Best Convex Lower Approximations of the l 0 Pseudonorm on Unit Balls

Thomas Bittar, Jean-Philippe Chancelier, Michel de Lara

Best Convex Lower Approximations of the l_0 Pseudonorm on Unit Balls

Thomas Bittar and Jean-Philippe Chancelier and Michel De Lara,
CERMICS, École des Ponts, Marne-la-Vallée, France

May 31, 2021

Abstract

Whereas the norm of a vector measures amplitude (and is a 1-homogeneous function), sparsity is measured by the 0-homogeneous l0 pseudonorm, which counts the number of nonzero components. We propose a family of conjugacies suitable for the analysis of 0-homogeneous functions. These conjugacies are derived from couplings between vectors, given by their scalar product divided by a 1-homogeneous normalizing factor. With this, we characterize the best convex lower approximation of a 0-homogeneous function on the unit "ball" of a normalization function (i.e. a norm without the requirement of subadditivity). We do the same with the best convex and 1-homogeneous lower approximation. In particular, we provide expressions for the tightest convex lower approximation of the l0 pseudonorm on any unit ball, and we show that the tightest norm which minorizes the l0 pseudonorm on the unit ball of any lp-norm is the l1-norm. We also provide the tightest convex lower convex approximation of the l0 pseudonorm on the unit ball of any norm.

Key words: l0 pseudonorm, convexity, Capra conjugacy, generalized top-k and k-support norms, sparsity inducing norm.

AMS classification: 46N10, 49N15, 46B99, 52A41, 90C46

1 Introduction

The *counting function*, also called *cardinality function* or ℓ_0 *pseudonorm*, counts the number of nonzero components of a vector in $\mathbb{R}^d$. It is used in sparse optimization, either as objective function or in the constraints, to obtain solutions with few nonzero entries. However, the mathematical expression of the ℓ_0 pseudonorm makes it difficult to handle as such in optimization problems on $\mathbb{R}^d$. This is why most of the literature on sparse optimization resorts to *substitute* or *surrogate* problems, obtained either from *estimates* (inequalities) for the ℓ_0 pseudonorm, or from *alternative* sparsity-inducing terms (especially suitable norms) [2]. We follow this approach, but using (and extending) the so-called Capra (Constant Along Primal RAys) couplings and conjugacies introduced in [3, 4].

The paper is organized as follows. In Section 2, we introduce a new Capra-coupling that extends the definition in [3]. We establish expressions for Capra-conjugates and Capra-subdifferentials of 0-homogeneous functions. Then, in §3.1, we manage to obtain convex lower bounds for general 0-homogeneous functions, thanks to this new coupling. We specialize these results for the ℓ_0 pseudonorm in §3.2 and we obtain, in particular, that the tightest norm below ℓ_0 on any ℓ_p-unit ball ($p \in [1, \infty]$) is the ℓ_1-norm, hence justifying the use of the ℓ_1-norm as a sparsity-inducing term. We also provide the tightest convex lower convex approximation of the l0 pseudonorm on the unit ball of any norm.

2 Capra-conjugacies for 0-homogeneous functions

In §2.1, we recall definitions related to homogeneous functions on $\mathbb{R}^d$, and we introduce a new Capra-coupling between vectors of $\mathbb{R}^d$. This Capra coupling is suited for the analysis of 0-homogeneous functions, for which we provide the expression of Capra-conjugates and Capra-subdifferential in §2.2.

We work on the Euclidean space $\mathbb{R}^d$ (where d is a positive integer), equipped with the scalar product $\langle \cdot, \cdot \rangle$. We use the notation $[\![j,k]\!] = \{j, j+1, \dots, k-1, k\}$ for any pair of integers such that $j \leq k$. As we manipulate functions with values in $\overline{\mathbb{R}} = [-\infty, +\infty]$, we adopt the Moreau *lower and upper additions* [7] that extend the usual addition with $(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = -\infty$ or with $(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = +\infty$. For any subset $W \subset \mathbb{R}^d$, $\delta_W : \mathbb{R}^d \to \overline{\mathbb{R}}$ denotes the *characteristic function* of the set W: $\delta_W(w) = 0$ if $w \in W$, and $\delta_W(w) = +\infty$ if $w \notin W$.

2.1 Definitions

Definition 1 *We say that a function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ is*

(1a) 0-homogeneous *if* $f(\rho x) = f(x)\ ,\ \forall \rho \in \mathbb{R} \setminus \{0\}\ ,\ \forall x \in \mathbb{R}^d\ ,$

(1b) positively 1-homogeneous *if* $f(\rho x) = \rho f(x)\ ,\ \forall \rho \in \mathbb{R}_+ \setminus \{0\}\ ,\ \forall x \in \mathbb{R}^d\ ,$

(1c) absolutely 1-homogeneous *if* $f(\rho x) = |\rho| f(x)\ ,\ \forall \rho \in \mathbb{R} \setminus \{0\}\ ,\ \forall x \in \mathbb{R}^d\ .$

Example 1 *An example of 0-homogeneous function is the* pseudonorm $\ell_0 : \mathbb{R}^d \to [\![0, d]\!]$ *defined by*

$$\ell_0(x) = \textit{number of nonzero components of } x\ ,\ \forall x \in \mathbb{R}^d\ . \tag{2}$$

For any $p \in]0, \infty[$, we define $\ell_p(x) = \left(\sum_{i=1}^d |x_i|^p\right)^{\frac{1}{p}}$, as well as $\ell_\infty(x) = \max_{i \in [\![1,d]\!]} |x_i|$. All these functions ℓ_p are absolutely 1-homogeneous (and therefore also positively 1-homogeneous). When $p \in [1, \infty]$, ℓ_p is a convex function which is the well-known ℓ_p-norm $\|\cdot\|_p$. For $p \in]0, 1[$, ℓ_p is not convex anymore and is only a normalization function *(see below) as it lacks the subadditivity property.*

Definition 2 *A function $\nu : \mathbb{R}^d \to \mathbb{R}_+$ is said to be a* normalization function *if it is a nonnegative, absolutely 1-homogeneous and such that, for any $x \in \mathbb{R}^d$, we have that $\nu(x) = 0$ if and only if $x = 0$. We introduce the subsets, abusively called unit "sphere" and "ball",*

$$\mathbb{S}_\nu = \left\{x \in \mathbb{R}^d \,\middle|\, \nu(x) = 1\right\}\ ,\ \mathbb{S}_\nu^{(0)} = \mathbb{S}_\nu \cup \{0\}\ ,\ \mathbb{B}_\nu = \left\{x \in \mathbb{R}^d \,\middle|\, \nu(x) \leq 1\right\}\ , \tag{3}$$

and the primal normalization mapping[1] *$n_\nu : \mathbb{R}^d \to \mathbb{S}_\nu^{(0)}$ by*

$$n_\nu : x \in \mathbb{R}^d \mapsto \begin{cases} \frac{x}{\nu(x)}\ , & x \neq 0\ , \\ 0\ , & \textit{else.} \end{cases} \tag{4}$$

A normalization function satisfies the same properties as a norm except subadditivity. Thus, the unit "ball" $\mathbb{B}_\nu$ in (3) is not necessarily convex. When $\nu = |||\cdot|||$ is a norm on $\mathbb{R}^d$ (resp. $\nu = ||\cdot||_p$ is the ℓ_p norm, for $p \in [1, +\infty]$), we denote by $\mathbb{B}$ (resp. $\mathbb{B}_p$) the unit ball:

$$\mathbb{B} = \left\{x \in \mathbb{R}^d \,\middle|\, |||x||| \leq 1\right\}\ ,\ \mathbb{B}_p = \left\{x \in \mathbb{R}^d \,\middle|\, ||x||_p \leq 1\right\}\ . \tag{5}$$

[1] We distinguish the normalization *function* with codomain $\mathbb{R}_+$ from the normalization *mapping* with codomain $\mathbb{S}_\nu^{(0)}$. Indeed, adopting usage in mathematics, we follow Serge Lang and use "function" only to refer to mappings in which the codomain is a set of numbers (i.e. a subset of $\mathbb{R}$ or $\mathbb{C}$), and reserve the term mapping for more general codomains.

Capra-couplings and conjugacies We introduce a new coupling, which extends [3, Definition 8], where the normalizing factor was a norm, whereas it is more general in the definition below.

Definition 3 *Let $\nu : \mathbb{R}^d \to \mathbb{R}_+$ be a normalization function. The* constant along primal rays coupling *$\dot{c} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, or* Capra*, between $\mathbb{R}^d$ and itself, associated with ν, is the function*

$$\dot{c} : (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto \langle n_\nu(x),\, y\rangle = \begin{cases} \frac{\langle x,\, y\rangle}{\nu(x)} \,, & x \neq 0 \,, \\ 0 \,, & \textit{else.} \end{cases} \tag{6}$$

The coupling Capra has the property of being constant along primal rays, hence the acronym Capra (Constant Along Primal RAys). This is a special case of a one-sided linear coupling as introduced in [4]. We review concepts and notations related to Fenchel-Moreau conjugacies [12, 11, 6]. The classical Fenchel conjugacy $\star$ is outlined in Appendix A.

Definition 4 *The* $\dot{c}$-Fenchel-Moreau conjugate *of a function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$, with respect to the coupling $\dot{c}$, is the function $f^{\dot{c}} : \mathbb{R}^d \to \overline{\mathbb{R}}$ defined by*

$$f^{\dot{c}}(y) = \sup_{x \in \mathbb{R}^d} \Big(\dot{c}(x, y) \dot{+} \big(-f(x)\big) \Big) \,, \quad \forall y \in \mathbb{R}^d \,. \tag{7a}$$

With $\star'$ the Fenchel conjugate as defined in (12b)*, we denote*

$$f^{\dot{c}\star'} = \big(f^{\dot{c}}\big)^{\star'} \,. \tag{7b}$$

The $\dot{c}$-Fenchel-Moreau biconjugate *of a function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$, with respect to the coupling $\dot{c}$, is the function $f^{\dot{c}\dot{c}'} : \mathbb{R}^d \to \overline{\mathbb{R}}$ defined by*[2]

$$f^{\dot{c}\dot{c}'}(x) = \sup_{y \in \mathbb{R}^d} \Big(\dot{c}(x, y) \dot{+} \big(-f^{\dot{c}}(y)\big) \Big) \,, \quad \forall x \in \mathbb{R}^d \,. \tag{7c}$$

The Fenchel-Moreau biconjugate of a function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ satisfies $f^{\dot{c}\dot{c}'} \leq f$. We define the Capra-*subdifferential* of a function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ at $x \in \mathbb{R}^d$ by [1]

$$\partial_{\dot{c}} f(x) = \big\{ y \in \mathbb{R}^d \,\big|\, f^{\dot{c}}(y) = \dot{c}(x, y) \dot{+} \big(-f(x)\big) \big\} \tag{8}$$

2.2 Capra-conjugates and subdifferentials of 0-homogeneous functions

We now provide expressions for the Capra-conjugates and subdifferentials of 0-homogeneous functions.

Proposition 5 *Let $\nu : \mathbb{R}^d \to \mathbb{R}_+$ be a normalization function and $\dot{c}$ be the associated coupling* (6)*. For any 0-homogeneous function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$, we have that (with $\star$ the Fenchel conjugate as defined in* (12a)*),*

$$f^{\dot{c}} = \big(f \dot{+} \delta_{\mathbb{B}_\nu}\big)^\star = \big(f \dot{+} \delta_{\mathbb{S}_\nu^{(0)}}\big)^\star \,. \tag{9}$$

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

$$f^{\dot{c}}(y) = \sup_{x \in \mathbb{R}^d} \Big(\dot{c}(x, y) \dot{+} \big(-f(x)\big) \Big) = \sup_{x' \in \mathbb{R}^d} \Big(\big\langle n_\nu(x'),\, y \big\rangle \dot{+} \big(-f\big(n_\nu(x')\big)\big) \Big)$$

by definition (6) of $\dot{c}(x, y)$, by definition (4) of the normalization mapping n_ν and by 0-homogeneity (1a) of the function f (including the case $x = 0$)

$$= \sup_{s \in \mathbb{S}_\nu^{(0)}} \Big(\langle s,\, y \rangle \dot{+} \big(-f(s)\big) \Big)$$

[2] With the coupling $\dot{c}$, we associate the *reverse coupling* $\dot{c}'$ defined by $\dot{c}'(y, x) = \dot{c}(x, y)$ for $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, hence the notation $f^{\dot{c}\dot{c}'}$.

by definition (3) of $\mathbb{S}_\nu^{(0)}$ and as $n_\nu(\mathbb{R}^d) \subset \mathbb{S}_\nu^{(0)} = n_\nu(\mathbb{S}_\nu^{(0)}) \subset n_\nu(\mathbb{R}^d)$ using the positive homogeneity of ν and the property that $\nu(x) \neq 0$ if $x \neq 0$

$$\begin{aligned} &= \sup_{x \in \mathbb{R}^d} \Big(\langle x,\, y \rangle \dot{+} \big(-\big(f \dot{+} \delta_{\mathbb{S}_\nu^{(0)}} \big)(x) \big) \Big) \\ &= \big(f \dot{+} \delta_{\mathbb{S}_\nu^{(0)}} \big)^\star(y) \,. \qquad \text{(by (12a))} \end{aligned}$$

We have shown that $f^{\dot{\subset}} = \big(f \dot{+} \delta_{\mathbb{S}_\nu^{(0)}} \big)^\star$. We now prove that $\big(f \dot{+} \delta_{\mathbb{B}_\nu} \big)^\star = \big(f \dot{+} \delta_{\mathbb{S}_\nu^{(0)}} \big)^\star$.

On the one hand, as $\mathbb{B}_\nu \supset \mathbb{S}_\nu^{(0)}$ implies that $f \dot{+} \delta_{\mathbb{B}_\nu} \leq f \dot{+} \delta_{\mathbb{S}_\nu^{(0)}}$, we deduce that $\big(f \dot{+} \delta_{\mathbb{B}_\nu} \big)^\star \geq \big(f \dot{+} \delta_{\mathbb{S}_\nu^{(0)}} \big)^\star$ by taking the order-reversing Fenchel conjugate (12a). On the other hand, we fix $y \in \mathbb{R}^d$ and we show that, for any $x \in \mathbb{B}_\nu$, there exists $s \in \mathbb{S}_\nu^{(0)}$ such that $\langle x,\, y \rangle \leq \langle s,\, y \rangle$ and $f(x) = f(s)$. Indeed, it suffices to take $s = 0$ when $x = 0$, $s = x/\nu(x)$ when $x \neq 0$ and $\langle x,\, y \rangle \geq 0$, and $s = -x/\nu(-x)$ when $x \neq 0$ and $\langle x,\, y \rangle \leq 0$ (by using $\nu(x) = \nu(-x) \leq 1$ as $x \in \mathbb{B}_\nu$ and the 0-homogeneity (1a) of the function f). We deduce that

$$\big(f \dot{+} \delta_{\mathbb{B}_\nu} \big)^\star(y) = \sup_{x \in \mathbb{B}_\nu} \Big(\langle x,\, y \rangle \dot{+} \big(-f(x) \big) \Big) \leq \sup_{s \in \mathbb{S}_\nu^{(0)}} \Big(\langle s,\, y \rangle \dot{+} \big(-f(s) \big) \Big) = \big(f \dot{+} \delta_{\mathbb{S}_\nu^{(0)}} \big)^\star(y) \,.$$

This ends the proof. □

Whereas Proposition 5 relates the Capra-conjugate of 0-homogeneous functions with the classical Fenchel-Moreau conjugate, in Proposition 6, we relate the Capra-subdifferential with the well-known Rockafellar-Moreau subdifferential.

Proposition 6 *Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be any function. For all $x \in \mathbb{R}^d$, the* Capra*-subdifferential $\partial_{\dot{\subset}} f(x)$, as in* (8)*, is a closed convex set. Moreover, if the function f is 0-homogeneous, we have that*

$$\begin{cases} \partial_{\dot{\subset}} f(0) = \partial(f \dot{+} \delta_{\mathbb{B}_\nu})(0) \,, & \text{(10a)} \\ s \in \mathbb{S}_\nu \text{ and } f^{\dot{\subset}\star'}(s) = f(s) \implies \partial_{\dot{\subset}} f(s) = \partial f^{\dot{\subset}\star'}(s) \,, & \text{(10b)} \end{cases}$$

where $\mathbb{B}_\nu$ is defined in (3) *for the normalization function ν that generates the coupling $\dot{\subset}$, and $\partial(f \dot{+} \delta_{\mathbb{B}_\nu})$ and $\partial f^{\dot{\subset}\star'}$ are (Rockafellar-Moreau) subdifferentials as in* (13)*.*

Proof. We prove that $\partial_{\dot{\subset}} f(x)$ is a closed convex set. Let $x \in \mathbb{R}^d$ and first suppose that $f(x) = -\infty$. Using (7a) and (8), we can check that $\partial_{\dot{\subset}} f(x) = \mathbb{R}^d$, which is closed and convex. In the case where $f(x) = +\infty$, we have that $\partial_{\dot{\subset}} f(x) = \emptyset$ if f is not identically $+\infty$, and that $\partial_{\dot{\subset}} f(x) = \mathbb{R}^d$ otherwise; in either cases, the Capra-subdifferential is closed and convex. Now, suppose that $f(x) \in \mathbb{R}$. By definition (7a) of $f^{\dot{\subset}}$, the Capra-subdifferential (8) can be written as $\partial_{\dot{\subset}} f(x) = \big\{ y \in \mathbb{R}^d \,\big|\, f^{\dot{\subset}}(y) \leq \dot{\subset}(x, y) \dot{+} \big(-f(x) \big) \big\}$ where the function $f^{\dot{\subset}} = \big(f \dot{+} \delta_{\mathbb{S}_\nu^{(0)}} \big)^\star$ is a Fenchel conjugate by (9) hence is closed convex (see the background material in Appendix A), and the function $g_x : y \mapsto \dot{\subset}(x, y) \dot{+} \big(-f(x) \big) = \dot{\subset}(x, y) - f(x)$ is affine. As a consequence, the set of points where $f^{\dot{\subset}} \leq g_x$, which is exactly $\partial_{\dot{\subset}} f(x)$, is a closed convex set.

We prove (10a) as follows:

$$y \in \partial(f \dot{+} \delta_{\mathbb{B}_\nu})(0) \iff \big(f \dot{+} \delta_{\mathbb{B}_\nu} \big)^\star(y) = \langle 0,\, y \rangle \dot{+} \big(-(f \dot{+} \delta_{\mathbb{B}_\nu})(0) \big)$$

by definition (13) of the (Rockafellar-Moreau) subdifferential of a function

$$\begin{aligned} &\iff f^{\dot{\subset}}(y) = \langle 0,\, y \rangle \dot{+} \big(-f(0) \big) && \text{(by Proposition 5)} \\ &\iff f^{\dot{\subset}}(y) = \dot{\subset}(0, y) \dot{+} \big(-f(0) \big) && \text{(since since } \dot{\subset}(0, y) = 0 \text{ by (6))} \\ &\iff y \in \partial_{\dot{\subset}} f(0) \,. && \text{(by definition (8) of the Capra-subdifferential)} \end{aligned}$$

Letting $s \in \mathbb{S}_\nu$ be such that $f^{\dot{\varsigma}\star'}(s) = f(s)$, we prove (10b) as follows:

$$\begin{aligned} y \in \partial f^{\dot{\varsigma}\star'}(s) &\iff \big(f^{\dot{\varsigma}\star'}\big)^\star(y) = \langle s,\, y\rangle \mathbin{\dot{+}} \big(-f^{\dot{\varsigma}\star'}(s)\big) && \text{(by (13))} \\ &\iff f^{\dot{\varsigma}}(y) = \langle s,\, y\rangle \mathbin{\dot{+}} \big(-f^{\dot{\varsigma}\star'}(s)\big) \end{aligned}$$

because the function $f^{\dot{\varsigma}} = \big(f \mathbin{\dot{+}} \delta_{\mathbb{S}_\nu^{(0)}}\big)^\star$ is a Fenchel conjugate by (9), hence is closed convex, and therefore equal to its Fenchel biconjugate $f^{\dot{\varsigma}\star'\star}$, (see the background material in Appendix A)

$$\iff f^{\dot{\varsigma}}(y) = \dot{\varsigma}(s, y) \mathbin{\dot{+}} \big(-f^{\dot{\varsigma}\star'}(s)\big)$$

by definition (6) of $\dot{\varsigma}(s, y)$ as $s \in \mathbb{S}_\nu$ hence $\nu(x) = 1$ by (3),

$$\begin{aligned} &\iff f^{\dot{\varsigma}}(y) = \dot{\varsigma}(s, y) \mathbin{\dot{+}} \big(-f(s)\big) && \text{(by assumption that } f^{\dot{\varsigma}\star'}(s) = f(s)\text{)} \\ &\iff y \in \partial_{\dot{\varsigma}} f(s)\ . && \text{(by definition (8) of the CAPRA-subdifferential)} \end{aligned}$$

This ends the proof. □

Now, we are going to show how CAPRA-couplings are a suitable tool to obtain lower convex approximations of 0-homogeneous functions.

3 Best convex lower approximations of 0-homogeneous functions

In §3.1, we identify — in terms of CAPRA-conjugacy introduced in Section 2 — the best lower approximation of 0-homogeneous functions by convex and by positively 1-homogeneous convex functions. We apply these results to the ℓ_0 pseudonorm in §3.2.

3.1 General result

We prove that the best convex lower approximation of a 0-homogeneous function can be expressed in term of its CAPRA-conjugate (7b). We also prove that the best positively 1-homogeneous closed convex lower approximation of a 0-homogeneous function can be expressed in term of its CAPRA-subdifferential (8). We recall that, for any subset $Y \subset \mathbb{R}^d$, $\sigma_Y : \mathbb{R}^d \to \overline{\mathbb{R}}$ denotes the *support function of the subset* Y: $\sigma_Y(x) = \sup_{y\in Y} \langle x,\, y\rangle$, for any $x \in \mathbb{R}^d$.

The proof of the following theorem relies on results given in Appendix B.

Theorem 7 *Let $\nu : \mathbb{R}^d \to \mathbb{R}_+$ be a normalization function, with unit "ball" $\mathbb{B}_\nu$ defined in* (3)*, and with associated* CAPRA*-coupling $\dot{\varsigma}$ in* (6)*. Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be a 0-homogeneous function. Then,*

1. *the function $f^{\dot{\varsigma}\star'}$ is the tightest closed convex function below f on the unit "ball" $\mathbb{B}_\nu$,*
2. *if $f(0) = 0$, the function $\sigma_{\partial_{\dot{\varsigma}} f(0)}$ is the tightest closed convex positively 1-homogeneous function below f on the unit "ball" $\mathbb{B}_\nu$.*

Proof. From Proposition 10 with $W = \mathbb{B}_\nu$, the tightest closed convex function below f on the unit "ball" $\mathbb{B}_\nu$ is $\big(f \mathbin{\dot{+}} \delta_{\mathbb{B}_\nu}\big)^{\star\star'}$. As the function f is 0-homogeneous, by Proposition 5, we have that $f^{\dot{\varsigma}} = \big(f \mathbin{\dot{+}} \delta_{\mathbb{B}_\nu}\big)^\star$. By taking the Fenchel conjugate (12b), we deduce that $\big(f \mathbin{\dot{+}} \delta_{\mathbb{B}_\nu}\big)^{\star\star'} = f^{\dot{\varsigma}\star'}$.

From Proposition 12 with $W = \mathbb{B}_\nu$, the tightest positively 1-homogeneous closed convex function below f on the unit "ball" $\mathbb{B}_\nu$ is $\sigma_{\partial(f \dot{+} \delta_{\mathbb{B}_\nu})(0)}$. As the function f is 0-homogeneous, by Proposition 6, we have $\partial(f \mathbin{\dot{+}} \delta_{\mathbb{B}_\nu})(0) = \partial_{\dot{\varsigma}} f(0)$ and therefore $\sigma_{\partial(f \dot{+} \delta_{\mathbb{B}_\nu})(0)} = \sigma_{\partial_{\dot{\varsigma}} f(0)}$, which gives the desired result. □

3.2 Application to the ℓ_0 pseudonorm

As an application, we study the particular case of the ℓ_0 pseudonorm. For any $x \in \mathbb{R}^d$ and subset $K \subset \{1,\dots,d\}$, we denote by $x_K \in \mathbb{R}^d$ the vector which coincides with x, except for the components outside of K that vanish. Let $|||\cdot|||$ be a norm on $\mathbb{R}^d$, called the *source norm*. For any subset $K \subset [\![1,d]\!]$, we define the subspace $\mathcal{R}_K = \left\{x \in \mathbb{R}^d \,\middle|\, x_j = 0\ ,\ \forall j \notin K\right\}$ of $\mathbb{R}^d$. We introduce the *K-restriction norm* $|||\cdot|||_K$, defined by $|||x|||_K = |||x|||$, for any $x \in \mathcal{R}_K$. We also define the $(K,\star)$-norm $|||\cdot|||_{K,\star}$, which is the norm $\left(|||\cdot|||_K\right)_\star$, given by the dual norm (on the subspace $\mathcal{R}_K$) of the restriction norm $|||\cdot|||_K$ to the subspace $\mathcal{R}_K$ (first restriction, then dual). Following [3, Definition 3]), for $k \in [\![1,d]\!]$, we call *generalized coordinate-k norm* the norm $|||\cdot|||_{(k)}^{\mathcal{R}}$ whose dual norm is the *generalized dual coordinate-k norm*, denoted by $|||\cdot|||_{(k),\star}^{\mathcal{R}}$, with expression $|||y|||_{(k),\star}^{\mathcal{R}} = \sup_{|K|\leq k} |||y_K|||_{K,\star}$ for all $y \in \mathbb{R}^d$. We denote by $\mathbb{B}_{(k)}^{\mathcal{R}}$ and $\mathbb{B}_{(k),\star}^{\mathcal{R}}$ the respective unit balls. In Table 1, we give examples of coordinate-k and dual coordinate-k norms in the case of the ℓ_p source norms. We recall that, for $p \in [1,\infty]$, the dual norm of the ℓ_p-norm $\|\cdot\|_p$ is the ℓ_q-norm $\|\cdot\|_q$, where q is such that $1/p + 1/q = 1$ (with the extreme cases $q = \infty$ when $p = 1$, and $q = 1$ when $p = \infty$).

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

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

Best convex lower approximation of the ℓ_0 pseudonorm on a unit ball Let $|||\cdot|||$ be a source norm on $\mathbb{R}^d$ and $\varphi : [\![0,d]\!] \to \overline{\mathbb{R}}$ be a function. From Theorem 7, the function $\left((\varphi \circ \ell_0)^{\dot{\varsigma}}\right)^{\star'}$ is the tightest closed convex function below $\varphi \circ \ell_0$ on the unit ball $\mathbb{B}$, defined in (5). We refer the reader to [3, Proposition 12], where several expressions of the function $\left((\varphi \circ \ell_0)^{\dot{\varsigma}}\right)^{\star'}$ are provided. In particular, it is shown that it is the largest closed convex function below the integer valued function $\mathbb{B}_{(j)}^{\mathcal{R}} \backslash \mathbb{B}_{(j-1)}^{\mathcal{R}} \ni x \mapsto \varphi(j)$ for $l \in [\![1,d]\!]$, and $x \in \mathbb{B}_{(0)}^{\mathcal{R}} = \{0\} \mapsto \varphi(0)$, the function being infinite outside $\mathbb{B}_{(d)}^{\mathcal{R}} = \mathbb{B}$. In dimension 2, this is seen in Figure 1 which depicts the tightest closed convex function below ℓ_0 on the Euclidean unit ball on $\mathbb{R}^2$: the function goes up from zero to the value 1 on the border (sphere) of the lozenge unit ball $\mathbb{B}_1 = \mathbb{B}_{(1)}^{\mathcal{R}}$ of the ℓ_1-norm; then, from the lozenge to the value 2 on the border (sphere) of the round unit ball $\mathbb{B}_2 = \mathbb{B}_{(2)}^{\mathcal{R}}$ of the ℓ_2-norm; the function is discontinuous on the four "sparse" points on the unit Euclidean circle, and takes the value $+\infty$ outside the unit disk.

On the square unit ball $\mathbb{B}_\infty$ of the ℓ_∞-norm, we obtain that the best convex lower approximation on $\mathbb{B}_\infty$ is the ℓ_1-norm. Indeed, if $\dot{\varsigma}$ is the CAPRA-coupling (6) associated with the ℓ_∞-norm, this best approximation

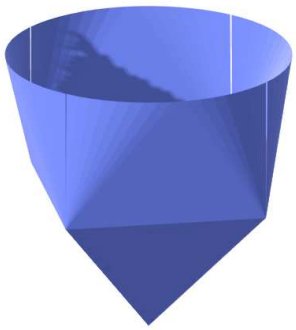

Figure 1: Tightest closed convex function below the ℓ_0 pseudonorm on the Euclidean unit ball on $\mathbb{R}^2$

is $\ell_0^{\dot{\zeta}\star'}$ by Theorem 7. Now, using [3, Proposition 12], we get that $\ell_0^{\dot{\zeta}\star'} = \left(\sup_{j\in[\![0,d]\!]} \left[\|\cdot\|_{(j),\star}^{\mathcal{R}} - j\right]\right)^{\star'} = \left(\sup_{j\in[\![0,d]\!]} \left[||\cdot||_{1,j}^{\mathrm{tn}} - j\right]\right)^{\star'}$ (see Table 1). As $\sup_{j\in[\![0,d]\!]} \left[||y||_{1,j}^{\mathrm{tn}} - j\right] = \sum_{i=1}^{d}(1-|y_i|)\mathbf{1}_{|y_i|\geq 1}$, we get that

$$\forall x \in \mathbb{R}^d\ ,\ \ell_0^{\dot{\zeta}\star'}(x) = \sum_{i=1}^{d} \sup_{y_i\in\mathbb{R}} \left(x_i y_i + (1-|y_i|)\mathbf{1}_{|y_i|\geq 1}\right) = \begin{cases} ||x||_1\ , & x \in \mathbb{B}_\infty\ , \\ +\infty\ , & \text{otherwise.} \end{cases}$$

Best norm lower approximation of the ℓ_0 pseudonorm on a unit ball We compute the expression of the best lower approximation of the ℓ_0 pseudonorm with a norm. Then, we show that for the ℓ_p source norm, with $p \in [1,\infty]$, the tightest norm below ℓ_0 on the unit ball $\mathbb{B}_p$ is the ℓ_1-norm.

Proposition 8 *Let $\|\cdot\|$ be a source norm on $\mathbb{R}^d$, with associated sequence $\left\{\|\cdot\|_{(j)}^{\mathcal{R}}\right\}_{j\in[\![1,d]\!]}$ of coordinate-k norms and sequence $\left\{\|\cdot\|_{\star(j)}^{\mathcal{R}}\right\}_{j\in[\![1,d]\!]}$ of dual coordinate-k norms and their respective unit balls $\left\{\mathbb{B}_{(j),\star}^{\mathcal{R}}\right\}_{j\in[\![1,d]\!]}$. Let $\varphi : [\![0,d]\!] \to \mathbb{R}_+ \cup \{+\infty\}$ be a function such that $\varphi(j) > \varphi(0) = 0$ for all $j \in [\![1,d]\!]$, and with $\varphi(j) < +\infty$ for at least one $j \in [\![1,d]\!]$. Then, there exists a norm $\|\cdot\|_{(\varphi)}^{\mathcal{R}}$ on $\mathbb{R}^d$, characterized by its dual norm $\|\cdot\|_{(\varphi),\star}^{\mathcal{R}}$, which has unit ball $\mathbb{B}_{(\varphi),\star}^{\mathcal{R}} = \bigcap_{j\in[\![1,d]\!]} \varphi(j)\mathbb{B}_{(j),\star}^{\mathcal{R}}$, and $\|\cdot\|_{(\varphi)}^{\mathcal{R}} = \sigma_{\mathbb{B}_{(\varphi),\star}^{\mathcal{R}}}$, where by convention $+\infty\mathbb{B}_{(j),\star}^{\mathcal{R}} = \mathbb{R}^d$. The norm $\|\cdot\|_{(\varphi)}^{\mathcal{R}}$ is the tightest norm below $\varphi\circ\ell_0$ on the unit ball $\mathbb{B}$, that is,*

$$\underbrace{\|x\|_{(\varphi)}^{\mathcal{R}}}_{\text{tightest norm}} \leq \varphi\big(\ell_0(x)\big)\ ,\ \forall x \in \mathbb{B}\ . \tag{11}$$

Proof. From [3, Proposition 14], we have that $\partial_{\dot{\zeta}}(\varphi\circ\ell_0)(0) = \mathbb{B}_{(\varphi),\star}^{\mathcal{R}}$. With Theorem 7, we deduce that $\sigma_{\mathbb{B}_{(\varphi),\star}^{\mathcal{R}}} = \|\cdot\|_{(\varphi)}^{\mathcal{R}}$ is the tightest closed convex positively 1-homogeneous function below $\varphi\circ\ell_0$ on the unit ball $\mathbb{B}$. We can easily check that $\sigma_{\mathbb{B}_{(\varphi),\star}^{\mathcal{R}}}$ indeed defines a norm. The norm $\|\cdot\|_{(\varphi)}^{\mathcal{R}}$ was introduced in [3, Proposition 15] under the slightly stronger assumption that $\varphi(j) < +\infty$ for all $j \in [\![1,d]\!]$. □

Proposition 9 *(Application with the ℓ_p source norm). Let $\|\cdot\| = ||\cdot||_p$, with $p \in [1,+\infty]$, be the source norm and let $\varphi : [\![0,d]\!] \to \mathbb{R}_+ \cup \{+\infty\}$ with $\varphi(j) > \varphi(0) = 0$ for all $j \in [\![1,d]\!]$ and $\varphi(j) < +\infty$ for at least one $j \in [\![1,d]\!]$. We also assume that either, for $p \in]1,+\infty]$ and q such that $1/p + 1/q = 1$, the function $[\![1,d]\!] \ni j \mapsto \varphi(j)^q/j$ is nondecreasing, or, for $p = 1$ and $q = +\infty$, the function φ is nondecreasing. Then, the tightest norm below $\varphi\circ\ell_0$ on the unit ball $\mathbb{B}_p$ is $\|\cdot\|_{(\varphi)}^{\mathcal{R}} = \varphi(1)||\cdot||_1$. In particular, the ℓ_1-norm is the tightest norm below ℓ_0 on the unit ball $\mathbb{B}_p$ of any ℓ_p-norm, for $p \in [1,\infty]$.*

Proof. From Proposition 8, the tightest norm below $\varphi \circ \ell_0$ on the unit ball $\mathbb{B}_p$ is $\|\!|\cdot|\!\|^{\mathcal{R}}_{(\varphi)}$ and, for all $y \in \mathbb{R}^d$, we have that $\|\!|y|\!\|^{\mathcal{R}}_{(\varphi),\star} = \sup_{j\in[\![1,d]\!]} \frac{\|\!|y|\!\|^{\mathcal{R}}_{(j),\star}}{\varphi(j)}$ (see [3, Proposition 15]). The source norm is $||\cdot||_p$. If $p = 1$, then from Table 1, $\|\!|y|\!\|^{\mathcal{R}}_{(j),\star} = ||y||^{\mathrm{tn}}_{\infty,j} = ||y||_\infty$. We deduce that $\|\!|y|\!\|^{\mathcal{R}}_{(\varphi),\star} = \sup_{j\in[\![1,d]\!]} \frac{||y||_\infty}{\varphi(j)} = \frac{||y||_\infty}{\varphi(1)}$ as φ is nondecreasing. For $p > 1$, from Table 1, we get that $\|\!|y|\!\|^{\mathcal{R}}_{(j),\star} = ||y||^{\mathrm{tn}}_{q,j} = \left(\sum_{l=1}^{j} |y_{\tau(l)}|^q\right)^{\frac{1}{q}}$. Using that $j \mapsto \varphi(j)^q/j$ is nondecreasing, it can be computed that $j \mapsto ||y||^{\mathrm{tn}}_{q,j}/\varphi(j)$ is nonincreasing as follows:

$$\left(\frac{||y||^{\mathrm{tn}}_{q,j+1}}{\varphi(j+1)}\right)^q = \frac{\left(\sum_{l=1}^{j} |y_{\tau(l)}|^q\right) + |y_{\tau(j+1)}|^q}{\varphi(j+1)^q} \leq (1+\frac{1}{j})\frac{\left(\sum_{l=1}^{j} |y_{\tau(l)}|^q\right)}{\varphi(j+1)^q} \leq \frac{\left(\sum_{l=1}^{j} |y_{\tau(l)}|^q\right)}{\varphi(j)^q} = \left(\frac{||y||^{\mathrm{tn}}_{q,j}}{\varphi(j)}\right)^q .$$

Therefore, $\|\!|y|\!\|^{\mathcal{R}}_{(\varphi),\star} = \sup_{j\in[\![1,d]\!]} \frac{\|\!|y|\!\|^{\mathcal{R}}_{(j),\star}}{\varphi(j)} = \frac{||y||^{\mathrm{tn}}_{q,1}}{\varphi(1)} = \frac{||y||_\infty}{\varphi(1)}$, as $||y||^{\mathrm{tn}}_{q,1} = |y_{\tau(1)}| = ||y||_\infty$. Thus, for $p \in [1,\infty]$, we have obtained $\|\!|y|\!\|^{\mathcal{R}}_{(\varphi),\star} = \frac{||y||_\infty}{\varphi(1)}$ from which we deduce that $\|\!|\cdot|\!\|^{\mathcal{R}}_{(\varphi)} = \varphi(1)||\cdot||_1$. The last statement of the theorem follows by taking φ being the identity mapping. This concludes the proof. □

Fazel [5, Theorem 1, §5.1.4] shows that the best convex lower approximation of the rank function on the spectral norm unit ball is given by the nuclear norm. By considering singular values, we easily deduce that the best convex lower approximation of the ℓ_0 pseudonorm on the ℓ_∞ unit ball is given by the ℓ_1-norm. Thus, in a sense, Proposition 9 generalizes the result of Fazel.

4 Conclusion

In this paper, we have extended the Capra couplings and conjugacies introduced in [3, 4]. Indeed, they now depend on a given normalization function (i.e. a norm without the requirement for subadditivity). With this new coupling, we are able to provide expressions for the best convex (and positively 1-homogeneous) lower approximations of 0-homogeneous functions on the unit "ball" of the normalization function. As an application, we show that the best norm lower approximation of the ℓ_0 pseudonorm on any ℓ_p unit ball, $p \in [1, +\infty]$, is the ℓ_1 norm, thus strengthening theoretical grounds for the use of the ℓ_1 norm as a sparsity-inducing term in optimization problems. We have also provided the tightest convex lower convex approximation of the l0 pseudonorm on the unit ball of any norm.

A Background on the Fenchel conjugacy

We review concepts and notations related to the Fenchel conjugacy (we refer the reader to [8]). For any function $h : \mathbb{R}^d \to \overline{\mathbb{R}}$, its *epigraph* is $\mathrm{epi}h = \left\{(w,t) \in \mathbb{R}^d \times \mathbb{R} \,\middle|\, h(w) \leq t\right\}$, its *effective domain* is $\mathrm{dom}h = \left\{w \in \mathbb{R}^d \,\middle|\, h(w) < +\infty\right\}$. A function $h : \mathbb{R}^d \to \overline{\mathbb{R}}$ is said to be *convex* if its epigraph is a convex set, *proper* if it never takes the value $-\infty$ and that $\mathrm{dom}h \neq \emptyset$, *lower semi continuous (lsc)* if its epigraph is closed, *closed* if it is either lsc and nowhere having the value $-\infty$, or is the constant function $-\infty$ [8, p. 15]. Closed convex functions are the two constant functions $-\infty$ and $+\infty$ united with all proper convex lsc functions.[3] It is proved that the Fenchel conjugacy (indifferently $f \mapsto f^\star$ or $g \mapsto g^{\star'}$ as below) induces a one-to-one correspondence between the closed convex functions on $\mathbb{R}^d$ and themselves [8, Theorem 5]. For any functions $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^d \to \overline{\mathbb{R}}$, we denote[4]

$$f^\star(y) = \sup_{x\in\mathbb{R}^d} \Big(\langle x,\, y\rangle \dot{+} \big(-f(x)\big)\Big) \,,\ \ \forall y \in \mathbb{R}^d \,, \tag{12a}$$

$$g^{\star'}(x) = \sup_{y\in\mathbb{Y}} \Big(\langle x,\, y\rangle \dot{+} \big(-g(y)\big)\Big) \,,\ \ \forall x \in \mathbb{R}^d \,, \tag{12b}$$

$$f^{\star\star'}(x) = \sup_{y\in\mathbb{R}^d} \Big(\langle x,\, y\rangle \dot{+} \big(-f^\star(y)\big)\Big) \,,\ \ \forall x \in \mathbb{R}^d \,. \tag{12c}$$

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

[4] In convex analysis, one does not use the notation $\star'$ in (12b) and $\star\star'$ in (12c), but simply $\star$ and $\star\star$. We use $\star'$ and $\star\star'$ to be consistent with the notation (7c) for general conjugacies.

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

$$f^{\star\star'}(x) \leq f(x) \ , \ \ \forall x \in \mathbb{R}^d \ . \tag{12d}$$

In [10, p. 214-215], the notions[5] of (Moreau) subgradient and of (Rockafellar) subdifferential are defined for a convex function. Following the definition of the subdifferential of a function with respect to a duality in [1], we define the *(Rockafellar-Moreau) subdifferential* $\partial f(x)$ of a function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ at $x \in \mathbb{R}^d$ by

$$\partial f(x) = \big\{ y \in \mathbb{R}^d \,\big|\, f^\star(y) = \langle x,\, y\rangle \dot{+} \big(-f(x)\big) \big\} \ . \tag{13}$$

When the function f is proper convex and $x \in \mathrm{dom} f$, we recover the classic definition.

B Best convex lower approximations of a function on a subset

We compute the best convex (and convex positively 1-homogeneous) lower approximations of a general function on a (not necessarily convex) subset $W \subset \mathbb{R}^d$.

Proposition 10 *For any subset $W \subset \mathbb{R}^d$ and any function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$, the best closed convex lower approximation $\hat{f}$ of f on W is given by $\hat{f} = \big(f \dot{+} \delta_W\big)^{\star\star'}$.*

Proof. For any subset $W \subset \mathbb{R}^d$ and any function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$, we define the set of functions

$$\mathcal{C}_W^f = \big\{ \text{closed convex function } h : \mathbb{R}^d \to \overline{\mathbb{R}} \,\big|\, h(x) \leq f(x) \ , \ \ \forall x \in W \big\}. \tag{14}$$

As the set $\mathcal{F} = \big\{\text{function } f : \mathbb{R}^d \to \overline{\mathbb{R}}\big\}$ endowed with the partial order $\leq$ is a complete lattice, the subset $\mathcal{C}_W^f \subset \mathcal{F}$ has a (unique) supremum $\hat{f} = \bigvee \mathcal{C}_W^f$. As the set of closed convex functions is stable by pointwise supremum, the function $\hat{f} : \mathbb{R}^d \to \overline{\mathbb{R}}$ is given, for all $x \in W$, by $\hat{f}(x) = \sup \big\{ h(x) \,\big|\, h \in \mathcal{C}_W^f \big\}$.

By (12d), it is easily established that $\big(f \dot{+} \delta_W\big)^{\star\star'} \in \mathcal{C}_W^f$. For any $h \in \mathcal{C}_W^f$, we have $h \dot{+} \delta_{\overline{\mathrm{co}}W} \in \mathcal{C}_W^f$ and $h \dot{+} \delta_{\overline{\mathrm{co}}W} \leq \big(f \dot{+} \delta_W\big)^{\star\star'}$, as $\delta_{\overline{\mathrm{co}}W} \leq \delta_W$ where $\overline{\mathrm{co}}W$ denotes the closed convex hull of the set W. As $h \leq h \dot{+} \delta_{\overline{\mathrm{co}}W}$, we conclude that $\hat{f} = \bigvee \mathcal{C}_W^f \leq \big(f \dot{+} \delta_W\big)^{\star\star'}$, hence that $\hat{f} = \big(f \dot{+} \delta_W\big)^{\star\star'}$ since $\big(f \dot{+} \delta_W\big)^{\star\star'} \in \mathcal{C}_W^f$. □

Remark 11 *We cannot replace δ_W by $\delta_{\overline{\mathrm{co}}W}$ in the statement of Proposition 10, although $\delta_{\overline{\mathrm{co}}W}$ is involved in the proof. For instance, consider $f = |\cdot|$ being the absolute value function defined on $\mathbb{R}$ and $W =]-\infty, -1] \cup [1, +\infty[$, hence $\overline{\mathrm{co}}W = \mathbb{R}$. We can easily check that $\big(f \dot{+} \delta_W\big)^{\star\star'}(x) = |x|$ if $x \in W$ and $\big(f \dot{+} \delta_W\big)^{\star\star'}(x) = 1$ if $x \in \mathbb{R} \setminus W$, that is, $\hat{f} = \max\{1, |\cdot|\}$. On the other hand, we have $\big(f \dot{+} \delta_{\overline{\mathrm{co}}W}\big)^{\star\star'} = f^{\star\star'} = |\cdot|$, and therefore $\big(f \dot{+} \delta_{\overline{\mathrm{co}}W}\big)^{\star\star'} = |\cdot| \neq \max\{1, |\cdot|\} = \big(f \dot{+} \delta_W\big)^{\star\star'}$.*

Now, we give the best approximation of a general function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ with a closed convex positively 1-homogeneous function.

Proposition 12 *For any subset $W \subset \mathbb{R}^d$ with $0 \in W$ and any function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ such that $f(0) = 0$, the best closed convex, positively 1-homogeneous approximation $\tilde{f}$ of f on W is given by $\tilde{f} = \sigma_{\partial(f \dot{+} \delta_W)(0)}$.*

Proof. For any subset $W \subset \mathbb{R}^d$ and any function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ such that $f(0) = 0$, we define the set of functions

$$\mathcal{H}_W^f = \big\{ \text{closed convex, positively 1-homogeneous function } h : \mathbb{R}^d \to \overline{\mathbb{R}} \,\big|\, h(x) \leq f(x) \ , \ \ \forall x \in W \big\} \ . \tag{15}$$

As the set $\mathcal{F} = \big\{\text{function } f : \mathbb{R}^d \to \overline{\mathbb{R}}\big\}$ endowed with the partial order $\leq$ is a complete lattice, the subset $\mathcal{H}_W^f \subset \mathcal{F}$ has a (unique) supremum $\tilde{f} = \bigvee \mathcal{H}_W^f$. As the set of closed convex, positively 1-homogeneous functions is stable by pointwise supremum, the function $\tilde{f} : \mathbb{R}^d \to \overline{\mathbb{R}}$ is given, for all $x \in W$, by $\tilde{f}(x) = \sup \big\{ h(x) \,\big|\, h \in \mathcal{H}_W^f \big\}$. From [9,

[5] See the historical note in [9, p. 343].

Theorem 8.24], the proper closed convex positively 1-homogeneous functions can be identified with the support functions of nonempty closed convex subsets of $\mathbb{R}^d$. Thus, we describe the functions of $\mathcal{H}_W^f$ by means of support functions. Let $Y \subset \mathbb{R}^d$ be a nonempty closed convex set such that $\sigma_Y(x) \leq f(x)$, for all $x \in W$. We have

$$\begin{aligned} y \in Y &\implies \langle x,\, y\rangle \leq f(x)\ ,\ \forall x \in W && \text{(by definition of the support function } \sigma_Y) \\ &\iff \langle x,\, y\rangle \dot{+} \big(-f(x)\big) \leq 0 = \langle 0,\, y\rangle \dot{+} \big(-f(0)\big)\ ,\ \forall x \in W \end{aligned}$$

using property of the Moreau lower addition [7], and the assumption that $f(0) = 0$

$$\begin{aligned} &\iff \langle x,\, y\rangle \dot{+} \big(-(f \dot{+} \delta_W)(x)\big) \leq \langle 0,\, y\rangle \dot{+} \big(-(f \dot{+} \delta_W)(0)\big)\ ,\ \forall x \in \mathbb{R}^d\ , && \text{(as } 0 \in W) \\ &\iff y \in \partial\big(f \dot{+} \delta_W\big)(0) \end{aligned}$$

by definition (13) of the (Rockafellar-Moreau) subdifferential. Thus, we have obtained that $Y \subset \partial\big(f \dot{+} \delta_W\big)(0)$ from which we deduce the inequality $\sigma_Y \leq \sigma_{\partial(f \dot{+} \delta_W)(0)}$. We get that $\tilde{f} \leq \sigma_{\partial(f \dot{+} \delta_W)(0)}$, as the subset $\big\{\sigma_Y, Y \subset \mathbb{R}^d \text{ nonempty closed convex} \,\big|\, \sigma_Y$ is equal to $\mathcal{H}_W^f$ in (15).

On the other hand, using the previous equivalences, if $y \in \partial\big(f \dot{+} \delta_W\big)(0)$, we get that $\langle x,\, y\rangle \leq f(x)$ for all $x \in W$. Therefore, we obtain that $\sigma_{\partial(f \dot{+} \delta_W)(0)}(x) = \sup_{y \in \partial(f \dot{+} \delta_W)(0)} \langle x,\, y\rangle \leq f(x)$ for all $x \in W$, that is, $\sigma_{\partial(f \dot{+} \delta_W)(0)} \in \mathcal{H}_W^f$. We conclude that $\tilde{f} = \sigma_{\partial(f \dot{+} \delta_W)(0)}$. □

References

[1] M. Akian, S. Gaubert, and V. Kolokoltsov. Invertibility of functional Galois connections. *Comptes Rendus Mathematique*, 335(11):883–888, 2002.

[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] J.-P. Chancelier and M. De Lara. Constant along primal rays conjugacies and the l_0 pseudonorm. *Optimization*, 0(0):1–32, 2020.

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

[5] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, Stanford University, 2002.

[6] J. E. Martínez-Legaz. Generalized convex duality and its economic applications. In S. S. Hadjisavvas N., Komlósi S., editor, *Handbook of Generalized Convexity and Generalized Monotonicity. Nonconvex Optimization and Its Applications*, volume 76, pages 237–292. sv, 2005.

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

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

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

[10] T. R. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, N.J., 1970.

[11] A. Rubinov. *Abstract convexity and global optimization*, volume 44 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 2000.

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