Deforming $||.||$ into $||.||_\infty$ via Polyhedral Norms: A Pedestrian Approach

Manlio Gaudioso, Jean-Baptiste Hiriart-Urruty

Deforming $\|.\|_1$ into $\|.\|_\infty$ via polyhedral norms: a pedestrian approach

Manlio Gaudioso[1] and Jean-Baptiste Hiriart-Urruty[2]

Abstract. We consider, and study with elementary mathematics from Calculus, the polyhedral norms $\|x\|_{(k)}$ = sum of the k largest among the $|x_i|$'s. Besides their basic properties, we provide various expressions of the unit balls associated with them, and determine all the facets and vertices of these balls. We do the same with the dual norm $\|.\|_{(k)}^*$ of $\|.\|_{(k)}$. The study of these polyhedral norms is motivated, among other reasons, by the necessity of handling sparsity in some modern optimization problems, as it is explained at the end of the paper.

2010 MSC. 52A51, 52B05, 15Axx

Keywords. Polyhedral norms in $\mathbb{R}^n$. Unit balls of norms. Dual norms. Polytopes. Vertices of polytopes. Facets of polytopes.

Introduction

Ask a student about examples of norms in $\mathbb{R}^n$... Very likely he will answer with the usual Euclidean norm $\|.\|_2$ or, with $p \geqslant 1$, the general ℓ_p-norms $\|.\|_p$ defined as follows:

$$\mathbb{R}^n \ni x = (x_1, x_2, ..., x_n) \mapsto \|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

Drawing the unit balls $B_p(0,1) = \left\{ x : \|x\|_p \leqslant 1 \right\}$ in two or three dimensions, seeing how they change when p increases, are interesting and standard exercises. One should however not hide the difficulty of a numerical calculation of $\|x\|_p$, for large p, due to the opposition between the powers p and $1/p$ in the expression $(\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$. The case $p = 1$ is a little apart since it is the only case where $\|.\|_p$ is *polyhedral* (one also says *polytopal*): $\|.\|_1$ is the maximum of a finite number of linear forms, the associated unit ball

[1]Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica, Università della Calabria, 87036 Rende (CS), Italia.
E-mail: gaudioso@dimes.unical.it

[2]Institut de Mathématiques, Université Paul Sabatier, 31062 Toulouse, France.
E-mail: jbhu@math.univ-toulouse.fr

$B_1(0,1)$ is a polytope. Considering the limiting case, $p \to +\infty$, is also the opportunity of an interesting exercise: for all $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$,

$$\lim_{p \to +\infty} \|x\|_p = \max_{i=1,\dots,n} |x_i| \,.$$

This result is precisely the reason why the notation $\|x\|_\infty$ is used for $\max_{i=1,\dots,n} |x_i|$. Indeed, $\|.\|_\infty$ is another norm, *polyhedral* like $\|.\|_1$, called the max norm or even the TCHEBYCHEV[3] norm. When p increases, $\|x\|_p$ decreases, so that we have the following string of inclusions between associated unit balls: whenever $q > p$,

$$B_1(0,1) \subset ... \subset B_p(0,1) \subset B_q(0,1)... \subset B_\infty(0,1).$$

All the "intermediate" norms, that are corresponding to $p \in (1, +\infty)$, are "*smooth*" ones, in the sense that the boundaries of $B_p(0,1)$ are smooth surfaces. Another point: calculating the dual norm of $\|.\|_p$ leads to the norm $\|.\|_q$, where p and q are related by the equation $1/p + 1/q = 1$ (with its extension $1/1 + 1/\infty = 1$), so that there is nothing new under the sun.

At this stage, the student we imagined at the start of the introduction to be asked about the norms in $\mathbb{R}^n$, could think that the essentials has been said, or even that is the end of the story. Actually, there is an infinity of ways to interpolate or deform $\|.\|_1$ into $\|.\|_\infty$; one of them is very interesting for the structure and properties of obtained norms, all polyhedral. Their unit balls therefore are all polytopes. The objective of this paper is to study them (and their dual versions) with basic analysis and algebra tools from the undergraduate level. This is again an opportunity to see how mathematical aspects from various fields like linear algebra, convex geometry, combinatorics, real analysis blend harmoniously.

1. Basic definitions

For an integer k lying between 1 and n, we consider the following real-valued function N_k defined on $\mathbb{R}^n$:

$$N_k(x_1, x_2, ..., x_n) = \max \left\{ |x_{i_1}| + |x_{i_2}| + ... + \ |x_{i_k}| : 1 \leqslant i_1 < ... < i_k \leqslant n \right\}. \tag{1}$$

Clearly, N_1 is the $\|.\|_\infty$ norm, while N_n is the $\|.\|_1$ norm. Actually, N_k is also a norm, "intermediate" between them: $\|.\|_\infty \leqslant N_k \leqslant \|.\|_1$; it was introduced in [10] as a tool to solve linear approximation problems. To prove that N_k is a norm, the only axiom whose verification requires some reasoning

[3] This is an opportunity to recall that we commemorate in 2021 the birth of this eminent mathematician 200 years ago.

is the triangle inequality. For that purpose, consider a k-uple $i_1 < ... < i_k$ for which

$$N_k(x_1+y_1,...,x_n+y_n) = (|x_{i_1}+y_{i_1}|) + (|x_{i_2}+y_{i_2}|) + ... + \ (|x_{i_k}+y_{i_k}|)\,.$$

Applying the triangle inequality $|x_{i_\ell}+y_{i_\ell}| \leqslant |x_{i_\ell}| + |y_{i_\ell}|$ for every $\ell = 1, 2, ..., k$, one gets

$$\begin{aligned} N_k(x_1+y_1,...,x_n+y_n) &\leqslant (|x_{i_1}| + |x_{i_2}| + ... + \ |x_{i_k}|) \\ &\quad + (|y_{i_1}| + |y_{i_2}| + ... + \ |y_{i_k}|) \\ &\leqslant N_k(x_1,x_2,...,x_n) + N_k(y_1,y_2,...,y_n). \end{aligned}$$

From now on, we use the following notations: $\|.\|_{(k)}$ for N_k, and $B_{(k)}$ for the (closed) unit ball associated with $\|.\|_{(k)}$. In order to avoid confusion and keep old habits, we continue to denote $B_\infty(0,1)$ (resp. $B_1(0,1)$) the unit ball associated with $\|.\|_\infty$ (resp. with $\|.\|_1$).

2. First properties of $\|.\|_{(k)}$ and $B_{(k)}$

2.1 Polyhedral norms $\|.\|_{(k)}$

Since $|x_{i_1}| + |x_{i_2}| + ...+ \ |x_{i_k}| = \max_{\varepsilon_i \in \{-1,1\}} (\varepsilon_1 x_{i_1} + \varepsilon_2 x_{i_2} + ... + \ \varepsilon_k x_{i_k})$ and since there are $\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$ choices for the k-uples $i_1 < ... < i_k$, $\|.\|_{(k)}$ is the maximum of $H(n,k) = 2^k \begin{pmatrix} n \\ k \end{pmatrix}$ linear forms, that is:

$$\|(x_1,x_2,...,x_n)\|_{(k)} = \max_{\substack{1 \leqslant i_1 < ... < i_k \leqslant n \\ \varepsilon_i \in \{-1,1\}}} (\varepsilon_1 x_{i_1} + \varepsilon_2 x_{i_2} + ... + \ \varepsilon_k x_{i_k}). \qquad (2)$$

$\|.\|_{(k)}$ is therefore a *polyhedral* norm. The string of inequalities

$$\|.\|_\infty = \|.\|_{(1)} \leqslant ... \|.\|_{(k)} \leqslant \|.\|_{(k+1)} \cdots \leqslant \|.\|_{(n)} = \|.\|_1 \qquad (3)$$

expresses that $\left\{\|.\|_{(k)}\right\}_{k=1,..,n}$ is an *increasing sequence of polyhedral norms interpolating from* $\|.\|_\infty = \|.\|_{(1)}$ *to* $\|.\|_{(n)} = \|.\|_1$.

Note incidentally that all the $H(n,k)$ linear forms appearing in the right-hand side of the formula (2) are relevant, none of them can be removed without affecting the function $\|.\|_{(k)}$.

Just note that $H(n,1) = 2n$ and $H(n,n) = 2^n$.

2.2 Polyhedral unit balls $B_{(k)}$

According to (2), the (closed) unit ball for $\|.\|_{(k)}$, $B_{(k)} = \left\{x : \|x\|_{(k)} \leqslant 1\right\}$, is defined via $H(n,k)$ linear inequalities

$$\varepsilon_1 x_{i_1} + \varepsilon_2 x_{i_2} + ... + \ \varepsilon_k x_{i_k} \leqslant 1. \qquad (4)$$

$B_{(k)}$ is therefore a *convex polyhedral* set (we use the wording "a *polytope*"). The string of inclusions

$$B_1(0,1) = B_{(n)} \subset ... \subset B_{(k+1)} \subset B_{(k)} ... \subset B_{(1)} = B_\infty(0,1) \tag{5}$$

expresses that $\left\{B_{(k)}\right\}_{k=1,..,n}$ is an *increasing sequence (in the sense of inclusion) of polytopes deforming* $B_1(0,1)$ *into* $B_\infty(0,1)$.

$B_1(0,1) = B_{(n)}$ is the well-known *cross-polytope,* defined via 2^n linear inequalities

$$\varepsilon_1 x_1 + \varepsilon_2 x_2 + ... + \ \varepsilon_n x_n \leqslant 1, \quad (\varepsilon_i \in \{-1,1\} \text{ for all } i = 1,...,n)\,,$$

with its $2n$ vertices $\pm e_i = (0,0,...,\pm 1,...,0)$ (for $i = 1,...,n)$.

$B_{(1)} = B_\infty(0,1)$ is the *n-dimensional hypercube* $[-1,1]^n$, defined via $2n$ linear inequalities

$$\varepsilon_i x_i \leqslant 1, (\varepsilon_i \in \{-1,1\} \text{ for all } i = 1,...,n)\,,$$

with 2^n vertices $(\pm 1, \pm 1, ..., \pm 1, ..., \pm 1)$.

2.3 The special case of $n = 3$ or 4 and $k = 2$

For $n = 3$, the only "intermediate" norm is, for $k = 2$,

$$\|(x,y,z)\|_{(2)} = \max \ \left(|x| + |y|\,, |y| + |z|\,, |x| + |z|\right).$$

Its unit ball $B_{(2)}$ is the so-called *rhombic dodecahedron* or *granatahedron* (rhombic because all the facets are rhombuses (that is, diamond shaped polygons, from Greek *rhombos*)) ; it has exactly $f_0 = 14$ vertices, $f_1 = 24$ edges, $f_2 = 12$ facets[4]. According to its definition, *i.e.,* $(|x| + |y| \leqslant 1, |y| + |z| \leqslant 1, |x| + |z| \leqslant 1)$, it can be viewed as the intersection of three mutually orthogonal cylinders with square sections; see the picture below by A. Esculier.

[4]For *faces* of a convex set C, we follow the terminology from [7, p. 42 – 46]: 0-dimensional faces are called *vertices* (or *extreme points*) of C; 1-dimensional faces are called *edges* of C; and so until $(n-1)$-dimensional faces called *facets* of C.

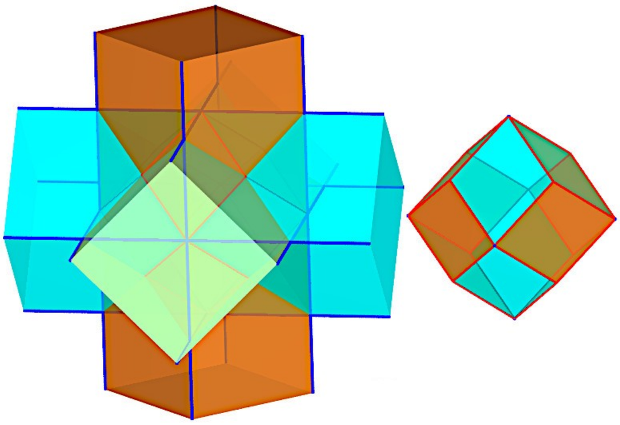

Figure 1. The unit ball $B_{(2)}$ (right) as the intersection of 3 orthogonal cylinders (left).

More on this polytope can be found on the website MATHCURVE.COM by R. FERREOL.

We ask sometimes our students in Calculus to draw the part of $B_{(2)}$ which is on the positive orthant of $\mathbb{R}^3$, that is $B_{(2)}^{+} = \{(x, y, z) \in B_{(2)}, x \geqslant 0, y \geqslant 0, z \geqslant 0\}$; they have difficulties... It is a polytope with vertices $(0, 0, 0)$, $(0, 1, 0)$, $(1, 0, 0)$, $(0, 0, 1)$ and $(1/2, 1/2, 1/2)$.

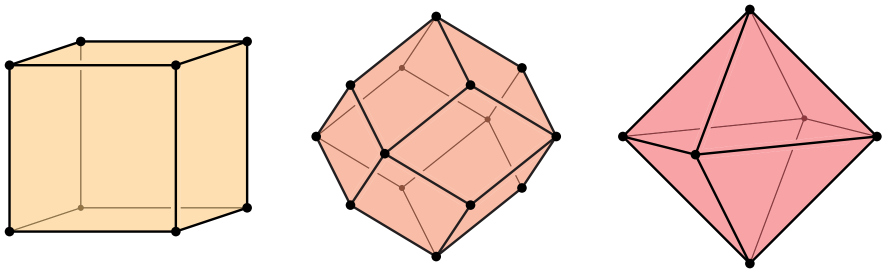

Figure 2. The unit ball $B_{(2)}$ between the hypercube and the cross-polytope.

For $n = 4$, the "intermediate" norm $\|(x, y, z, t)\|_{(2)}$ is also of interest. The associated unit ball $B_{(2)}$ is the so-called *24-cell polytope* (or *icositetrachoron* or *hypergranatohedron),* whose visual aspect (*i.e.*, projections on 3-dimensional spaces) can easily be found on websites; it has exactly $f_0 = 24$ vertices, $f_1 = 96$ edges, $f_2 = 96$ two-dimensional faces (also called ridges), and $f_3 = 24$ facets.

Just note on these two examples the illustration of EULER's formula (for polytopes in $\mathbb{R}^3$) : $f_0 - f_1 + f_2 = 2$ and EULER-POINCARÉ formula (for polytopes in $\mathbb{R}^4$) $f_0 - f_1 + f_2 - f_3 = 0$.

2.4 HAUSDORFF distances between $B_1(0,1)$ and $B_{(k)}$, between $B_{(k)}$ and $B_\infty(0,1)$

When a compact convex set C is included in another compact convex set D, the so-called HAUSDORFF (or POMPEIU-HAUSDORFF) *excess* of D over C, or HAUSDORFF *distance* between C and D, is

$$\Delta_H(C,D) = \max_{x\in D} d_C(x), \tag{6}$$

where $d_C(x)$ denotes the distance from x to the set C, that is $\min_{y\in C}\|x-y\|$. Here, only the usual Euclidean distance $\|.\|$ is invoked.

Consider therefore $C = B_1(0,1)$ and $D = B_{(k)}$. For symmetry reasons, the maximum in (6) is achieved for $\overline{x} = (\frac{1}{k}, \frac{1}{k}, ..., \frac{1}{k})$, while $d_{B_1(0,1)}(\overline{x}) = \left\|\overline{x} - (\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})\right\|$. This is easy to accept with the expression, to be seen later (formula (14)), of $B_{(k)}$ as the convex hull of $B_1(0,1) \cup B_\infty(0, \frac{1}{k})$. An alternate argument, used for example in [7, Example 1.3.4] for getting at $\Delta_H(B_1(0,1), B_\infty(0,1)) = \frac{n-1}{\sqrt{n}}$, is to use an expression of $\Delta_H(C,D)$ via the support functions σ_C of C and σ_D of D ([7, Theorem 3.3.6]): $\Delta_H(C,D)$ is the maximum of $(\sigma_D(d) - \sigma_C(d))$ over unit vectors d [5]. Anyway

$$\Delta_H(B_1(0,1), B_{(k)}) = \frac{n-k}{k\sqrt{n}}. \tag{7}$$

Similarly, $\Delta_H(B_{(k)}, B_\infty(0,1)) = \left\|(1,1,...,1) - (\frac{1}{k}, \frac{1}{k}, ..., \frac{1}{k})\right\|$, that is

$$\Delta_H(B_{(k)}, B_\infty(0,1)) = \sqrt{n}\left(\frac{k-1}{k}\right). \tag{8}$$

Indeed

$$\Delta_H(B_1(0,1), B_\infty(0,1)) = \Delta_H(B_1(0,1), B_{(k)}) + \Delta_H(B_{(k)}, B_\infty(0,1)), \tag{9}$$

which is fairly easy to understand, even "visually".

3. Norms, gauges, support functions. Applications to $\|.\|_{(k)}$ and its dual $\|.\|_{(k)}^*$

3.1 Some recalls

[5]As it will be explained in detail later, $\sigma_{B(0,1)}(d) = \|d\|_\infty$ and $\sigma_{B_{(k)}}(d) = \max\left(\frac{\|d\|_1}{k}, \|d\|_\infty\right)$.

Norms are special examples of (finite) *positive sublinear functions*, that are positively homogeneous positive convex functions, studied in detail in [7, chapter C].

The *gauge* γ_C of a compact convex set $C \subset \mathbb{R}^n$ containing the origin in its interior is the function $\mathbb{R}^n \ni x \mapsto \gamma_C(x) = \inf(\lambda > 0 : x \in \lambda C)$; we recover C by taking the sublevel-set at level 1 of γ_C, that is $C = \{x : \gamma_C(x) \leqslant 1\}$. A norm $\|.\|$ is indeed a gauge function, the gauge function of its unit ball $B = \{x : \|x\| \leqslant 1\}$.

The *support function* σ_C of a compact convex set $D \subset \mathbb{R}^n$ is defined as $\mathbb{R}^n \ni s \mapsto \sigma_D(x) = \max_{s \in D} \langle s, x\rangle$; here, we recover D by collecting the slopes s of all linear functions minorizing σ_D, that is $D = \{s : \langle s, x\rangle \leqslant \sigma_D(x) \text{ for all } x\}$. A norm $\|.\|$ is also a support function, that of

$$B^* = \{s : \langle s, x\rangle \leqslant 1 \text{ for all } x \text{ in the unit ball } B \text{ of } \|.\|\}.$$

This set B^* is called the *polar set* of B, it is also denoted as B° in the literature. It is actually the unit ball of another norm, called *dual norm* of $\|.\|$, denoted as $\|.\|^*$, and defined as

$$\|s\|^* = \sup_{x \in B} \langle s, x\rangle. \tag{10}$$

All these correspondences are explained in [7, pages 146 − 151]. The following scheme clarifies and summarizes everything.

Norm $\longrightarrow$	Take the sublevel-set at level 1 $\longrightarrow$	**Unit ball** B
Take the support function $\uparrow$	$\swarrow$ *Polarity between unit balls* $\nearrow$	$\downarrow$ Take the support function
Unit ball B^* $\longleftarrow$	Take the sublevel-set at level 1	$\longleftarrow$ **Dual norm**

The game ends here since $(B^*)^* = B$ and $(\|.\|^*)^* = \|.\|$.

3.2 Applications to $\|.\|_{(k)}$

The dual norm of $\|.\|_1$ is $\|.\|_\infty$, the dual norm of $\|.\|_\infty$ is $\|.\|_1$. Stated in terms of balls, the polar set of $B_\infty(0,1)$ is $B_1(0,1)$, the polar set of $B_1(0,1)$ is $B_\infty(0,1)$. Expressed in terms of support functions, the support function of $B_\infty(0,1)$ is $\|.\|_1$, the support function of $B_1(0,1)$ is $\|.\|_\infty$. *So, what about*

$\|\cdot\|_{(k)}$? The question which naturally arises is:

- *$\|\cdot\|_{(k)}$ is the support function of what polytope?*
or, equivalently,
- *what is the polar polytope of $B_{(k)}$?*
or, equivalently,
- *what is the dual norm $\|\cdot\|_{(k)}^*$ of $\|\cdot\|_{(k)}$?*

Different paths can be followed for answering these questions... We choose one of them.

Theorem 1. *Let $\Pi_{(k)}$ be the polytope in $\mathbb{R}^n$ defined as the set of all $(\alpha_1, \alpha_2, .., \alpha_n)$ satisfying the following inequalities:*

$$\begin{cases} -1 \leqslant \alpha_i \leqslant 1 \text{ for all } i = 1, ..., n \\ \sum_{i=1}^n |\alpha_i| \leqslant k. \end{cases} \tag{11}$$

Then, its support function is precisely $\|\cdot\|_{(k)}$.

Here are immediate consequences of Theorem 1 and some additional observations.

- Since (11) is sumarized as $\left(\|\alpha\|_\infty \leqslant 1 \text{ and } \frac{\|\alpha\|_1}{k} \leqslant 1\right)$, according to what has been explained in § 3.1, the polar set of the unit ball $B_{(k)}$ for $\|\cdot\|_{(k)}$ is

$$B_{(k)}^* = kB_1(0,1) \cap B_\infty(0,1). \tag{12}$$

- Since the polytope defined in (12) is the unit ball for the dual norm $\|\cdot\|_{(k)}^*$ of $\|\cdot\|_{(k)}$ (see again § 3.1), we have:

$$\|\cdot\|_{(k)}^* = \max\left\{\frac{\|\cdot\|_1}{k}, \|\cdot\|_\infty\right\}. \tag{13}$$

This formula (13) covers the two well-known “extreme” cases, that are for $k = 1$ and $k = n$. Indeed, since $\|\cdot\|_\infty \leqslant \|\cdot\|_1$, we get that

$$\|\cdot\|_\infty^* = \|\cdot\|_{(1)}^* = \max\left\{\frac{\|\cdot\|_1}{k}, \|\cdot\|_\infty\right\} = \|\cdot\|_1.$$

Similarly, since $\|\cdot\|_1 \leqslant n\|\cdot\|_\infty$, we get that

$$\|\cdot\|_1^* = \|\cdot\|_{(n)}^* = \max\left\{\frac{\|\cdot\|_1}{n}, \|\cdot\|_\infty\right\} = \|\cdot\|_\infty.$$

- By playing with simple calculus rules on polarity dealing with compact convex sets C and D containing the origin in their interior, like (see [3, §6]):

$$\begin{aligned}(C\cap D)^{*} &= \operatorname{co}(C^{*}\cup D^{*}),\\ (\operatorname{co}(C\cup D))^{*} &= C^{*}\cap D^{*},\\ (rC)^{*} &= \frac{1}{r}C^{*} \text{ whenever } r>0,\end{aligned}$$

where $\operatorname{co}(S)$ stands for the convex hull of S, we get from (12) an alternate expression of the unit ball $B_{(k)}$ of $\|.\|_{(k)}$:

$$B_{(k)} = \operatorname{co}(B_1(0,1)\cup B_\infty(0,\frac{1}{k})). \tag{14}$$

This alternate formulation paves the way to the definition of a norm $\|.\|_{(k)}$ when k is no more an integer: for $1\leqslant k\leqslant n$, $\|.\|_{(k)}$ is the polyhedral norm whose unit ball is the polytope $B_{(k)}$ such as defined in (14). In doing so, we have a "continuous family of polyhedral norms" $\left\{\|.\|_{(k)}\right\}_{1\leqslant k\leqslant n}$, decreasingly interpolating from $\|.\|_\infty$ (which is $\|.\|_{(1)}$) to $\|.\|_1$(which is $\|.\|_{(n)}$), like in the "discrete" case (5). This is what has been studied in the recent work [4].

- We have another string of inclusions, a "dual string" to the one displayed in (5):

$$B_\infty(0,1) = B^*_{(n)} \supset ... \supset B^*_{(k+1)} \supset B^*_{(k)} ... \supset B^*_{(1)} = B_1(0,1), \tag{5^*}$$

expressing that $\left\{B^*_{(k)}\right\}_{k=1,..,n}$ is an *decreasing sequence (in the sense of inclusion) of polytopes deforming* $B_\infty(0,1)$ *into* $B_1(0,1)$.

- There are still other expressions of $\|.\|_{(k)}$, one of them being in terms of $\|.\|_1$ and $\|.\|_\infty$ alone via the so-called *infimal convolution* (a sort of mixture) of convex functions. This operation on convex functions is a very basic one in Convex Analysis, as important as the mere addition of functions (*cf.* [7]). Given two convex functions f and g, the infimal convolution of the convex functions f and g is a new convex function, denoted as $f\diamond g$, defined as

$$x\mapsto (f\diamond g)(x) = \inf_{u+v=x}\{f(u)+g(v)\}. \tag{15}$$

When both f and g are sublinear functions, like norms, $f\diamond g$ is also the convex envelope of the function $\min(f,g)$ ([7, Proposition 1.3.2 in Chapter C]). The construction in (15) is the one used in Functional Analysis to define a new norm "mixing" or "interpolating" two other ones. So, to say things shortly, using techniques from Convex Analysis, like the LEGENDRE-FENCHEL transformation on sublinear functions, one gets at

$$\|.\|_{(k)} \text{ is the convex envelope of } \min\left(\|.\|_1, k\,\|.\|_\infty\right), \tag{16}$$

$$\left\{\begin{array}{c} \|.\|_{(k)} = \|.\|_1 \diamond k\,\|.\|_\infty\,, \\ \text{that is to say, } \|x\|_{(k)} = \inf_{u+v=x}\left\{\|u\|_1 + k\,\|v\|_\infty\right\}. \end{array}\right. \tag{17}$$

The formulation (17) was observed and proved in ([2, Proposition IV.1.5])[6].

- In a way similar to what has been carried out in § 2.4, we measure the HAUSDORFF distance between $B^*_{(k)}$ and $B_\infty(0,1)$, and that between $B_1(0,1)$ and $B^*_{(k)}$; indeed

$$\Delta_H(B^*_{(k)}, B_\infty(0,1)) = \frac{n-k}{\sqrt{n}}; \tag{7*}$$

$$\Delta_H(B_1(0,1), B^*_{(k)}) = \frac{k-1}{\sqrt{n}}. \tag{8*}$$

Also,

$$\Delta_H(B_1(0,1), B^*_{(k)}) + \Delta_H(B^*_{(k)}, B_\infty(0,1)) = \Delta_H(B_1(0,1), B_\infty(0,1)). \tag{9*}$$

Proof of Theorem 1. We intend to prove that, for all $x \in \mathbb{R}^n$,

$$\sup_{\alpha \in \Pi_{(k)}} \langle \alpha, x\rangle = \|x\|_{(k)}\,.$$

We provide a self-contained proof, using elementary techniques from Calculus.

Let $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$. We consider $1 < k < n$. Without loss of generality, we may suppose that

$$|x_1| \geqslant |x_2| \geqslant ... \geqslant |x_k| \geqslant |x_{k+1}| \geqslant ... \geqslant |x_n|\,.$$

Let $\alpha = (\alpha_1, \alpha_2, .., \alpha_n) \in \Pi_{(k)}$, that is satisfying the inequalities in (11). We firstly intend to prove that $\langle \alpha, x\rangle = \sum_{i=1}^n \alpha_i x_i \leqslant \sum_{i=1}^k |x_i| = \|x\|_{(k)}$.

[6]We even can prove that the infimal convolution is *exact* at all $x \in \mathbb{R}^n$, that is to say: the infimum in (17) is achieved. This is a property of utmost importance in treating properties of the inf-convoluted function. For that, consider, without loss of generality, that $x_1 \geqslant x_2 \geqslant ... \geqslant x_k \geqslant x_{k+1} \geqslant ... \geqslant x_n > 0$. Then, for $u_x = (x_1 - x_k, x_2 - x_k, ..., x_k - x_k, 0, .., 0)$ and $v_x = (x_k, x_k, ..., x_k, x_{k+1}, .., x_n)$, we note that $\|u_x\|_1 + k\,\|v_x\| = (\|.\|_1 \diamond k\,\|.\|_\infty)(x)$.

For that, we begin by noticing that

$$\sum_{i=1}^{k}(1-|\alpha_i|) \geqslant \sum_{i=k+1}^{n}|\alpha_i|\,, \text{ because } \sum_{i=1}^{n}|\alpha_i| \leqslant k. \tag{18}$$

Furthermore,

$$\sum_{i=k+1}^{n}\alpha_i x_i \leqslant \sum_{i=k+1}^{n}|\alpha_i|\,.\,|x_i| \leqslant \left(\sum_{i=k+1}^{n}|\alpha_i|\right)|x_k| \tag{19}$$

$$\sum_{i=k+1}^{n}\alpha_i x_i \leqslant \left(\sum_{i=1}^{k}(1-|\alpha_i|)\right)|x_k|\,, \text{ because of (18)}$$

$$\sum_{i=k+1}^{n}\alpha_i x_i \leqslant \sum_{i=1}^{k}(1-|\alpha_i|)\,|x_i| \tag{20}$$

because $|x_k| \leqslant |x_i|$ for all $i = 1, ..., k$.

Consequently,

$$\begin{aligned}
\sum_{i=1}^{n}\alpha_i x_i &= \sum_{i=1}^{k}\alpha_i x_i + \sum_{i=k+1}^{n}\alpha_i x_i \\
&\leqslant \sum_{i=1}^{k}|\alpha_i|\,|x_i| + \sum_{i=k+1}^{n}\alpha_i x_i \\
&\leqslant \sum_{i=1}^{k}|\alpha_i|\,|x_i| + \sum_{i=1}^{k}(1-|\alpha_i|)\,|x_i| \\
&\quad \text{because of (20),} \\
&\leqslant \sum_{i=1}^{k}|x_i|\,.
\end{aligned}$$

We therefore have proved that $\sup_{\alpha\in\Pi_{(k)}}\langle\alpha, x\rangle \leqslant \|x\|_{(k)}\,.$
Consider now a specific $\overline{\alpha} \in \Pi_{(k)}$ with

$$\left\{\begin{array}{c}
\overline{\alpha}_i = 0 \text{ for } i = k+1, ..., n \\
\overline{\alpha}_i = 1 \text{ or } -1 \text{ for } i = 1, ..., k, \\
\text{according to whether } x_i \geqslant 0 \text{ or } x_i \geqslant 0.
\end{array}\right.$$

Thus, $\langle\overline{\alpha}, x\rangle = \sum_{i=1}^{n}\overline{\alpha}_i x_i = \sum_{i=1}^{k}|x_i| = \|x\|_{(k)}$. $\boxdot$

3.3 The special case of $n = 3$ or 4 and $k = 2$

For $n = 3$ and $k = 2$,

$$\|(x,y,z)\|^*_{(2)} = \max\left(\frac{|x|+|y|+|z|}{2}, |x|, |y|, |z|\right). \tag{21}$$

Its unit ball $B^*_{(2)}$ is the so-called *cuboctahedron* or *heptaparallelohedron* (or even *dymaxion* by some architects); it has exactly $f_0 = 12$ vertices (of Cartesian coordinates $(\pm 1, \pm 1, 0)$, with permutations, $f_1 = 24$ edges, $f_2 = 14$ facets (8 triangles and 6 squares) ; see the picture below by L. Pournin.

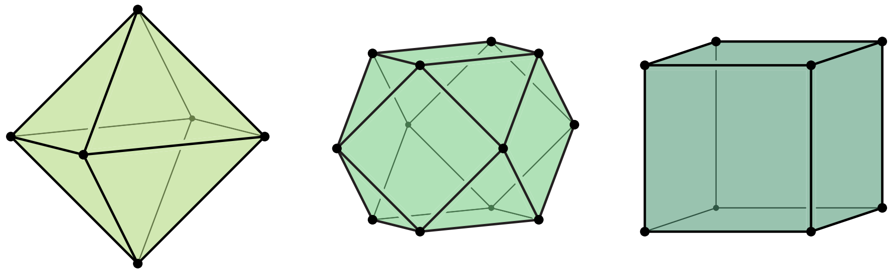

Figure 3. The unit ball $B^*_{(2)}$ between the cross-polytope and the hypercube.

More on this polytope can be found on the website MATHCURVE.COM by R. Ferreol.

For $n = 4$, something very interesting happens: $B^*_{(2)}$ has the same number of vertices, edges, ridges and facets as the 24-cell polytope $B_{(2)}$. The reason is that one can transform $B_{(2)}$ into $B^*_{(2)}$ via a simple affine transformation in $\mathbb{R}^4$ (a rotation followed by a dilation).

4. Extreme points, facets, of $B_{(k)}$ and of its dual $B^*_{(k)}$

Due the "polarity relation" between vertices and facets of a polytope and of its polar polytope ([3, Theorem 9.1 and Theorem 9.8]), once we have the number of facets (resp. of vertices) of $B_{(k)}$, we have the number of vertices (resp. of facets) of $B^*_{(k)}$. There are several paths of getting at them. We choose one, the shortest one we believe; we pull the right thread from the spool, and everything unwinds.

We begin with facets of $B_{(k)}$ and vertices of $B^*_{(k)}$.

Theorem 2. - *$B_{(k)}$ has exactly $H(n,k) = 2^k \begin{pmatrix} n \\ k \end{pmatrix}$ facets.*

- $B^*_{(k)}$ has exactly $H(n,k) = 2^k \begin{pmatrix} n \\ k \end{pmatrix}$ vertices. They are $\overline{\alpha} = (\overline{\alpha}_1, \overline{\alpha}_2, ..., \overline{\alpha}_n)$ in $\mathbb{R}^n$ in which all the coordinates $\overline{\alpha}_i$ are zero except k of them which are ± 1.

Proof. We have observed from the beginning (see § 2.2) that $B_{(k)}$ is defined via $H(n,k)$ linear inequalities

$$\varepsilon_1 x_{i_1} + \varepsilon_2 x_{i_2} + ... + \ \varepsilon_k x_{i_k} \leqslant 1,$$

built up from k-uples $i_1 < ... < i_k$ and $\varepsilon_i \in \{-1, 1\}$. None of these inequalities can be removed without affecting $B_{(k)}$. So, when we have a representation like this, say, for C,

$$\langle v^i, x \rangle \leqslant 1 \text{ for all } i = 1, 2, ..., \ell,$$

C has ℓ facets, the polar set C^* of C is $\operatorname{co}\{v^i : i = 1, 2, ..., \ell\}$, and all the v^i's are vertices of C^* ([3, Theorem 9.1]).

Actually, we have seen that in another way of proving Theorem 1: the polar set $B^*_{(k)}$, whose support function is $\|.\|_{(k)}$, is the polytope $\Pi_{(k)}$ evoked in Theorem 1; its vertices are all the $\overline{\alpha} = (\overline{\alpha}_1, \overline{\alpha}_2, ..., \overline{\alpha}_n) \in \mathbb{R}^n$ in which all the coordinates $\overline{\alpha}_i$ are zero except k of them which are ε_i. $\boxdot$

Now it's the turn of facets of $B^*_{(k)}$ and vertices of $B_{(k)}$. We already know the situation for the two "extreme" k: the cross-polytope $B^*_{(1)} = B_1(0,1)$ has 2^n facets and $2n$ vertices; the hypercube $B^*_{(n)} = B_\infty(0,1)$ has $2n$ facets and 2^n vertices.

Theorem 3. Let $1 < k < n$. Then:
- *$B^*_{(k)}$ has exactly $2n + 2^n$ facets.*
- *$B_{(k)}$ has exactly $H(n,k) = 2n + 2^n$ vertices. They are* $(0, 0, ..., \pm 1, ..., 0)$ *and* $\left(\pm\frac{1}{k}, \pm\frac{1}{k}, ..., \pm\frac{1}{k}\right)$*, with their permutated versions.*

Proof. We start with the formulation of $B^*_{(k)}$ seen in (12):

$$B^*_{(k)} = kB_1(0,1) \cap B_\infty(0,1).$$

Because $1 < k < n$, the intersection operation cannot be removed above. Hence, $B^*_{(k)}$ is defined via a conjunction of two series of (irredundant) linear inequalities: the ones defining $kB_1(0,1)$ (there are 2^n) and the ones defining $B_\infty(0,1)$ (there are $2n$). In an explicit format, they are:

$$\begin{aligned} \left\langle \left(\pm\frac{1}{k}, \pm\frac{1}{k}, ..., \pm\frac{1}{k}\right), x \right\rangle &\leqslant 1, \\ \langle (0, 0, ..., \pm 1, ..., 0), x \rangle &\leqslant 1. \end{aligned}$$

Accordingly, we get at all the vertices of $\left(B^*_{(k)}\right)^* = B_{(k)}$.

Remark 4.1. There are three other ways to find the vertices of $B_{(k)}$ (or $B^*_{(k)}$). Let us briefly present them for $B_{(k)}$.

- *First way* (a usual one). A way to prove that a given set of points $\{x^1, x^2, ..., x^p\}$ provides the vertices of a polytope C is: to show that every element in C is a convex combination of these points, and that none of the points is a convex combination of the others ([3, Theorem 7.2]). Here, due to the representation $B_{(k)} = \mathrm{co}(B_1(0,1) \cup B_\infty(0, \frac{1}{k}))$ (*cf.* (14)), the process can be carried out with the points $x^i = (0, 0, ..., \pm 1, ..., 0)$ and $\left(\pm\frac{1}{k}, \pm\frac{1}{k}, ..., \pm\frac{1}{k}\right)$, with their permutated versions.

- *Second way* (based on Linear Algebra). By using a representation of $B_{(k)}$ in the form $Ax \leqslant b$ like it is done in Linear Programming. Let therefore C be a polyhedron in $\mathbb{R}^n$ described as follows:

$$C = \{x \in \mathbb{R}^n : Ax \leqslant b\}, \tag{22}$$

where $A \in \mathcal{M}_{m,n}(\mathbb{R})$, $m \geqslant n$, none of the row vectors a_i is null, and $b \in \mathbb{R}^m$. For a nonempty subset I of $\{1, 2, ..., m\}$ (with ℓ elements for example), we denote

$$\begin{cases} A_I \text{ the matrix extracted from } A \text{ by keeping only the rows } i \in I \\ \qquad\qquad (\text{hence } A_I \in \mathcal{M}_{\ell,n}(\mathbb{R})); \\ b_I \text{ the vector extracted from } b \text{ by keeping only the coordinates} \\ \qquad\qquad \text{corresponding to } i \in I \quad (\text{hence } b_I \in \mathbb{R}^\ell). \end{cases}$$

Let $\overline{x}$ be on the boundary of C; we denote by $I(\overline{x})$ the set of indices $i \in \{1, 2, ..., m\}$ corresponding to the so-called active inequality constraints at $\overline{x}$, that is

$$I(\overline{x}) = \{i : \langle a_i, \overline{x}\rangle = b_i\}.$$

Then, $\overline{x}$ is a vertex of C if and only if the rank of $A_{I(\overline{x})}$ equals n.

The method is a bit heavy to apply in the case of $C = B_{(k)}$ in $\mathbb{R}^n$ for large n and k. An interesting exercise is however to do that when $n = 3$ and $k = 2$. Then the 14 vertices of $B_{(2)}$ in $\mathbb{R}^3$ are detected. See below.

At a vertex $\overline{x}$ of C, one has $\mathrm{Card}\ I(\overline{x}) \geqslant n$; when $\mathrm{Card}\ I(\overline{x}) > n$, $\overline{x}$ is called a degenerate vertex.

The evoked result is a beautiful example of an interplay between Linear Algebra and Geometry of convex sets. Let us see how it applies when $n = 3$

and $k = 2$. Our $C = B_{(2)}$ is expressed like in (22) with

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & -1 \\ -1 & 0 & 1 \\ -1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & -1 & -1 \end{bmatrix} \in \mathcal{M}_{12,3}(\mathbb{R}),$$

and $b = (1, 1, ..., 1)^\intercal \in \mathbb{R}^{12}$.

Take for example $\overline{x} = (1, 0, 0)$. They are 6 points of that type. We have $I(\overline{x}) = \{1, 2, 5, 6\}$, so $\ell = 4$, and

$$A_{I(\overline{x})} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & -1 \end{bmatrix} \in \mathcal{M}_{4,3}(\mathbb{R}).$$

Indeed, $A_{I(\overline{x})}$ is of rank 3. Hence, $\overline{x} = (1, 0, 0)$ is a vertex of $B_{(2)}$, a degenerate one (because Card $I(\overline{x}) = 4 > 3 = n$).

Take now $\overline{x} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. They are 8 points of that type. We have $I(\overline{x}) = \{1, 5, 9\}$, so $\ell = 3$, and

$$A_{I(\overline{x})} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \in \mathcal{M}_{3,3}(\mathbb{R}).$$

Indeed, $A_{I(\overline{x})}$ is of rank 3. Hence, $\overline{x} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ is a vertex of $B_{(2)}$, a nondegenerate one (because Card $I(\overline{x}) = 3 = n$).

- *Third way* (a more advanced one). We know that the support function of $B_{(k)}$ is the nonsmooth convex function $d \mapsto \sigma_{B_{(k)}}(d) = \max\left\{\frac{\|d\|_1}{k}, \|d\|_\infty\right\}$. The subdifferential, in the sense of Convex Analysis, of $\sigma_{B_{(k)}}$ at $d \neq 0$ is exactly the face of $B_{(k)}$ exposed by the direction d (see [7, Chapter D]). So, having this exposed face reduced to a singleton (*i.e.*, a vertex of $B_{(k)}$) amounts to the differentiability of $\sigma_{B_{(k)}}$ at $d \neq 0$. It therefore remains to collect all the gradients of the function $d \mapsto \sigma_{B_{(k)}}(d) = \max\left\{\frac{\|d\|_1}{k}, \|d\|_\infty\right\}$ whenever they exist. When $1 < k < n$, the two functions $\frac{\|.\|_1}{k}$ and $\|.\|_\infty$ have to be taken into account in the "max expression" of $\sigma_{B_{(k)}}$ above: their gradients yield $\left(\pm\frac{1}{k}, \pm\frac{1}{k}, ..., \pm\frac{1}{k}\right)$ and $(0, 0, ..., \pm 1, ..., 0)$, with permutations.

Remark 4.2. It is a bit surprising that the number of vertices of $B_{(k)}$ does not depend on k... One does not see that, at the first glance, in the definitions given in § 2.2; the intuition for that is more supported by the expression (14) of $B_{(k)}$.

For more on the combinatorial and geometric properties of the polytopes $B_{(k)}$, like their k-dimensional faces, their volume and the volume of their boundary, see the full-fledged research paper [4].

5. Links with the search of sparse solutions in optimization problems

In several areas of Applied Mathematics, one has to bound, to control, to optimize, etc. the largest or the sum of a sample of (positive) data $x_1, x_2, ..., x_n$; but it also happens that one has to deal with the sum of the k largest among these x_i's. This occurs in Numerical Analysis, Statistics, Optimization. We precisely focus here on one of these topics, namely the search of sparse solutions in optimization problems. A very recent texbook on sparse solutions of undetermined linear systems and their applications is [9].

Indeed, in various applications of modern Optimization, one is faced with the so-called *sparsity* constraint on solutions. This happens in data science and machine learning, mathematical imaging (in Astronomy for example), Statistics, but not solely. A measure of sparsity of a solution vector $x = (x_1, x_2, ..., x_n)$ in $\mathbb{R}^n$ is the number of non-zero components x_i of x. More specifically, either in the objective function or in the functions defining the constraints of the optimization problem, one has to deal with

$$x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n \mapsto \mathrm{Card}\,\{i : x_i \neq 0\}\,. \tag{23}$$

Various namings and notations are used in the literature for this function: cardinality function, counting function, $nnz(x)$ (the number of non-zeros in x), even pseudo-norm $\|.\|_0$. This last naming and notation are somehow misleading: $\|.\|_0$ is not a norm, not even a quasi-norm... The notation could let us think that $\|x\|_0$ is the limit of the usual ℓ_p-norm of x, $\|x\|_p$, when $p > 0$ tends to 0. This is not the case, we however note that $\|x\|_0$ is the limit of $\left(\|x\|_p\right)^p$ when $p > 0$ tends to 0. Since this notation $\|.\|_0$ is very much spread in the literature, we agree to use it in this section.

A weakness of $\|.\|_0$ is its zero-homogeneity: components like $x_i = 10^{-6}$ and $x_i = 10^5$ contribute in the same way (by 1) to the number $\|x\|_0$; so, one has to bound the considered vectors x in some way or another. To alleviate the wording, we say that a vector $x \in \mathbb{R}^n$ is *k-sparse* whenever $\|x\|_0 \leqslant k$.

The questions raised here are: *what are the relations between* $\|.\|_0$ *and* $\|.\|_{(k)}$ *or* $B_{(k)}$*?*

One sometimes reads in papers that "$\|.\|_1$ is the best relaxed convex form (that is to say, the convex envelope) of $\|.\|_0$". This is wrong, since the convex envelope of the $\|.\|_0$ function on $\mathbb{R}^n$ is just the (everywhere) zero function... To get something of interest, one has to restrict to balls defined by $\|.\|_\infty$ norms. What is behind the alluded to statement is the following relaxation result by M. Fazel (PhD thesis, Stanford University, 2002): *the convex envelope of* $\|.\|_0$ *on the ball* $\{x : \|.\|_\infty \leqslant R\}$ *is the function* $\frac{1}{R}\|.\|_1$ *(restricted to the same ball).* Actually, the result remains true for the quasiconvex envelope (an operation consisting in convexifying all the sublevel-sets of the original function), see [8] and references therein.

Here are the answers to raised questions, in two forms:

- We have, from [8, Theorem 1]:

$$\begin{aligned} \operatorname{co}\{x : \|x\|_0 \leqslant k,\ \|x\|_\infty \leqslant 1\} &= \{\|x\|_1 \leqslant k,\ \|x\|_\infty \leqslant 1\} \\ &= B^*_{(k)}. \end{aligned} \tag{24}$$

In other words, *the convex hull of the set of bounded (by 1) k-sparse vectors x is exactly the unit ball of the dual norm* $\|.\|^*_{(k)}$ *of* $\|.\|_{(k)}$.

There is another "k-norm" which has recently been introduced (in [1]) for sparse prediction problems, it is basically defined via its unit ball:

$$C_{(k)} = \operatorname{co}\{x : \|x\|_0 \leqslant k,\ \|x\|_2 \leqslant 1\}. \tag{25}$$

The difference with (24) is the use of the smooth (Euclidean) norm $\|.\|_2$ instead of the polyhedral norm $\|.\|_\infty$ for bounding k-sparse vectors. Hence, the norm whose unit ball is $C_{(k)}$, called "k-support norm" in [1], is no more polyhedral, we could qualify it as "semismooth".

- We clearly have that $\|.\|_{(k)} \leqslant \|.\|_{(\ell)}$ whenever $\ell > k$. In such a case, it is easy to check (and was observed in [6]) that

$$\|x\|_{(k)} - \|x\|_{(\ell)} = 0 \text{ (or } \geqslant 0) \Longleftrightarrow \|x\|_0 \leqslant k. \tag{26}$$

In particular, since $\|.\|_{(n)} = \|.\|_1$,

$$\|x\|_{(k)} - \|x\|_1 = 0 \text{ (or } \geqslant 0) \Longleftrightarrow \|x\|_0 \leqslant k. \tag{27}$$

Said in words, *k-sparse vectors are exactly those on which two norms like* $\|.\|_{(k)}$ *coincide.* We therefore can substitute the equality constraint $c(x) =$

$\|x\|_{(k)} - \|x\|_1 = 0$ for the "k-sparsity constraint" $\|x\|_0 \leqslant k$. This is not a relaxation of the sparsity constraint, but an equivalent reformulation indeed.The advantage is that c is the difference of two polyhedral norms (hence convex functions), whose subdifferentials in the sense of convex minimization are amenable to numerical computation. This was the objective in [6], and in [5] for feature selection in SVM.

6. Conclusion

In this paper, we carried out a pedagogical approach of a collection of polyhedral norms interpolating $\|.\|_1$ into $\|.\|_\infty$, namely the norms $\|.\|_{(k)}$ and their dual ones $\|.\|_{(k)}^*$. We determined all the facets and vertices of the unit balls associated with them. A motivation for that study was the necessity of handling the so-called sparsity constraint in modern optimization problems (coming from data science and machine learning).

Everything has been done with mathematical knowledge acquired at the undergraduate level.

References

1. A. Argyriou, R. Foygel and N. Srebro. *Sparse prediction with the k-support norm.* Advances in Neural Information Processing Systems 25 (2012), 1466 – 1474.

2. R. Bhatia. *Matrix analysis.* Graduate Texts in Mathematics. Springer (1983).

3. A. Brondsted. *An introduction to convex polytopes.* Graduate Texts in Mathematics. Springer (1996).

4. A. Deza, J.-B. Hiriart-Urruty and L. Pournin, *Polytopal balls arising in optimization.* To appear in Contributions to Discrete Mathematics.

5. M. Gaudioso, E. Gorgone and J.-B. Hiriart-Urruty. *Feature selection in SVM via polyhedral k-norm.* Optimization Letters 14 (2019), 19 – 36.

6. J. Gotoh, A. Takeda and K. Tono. *DC formulations and algorithms for sparse optimization problems.* Math. Programming 169 (2018), 141 – 176.

7. J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis.* Grundlehren Text Editions Springer (2001).

8. J.-B. Hiriart-Urruty and Hai Yen Le. *Convexifying the set of matrices of bounded rank: applications to the quasiconvexification and convexification of the rank function.* Optimization Letters 6 (2012), 841 – 849.

9. M.-J. Lai and Y. Wang. *Sparse solutions of undetermined linear systems and their applications.* SIAM (2021).

10. G. A. Watson, *Linear best approximation using a class of polyhedral norms.* Numerical Algorithms 2 (1992), 321 − 336.