{Y}$, hence its continuity points are $\mathrm{cont}\big(\sigma_{-\theta(W)}\big) = \mathbb{Y}$. As $\mathrm{dom}\big((n \rhd f)^\star\big) \neq \emptyset$, we deduce that $\mathrm{cont}\big(\sigma_{-\theta(W)}\big) \cap \mathrm{dom}\big((n \rhd f)^\star\big) = \mathrm{dom}\big((n \rhd f)^\star\big) \neq \emptyset$.

Thus, the conditions for a Fenchel-Rockafellar equality are satisfied [3, Prop. 15.13] and we obtain that

$$\sup_{y\in\mathbb{Y}} \Big(\big(-\big(\theta \rhd h\big)^\star(y)\big) \dot{+} \big(-\sigma_{-\theta(W)}(y)\big)\Big) = \min_{x\in\mathbb{X}} \Big(\big(\theta \rhd h\big)^{\star\star'}(x) \dot{+} \delta_{\overline{\mathrm{co}}(-\theta(W))}\Big) = \min_{x\in\overline{\mathrm{co}}(-\theta(W))} \big(\theta \rhd h\big)^{\star\star'}(x)\ .$$

The set $\overline{\mathrm{co}}(-\theta(W))$ is closed convex by definition, and is weakly bounded as the finite union of weakly bounded sets by (56).

This ends the proof. □

3 Lower bound convex programs for exact sparse optimization

In this section, we consider minimization problems under the constraint that the l_0 pseudonorm is less than a given integer. In §3.1, we introduce and recall the main properties of the so-called coupling Caprac [4]. Then, in §3.2, we show how to obtain lower bounds for exact sparse optimization problems.

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

3.1 Constant along primal rays coupling (Caprac) and conjugacy

To provide lower bounds, we make use of a suitable conjugacy (not the Fenchel one) induced by a novel coupling Caprac. This coupling has the property of being constant along primal rays, like the l_0 pseudonorm. The material here is mostly taken from the companion paper [4].

3.1.1 Constant along primal rays coupling and conjugacy

In [4], we have introduced and studied a novel coupling, defined as follows on the Euclidian space $\mathbb{R}^d$.

Definition 6 *We define the* (Euclidian) coupling Caprac ç *between* $\mathbb{R}^d$ *and* $\mathbb{R}^d$ *by*

$$\forall y \in \mathbb{R}^d \ , \ \begin{cases} ç(x,y) & = \dfrac{\langle x \, , y \rangle}{\|x\|} \ , \ \forall x \in \mathbb{R}^d \backslash \{0\} \ , \\ ç(0,y) & = 0. \end{cases} \tag{14}$$

The coupling Caprac has the property of being *constant along primal rays*, hence the acronym Caprac. We introduce the *Euclidian unit sphere*

$$S = \left\{ x \in \mathbb{R}^d \mid \|x\| = 1 \right\} \ , \tag{15}$$

and the *normalization mapping* n

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

With these notations, the Caprac coupling (14) is a special case of one-sided linear coupling c_n, as in (10) with $\theta = n$, the Fenchel coupling after primal normalization:

$$ç(x,y) = c_n(x,y) = \langle n(x) \, , y \rangle \ , \ \forall x \in \mathbb{R}^d \ , \ \forall y \in \mathbb{R}^d \ .$$

Here are expressions for the Caprac conjugates in function of the Fenchel conjugate.

Proposition 7 *For any function* $g : \mathbb{R}^d \to \overline{\mathbb{R}}$*, the* ç′*-Fenchel-Moreau conjugate is given by*

$$g^{ç'} = g^\star \circ n \ . \tag{16a}$$

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

$$f^{ç} = \big(n \rhd f \big)^\star \ , \tag{16b}$$

where the infimal postcomposition (9) *has the expression*

$$\big(n \rhd f \big)(x) = \inf \left\{ f(x') \mid n(x') = x \right\} = \begin{cases} \inf_{\lambda > 0} f(\lambda x) & \text{if } x \in S \cup \{0\} \\ +\infty & \text{if } x \notin S \cup \{0\} \end{cases} \tag{16c}$$

and the ç*-Fenchel-Moreau biconjugate is given by*

$$f^{ç ç'} = \big(f^{ç} \big)^\star \circ n = \big(n \rhd f \big)^{\star \star'} \circ n \ . \tag{16d}$$

3.1.2 Caprac conjugates and biconjugates related to the l_0 pseudonorm

Now, we will give formulas for conjugates and biconjugates of functions related to the l_0 pseudonorm.

First, we recall the definitions of the 2-k-symmetric gauge norm and the k-support norm. For any $x \in \mathbb{R}^d$ and $K \subset \{1, \ldots, d\}$, we denote by $x_K \in \mathbb{R}^d$ the vector which coincides with x, except for the components outside of K that vanish: x_K is the orthogonal projection of x onto the subspace $\mathbb{R}^K \times \{0\}^{-K} \subset \mathbb{R}^d$. Here, following notation from Game Theory, we have denoted by $-K$ the complementary subset of K in $\{1, \ldots, d\}$: $K \cup (-K) = \{1, \ldots, d\}$ and $K \cap (-K) = \emptyset$. In what follows, $|K|$ denotes the cardinal of the set K and the notation $\sup_{|K| \leq k}$ is a shorthand for $\sup_{K \subset \{1,\ldots,d\}, |K| \leq k}$ (the same holds for $\sup_{|K|=k}$).

Definition 8 *Let $x \in \mathbb{R}^d$. For $k \in \{1, \ldots, d\}$, we denote by $\|x\|_{(k)}^{\mathrm{sgn}}$ the maximum of $\|x_K\|$ over all subsets $K \subset \{1, \ldots, d\}$ with cardinal (less than) k:*

$$\|x\|_{(k)}^{\mathrm{sgn}} = \sup_{|K| \leq k} \|x_K\| = \sup_{|K|=k} \|x_K\| \,. \tag{17}$$

Thus defined, $\|\cdot\|_{(k)}^{\mathrm{sgn}}$ is a norm, the 2-k-symmetric gauge norm *[5]. Its dual norm (see Proposition 26) is called* k-support norm [2]*, denoted by $\|\cdot\|_{(k)}^{\mathrm{sn}}$:*

$$\|\cdot\|_{(k)}^{\mathrm{sn}} = \left(\|\cdot\|_{(k)}^{\mathrm{sgn}}\right)_\star \,. \tag{18}$$

Second, we recall the definition of the l_0 pseudonorm. We define the *support* of a vector in $\mathbb{R}^d$ by

$$\mathrm{supp}(x) = \left\{ j \in \{1, \ldots, d\} \mid x_j > 0 \right\} \,, \quad \forall x \in \mathbb{R}^d \,.$$

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

$$\ell_0(x) = |\mathrm{supp}(x)| \,, \quad \forall x \in \mathbb{R}^d \,. \tag{19}$$

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

$$\ell_0^{\leq k} = \left\{ x \in \mathbb{R}^d \mid \ell_0(x) \leq k \right\} \,, \quad \forall k \in \{0, 1, \ldots, d\} \,. \tag{20}$$

Third, we present the main result of [4]. The l_0 pseudonorm in (19), the characteristic function $\delta_{\ell_0^{\leq k}}$ of its level set and the symmetric gauge norms in (17) are related by the following conjugate formulas.

Theorem 9 *Let $k \in \{0, 1, \dots, d\}$. We have that:*

$$\delta^{-\dot{\complement}}_{\ell_0^{\leq k}} = \delta^{\dot{\complement}}_{\ell_0^{\leq k}} = \|\cdot\|^{\mathrm{sgn}}_{(k)} \,, \tag{21a}$$

$$\delta^{\dot{\complement}\dot{\complement}'}_{\ell_0^{\leq k}} = \delta_{\ell_0^{\leq k}} \,, \tag{21b}$$

$$\ell_0^{\dot{\complement}} = \sup_{l=0,1,\dots,d} \Big[\|\cdot\|^{\mathrm{sgn}}_{(l)} - l \Big] \,, \tag{21c}$$

$$\ell_0^{\dot{\complement}\dot{\complement}'} = \ell_0 \,, \tag{21d}$$

with the convention, in (21a) *and in* (21c)*, that* $\|\cdot\|^{\mathrm{sgn}}_{(0)} = 0$.

3.2 Lower bound convex program for exact sparse optimization

With the Caprac-conjugacy recalled in §3.1, we now show how to obtain lower bounds for exact sparse optimization problems, that are concave maximization programs. In addition, under a mild assumption, we will show that this lower bound coincides with a convex minimization program on the unit ball of the k-support norm, recalled in Definition 8.

Theorem 10 *Let $k \in \{0, 1, \dots, d\}$. For any function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$, we have the following lower bound*

$$\sup_{y \in \mathbb{R}^d} \Big(- \big(n \rhd f\big)^\star(y) - \|y\|^{\mathrm{sgn}}_{(k)} \Big) \leq \inf_{\ell_0(x) \leq k} f(x) \,, \tag{22a}$$

where the dual problem to the left consists in the maximization of a usc concave function.

If, in addition, the convex lsc function $\big(n \rhd f\big)^\star$ is proper, the above lower bound has the alternative primal expression

$$\min_{\|x\|^{\mathrm{sn}}_{(k)} \leq 1} \big(n \rhd f\big)^{\star\star'}(x) = \sup_{y \in \mathbb{R}^d} \Big(- \big(n \rhd f\big)^\star(y) - \|y\|^{\mathrm{sgn}}_{(k)} \Big) \leq \inf_{\ell_0(x) \leq k} f(x) \,, \tag{22b}$$

where the primal problem to the left consists in the minimization of a lsc convex function on the unit ball of the k-support norm.

Proof. From the Inequality (6d), where we use the expression (16b) for $f^{\dot{\complement}}$ and the expression (21a) for $\delta^{-\dot{\complement}}_{\ell_0^{\leq k}}$, we deduce Inequality (22a). Because the norm $\|\cdot\|^{\mathrm{sgn}}_{(k)}$ is convex lsc and has full effective domain $\mathbb{R}^d$, and because the convex lsc function $\big(n \rhd f\big)^\star$ is proper, we deduce that $\mathrm{cont}(\|\cdot\|^{\mathrm{sgn}}_{(k)}) \cap \mathrm{dom}\big((n \rhd f)^\star\big) = \mathrm{dom}\big((n \rhd f)^\star\big) \neq \emptyset$. Thus, the conditions for a Fenchel-Rockafellar equality are satisfied [3, Prop. 15.13] and we obtain that

$$\min_{\|x\|^{\mathrm{sn}}_{(k)} \leq 1} \big(n \rhd f\big)^{\star\star'}(x) = \sup_{y \in \mathbb{R}^d} \Big(- \big(n \rhd f\big)^\star(y) - \|y\|^{\mathrm{sgn}}_{(k)} \Big) \,.$$

This equation, combined with Equation (22a), gives Equation (22b). This ends the proof. □

As an application, we consider the least squares regression sparse optimization problem.

Proposition 11 *Letting A be a matrix with d rows and p columns, and $z \in \mathbb{R}^p$, we have*

$$\|z\|^2 + \sup_{y\in\mathbb{R}^d} \Big(- \Big[\sup_{x\in S} \big(\langle x\,,y\rangle + \frac{\langle z\,,Ax\rangle^2}{\|Ax\|^2}\mathbb{I}_{\langle z\,,Ax\rangle>0}\big)\Big]_+ - \|y\|^{\mathrm{sgn}}_{(k)}\Big)$$
$$= \|z\|^2 + \min_{\|x\|^{\mathrm{sn}}_{(k)}\leq 1} \Big(- \frac{\langle z\,,A\,\cdot\rangle^2}{\|A\cdot\|^2}\mathbb{I}_{\langle z\,,A\,\cdot\rangle>0} \dot{+}\, \delta_S\Big)^{\star\star'}(x) \leq \inf_{\ell_0(x)\leq k} \|z-Ax\|^2\,. \quad (23)$$

Proof. Let f be the function defined by $f(x) = \|z - Ax\|^2$, for all $x \in \mathbb{R}^d$. A straightforward calculation gives

$$\inf_{\lambda>0} f(\lambda x) = \|z\|^2 - \frac{\langle z\,,Ax\rangle^2}{\|Ax\|^2}\mathbb{I}_{\langle z\,,Ax\rangle>0}\,,\ \ \forall x\in\mathbb{R}^d\,. \quad (24a)$$

Therefore, using (16c), we obtain that, for all $y \in \mathbb{R}^d$,

$$\big(n \rhd f\big)^\star(y) = \sup_{x\in S\cup\{0\}} \Big(\langle x\,,y\rangle - \inf_{\lambda>0} f(\lambda x)\Big) = \Big[\sup_{x\in S}\big(\langle x\,,y\rangle + \frac{\langle z\,,Ax\rangle^2}{\|Ax\|^2}\mathbb{I}_{\langle z\,,Ax\rangle>0}\big)\Big]_+ - \|z\|^2\,. \quad (24b)$$

Then, inserting the expression (24b) of $\big(n \rhd f\big)^\star$ in Inequality (22a) yields the first part of Equation (23), namely

$$\|z\|^2 + \sup_{y\in\mathbb{R}^d} \Big(- \Big[\sup_{x\in S} \big(\langle x\,,y\rangle + \frac{\langle z\,,Ax\rangle^2}{\|Ax\|^2}\mathbb{I}_{\langle z\,,Ax\rangle>0}\big)\Big]_+ - \|y\|^{\mathrm{sgn}}_{(k)}\Big) \leq \inf_{\ell_0(x)\leq k} \|z-Ax\|^2\,.$$

Now, since the function $\big(n \rhd f\big)^\star$ is easily seen to be proper by (24b), we can use Theorem 10 and Equation (22b) gives

$$\begin{aligned}
\min_{\|x\|^{\mathrm{sn}}_{(k)}\leq 1} \Big(\Big[\sup_{x\in S} \big(\langle x\,,y\rangle + \frac{\langle z\,,Ax\rangle^2}{\|Ax\|^2}\mathbb{I}_{\langle z\,,Ax\rangle>0}\big)\Big]_+\Big)^{\star'} - \|z\|^2 &= \min_{\|x\|^{\mathrm{sn}}_{(k)}\leq 1} \big(n\rhd f\big)^{\star\star'}(x) \qquad \text{(by (24b))} \\
&= \sup_{y\in\mathbb{R}^d} \Big(- \big(n\rhd f\big)^\star(y) - \|y\|^{\mathrm{sgn}}_{(k)}\Big) \\
&\leq \inf_{\ell_0(x)\leq k} f(x)\,, \qquad \text{(by (22b))}
\end{aligned}$$

which is the second part of Equation (23). This ends the proof. □

4 Lower bound convex programs for generalized sparse optimization

In §4.1 we formally define generalized sparse optimization (GSO) as optimization over the union of a finite family of subsets. Then, we provide lower bound convex minimization programs over the unit ball of some norms. In §4.2, we present a systematic way to design such norms, with results on how to build up a (global) norm from (local) sets or norms. Finally, in §4.3, we wrap up the results from the both previous subsections and provide our main result. Thus doing, we hint at how we encompass most of the sparsity inducing norms used in machine learning.

4.1 Lower bound convex programs for generalized sparse optimization

Let $\mathbb{W}$ be a set, let $\mathbb{J}$ be a finite set and let $\{W_j\}_{j\in\mathbb{J}}$ be a family of subsets of $\mathbb{W}$. This family captures *sparsity*, where the finite set $\mathbb{J}$ of indices reflects the combinatorial nature of the optimization problem.

For any function $h : \mathbb{W} \to \overline{\mathbb{R}}$, the *generalized sparse optimization (GSO)* problem is[3]

$$\inf_{w\in\bigcup_{j\in\mathbb{J}} W_j} h(w) \ . \tag{25}$$

As the problem (25) is a special case of (12) — with constraint given by the belonging of possible solutions to the set $W = \bigcup_{j\in\mathbb{J}} W_j$ — the following Proposition is a straightforward application of Corollary 5.

Proposition 12 *Let $\mathbb{X} = \mathbb{Y}$ be a Hilbert space. Let $\mathbb{J}$ be a finite set. Let $\mathbb{W}$ be a set.*

1. *Let $\{W_j\}_{j\in\mathbb{J}}$ be a family of subsets of $\mathbb{W}$.*
2. *Let $\{\theta_j\}_{j\in\mathbb{J}}$ be a family of mappings $\theta_j : W_j \to \mathbb{X}$ such that*
   (a) *the family $\{\theta_j\}_{j\in\mathbb{J}}$ is compatible with the family $\{W_j\}_{j\in\mathbb{J}}$, in the sense that $w \in W_j \cap W_{j'} \Rightarrow \theta_j(w) = \theta_{j'}(w)$, $\forall (j, j') \in \mathbb{J}^2$,*
   (b) *every set $-\theta_j(W_j)$ is weakly bounded, for every $j \in \mathbb{J}$.*
3. *Let $h : \mathbb{W} \to \overline{\mathbb{R}}$ be a function such that every function $\big(\theta_j \rhd h\big)^\star$ is proper, for every $j \in \mathbb{J}$, and $\bigcap_{j\in\mathbb{J}} \mathrm{dom}\big(\theta_j \rhd h\big)^\star \neq \emptyset$.*

Then, we have the lower bound

$$\min_{x\in\overline{\mathrm{co}}\left(-\bigcup_{j\in\mathbb{J}}\theta_j(W_j)\right)} \Big(\sup_{j\in\mathbb{J}} \big(\theta_j \rhd h\big)^\star\Big)^\star(x) \leq \inf_{w\in\bigcup_{j\in\mathbb{J}} W_j} h(w) \ . \tag{26}$$

Proof. By item 2a, we can define the mapping

$$\theta : \bigcup_{j\in\mathbb{J}} W_j \to \mathbb{X} \ \text{ by } \ w \in W_j \Rightarrow \theta(w) = \theta_j(w) \ .$$

By item 2b, as every set $-\theta_j(W_j)$ is weakly bounded, for every $j \in \mathbb{J}$, the finite union $\bigcup_{j\in\mathbb{J}} -\theta_j(W_j) = -\theta\big(\bigcup_{j\in\mathbb{J}} W_j\big)$ is weakly bounded, by item 2 in Proposition 18.

From the definition (9) of the infimal postcomposition, we get that

$$\big(\theta \rhd h\big)(x) = \inf\big\{h(w') \ \big| \ w' \in \mathbb{W}, \exists j \in \mathbb{J}, \theta_j(w') = x\big\} = \inf_{j\in\mathbb{J}} \big(\theta_j \rhd h\big)(x) \ .$$

[3]The function $h : \mathbb{W} \to \overline{\mathbb{R}}$ needs only be known on $\mathrm{dom}h \cap \Big(\bigcup_{j\in\mathbb{J}} W_j\Big)$.

Therefore, $\big(\theta \rhd h\big)^\star = \sup_{j\in\mathbb{J}} \big(\theta_j \rhd h\big)^\star$, as conjugacies, being dualities, turn infima into suprema. By item 3, the mapping $\big(\theta \rhd h\big)^\star$ is proper.

To conclude, we apply Corollary 5, with $W = \bigcup_{j\in\mathbb{J}} W_j$ and $\overline{\mathrm{co}}\big(-\theta(W)\big) = \overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} -\theta(W_j)\Big) = \overline{\mathrm{co}}\big(-\bigcup_{j\in\mathbb{J}} \theta_j(W_j)\big)$. □

Going on, we provide conditions under which the lower bound (26) is a convex minimization program over the unit ball of a norm (that will be detailed in §4.2).

Proposition 13 *Let $\mathbb{X} = \mathbb{Y}$ be a Hilbert space. Let $\mathbb{J}$ be a finite set. Let $\mathbb{W}$ be a set.*

1. *Let $\{W_j\}_{j\in\mathbb{J}}$ be a family of two by two disjoint subsets of $\mathbb{W}$.*
2. *Let $\{\theta_j\}_{j\in\mathbb{J}}$ be a family of mappings $\theta_j : W_j \to \mathbb{X}$, such that*
   (a) *the following joint full sum condition is satisfied*
   $$\sum_{j\in\mathbb{J}} \mathrm{span}\big(\theta_j(W_j)\big) = \mathbb{X}\,, \tag{27}$$
   (b) *every subset $\theta_j(W_j)$ of $\mathbb{X}$ is symmetric and weakly bounded, for every $j \in \mathbb{J}$.*
3. *Let $h : \mathbb{W} \to \overline{\mathbb{R}}$ be a function such that every function $\big(\theta_j \rhd h\big)^\star$ is proper, for every $j \in \mathbb{J}$, and $\bigcap_{j\in\mathbb{J}} \mathrm{dom}\big(\theta_j \rhd h\big)^\star \neq \emptyset$.*

Then, there exists a norm $|||\cdot|||$, with unit ball $\overline{\mathrm{co}}\big(-\bigcup_{j\in\mathbb{J}} \theta_j(W_j)\big)$, such that we have the lower bound
$$\min_{|||x|||\leq 1} \Big(\sup_{j\in\mathbb{J}} \big(\theta_j \rhd h\big)^\star\Big)^\star(x) \leq \inf_{w\in\bigcup_{j\in\mathbb{J}} W_j} h(w)\,. \tag{28}$$

Proof. First, we use Proposition 12 to obtain the lower bound (26). For this purpose, we check its three assumptions (item 2a, item 2b, item 3) one by one.

- Because the family $\{W_j\}_{j\in\mathbb{J}}$ is made of two by two disjoint subsets of $\mathbb{W}$, the compatibility condition of item 2a, in Proposition 12 is satisfied.
- As every subset $\theta_j(W_j)$ of $\mathbb{X}$ is symmetric and weakly bounded for every $j \in \mathbb{J}$, by item 2b here, we deduce that item 2b of Proposition 12 is satisfied.
- Item 3 of Proposition 12 coincides with item 3 here.

Therefore, we obtain the lower bound (26).

Second, there remains to prove that there exists a norm $|||\cdot|||$ with unit ball $\overline{\mathrm{co}}\big(-\bigcup_{j\in\mathbb{J}} \theta_j(W_j)\big)$. Now, this is a straightforward application of Theorem 14 below, with $V_j = \theta_j(W_j)$, for every $j \in \mathbb{J}$, and by item 2a of the second assumption of this Proposition.

This ends the proof. □

4.2 Building up a (global) norm from (local) norms

Proposition 13 claims the existence of a norm. Here, we show how we can obtain a global norm on a Hilbert space, first from subsets in Proposition 14, second from local norms defined on closed subspaces in Proposition 15.

Proposition 14 *Let $\mathbb{V}$ be a Hilbert space. Let $\mathbb{J}$ be a finite set.*

Let $\{V_j\}_{j\in\mathbb{J}}$ be a family of subsets of $\mathbb{V}$ that are all symmetric and weakly bounded, that is,

$$-V_j = V_j \ , \ \ \mathrm{bar}V_j = \mathbb{V} \ , \ \ \forall j \in \mathbb{J} \ , \tag{29}$$

and that jointly satisfy the full sum condition

$$\sum_{j\in\mathbb{J}} \mathrm{span}V_j = \mathbb{V} \ . \tag{30}$$

Then, there is a (unique) norm $|||\cdot|||$ on $\mathbb{V}$ with unit ball $\overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} V_j\Big)$. Moreover, the norm $|||\cdot|||$ admis a dual norm $|||\cdot|||_\star$ with unit ball $\Big(\bigcup_{j\in\mathbb{J}} V_j\Big)^\odot$. The norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are given by

$$|||\cdot||| = \sigma_{\left(\bigcup_{j\in\mathbb{J}} V_j\right)^\odot} \ \ \textit{and} \ \ |||\cdot|||_\star = \sigma_{\overline{\mathrm{co}}\left(\bigcup_{j\in\mathbb{J}} V_j\right)} \ , \tag{31a}$$

and their respective unit balls are

$$B_{|||\cdot|||} = \overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} V_j\Big) \ \ \textit{and} \ \ B_{|||\cdot|||_\star} = \big(\bigcup_{j\in\mathbb{J}} V_j\big)^\odot \ . \tag{31b}$$

The topologies defined by the norm $|||\cdot|||$ and by the dual norm $|||\cdot|||_\star$ are both weaker (contain less open sets) than the Hilbertian topology.

Proof. We prove that the closed convex set

$$B = \overline{\mathrm{co}}\big(\bigcup_{j\in\mathbb{J}} V_j\big) \tag{32}$$

satisfies the following conditions of Proposition 24, namely

$$-B = B \ , \ \ \mathrm{bar}B = \mathbb{V} \ , \ \ \mathrm{cone}B = \mathbb{V} \ .$$

It is clear that $-B = B$ since $-V_j = V_j$ for all $j \in \mathbb{J}$ by (29).

We show that $\mathrm{bar}B = \mathbb{V}$:

$$\begin{aligned}
\mathrm{bar}B &= \mathrm{bar}\Big(\overline{\mathrm{co}}\big(\bigcup_{j\in\mathbb{J}} V_j\big)\Big) && \text{(by (32))}\\
&= \mathrm{bar}\big(\bigcup_{j\in\mathbb{J}} V_j\big) && \text{(by (55a))}\\
&= \bigcap_{j\in\mathbb{J}} \mathrm{bar}V_j && \text{(by (56))}\\
&= \bigcap_{j\in\mathbb{J}} \mathbb{V} && \text{(by (29))}\\
&= \mathbb{V}\,.
\end{aligned}$$

There remains to show that $\mathrm{cone}B = \mathbb{V}$:

$$\begin{aligned}
\mathrm{cone}B &= \mathrm{span}B && \text{(by [3, Prop. 6.4] as } B \text{ is nonempty convex and symmetric)}\\
&= \mathrm{span}\Big(\overline{\mathrm{co}}\big(\bigcup_{j\in\mathbb{J}} V_j\big)\Big) && \text{(by (32))}\\
&\supset \mathrm{span}\Big(\mathrm{co}\big(\bigcup_{j\in\mathbb{J}} V_j\big)\Big)\\
&= \mathrm{span}\big(\bigcup_{j\in\mathbb{J}} V_j\big)\\
&= \sum_{j\in\mathbb{J}} \mathrm{span}V_j\\
&= \mathbb{V}\,. && \text{(by (30))}
\end{aligned}$$

We have proved that the closed convex set B in (32) satisfies the conditions of Proposition 24. We conclude that $\sigma_{B^\odot}$ is a norm $|||\cdot|||$ on $\mathbb{V}$ with unit ball B, and that it admits the dual norm $|||\cdot|||_\star = \sigma_B$, with unit ball $B^\odot$. This gives (31).

In addition, the topologies defined by the norm $|||\cdot|||$ and by the dual norm $|||\cdot|||_\star$ are both weaker (contain less open sets) than the Hilbertian topology, by Proposition 19 since both unit balls B and $B^\odot$ are closed by construction.

This ends the proof. □

Here, we show how we can obtain a global norm on a Hilbert space, from local norms defined on closed subspaces. With this formulation, we are able to give expressions of the global norm and of its dual norm as convolution and supremum of local norms and dual norms.

Proposition 15 *Let $\mathbb{V}$ be a Hilbert space. Let $\mathbb{J}$ be a finite set.*

- *Let $\{\mathbb{V}_j\}_{j\in\mathbb{J}}$ be a family of closed subspaces of the Hilbert space $\mathbb{V}$, will full sum, that is, such that*

$$\sum_{j\in\mathbb{J}} \mathbb{V}_j = \mathbb{V}\,. \tag{33}$$

- *Let $\{|||\cdot|||_j\}_{j\in\mathbb{J}}$ be a family of (local) norms on the closed subspaces $\{\mathbb{V}_j\}_{j\in\mathbb{J}}$, such that, for every $j\in\mathbb{J}$, the norm $|||\cdot|||_j$ is equivalent to the restriction to $\mathbb{V}_j$ of the Hilbertian norm $\|\cdot\|$ on $\mathbb{V}$. We define, for every $j\in\mathbb{J}$, the (local) unit ball*

$$B_j = \{v\in\mathbb{V}_j \mid |||v|||_j \leq 1\} \subset \mathbb{V}_j\ ,\ \ \forall j\in\mathbb{J}\ , \tag{34}$$

Then, there is a (unique) norm $|||\cdot|||$ on $\mathbb{V}$ with unit ball $\overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} B_j\Big)$ and it admis a dual norm $|||\cdot|||_\star$ with unit ball $\Big(\bigcup_{j\in\mathbb{J}} B_j\Big)^\odot$. Moreover, the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ are equivalent to the Hilbertian norm $\|\cdot\|$, and have the following expressions.

1. *The norm $|||\cdot|||$ can be expressed as a convolution of the local norms $\{|||\cdot|||_j\}_{j\in\mathbb{J}}$:*

$$|||\cdot||| = \square_{j\in\mathbb{J}} \big(|||\cdot|||_j + \delta_{\mathbb{V}_j}\big)\ , \tag{35a}$$

$$|||v||| = \inf_{v^j\in\mathbb{V}_j,\sum_{j\in\mathbb{J}} v^j = v} \sum_{j\in\mathbb{J}} |||v^j|||_j\ ,\ \ \forall v\in\mathbb{V}\ . \tag{35b}$$

2. *For each $j\in\mathbb{J}$, the local norm $|||\cdot|||_j$ on $\mathbb{V}_j$ admits a local dual norm $|||\cdot|||_{j,\star}$ on $\mathbb{V}_j$, and the dual norm $|||\cdot|||_\star$, of the norm $|||\cdot|||$ can be expressed as a supremum of the local dual norms $\{|||\cdot|||_{j,\star}\}_{j\in\mathbb{J}}$:*

$$|||\cdot|||_\star = \sup_{j\in\mathbb{J}} \big(|||\cdot|||_{j,\star}\circ\pi_j\big)\ , \tag{36a}$$

$$|||v'|||_\star = \sup_{j\in\mathbb{J}} |||\pi_j(v')|||_{j,\star}\ ,\ \ \forall v'\in\mathbb{V}\ , \tag{36b}$$

where, for every $j\in\mathbb{J}$, we introduce the orthogonal projection mapping onto the closed subspace $\mathbb{V}_j$

$$\pi_j : \mathbb{V}\to\mathbb{V}_j \quad \textit{such that} \quad \pi_j(v)\in\mathbb{V}_j\ ,\ \ v-\pi_j(v)\perp\mathbb{V}_j\ ,\ \ \forall v\in\mathbb{V}\ . \tag{37}$$

Proof. First, we establish two useful properties of the local unit balls B_j in (34). By assumption, for every $j\in\mathbb{J}$, $|||\cdot|||_j$ is a norm on $\mathbb{V}_j$ which is equivalent to the restriction to $\mathbb{V}_j$ of the Hilbertian norm $\|\cdot\|$ on $\mathbb{V}$. Therefore, for every $j\in\mathbb{J}$, every local unit ball B_j is

- bounded (for the Hilbertian norm $\|\cdot\|$), by Proposition 19 because there exists $m_j>0$ such that $m_j\|\cdot\|\leq|||\cdot|||_j$ on $\mathbb{V}_j$, hence weakly bounded by (57),
- closed in $\mathbb{V}_j$ (for the relative Hilbertian topology of $\mathbb{V}_j$), by Proposition 19 because there exists $M_j>0$ such that $|||\cdot|||_j\leq M_j\|\cdot\|$ on $\mathbb{V}_j$.

Second, we prove that there is a (unique) norm $|||\cdot|||$ on $\mathbb{V}$ with unit ball $\overline{\mathrm{co}}\big(\bigcup_{j\in\mathbb{J}} B_j\big)$ and it admis a dual norm $|||\cdot|||_\star$ with unit ball $\big(\bigcup_{j\in\mathbb{J}} B_j\big)^{\odot}$. For this purpose, it suffices to show that the family $\{B_j\}_{j\in\mathbb{J}}$ of local unit balls, defined in (34), satisfies the assumptions of Proposition 14.

Now, for every $j \in \mathbb{J}$, every local unit ball B_j is symmetric, and, by (57), is also weakly bounded since it is bounded. Thus, the assumptions (29) are satisfied. There remains to prove the full sum condition (30). But it follows from an easily proven property of a unit ball — namely, that $\mathrm{span}B_j = \overline{\mathrm{span}}B_j = \mathbb{V}_j$, for every $j \in \mathbb{J}$ — from which we get $\sum_{j\in\mathbb{J}} \mathrm{span}B_j = \sum_{j\in\mathbb{J}} \mathbb{V}_j = \mathbb{V}$ by (33).

Moreover, Proposition 14 establishes that the topologies defined by the norm $|||\cdot|||$ and by the dual norm $|||\cdot|||_\star$ are both weaker (contain less open sets) than the Hilbertian topology. Yet, by Proposition 19, the topology induced by the norm $|||\cdot|||$ is stronger than the Hilbertian topology, because the unit ball $B_{|||\cdot|||} = \overline{\mathrm{co}}\big(\bigcup_{j\in\mathbb{J}} B_j\big)$ is bounded, as finite union of bounded local unit balls. Therefore, the norm $|||\cdot|||$ is equivalent to the Hilbertian norm $\|\cdot\|$. We conclude this part with Proposition 25 that asserts that the dual norm $|||\cdot|||_\star$ is then also equivalent to the Hilbertian norm $\|\cdot\|$.

Third, we prove the two items (but in reverse order).

2. By Proposition 25 (applied on each of the Hilbert closed subspace $\mathbb{V}_j$), for each $j \in \mathbb{J}$ the local norm $|||\cdot|||_j$ on $\mathbb{V}_j$ admits a local dual norm $|||\cdot|||_{j,\star}$ on $\mathbb{V}_j$. Indeed, we have seen that every local unit ball B_j is weakly bounded, hence weakly bounded on $\mathbb{V}_j$ by (57), and closed in $\mathbb{V}_j$ (for the Hilbertian relative topology).

   We prove (36). For this purpose, let $B_j^{\odot_j} = \{v' \in \mathbb{V}_j \mid \langle v\,, v'\rangle \leq 1\,,\ \forall v \in B_j\}$ denote the polar of the set B_j in $\mathbb{V}_j$, as in (54). We have

$$|||v|||_j = \sigma_{B_j^{\odot_j}}(v)\,,\ \forall v \in \mathbb{V}_j\,, \qquad \text{by (63b)}, \tag{38a}$$

$$|||v'|||_{j,\star} = \sigma_{B_j}(v')\,,\ \forall v' \in \mathbb{V}_j\,, \qquad \text{by (63b)}, \tag{38b}$$

$$B_j^{\odot} = B_j^{\odot_j} + \mathbb{V}_j^{\perp}\,, \qquad \text{as easily deduced from the definition (54) of a polar set} \tag{38c}$$

$$\sigma_{B_j^{\odot}} = \sigma_{B_j^{\odot_j}} \dot{+} \sigma_{\mathbb{V}_j^{\perp}} = |||\cdot|||_j \dot{+} \delta_{\mathbb{V}_j}\,, \tag{38d}$$

   by (38a) and by a property of the support function of a vector space.

   For all $v' \in \mathbb{V}$, we have

$$\begin{aligned}\sigma_{V_j}(v') &= \sup_{v\in V_j} \big(\,\langle v\,, \pi_j(v')\rangle + \langle v\,, v' - \pi_j(v')\rangle\,\big) \quad \text{(where the mapping } \pi_j \text{ is defined in (37))}\\ &= \sup_{v\in V_j} \langle v\,, \pi_j(v')\rangle \\ &\qquad\qquad \text{(by property of the orthogonal projection mapping (37) since } V_j \subset \mathbb{V}_j)\\ &= \sigma_{V_j}\big(\pi_j(v')\big)\,. \end{aligned} \tag{39}$$

Now, we are ready to prove (36):

$$
\begin{aligned}
|||\cdot|||_\star &= \sigma_{\overline{\mathrm{co}}\left(\bigcup_{j\in\mathbb{J}} B_j\right)} && \text{(by (31a))}\\
&= \sigma_{\bigcup_{j\in\mathbb{J}} B_j}\\
&= \sup_{j\in\mathbb{J}} \sigma_{B_j}\\
&= \sup_{j\in\mathbb{J}} \sigma_{B_j}\circ\pi_j && \text{(by (39))}\\
&= \sup_{j\in\mathbb{J}} \left(|||\cdot|||_{j,\star}\circ\pi_j\right) . && \text{(by (38b))}
\end{aligned}
$$

1. We prove (35).

   For this purpose, we start by showing that $0 \in \bigcap_{j\in\mathbb{J}} \mathrm{cont}\left(\delta_{B_j^\odot}\right)$. We have proven at the beginning that every local unit ball B_j is bounded. Therefore, $\bigcup_{j\in\mathbb{J}} B_j$ is bounded because the set $\mathbb{J}$ is finite. Letting $M > 0$ be such that $\bigcup_{j\in\mathbb{J}} B_j \subset MB_{\|\cdot\|}$, we get that

$$
\begin{aligned}
\bigcup_{j\in\mathbb{J}} B_j \subset MB_{\|\cdot\|} &\Rightarrow \left(MB_{\|\cdot\|}\right)^\odot \subset \Big(\bigcup_{j\in\mathbb{J}} B_j\Big)^\odot\\
&\Rightarrow \frac{1}{M} B_{\|\cdot\|_\star} \subset \Big(\bigcup_{j\in\mathbb{J}} B_j\Big)^\odot \quad \text{(by (62) and the definition (54) of a polar set)}\\
&\Rightarrow \frac{1}{M} B_{\|\cdot\|} \subset \Big(\bigcup_{j\in\mathbb{J}} B_j\Big)^\odot\\
&\qquad \text{(because the dual norm } \|\cdot\|_\star \text{ of the Hilbertian norm is the Hilbertian norm)}\\
&\Rightarrow 0 \in \mathrm{int}\Big(\bigcup_{j\in\mathbb{J}} B_j\Big)^\odot = \mathrm{int}\bigcap_{j\in\mathbb{J}} B_j^\odot ,
\end{aligned}
$$

   where the interior is with respect to the Hilbertian topology. Now, as it is always true that $\mathrm{int}\bigcap_{j\in\mathbb{J}} B_j^\odot \subset \bigcap_{j\in\mathbb{J}} \mathrm{int} B_j^\odot$, we get that $0 \in \bigcap_{j\in\mathbb{J}} \mathrm{int} B_j^\odot$. Finally, it is easily seen that $\mathrm{cont}\left(\delta_{B_j^\odot}\right) = \mathrm{int} B_j^\odot$, for every $j \in \mathbb{J}$. We conclude that $0 \in \bigcap_{j\in\mathbb{J}} \mathrm{cont}\left(\delta_{B_j^\odot}\right)$.

Now, we are ready to prove (35):

$$
\begin{aligned}
|||\cdot||| &= \sigma_{B^\odot} && \text{(by (31) and the definition (32) of } B\text{)}\\
&= \delta^\star_{B^\odot}\\
&= \delta^\star_{\bigcap_{j\in\mathbb{J}} B_j^\odot} && \text{(as } B^\odot = \Big(\overline{\mathrm{co}}\big(\bigcup_{j\in\mathbb{J}} B_j\big)\Big)^\odot = \big(\bigcup_{j\in\mathbb{J}} B_j\big)^\odot = \bigcap_{j\in\mathbb{J}} B_j^\odot\text{)}\\
&= \Big(\sum_{j\in\mathbb{J}} \delta_{B_j^\odot}\Big)^\star && \text{(as } \delta_{\bigcap_{j\in\mathbb{J}} B_j^\odot} = \textstyle\sum_{j\in\mathbb{J}} \delta_{B_j^\odot}\text{)}\\
&= \square_{j\in\mathbb{J}}\, \delta^\star_{B_j^\odot}\\
&\quad \text{(because } 0 \in \textstyle\bigcap_{j\in\mathbb{J}} \mathrm{cont}\big(\delta_{B_j^\odot}\big) \text{ as shown above and using [3, Prop. 15.3-15.5])}\\
&= \square_{j\in\mathbb{J}}\, \sigma_{B_j^\odot}\\
&= \square_{j\in\mathbb{J}}\, \big(|||\cdot|||_j + \delta_{\mathbb{V}_j}\big)\,. && \text{(by (38d))}
\end{aligned}
$$

This ends the proof. □

4.3 Design of norms for lower bound convex programs for GSO

Finally, we consider a generalized sparse optimization problem on a Hilbert space and we present a systematic way to design norms, that mixes the formulations and results of Proposition 13 and Proposition 15. In what follows, *sparsity* is captured by a finite family $\{W_j\}_{j\in\mathbb{J}}$ of subsets of a Hilbert space $\mathbb{V}$, whereas *amplitude* is measured by a family $\{|||\cdot|||_j\}_{j\in\mathbb{J}}$ of (local) norms on every closed subspace $\overline{\mathrm{span}}W_j$.

Theorem 16 *Let $\mathbb{V}$ be a Hilbert space. Let $\mathbb{J}$ be a finite set.*

1. *Let $\{W_j\}_{j\in\mathbb{J}}$ be a family of two by two disjoint symmetric subsets of $\mathbb{V}$ such that the closed subspaces*

$$\mathbb{V}_j = \overline{\mathrm{span}}W_j\,,\ \ \forall j\in\mathbb{J} \tag{40}$$

generate a full sum as follows

$$\sum_{j\in\mathbb{J}} \mathbb{V}_j = \mathbb{V}\,. \tag{41}$$

2. *Let $\{|||\cdot|||_j\}_{j\in\mathbb{J}}$ be a family of (local) norms such that, for every $j\in\mathbb{J}$, $|||\cdot|||_j$ is a norm on $\mathbb{V}_j$, which is equivalent to the restriction to $\mathbb{V}_j$ of the Hilbertian norm $\|\cdot\|$ on $\mathbb{V}$. We denote the (local) unit balls and spheres by*

$$B_j = \{w\in\mathbb{V}_j \mid |||w|||_j \le 1\} \subset \mathbb{V}_j\,,\ \ \forall j\in\mathbb{J}\,, \tag{42a}$$

$$S_j = \{w\in\mathbb{V}_j \mid |||w|||_j = 1\} \subset \mathbb{V}_j\,,\ \ \forall j\in\mathbb{J}\,. \tag{42b}$$

3. *Let* $\{\theta_j\}_{j\in\mathbb{J}}$ *be a family of symmetric*[4] *mappings* $\theta_j : W_j \to \mathbb{V}$*, such that*

$$S_j \subset \overline{\theta_j(W_j)} \subset B_j \ . \tag{43}$$

4. *Let* $h : \mathbb{V} \to \overline{\mathbb{R}}$ *be a function such that every function* $\big(\theta_j \rhd h\big)^\star$ *is proper, for every* $j \in \mathbb{J}$*, and* $\bigcap_{j\in\mathbb{J}} \mathrm{dom}\big(\theta_j \rhd h\big)^\star \neq \emptyset$*.*

Then, there exists a norm $|||\cdot|||$*, with unit ball* $\overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} B_j\Big)$ *such that we have the lower bound*

$$\min_{|||w|||\leq 1} \Big(\sup_{j\in\mathbb{J}} \big(\theta_j \rhd h\big)^\star\Big)^\star(w) \leq \inf_{w\in\bigcup_{j\in\mathbb{J}} W_j} h(w) \ . \tag{44}$$

Moreover, expressions for the norm $|||\cdot|||$ *and for its dual norm can be found in Proposition 15.*

Proof. First, we use Proposition 12 to obtain the lower bound (26). For this purpose, we check its three assumptions (item 2a, item 2b, item 3) one by one.

- By item 1 here, the family $\{W_j\}_{j\in\mathbb{J}}$ is made of two by two disjoint subsets of $\mathbb{V}$. Therefore, item 2a of Proposition 12 is satisfied.
- For every $j \in \mathbb{J}$, every subset $\theta_j(W_j)$ is symmetric, because so are the subsets W_j (by item 1 here) and the mappings θ_j (by item 3 here). For every $j \in \mathbb{J}$, every subset $-\theta_j(W_j) = \theta_j(W_j)$ is weakly bounded, because, by (43), it is a subset of the ball (42a), which is bounded (by Proposition 19, as seen at the beginning of the proof of Proposition 13). Therefore, item 2b of Proposition 12 is satisfied.
- Item 3 of Proposition 12 coincides with item 4 here.

Second, we prove that the term $\overline{\mathrm{co}}\Big(-\bigcup_{j\in\mathbb{J}} \theta_j(W_j)\Big) = \overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} \theta_j(W_j)\Big)$ in the lower bound (26)

[4] that is, $\theta_j(-w) = \theta_j(w)$, for all $w \in W_j$ and for every $j \in \mathbb{J}$.

can be replaced by $\overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} B_j\Big)$. Indeed

$$\begin{aligned}
\overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} B_j\Big) &\supset \overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} \overline{\theta_j(W_j)}\Big) && \text{(by } \overline{\theta_j(W_j)} \subset B_j \text{ in (43))}\\
&= \overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} \theta_j(W_j)\Big) && \text{(as easily proven)}\\
&\supset \overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} S_j\Big) && \text{(by } S_j \subset \overline{\theta_j(W_j)} \text{ in (43))}\\
&= \overline{\mathrm{co}\Big(\bigcup_{j\in\mathbb{J}} S_j\Big)}\\
&= \overline{\mathrm{co}\Big(\bigcup_{j\in\mathbb{J}} \mathrm{co} S_j\Big)} && \text{(as easily proved)}\\
&= \overline{\mathrm{co}\Big(\bigcup_{j\in\mathbb{J}} B_j\Big)} && \text{(because } \mathrm{co}S_j = B_j, \forall j \in \mathbb{J})\\
&= \overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} B_j\Big)\ .
\end{aligned}$$

Third, by Proposition 15, there exists a norm $|||\cdot|||$, with unit ball $\overline{\mathrm{co}}\Big(\bigcup_{j\in\mathbb{J}} B_j\Big)$.
This ends the proof. □

A possible choice for the family $\{\theta_j\}_{j\in\mathbb{J}}$ of symmetric mappings $\theta_j : W_j \to \mathbb{V}$ is given by the normalization mappings

$$\forall j \in \mathbb{J}\ ,\ \ n_j : W_j \to S_j \cup \{0\}\ ,\ \ n_j(w) = \begin{cases} \frac{w}{|||w|||_j} & \text{if } w \in W_j\backslash\{0\}\ ,\\ 0 & \text{if } w = 0\ , \end{cases}$$

under the assumption (equivalent to (43)) that

$$\overline{\left\{ \frac{w}{|||w|||_j} \mid w \in W_j\backslash\{0\} \right\}} = \overline{n_j(W_j)} = S_j\ ,\ \ \forall j \in \mathbb{J}\ . \tag{45}$$

With these normalization mappings in Theorem 16, we recover the case developed in §3.2, where sparsity is exactly measured by the l_0 pseudonorm, with:

- finite set $\mathbb{J} = \{J \mid J \subset \{1,\dots,d\} \text{ and } |J| \le k\}$,
- subsets $W_J = \big\{w \in \mathbb{R}^d \mid \mathrm{supp}(w) = J\big\}$ of the Euclidian space $\mathbb{V} = \mathbb{R}^d$ (sparsity), for all $j \in \mathbb{J}$, with the convention that $W_\emptyset = \{0\}$,
- norms $|||w|||_J = \|w\|_2$ (amplitude).

Our framework encompasses the (latent) group Lasso norms [8, 7], with:

- finite set $\mathbb{J} = 2^{\{1,\dots,d\}} = \{J \mid J \subset \{1,\dots,d\}\}$,
- subsets $W_J = \{w \in \mathbb{R}^d \mid \mathrm{supp}(w) = J\}$ of the Euclidian space $\mathbb{V} = \mathbb{R}^d$ (sparsity), with the convention that $W_\emptyset = \{0\}$,
- norms $|||w|||_J = \frac{\|w\|_q}{F(J)^{1/q}}$, where $F : 2^{\{1,\dots,d\}} \to]0, +\infty]$ (amplitude).

In both cases, a global norm is inferred, by convolution, from local norms on $\overline{\mathrm{span}} W_J = \mathbb{R}^J \times \{0\}^{-J}$. The condition (45) holds indeed true as $n_J(W_J)$ is S_J minus a finite number of points (those on the axis), hence $\overline{n_J(W_J)} = S_J$. Theorem 16 provides support for the use of this type of norms in sparse optimization, and Proposition 15 displays a general method to construct large classes of norms.

5 Conclusion

In this paper, we have consider *exact* sparse optimization problems, that is, problems with combinatorial sparsity constraint. More precisely, we have focused on problems where sparsity is measured either by the nonconvex l_0 pseudonorm (and not by substitute penalty terms) or by the belonging of the solution to a finite union of given subsets.

In exact sparse optimization problems, where sparsity is measured by the l_0 pseudonorm, one looks for solution that have few nonzero components. It is well-known that the Fenchel biconjugate of the l_0 pseudonorm is zero, making it hopeless to replace the l_0 pseudonorm by its best lower convex lsc approximation. In the same vein, the highly nonconvex constraint that the l_0 pseudonorm be less than a given integer cannot be handled by the Fenchel conjugacy, because the conjugate of its characteristic function is identically $+\infty$.

In this paper, we have proposed to handle the l_0 pseudonorm, not by the Fenchel conjugacy, but by a suitable so-called Caprac conjugacy, as introduced in the companion paper [4]. By doing so, we have displayed a convex program that is a lower bound of the original combinatorial optimization problem. We insist that it is a lower bound, and not a substitute problem with substitute penalty terms. Thus doing, we keep track of the original nonconvex problem.

Going on, we have studied generalized sparse optimization, where the solution is looked for in a finite union of given subsets. We have identified suitable couplings, namely the one-sided linear couplings, and, thus equipped, we have been able to obtain more general results. Our main contribution is to provide a systematic way to design norms, and associated convex programs that are lower bounds for the original exact sparse optimization problem.

Acknowledgements. We want to thank Guillaume Obozinski, Yohann De Castro, Nathan De Lara, Antonin Chambolle, Miguel Angel Goberna and Marco Antonio Lopez Cerda for discussions on first versions of this work.

A Appendix

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

When we manipulate functions with values in $\overline{\mathbb{R}} = [-\infty, +\infty]$, we adopt the following Moreau *lower addition* or *upper addition*, depending on whether we deal with sup or inf operations. We follow [6]. In the sequel, u, v and w are any elements of $\overline{\mathbb{R}}$.

Moreau lower addition

The Moreau *lower addition* extends the usual addition with

$$(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = -\infty \ . \tag{46a}$$

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

$$u \leq u' \ , \ \ v \leq v' \Rightarrow u \dot{+} v \leq u' \dot{+} v' \ , \tag{46b}$$

$$(-u) \dot{+} (-v) \leq -(u \dot{+} v) \ , \tag{46c}$$

$$(-u) \dot{+} u \leq 0 \ , \tag{46d}$$

$$\sup_{a \in \mathbb{A}} f(a) \dot{+} \sup_{b \in \mathbb{B}} g(b) = \sup_{a \in \mathbb{A}, b \in \mathbb{B}} \big(f(a) \dot{+} g(b) \big) \ , \tag{46e}$$

$$\inf_{a \in \mathbb{A}} f(a) \dot{+} \inf_{b \in \mathbb{B}} g(b) \leq \inf_{a \in \mathbb{A}, b \in \mathbb{B}} \big(f(a) \dot{+} g(b) \big) \ , \tag{46f}$$

$$t < +\infty \Rightarrow \inf_{a \in \mathbb{A}} f(a) \dot{+} t = \inf_{a \in \mathbb{A}} \big(f(a) \dot{+} t \big) \ . \tag{46g}$$

Moreau upper addition

The Moreau *upper addition* extends the usual addition with

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

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

$$u \leq u' \ , \ \ v \leq v' \Rightarrow u \dot{+} v \leq u' \dot{+} v' \ , \tag{47b}$$

$$(-u) \dot{+} (-v) \geq -(u \dot{+} v) \ , \tag{47c}$$

$$(-u) \dot{+} u \geq 0 \ , \tag{47d}$$

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

$$\sup_{a \in \mathbb{A}} f(a) \dot{+} \sup_{b \in \mathbb{B}} g(b) \geq \sup_{a \in \mathbb{A}, b \in \mathbb{B}} \big(f(a) \dot{+} g(b) \big) \ , \tag{47f}$$

$$-\infty < t \Rightarrow \sup_{a \in \mathbb{A}} f(a) \dot{+} t = \sup_{a \in \mathbb{A}} \big(f(a) \dot{+} t \big) \ . \tag{47g}$$

Joint properties of the Moreau lower and upper addition

We obviously have that

$$u \mathbin{\dot{+}} v \leq u \mathbin{\dot{+}} v \; . \tag{48a}$$

The Moreau lower and upper additions are related by

$$-(u \mathbin{\dot{+}} v) = (-u) \mathbin{\dot{+}} (-v) \; , \;\; -(u \mathbin{\dot{+}} v) = (-u) \mathbin{\dot{+}} (-v) \; . \tag{48b}$$

They satisfy the inequality

$$(u \mathbin{\dot{+}} v) \mathbin{\dot{+}} w \leq u \mathbin{\dot{+}} (v \mathbin{\dot{+}} w) \; . \tag{48c}$$

with

$$(u \dot{+} v) \mathbin{\dot{+}} w < u \dot{+} (v \mathbin{\dot{+}} w) \iff \begin{cases} u = +\infty \;\text{ and }\; w = -\infty \; , \\ \;\text{ or} \\ u = -\infty \;\text{ and }\; w = +\infty \;\text{ and }\; -\infty < v < +\infty \; . \end{cases} \tag{48d}$$

Finally, we have that

$$u \mathbin{\dot{+}} (-v) \leq 0 \iff u \leq v \iff 0 \leq v \mathbin{\dot{+}} (-u) \; , \tag{48e}$$

$$u \mathbin{\dot{+}} (-v) \leq w \iff u \leq v \mathbin{\dot{+}} w \iff u \mathbin{\dot{+}} (-w) \leq v \; , \tag{48f}$$

$$w \leq v \mathbin{\dot{+}} (-u) \iff u \mathbin{\dot{+}} w \leq v \iff u \leq v \mathbin{\dot{+}} (-w) \; . \tag{48g}$$

A.2 Background on sets and functions

Let $\mathbb{W}$ be a set.

- The *effective domain* of a function $h : \mathbb{W} \to \overline{\mathbb{R}}$ is $\mathrm{dom} h = \{w \in \mathbb{W} \mid h(w) < +\infty\}$.
- The function $h : \mathbb{W} \to \overline{\mathbb{R}}$ is said to be *proper* when $\mathrm{dom} h \neq \emptyset$ and $\{w \in \mathbb{W} \mid h(w) = -\infty\} = \emptyset$.
- When $\mathbb{W}$ is a topological space, $\mathrm{cont} h$ denotes the *continuity points* of the function $h : \mathbb{W} \to \overline{\mathbb{R}}$.
- The *characteristic function* of a set $W \subset \mathbb{W}$ is the function δ_W defined by

$$\delta_W(w) = \begin{cases} 0 & \text{if } w \in W \; , \\ +\infty & \text{if } w \notin W \; . \end{cases} \tag{49}$$

A.2.1 Topological vector space

Let $\mathbb{W}$ be a topological vector space and let $W \subset \mathbb{W}$.

- The set W is *symmetric* if $-W = W$.
- The *conical hull* of W is the smallest cone in $\mathbb{W}$ that contains W, denoted by $\mathrm{cone} W$.

- The *convex hull* of W is the smallest convex set in $\mathbb{W}$ that contains W, denoted by co$W$.
- The *closed convex hull* of W is the smallest closed convex set in $\mathbb{W}$ that contains W, denoted by $\overline{\text{co}}W$.
- The *span* of W is the smallest subspace of $\mathbb{W}$ that contains W, denoted by span$W$.
- The *closed span* of W is the smallest closed subspace of $\mathbb{W}$ that contains W, denoted by $\overline{\text{span}}W$.

A.2.2 Dual system

We say that two vector spaces $\mathbb{X}$ and $\mathbb{Y}$ form a *dual system* [1, p. 211] if $\mathbb{X}$ and $\mathbb{Y}$ are equipped with a bilinear form $\langle\,,\rangle$, such that $\Big(\forall x \in \mathbb{X}\ ,\ \langle x\,, y\rangle = 0\Big) \Rightarrow y = 0$ and $\Big(\forall y \in \mathbb{Y}\ ,\ \langle x\,, y\rangle = 0\Big) \Rightarrow x = 0$. By default, the primal space $\mathbb{X}$ is equipped with the weak topology (of pointwise convergence), and the dual space $\mathbb{Y}$ with the weak* topology.

Definition 17 *Let $\mathbb{X}$ and $\mathbb{Y}$ be a dual system and let $X \subset \mathbb{X}$.*

- *The* support function *of X is defined by*

$$\sigma_X(y) = \sup_{x \in X} \langle x\,, y\rangle\ ,\ \ \forall y \in \mathbb{Y}\ . \tag{50}$$

- *The* barrier cone *of X is the effective domain of the support function σ_X:*

$$\text{bar}X = \left\{ y \in \mathbb{Y} \;\middle|\; \sup_{x \in X} \langle x\,, y\rangle < +\infty \right\} = \text{dom}\sigma_X\ . \tag{51}$$

- *The set X is said to be* weakly bounded *if $\sup_{x\in X} \langle x\,, y\rangle < +\infty$ for all $y \in \mathbb{Y}$:*

$$X \text{ is weakly bounded } \iff \text{bar}X = \mathbb{Y} \iff \text{dom}\sigma_X = \mathbb{Y}\ . \tag{52}$$

- *The* orthogonal set *of X is defined by*

$$X^{\perp} = \{ y \in \mathbb{Y} \mid \langle x\,, y\rangle = 0\ ,\ \ \forall x \in X \}\ . \tag{53}$$

- *The* polar set *of X is defined by*

$$X^{\odot} = \{ y \in \mathbb{Y} \mid \langle x\,, y\rangle \leq 1\ ,\ \ \forall x \in X \}\ . \tag{54}$$

We obtain symmetric definitions for $Y \subset \mathbb{Y}$.

We provide different properties of barrier cones and of weakly bounded sets.

Proposition 18

1. *The barrier cone* (51) *of* $X \subset \mathbb{X}$ *satisfies the following properties*

$$\mathrm{bar}X = \mathrm{bar}(\mathrm{co}X) = \mathrm{bar}(\overline{\mathrm{co}}X) \ , \tag{55a}$$

$$\mathrm{bar}X = \mathrm{cone}\big(X^{\odot}\big) \ . \tag{55b}$$

2. *Let* $\{X_j\}_{j\in\mathbb{J}}$ *be a family of subsets of* $\mathbb{X}$*. Then*

$$\mathbb{J} \text{ is a finite set } \Rightarrow \mathrm{bar}\Big(\bigcup_{j\in\mathbb{J}} X_j\Big) = \bigcap_{j\in\mathbb{J}} \mathrm{bar}X_j \ . \tag{56}$$

   As an application, if $\{X_j\}_{j\in\mathbb{J}}$ *is a finite family of weakly bounded subsets of* $\mathbb{X}$*, then the finite union* $\bigcup_{j\in\mathbb{J}} X_j$ *is weakly bounded.*

3. *If* $\mathbb{V}$ *is a Hilbert space, then bounded subsets of* $\mathbb{V}$ *are weakly bounded:*

$$V \subset \mathbb{V} \text{ is bounded } \Rightarrow \mathrm{bar}V = \mathbb{V} \ . \tag{57}$$

Proof.

1. Equation (55a) is a consequence of the definition (51) and of the property of support functions that $\sigma_X = \sigma_{\mathrm{co}X} = \sigma_{\overline{\mathrm{co}}X}$.

   The proof of (55b) follows easily from

$$\mathrm{bar}X = \bigcup_{\lambda>0} \Big\{ y \in \mathbb{Y} \ \Big| \ \sup_{x\in X} \langle x \, , y\rangle \leq \lambda \Big\} = \bigcup_{\lambda>0} \Big\{ y \in \mathbb{Y} \ \Big| \ \frac{y}{\lambda} \in X^{\odot} \Big\} = \mathrm{cone}\big(X^{\odot}\big) \ ,$$

   by the definition (54) of the polar set $X^{\odot}$.

2. The proof of (56) follows from the observation that $\sigma_{(\bigcup_{j\in\mathbb{J}} X_j)} = \max_{j\in\mathbb{J}} \sigma_{X_j}$ with a maximum since the set $\mathbb{J}$ is finite, and from the definition (51) that $\mathrm{bar}X = \mathrm{dom}\sigma_X$.

   As an application, if every X_j is weakly bounded, for every $j \in \mathbb{J}$, and $\mathbb{J}$ is finite, we get, by (56), that

$$\mathrm{bar}\Big(\bigcup_{j\in\mathbb{J}} X_j\Big) = \bigcap_{j\in\mathbb{J}} \mathrm{bar}X_j = \bigcap_{j\in\mathbb{J}} \mathbb{Y} = \mathbb{Y} \ ,$$

   and we conclude that the finite union $\bigcup_{j\in\mathbb{J}} X_j$ is weakly bounded by definition (51).

3. The proof of (57) follows from the observation that, in a Hilbert space, $\langle v \, , v'\rangle \leq \|v\|\|v'\|$, so that, for any $v' \in \mathbb{V}$, we have that $\sup_{v\in V} \langle v \, , v'\rangle \leq \big(\sup_{v\in V} \|v\|\big)\|v'\| < +\infty$, as $\sup_{v\in V} \|v\| < +\infty$ since V is bounded.

This ends the proof. □

A.3 Background on norms and dual norms

Here, we collect different results on norms, equivalent norms, and norms induced by support functions (in a dual system and in a Hilbert space).

For a norm $\|\cdot\|$ on a vector space $\mathbb{W}$, we denote the unit ball by

$$B_{\|\cdot\|} = \{w \in \mathbb{W} \mid \|w\| \leq 1\} \ . \tag{58}$$

A unit ball is always convex, symmetric and with full conical hull, that is, $\mathrm{cone} B_{|||\cdot|||} = \mathbb{W}$ (indeed any $w \in \mathbb{W}\backslash\{0\}$ can be written $w = \|w\| \frac{w}{\|w\|} \in \mathrm{cone} B_{\|\cdot\|}$, and $0 \in B_{\|\cdot\|} \subset \mathrm{cone} B_{\|\cdot\|}$).

A.3.1 Equivalent norms

We recall definition and characterizations of equivalent norms.

Proposition 19 *Let $\|\cdot\|^\sharp$ and $\|\cdot\|^\flat$ be two norms on a vector space $\mathbb{W}$. The following statements are equivalent.*

1. *There exists $M > 0$ such that $\|\cdot\|^\flat \leq M\|\cdot\|^\sharp$.*
2. *The topology of $\|\cdot\|^\sharp$ is richer (contains more open sets) than the topology of $\|\cdot\|^\flat$.*
3. *The function $\|\cdot\|^\flat : \left(\mathbb{W}, \|\cdot\|^\sharp\right) \to \overline{\mathbb{R}}$ is continuous.*
4. *The unit ball $B_{\|\cdot\|^\flat}$ is closed for the topology of $\|\cdot\|^\sharp$.*
5. *The unit ball $B_{\|\cdot\|^\flat}$ has nonempty interior for the topology of $\|\cdot\|^\sharp$.*
6. *$0 \in \mathrm{int}_{\|\cdot\|^\sharp} B_{\|\cdot\|^\flat}$.*
7. *The unit ball $B_{\|\cdot\|^\sharp}$ is bounded for the norm $\|\cdot\|^\flat$.*

Proof. The chain of implications (in both directions) from statements 1 to 4 is easy to prove. So is statement 3 $\Rightarrow$ statement 5.

Statement 7 is equivalent to the property that there exists $M > 0$ such that $B_{\|\cdot\|^\sharp} \subset M B_{\|\cdot\|^\flat}$, hence to statement 1.

Statement 6 is equivalent to the property that there exists $M > 0$ such that $\frac{1}{M} B_{\|\cdot\|^\sharp} \subset B_{\|\cdot\|^\flat}$, hence is equivalent to statement 7. Indeed, using [3, (6.6) p. 114], we have that the interior of a set D is $\mathrm{int}_{\|\cdot\|^\sharp} D = \left\{ w \in D \;\middle|\; (\exists \rho > 0) \quad \rho B_{\|\cdot\|^\sharp} \subset D - w \right\}$. With this, we prove that statement 5 implies statement 6 (the reverse is obvious). Let $w \in \mathrm{int}_{\|\cdot\|^\sharp} B_{\|\cdot\|^\flat} = \left\{ w \in B_{\|\cdot\|^\flat} \;\middle|\; (\exists \rho > 0) \quad \rho B_{\|\cdot\|^\sharp} \subset B_{\|\cdot\|^\flat} - w \right\}$, there exists $\rho > 0$ such that $\rho B_{\|\cdot\|^\sharp} \subset B_{\|\cdot\|^\flat} - w$. Now, choosing $\mu = 1/(1 + \|w\|^\flat)$, we get that $\mu\rho B_{\|\cdot\|^\sharp} \subset \mu(B_{\|\cdot\|^\flat} - w) \subset B_{\|\cdot\|^\flat}$, and thus $0 \in \mathrm{int}_{\|\cdot\|^\sharp} B_{\|\cdot\|^\flat}$.

This ends the proof. □

We easily deduce the following Proposition (and the definition of equivalent norms).

Proposition 20 *Let $\|\cdot\|^\sharp$ and $\|\cdot\|^\flat$ be two norms on a vector space $\mathbb{W}$. The following statements are equivalent.*

1. *There exist two positive numbers m and M, such that*

$$0 < m \leq M < +\infty \ \text{ and } \ m\|\cdot\|^\sharp \leq \|\cdot\|^\flat \leq M\|\cdot\|^\sharp \,. \tag{59}$$

2. *The topologies of $\|\cdot\|^\sharp$ and $\|\cdot\|^\flat$ are the same.*
3. *The unit ball $B_{\|\cdot\|^\flat}$ is closed for the topology of $\|\cdot\|^\sharp$, and bounded for the norm $\|\cdot\|^\sharp$.*
4. *The unit ball $B_{\|\cdot\|^\flat}$ is closed for the topology of $\|\cdot\|^\sharp$, and $0 \in \mathrm{int}_{\|\cdot\|^\sharp} B_{\|\cdot\|^\flat}$.*
5. *The unit ball $B_{\|\cdot\|^\sharp}$ is closed for the topology of $\|\cdot\|^\flat$, and bounded for the norm $\|\cdot\|^\flat$.*
6. *The unit ball $B_{\|\cdot\|^\sharp}$ is closed for the topology of $\|\cdot\|^\flat$, and $0 \in \mathrm{int}_{\|\cdot\|^\flat} B_{\|\cdot\|^\sharp}$.*

In any of these equivalent cases, we say that the norms $\|\cdot\|^\sharp$ and $\|\cdot\|^\flat$ are equivalent.

A.3.2 Dual norm in the dual system case

Let $\mathbb{X}$ and $\mathbb{Y}$ be a dual system, as recalled in §A.2. By default, the primal space $\mathbb{X}$ is equipped with the weak topology (of pointwise convergence), and the dual space $\mathbb{Y}$ with the weak* topology. In the paper, we will mostly consider the case where $\mathbb{X} = \mathbb{Y}$ is a Hilbert space, and the natural dual system it induces.

We study under which stronger and stronger assumptions the support function of a set is a norm.

Proposition 21 *Let $\mathbb{X}$ and $\mathbb{Y}$ be a dual system.*

1. *Let $X \subset \mathbb{X}$ be symmetric, weakly bounded and with full conical hull, that is, such that*

$$-X = X \ , \ \ \mathrm{bar}X = \mathbb{Y} \ , \ \ \mathrm{cone}X = \mathbb{X} \,. \tag{60}$$

   Then the support function σ_X is a norm on $\mathbb{Y}$, whose unit ball is $X^\odot$.

2. *Let $C \subset \mathbb{X}$ be closed, convex and containing 0. The following statements are equivalent.*
   (a) *The support function σ_C is a norm on $\mathbb{Y}$, whose unit ball is the polar set $C^\odot$.*
   (b) *The set C is symmetric, weakly bounded and with full conical hull.*
   (c) *The polar set $C^\odot$ is symmetric, weakly bounded and with full conical hull.*
   (d) *The support function $\sigma_{C^\odot}$ is a norm on $\mathbb{X}$, whose unit ball is C.*

Proof.

1. We prove item 1.

   First, as X is weakly bounded, that is, $\mathrm{bar}X = \mathbb{Y}$, we have that $\mathrm{dom}\sigma_X = \mathbb{Y}$ by (52), hence that $\sigma_X < +\infty$.

   Second, as X is symmetric, that is, $-X = X$, we have that $\sigma_X(y) = \sigma_X(-y)$, for all $y \in \mathbb{Y}$.

   Third, as X is symmetric (and nonempty since $\mathrm{cone}X = \mathbb{X}$), we deduce that $0 \in \mathrm{co}X$, hence that $\sigma_X(y) = \sigma_{\mathrm{co}X}(y) \geq 0$, for all $y \in \mathbb{Y}$.

   Fourth, we show that $\sigma_X(y) = 0 \Rightarrow y = 0$. Indeed, from $\sigma_X \geq 0$, we deduce that $\sigma_X(y) = 0 \iff y \in X^\perp$. Now, as X has full conical hull, that is, $\mathrm{cone}X = \mathbb{X}$, we deduce that $X^\perp = \big(\mathrm{cone}X\big)^\perp = \mathbb{X}^\perp = \{0\}$, hence $y = 0$.

   Finally, we conclude that σ_X is a norm since it is subadditive and 1-homogeneous, as it is a support function.

   The unit ball of the norm σ_X is $B_{\sigma_X} = \{y \in \mathbb{Y} \mid \sigma_X(y) \leq 1\} = X^\odot$ by definition (54) of the polar set of X.

2. We prove item 2.

   Since the set C is closed, convex and contains 0, we have $C^{\odot\odot} = C$ [1, Th. 5.103].

   - We prove that statement 2a implies statement 2b.

     The set C is symmetric because $\sigma_C(y) = \sigma_C(-y)$, for all $y \in \mathbb{Y}$, implies that $\sigma_C = \sigma_{-C}$, hence that $-C = C$ since C is closed and convex. The set C is weakly bounded because $\sigma_C < +\infty \iff \mathrm{dom}\sigma_C = \mathbb{Y} \iff \mathrm{bar}C = \mathbb{Y}$ by (52). The set C has full conical hull because $y \in \big(\mathrm{span}C\big)^\perp = C^\perp \Rightarrow \sigma_C(y) = 0 \Rightarrow y = 0$, hence $\mathrm{span}C = \mathbb{X}$; now, as the set C is convex and symmetric, we have that $\mathrm{span}C = \mathrm{cone}C$.
   - By item 1, statement 2b implies statement 2a.
   - We prove that statement 2b implies statement 2c.

     The conditions (60) give

     $$-(C^\odot) = (-C)^\odot = C^\odot \,, \tag{61a}$$

     by $-C = C$ and by definition (54) of the polar set,

     $$\mathrm{cone}(C^\odot) = \mathrm{bar}C = \mathbb{Y} \,, \tag{61b}$$

     by (55b) and the assumption that C weakly bounded,

     $$\mathrm{bar}(C^\odot) = \mathrm{cone}(C^{\odot\odot}) = \mathrm{cone}C = \mathbb{X} \,, \tag{61c}$$

     by $C^{\odot\odot} = C$ and since C has full conical hull.
   - Statement 2c implies statement 2b. Indeed, we use the shown property that statement 2b implies statement 2c, but with $C^\odot$ instead of C, where the polar set $C^\odot$ is closed convex and contains 0. Thus, we obtain statement 2b for $C^{\odot\odot}$, but we have seen that $C^{\odot\odot} = C$.

- Because the polar set $C^{\odot}$ is closed convex and contains 0 , we deduce that statement 2c is equivalent to statement 2d from the shown property that statement 2a is equivalent to statement 2b.

This ends the proof. □

Now, we define the dual norm.

Definition 22 *Let $\mathbb{X}$ and $\mathbb{Y}$ be a dual system. Let $|||\cdot|||$ be a norm on $\mathbb{X}$. If the support function $\sigma_{B_{|||\cdot|||}}$ is a norm (on $\mathbb{Y}$), it is called the* dual norm *of $|||\cdot|||$ and it is denoted by $|||\cdot|||_\star$.*

When a dual norm exists $|||\cdot|||_\star$, then, by item 1 in Proposition 21, its unit ball is the polar set of the original unit ball:

$$B_{|||\cdot|||_\star} = B^{\odot}_{|||\cdot|||} \,. \tag{62}$$

When both the norm $|||\cdot|||$ and the dual norm $|||\cdot|||_\star$ admit a dual norm, the norm $|||\cdot|||_{\star\star} = \big(|||\cdot|||_\star\big)_\star$ (on $\mathbb{X}$) is called the *bidual norm*. We provide a characterization of when a dual norm exists, and of when a bidual norm exists and coincides with the original norm.

Proposition 23 *Let $\mathbb{X}$ and $\mathbb{Y}$ be a dual system. Let $|||\cdot|||$ be a norm on $\mathbb{X}$.*

1. *The following statements are equivalent.*
   (a) *The norm $|||\cdot|||$ admits a dual norm.*
   (b) *The unit ball $B_{|||\cdot|||}$ is weakly bounded.*
2. *The following statements are equivalent.*
   (a) *The norm $|||\cdot|||$ admits a dual norm $|||\cdot|||_\star$, and the dual norm $|||\cdot|||_\star$ has $|||\cdot|||$ for dual norm ($|||\cdot|||_{\star\star} = |||\cdot|||$).*
   (b) *The unit ball $B_{|||\cdot|||}$ is weakly bounded and closed.*

   In that case, each norm is the dual norm of the other norm, the unit balls $B_{|||\cdot|||}$ and $B_{|||\cdot|||_\star}$ are polar to each other, that is,

$$B_{|||\cdot|||} = B^{\odot}_{|||\cdot|||_\star} \;\; \textit{and} \;\; B_{|||\cdot|||_\star} = B^{\odot}_{|||\cdot|||} \,, \tag{63a}$$

   and their support functions satisfy

$$|||\cdot||| = \sigma_{B_{|||\cdot|||_\star}} \;\; \textit{and} \;\; |||\cdot|||_\star = \sigma_{B_{|||\cdot|||}} \,. \tag{63b}$$

Proof.

1. We prove item 1.

   - We prove that statement 1a implies statement 1b.

     Indeed, if the norm $|||\cdot|||$ admits a dual norm, the support function $\sigma_{B_{|||\cdot|||}}$ satisfies $\sigma_{B_{|||\cdot|||}} < +\infty$. Therefore $\mathrm{dom}\sigma_{B_{|||\cdot|||}} = \mathbb{Y}$, meaning that the unit ball $B_{|||\cdot|||}$ is weakly bounded by (52).

   - We prove that statement 1b implies statement 1a

     Indeed, being a unit ball, $B_{|||\cdot|||}$ is convex, symmetric and with full conical hull. Moreover, it is also weakly bounded by assumption. We deduce from item 1 in Proposition 21 that the support function $\sigma_{B_{|||\cdot|||}}$ is a norm on $\mathbb{Y}$, whose unit ball is the polar set $B^{\odot}_{|||\cdot|||}$.

2. Item 2 is a straightforward consequence of item 2 in Proposition 21 with $C = B_{|||\cdot|||}$. Indeed, being a unit ball, $B_{|||\cdot|||}$ is convex, containing 0, symmetric and with full conical hull. Moreover, it is also closed and weakly bounded by assumption.

   The equations (63a)–(63b) are also a straightforward consequence of item 2 in Proposition 21.

This ends the proof. □

A.3.3 Dual norm in the Hilbert space case

Let $\mathbb{X} = \mathbb{Y}$ be a Hilbert space with scalar product $\langle\,,\rangle$, and induced *Hilbertian norm* $\|\cdot\| = \sqrt{\langle\cdot\,,\cdot\rangle}$ and Hilbertian topology. It is easy to see that the dual norm $\|\cdot\|_\star$ of the Hilbertian norm is the Hilbertian norm, that is, $\|\cdot\|_\star = \|\cdot\|$.

When we refer to notions attached to a dual system (support function, weakly bounded set), by default it is the natural dual system induced by the Hilbertian structure.

We study under which assumptions the support function of a set is a norm, and the topology that it induces.

Proposition 24 *Let $C \subset \mathbb{X}$ be closed, convex, symmetric, weakly bounded and with full conical hull (*cone$C = \mathbb{X}$*). Then,*

- *the support function σ_C is a norm on $\mathbb{Y}$, whose unit ball is the polar set $C^\odot$, and $C^\odot$ is closed, convex, symmetric, weakly bounded and with full conical hull,*
- *the support function $\sigma_{C^\odot}$ is a norm on $\mathbb{X}$, whose unit ball is C,*
- *each norm is the dual norm of the other norm,*
- *the topologies induced by both norms are both weaker than the Hilbertian topology.*

The assertions remain true with "weakly bounded" replaced by "bounded" in the two instances where it appears. In that case, the topologies induced by both norms are equivalent to the Hilbertian topology.

Proof. Being convex, the set C is closed in the weak topology,

By Proposition 23, the three first items hold true. We use the property that the set C is closed in the weak topology, and that the set $C^{\odot}$ is closed in the weak topology, hence is closed because it is convex (being a polar set).

Regarding the fourth item, the topologies defined by the norm and by the dual norm are both weaker (contain less open sets) than the Hilbertian topology, because, by construction, their unit balls are closed (for the Hilbertian topology). This results from Proposition 19.

If all the assumptions on $C \subset \mathbb{X}$ are true, except for weakly bounded replaced by bounded, then the three first items hold true because the bounded subset C is weakly bounded, as seen in (57). There remains to prove that the polar set $C^{\odot}$ is bounded. For this purpose, we denote by $|||\cdot|||$ the norm σ_C and we get

$$
\begin{aligned}
B_{|||\cdot|||} = C \text{ is closed } &\Rightarrow 0 \in \mathrm{int}C = \mathrm{int}B_{|||\cdot|||} && \text{(by Proposition 19)} \\
&\iff \exists m > 0\ ,\ \ mB_{\|\cdot\|} \subset C \\
&\Rightarrow C^{\odot} \subset \left(mB_{\|\cdot\|}\right)^{\odot} = \frac{1}{m} B_{\|\cdot\|_\star} \\
& && \text{(by (62) and the definition (54) of a polar set)} \\
&\Rightarrow C^{\odot} \subset \frac{1}{m} B_{\|\cdot\|} \\
& && \text{(because the dual norm } \|\cdot\|_\star \text{ of the Hilbertian norm is the Hilbertian norm)} \\
&\Rightarrow C^{\odot} \text{ is bounded.}
\end{aligned}
$$

We conclude that the topologies induced by both norms are equivalent to the Hilbertian topology, by Proposition 20 because their balls are closed and bounded.

This ends the proof. □

We provide assumptions under which a dual norm exists, and we precise the topologies that norm and dual norm induce.

Proposition 25 *Let $|||\cdot|||$ be a norm on a Hilbert space. If the norm $|||\cdot|||$ is equivalent to the Hilbertian norm $\|\cdot\|$ (or, equivalently, if its unit ball $B_{|||\cdot|||}$ is bounded and closed), then the norm $|||\cdot|||$ admits a dual norm $|||\cdot|||_\star$ which is equivalent to the Hilbertian norm $\|\cdot\|$ (or, equivalently whose unit ball $B_{|||\cdot|||_\star}$ is bounded and closed), and* (63a)–(63b) *hold true.*

Proof. If the norm $|||\cdot|||$ is equivalent to the Hilbertian norm $\|\cdot\|$, then the unit ball $B_{|||\cdot|||}$ is closed and bounded by Proposition 20. Therefore, by Proposition 24 with $C = B_{|||\cdot|||}$, we deduce that the norm $|||\cdot|||$ admits a dual norm $|||\cdot|||_\star$, that the topologies induced by both norms are equivalent to the Hilbertian topology, and that (63a)–(63b) hold true.

This ends the proof. □

A.3.4 Dual norm in the Euclidian case

Proposition 26 *Any norm on $\mathbb{R}^d$ admits a dual norm.*

Proof. We use Proposition 25, as all norms on $\mathbb{R}^d$ are equivalent to the Euclidian norm. □

References

[1] C. D. Aliprantis and K. C. Border. *Infinite dimensional analysis*. Springer-Verlag, Berlin, second edition, 1999.

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

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

[4] J.-P. Chancelier and M. De Lara. A suitable conjugacy for the l_0 pseudonorm, 2019. preprint.

[5] L. Mirsky. Symmetric Gauge Functions and Unitarily Invariant Norms. *The Quarterly Journal of Mathematics*, 11(1):50–59, 01 1960.

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

[7] G. Obozinski and F. Bach. A unified perspective on convex structured sparsity: Hierarchical, symmetric, submodular norms and beyond. working paper or preprint, Dec. 2016.

[8] G. Obozinski, L. Jacob, and J.-P. Vert. Group Lasso with Overlaps: the Latent Group Lasso approach. Research report, Oct. 2011.