Metrics and norms used for obtaining sparse solutions to underdetermined Systems of Linear Equations

Leoni Dalla and George K. Papageorgiou

April 24, 2012

Abstract

This paper focuses on defining a measure, appropriate for obtaining optimally sparse solutions to underdetermined systems of linear equations.[1] The general idea is the extension of metrics in n-dimensional spaces via the Cartesian product of metric spaces.

1 Introduction

In general topology, mathematicians have long ago defined measures that had seen minimum usage (if not at all) in applications. Later on, the development of Measure Theory was mandatory for the progression of applied mathematics and other sciences too. Along with the progress in computer sciences came the demand defining measures of unusual nature and uncovering the properties they obey.

In signal (image or sound) processing a usual problem that arises, is how to transfer a signal using a sparse (economical, but sufficient) representation [2, 3]. Given a specific matrix $\mathbf{A}$ of dimension $m \times n$, $m, n \in \mathbb{N}$ with $m < n$ (underdetermined) and a vector $\mathbf{b}$, find among all, the sparsest or (a less sparse) solution $\mathbf{x} \in \mathbb{R}^n$ of the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$. This is the simplest form of the problem, which means that the noise of the signal is not included (noiseless problem).

Since an undefined system of linear equations has infinite number of solutions, they need to be filtered, using additional functions, in order to obtain solutions of a certain type according to specific criteria. Functions that measure "energy", like the l_2 norm, are used in many occasions, yet measuring sparsity requires a measure of "sparsity", i.e. a different function [2, 3].

The optimization task is minimizing the number of nonzero coordinates of a vector in n-dimensions, i.e. finding a sparse representative of the signal. The number of nonzero

---

[1]The following work done, was within the completion of my master thesis titled "Algorithms for the computation of sparse solutions of undefined systems of equations" at the department of Mathematics, University of Athens which was assigned to me in association with the department of Informatics and Telecommunications, National and Kapodistrian University of Athens.

coordinates of a vector $\mathbf{x}$ is known to be the number of elements included in the set of nonzero values of a vector, which is called *support* of the vector, i.e. $supp\{\mathbf{x}\}$. Also, in recent work of *Donoho* and *Elad* the measure was "used" under the symbol of norm $\|\mathbf{x}\|_0 = \#\{i : x_i \neq 0\}$, but it is clear that it does not satisfy the norm properties [2, 3].

In the following paper, we begin posing some examples in everyday life, where different measures are needed in order to figure out distances. After a short review on metric spaces follows the definition of p-metrics in a Cartesian product space. The next section is of main interest, since we equip $\mathbb{R}^n$ with metrics and prove that the discrete metric could be obtained as a limit of a p-metric [2]. In addition, follows a review on norms and the correlation between norms and metrics. Finally, we conclude with a comparison of functions, on the quest for a convex one, suitable for optimization tasks.

2 Measures in everyday life

In everyday life people subconsciously use measures in order to figure out distances, albeit those measures are not always well defined. Given an arbitrary set of points, a matter of most concern is to measure the distances between those points. However, the measure that we use in every problem is different and depends on the scale we would like to use, as well as the structure of the setting.

The distance between Athens and New York is measured using the geodesic line between those points, i.e. the shortest route between two points on Earth's surface. (Fig. 1).
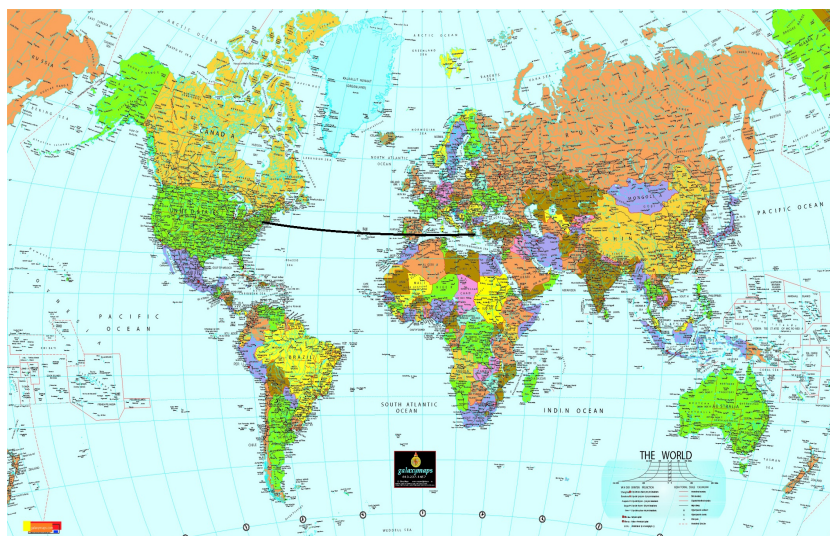

Figure 1: Geodesic distance between Athens - New York: 7920 km

In case of a road trip, the travel distance depends on the road's structure and does not coincide with shortest distance between those two points (towns) (Fig. 2).

Furthermore, the existence of distances that differ from our perception of the shortest path cannot pass unnoticed. The distance that a person has to travel in the area of
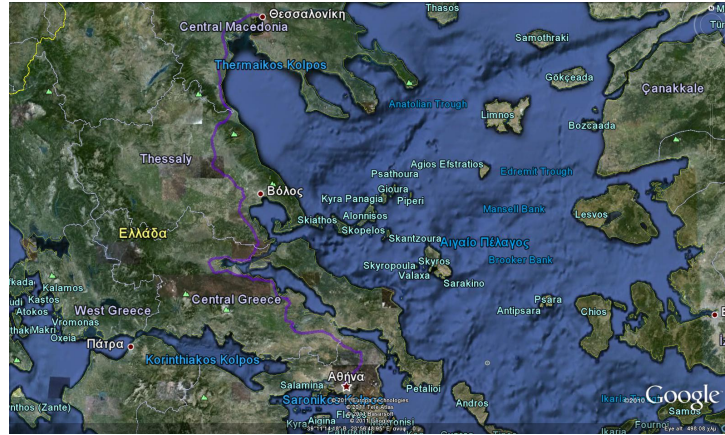
Figure 2: Travel distance between Athens - Thessaloniki: 502 km

Figure 3: Distance in the area of Manhattan, borough of New York City.

3

Manhattan (borough of New York City) in order to move from Times Square to the junction of 57th Street with 9th Avenue depends on the structure of the setting (Fig. 3).

Another measure, used in order to define distances between compact sets, e.g. the distance between two islands, is the Hausdorff distance (named after Felix Hausdorff) between the whole sets K and A defined

$$h(K, A) = max\{\max_{\kappa \in K} \min_{\alpha \in A} |\kappa - \alpha|, \max_{\alpha \in A} \min_{\kappa \in K} |\kappa - \alpha|\}.$$

The latter represents, e.g. the minimum distance one has to travel in order to move from any village of the island of Andros (or Kefallonia) to any village of the island of Kefallonia (or Andros) (Fig. 4).

Figure 4: Distance between Kefallonia - Andros

So far, we have seen cases where the concept of distance needs to be mathematically defined in order to understand, develop and solve problems arising from very different settings. Hence, we should define the means needed in order to measure in a wide variety of cases.

3   Metric spaces

Definition 1 Let X be an arbitrary nonempty set. **Metric**[2] (or distance) in X, is a map $\rho : X \times X \longrightarrow \mathbb{R}$ obeying the following properties:

1. $\rho(x, y) \geq 0$, $\forall x, y \in X$ and $\rho(x, y) = 0 \Leftrightarrow x = y$

2. $\rho(x, y) = \rho(y, x)$, $\forall x, y \in X$, (Symmetric property)

---

[2]Symbolize d, ρ or σ.

4

3. $\rho(x,y) \leq \rho(x,z) + \rho(z,y)$, $\forall x, y, z \in X$, (Triangular inequality)

The elements of the set are called **points**, the real nonnegative number $\rho(x,y)$ is called the **distance** between $x, y \in X$ and the pair (X, ρ) **metric space**.

Consequently, a set equipped with a metric, automatically obtains the structure of a topological space [3]. We now define the open and the closed ball of center $x_0 \in X$ and radius $r > 0$, notions necessary for the topological description of a metric space.

Definition 2 Let (X, d) be a metric space, $x_0 \in X$ and $r > 0$. The set $S_d(x_0, r) = \{x \in X : d(x, x_0) < r\}$ is called an **open ball** of center x_0 and radius r.

Definition 3 Let (X, d) be a metric space, $x_0 \in X$ and $r \geq 0$. The set $\widetilde{S}_d(x_0, r) = \{x \in X : d(x, x_0) \leq r\}$ is called a **closed ball** of center x_0 and radius r.

Definition 4 The set $A \subseteq X$ is called an **open set**, if for every $\alpha \in A$ there exists $r > 0$ such that $S_d(\alpha, r) \subseteq A$.

Definition 5 The set $B \subseteq X$ is called a **closed set**, if its complement $X \setminus B$ is an open set.

Definition 6 The set $\Gamma \subseteq X$ is **bounded** if there exists $x_0 \in X$ and $r > 0$, such that $\Gamma \subseteq S_d(x_0, r)$.

Examples of Metric spaces:

- The most common metrics to use in $\mathbb{R}^n$ are d_1, d_2 between its points $\mathbf{x} = (x_1, x_2, ..., x_n)$, $\mathbf{y} = (y_1, y_2, ..., y_n)$.

  The **metric d_1** (Manhattan metric) in $\mathbb{R}^n$ is defined as

  $$d_1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i|.$$

  Thus, in $\mathbb{R}^2$: $d_1(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + |x_2 - y_2|$ and the closed ball is respectively $\widetilde{S}_{d_1}(\mathbf{0}, r) = \{\mathbf{x} \in \mathbb{R}^2 : d_1(\mathbf{x}, \mathbf{0}) \leq r\}$, $\big($Fig. 5 - i$)$, with $r = 1$ $\big)$.

  The **metric d_2** (Euclidean metric) in $\mathbb{R}^n$ is defined as

  $$d_2(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}}.$$

  Thus, in $\mathbb{R}^2$: $d_2(\mathbf{x}, \mathbf{y}) = \left((x_1 - y_1)^2 + (x_2 - y_2)^2 \right)^{\frac{1}{2}}$ and the closed ball is respectively $\widetilde{S}_{d_2}(\mathbf{0}, r) = \{\mathbf{x} \in \mathbb{R}^2 : d_2(\mathbf{x}, \mathbf{0}) \leq r\}$, $\big($Fig. 5 - ii$)$, with $r = 1$ $\big)$.

---

[3]A topological space doesn't have to be a metric space.

- In every nonempty set X the **metric σ_0** (discrete metric) between points $x, y \in X$ is defined as

$$\sigma_0(x, y) = \begin{cases} 1, & \text{for } x \neq y, \\ 0, & \text{for } x = y. \end{cases} \tag{1}$$

Thus, the closed ball is $\widetilde{S}_{\sigma_0}(x_0, r) = \begin{cases} \{x_0\}, & 0 < r < 1 \\ X, & r \geq 1 \end{cases}$, (Fig. 5 - iii), with $X = (\alpha, \beta) \subseteq \mathbb{R}$).
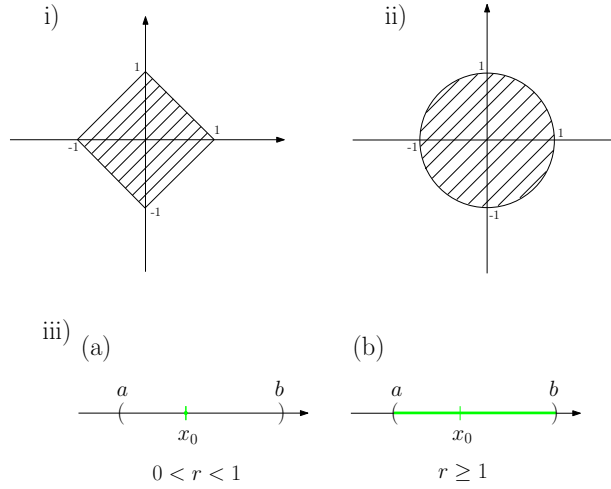

Figure 5: i) $\widetilde{S}_{d_1}(\mathbf{0}, 1) = \{\mathbf{x} \in \mathbb{R}^2 : d_1(\mathbf{x}, \mathbf{0}) \leq 1\}$. ii) The closed ball $\widetilde{S}_{d_2}(\mathbf{0}, 1) = \{\mathbf{x} \in \mathbb{R}^2 : d_2(\mathbf{x}, \mathbf{0}) \leq 1\}$ of the Euclidean metric in $\mathbb{R}^2$ is the unit circle. iii) (a) For $x_0 \in X = (a, b) \subseteq \mathbb{R}$ and $0 < r < 1$, $\widetilde{S}_{\sigma_0}(x_0, r) = \{x_0\}$. iii) ($b$) For $x_0 \in X = (a, b) \subseteq \mathbb{R}$ and $r \geq 1$ the closed ball is the entire set X, i.e. $\widetilde{S}_{\sigma_0}(x_0, r) = (a, b)$.

3.1 Cartesian product space

If the set X is arbitrary and not of a specific structure (e.g. vector space), the discrete metric (1) seems to be the only available choice.

Given the metric spaces (X_i, ρ_i), $i = 1, 2..., n$ we define the p-metrics $(p \geq 1)$[4] in the Cartesian product $X = X_1 \times X_2 \times ... \times X_n$ for $\mathbf{x} = (x_1, ..., x_n)$, $\mathbf{y} = (y_1, ..., y_n) \in X$:

$$d_{(\rho_1, \rho_2, ..., \rho_n;\ p)}(\mathbf{x}, \mathbf{y}) = \begin{cases} \left(\sum_{i=1}^{n} \left(\rho_i(x_i, y_i) \right)^p \right)^{\frac{1}{p}}, & \text{for } 1 \leq p < +\infty \\ \max\{\rho_i(x_i, y_i),\ i = 1, 2, ..., n\}, & \text{for } p = +\infty \end{cases} \tag{2}$$

In case $X_1 = X_2 = ... = X_n = Y$, i.e. $X = Y^n$ and $\rho_1 = \rho_2 = ... = \rho_n = \rho$, we denote $d_{(\rho;\ p)}(\cdot\ ,\ \cdot)$ instead of $d_{(\rho_1, \rho_2, ..., \rho_n;\ p)}(\cdot\ ,\ \cdot)$.

---

[4]For $0 < p < 1$ the triangular inequality does not hold, hence we do not define a metric.

Metrics in (2) are compatible with the ones that already exist in $X_1, X_2, ..., X_n$ according to the following sense. Let $(y_1, ..., y_n) \in X$ be an arbitrary fixed point. Coinciding $x_i \in X_i$ with $(y_1, ..., x_i, ..., y_n) \in X$ we have

$$d_{(\rho_1, \rho_2, ..., \rho_n;\ p)}\big((y_1, ..., x_i, ..., y_n),\ (y_1, ..., x'_i, ..., y_n)\big) = \rho_i(x_i, x'_i),\ \text{for } p \geq 1,$$

where i is the index corresponding to the metric space (X_i, ρ_i).

3.2 $\mathbb{R}^n$ $(n \geq 1)$ equipped with metrics

Due to the discrete nature of computers, our main interest is the set X of the metric space to be a vectored space or subspace. In most of the applications the space that appears is $\mathbb{R}^n$ or subsets of this space. The axiomatic foundation of the set $\mathbb{R}$ of real numbers, gives us the latitude to define the metric $\sigma_{|\cdot|}(x, y) = |x - y|$, $x,\ y \in \mathbb{R}$, where $|\cdot|$ stands for the absolute value of a real number. More generally we may take the metrics $\sigma_s(x, y) = |x-y|^s$, for $0 < s \leq 1$ ($\sigma_1 = \sigma_{|\cdot|}$). At this point it is important to consider that

$$\lim_{s \to 0^+} \sigma_s(x, y) = \sigma_0(x, y). \tag{3}$$

In case of the set $X \subseteq \mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R} = \mathbb{R}^n$, emerge the p-metrics resulting from $(\mathbb{R}, \sigma_{|\cdot|})$, $(\mathbb{R}, \sigma_s)$ for $0 < s < 1$ and $(\mathbb{R}, \sigma_0)$ for points $\mathbf{x} = (x_1, ..., x_n)$, $\mathbf{y} = (y_1, ..., y_n) \in \mathbb{R}^n$, respectively. Alternatively, a combination of σ-metrics is also possible in order to measure in a different way among subsets of X, however the latter choice lacks in practice. Analytically we use the following metrics:

Usual metrics in $\mathbb{R}^n$:

- Let $\mathbb{R}$ equipped with the metric $\sigma_{|\cdot|} = |x - y|$. It follows that $\mathbb{R}^n$ is equipped with the metric $d_{(\sigma_{|\cdot|};\ p)}$ which according to (2) leads to:

$$d_{(\sigma_{|\cdot|};\ p)}(\mathbf{x}, \mathbf{y}) = \begin{cases} \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}}, & \text{for } 1 \leq p < +\infty \\ \max\{|x_i - y_i|,\ i = 1, ..., n\}, & \text{for } p = +\infty \end{cases} \tag{4}$$

  Thus, for $p = 2$ we have the Euclidean metric in $\mathbb{R}^n$ (Fig. 6).

- Let $\mathbb{R}$ equipped with the metric $\sigma_s = |x - y|^s$ for $0 < s \leq 1$. It follows that $\mathbb{R}^n$ is equipped with the metric $d_{(\sigma_s;\ p)}$ that according to (2) leads to:

$$d_{(\sigma_s;\ p)}(\mathbf{x}, \mathbf{y}) = \begin{cases} \left(\sum_{i=1}^{n} \big(\sigma_s(x_i, y_i)\big)^p\right)^{\frac{1}{p}}, & \text{for } 1 \leq p < +\infty \\ \max\{\sigma_s(x_i, y_i),\ i = 1, ..., n\}, & \text{for } p = +\infty \end{cases} \tag{5}$$

  Specifically, for $p = 1$ we have:

$$d_{(\sigma_s;\ 1)}(\mathbf{x},\ \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i|^s,\ 0 < s \leq 1. \tag{6}$$
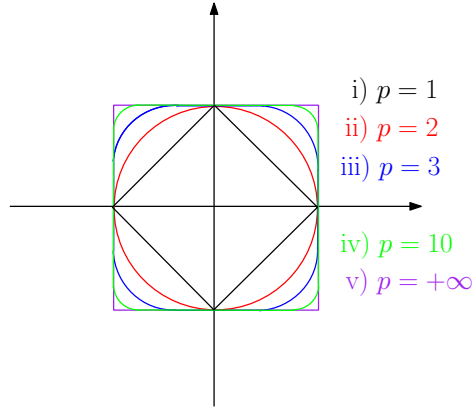
Figure 6: $\widetilde{S}_{d_{(\sigma_{|\cdot|};\ p)}}(\mathbf{0},1) = \{\mathbf{x} \in \mathbb{R}^2 :\ d_{(\sigma_{|\cdot|};\ p)}(\mathbf{x},\mathbf{0}) \le 1\}$ for different values of p. Notice that while p increases, the ball of our space tends to be $\widetilde{S}_{d_{(\sigma_{|\cdot|};\ +\infty)}}(\mathbf{0},1)$, whereas p decreases to 1, tends to be $\widetilde{S}_{d_{(\sigma_{|\cdot|};\ 1)}}(\mathbf{0},1)$.

Discrete metric in $\mathbb{R}^n$:

Let $\mathbb{R}$ equipped with the metric $\sigma_0(x,y) = \begin{cases} 1, & \text{for } x \ne y \\ 0, & \text{for } x = y \end{cases}$, thus $\mathbb{R}^n$ is equipped with the metric $d_{(\sigma_0;\ p)}$ that according to (2) results to:

$$d_{(\sigma_0;\ p)}(\mathbf{x},\mathbf{y}) = \begin{cases} \left(\sum_{i=1}^n \left(\sigma_0(x_i,y_i) \right)^p \right)^{\frac{1}{p}}, & \text{for } 1 \le p < +\infty \\ \max\{\sigma_0(x_i,y_i),\ i=1,...,n\}, & \text{for } p = +\infty \end{cases} \tag{7}$$

Hence considering the case $p = 1$ we have

$$d_{(\sigma_0;\ 1)}(\mathbf{x},\ \mathbf{y}) = \#\{i : x_i \ne y_i\}. \tag{8}$$

As $d_{(\sigma_s;\ 1)}$ is of most importance in sparse theory, we denote it as d_s, if not to be confused with any other metric and use the symbolism $\widetilde{S}_{d_s}$ for the closed ball respectively. Finally, combining equation (8), with both (3) and (6) we obtain:

$$\lim_{s \to 0^+} d_s(\mathbf{x},\mathbf{y}) = \lim_{s \to 0^+} \left(\sum_{i=1}^n |x_i - y_i|^s \right) = \#\{i : x_i \ne y_i\} = d_0(\mathbf{x},\mathbf{y}). \tag{9}$$

The final equation indicates the behaviour of closed balls. In (Fig. 7) it can be easily seen that in $\mathbb{R}^2$ and for $r \in (0,1)$ ($r = 0.5$) the balls $\widetilde{S}_s(\mathbf{0},r)$ decrease, i.e. for $0 < s' < s \le 1$ we have $\widetilde{S}_{s'}(\mathbf{0},r) \subset \widetilde{S}_s(\mathbf{0},r)$ and finally tend to be the ball $\widetilde{S}_0(\mathbf{0},r) = \bigcap_{0 < s \le 1} \widetilde{S}_s(\mathbf{0},r)$. For $r \in [1,2)$ ($r = 1.5$) a relation of subset does not exist between the balls $\widetilde{S}_{s'}(\mathbf{0},r)$ and $\widetilde{S}_s(\mathbf{0},r)$ for $s' < s$, however $\widetilde{S}_s(\mathbf{0},r)$ decrease and tends to coincide with the axis while $s \to 0^+$, i.e.

8

the ball $\widetilde{S}_0(\mathbf{0}, r)$. For $r \in [2, +\infty)$ $(r = 3)$ the balls increase and for $0 < s' < s \le 1$ we have $\widetilde{S}_{s'}(\mathbf{0}, r) \supset \widetilde{S}_s(\mathbf{0}, r)$ until they finally fill the whole space, while $\widetilde{S}_0(\mathbf{0}, r) = \bigcup\limits_{0 < s \le 1} \widetilde{S}_s(\mathbf{0}, r)$.

Equation (9) indicates a desirable measure of sparsity, defined as

$$d_0(\mathbf{x}, \mathbf{0}) = \#\{i : x_i \neq 0\}, \tag{10}$$

measures the number of nonzero coordinates[5] of a vector and belongs to the family of metrics.

Alternative metrics in $\mathbb{R}^n$:
Another measure constructed by a combination of different metrics (2) enables us to measure each subset differently. At its simplest form we state an example in $\mathbb{R}^2$. Let the metric space $(\mathbb{R}, \sigma_0) \times (\mathbb{R}, \sigma_{|\cdot|})$ and set $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2) \in \mathbb{R}^2$.

$$d_{(\sigma_0, \sigma_{|\cdot|};\ p)}(\mathbf{x}, \mathbf{y}) = \left((\sigma_0(x_1, y_1))^p + |x_2 - y_2|^p \right)^{1/p}, \ 1 \le p < +\infty$$

Thus, for $p = 1$:
$$d_{(\sigma_0, \sigma_{|\cdot|};\ 1)}(\mathbf{x}, \mathbf{y}) = \sigma_0(x_1, y_1) + |x_2 - y_2| \tag{11}$$

Consequently, the closed ball of center $\mathbf{0}$ and radius r are (Fig.8):

$$\widetilde{S}(\mathbf{0}, r) = \{\mathbf{x} \in \mathbb{R}^2 : \sigma_0(x_1, 0) + |x_2| \le r\} = \begin{cases} -r \le x_2 \le r, & x_1 = 0, \\ 1 - r \le x_2 \le r - 1, & x_1 \neq 0. \end{cases} \tag{12}$$

---

[5] Also called support of a vector and denoted as $supp\{\mathbf{x}\}$.

i) $\widetilde{S}_s(\mathbf{0}, r)$, $r = 0.5$ ii) $\widetilde{S}_s(\mathbf{0}, r)$, $r = 1.5$ iii) $\widetilde{S}_s(\mathbf{0}, r)$, $r = 3$

Figure 7: $\widetilde{S}_s(\mathbf{0}, r) = \{\mathbf{x} = (x, y) \in \mathbb{R}^2 : \sigma_s(\mathbf{x}, \mathbf{0}) \le r\}$, $0 \le s \le 1$: i) For $0 < r < 1$, $\widetilde{S}_0(\mathbf{0}, r) = \{(0,0)\}$. ii) For $1 \le r < 2$, $\widetilde{S}_0(\mathbf{0}, r) = \{\mathbf{x} = (x, y) \in \mathbb{R}^2 : x = 0$ or $y = 0\}$. iii) For $r \ge 2$, $\widetilde{S}_0(\mathbf{0}, r) = \mathbb{R}^2$.

10

$$\tilde{S}(\mathbf{0}, r)$$
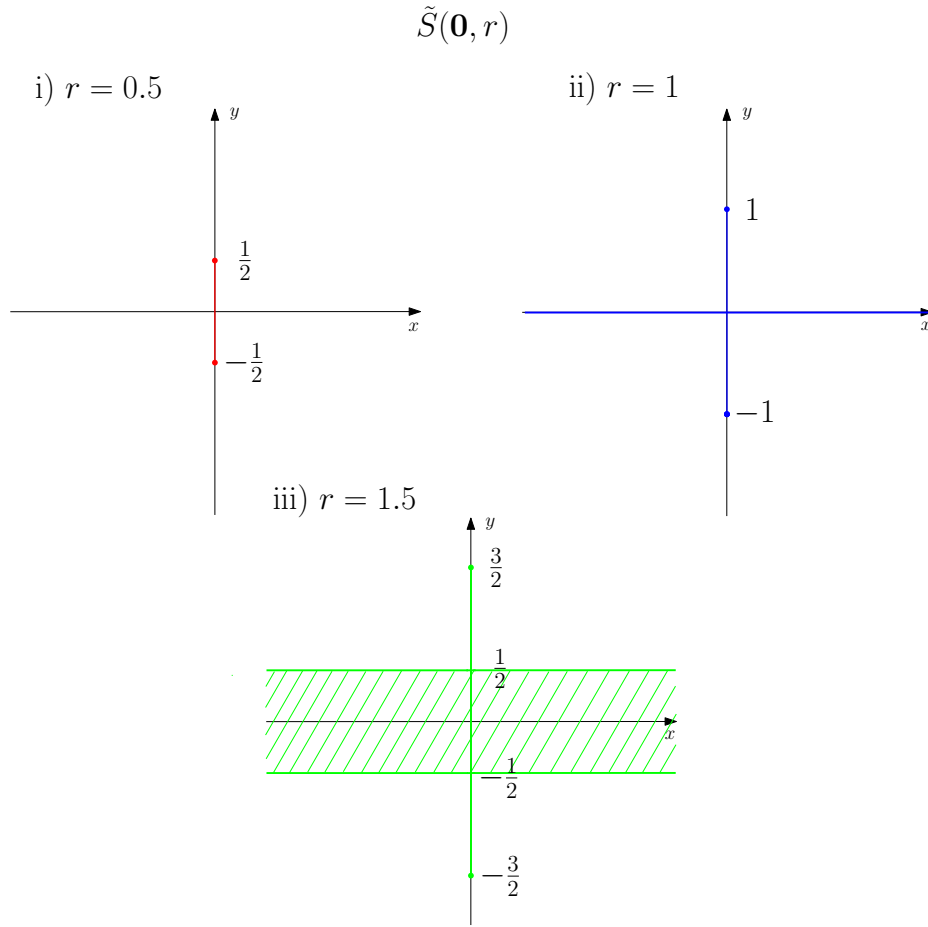
i) $r = 0.5$

ii) $r = 1$

iii) $r = 1.5$

Figure 8: $\widetilde{S}(\mathbf{0}, r) = \{\mathbf{x} = (x, y) \in \mathbb{R}^2 : \sigma_0(x, 0) + |y| \leq r\}$ i) For $0 < r = 0.5 < 1$, $\widetilde{S}(\mathbf{0}, 0.5) = \{-1/2 \leq y \leq 1/2, \ x = 0\}$. ii) For $r = 1$, $\widetilde{S}(\mathbf{0}, 1) = \{-1 \leq y \leq 1, \ x = 0 \ or \ y = 0, \ x \neq 0\}$. iii) For $r = 1.5 > 1$, $\widetilde{S}(\mathbf{0}, 1.5) = \{-3/2 \leq y \leq 3/2, \ x = 0 \ or \ -1/2 \leq y \leq 1/2, \ x \neq 0\}$.

4   Normed spaces

Definition 7 Vector (linear) space is called the trio $(X, +, \cdot)$, where X is a nonempty set, $+ \ : \ X \times X \longrightarrow X$ an inner operation (addition) and $\cdot \ : \ F^6 \times X \longrightarrow X$ an outer operation (scalar product) that obey the following properties:

1. $x + y = y + x, \ \forall x, y \in X$,

2. $(x + y) + z = x + (y + z), \ \forall x, y, z \in X$,

3. There exists $0 \in X$ such that $x + 0 = 0 + x = x, \ \forall x \in X$,

4. For all $x \in X$ there exists $-x \in X$ such that $x + (-x) = (-x) + x = 0$,

---

[6]The field $F = \mathbb{C}$ or $\mathbb{R}$.

5. $\lambda(x + y) = \lambda x + \lambda y, \ \forall x, y \in X$ and $\lambda \in F$,

6. $(\lambda + \mu)x = \lambda x + \mu x, \ \forall x \in X$ and $\lambda, \ \mu \in F$,

7. $\lambda(\mu x) = (\lambda \mu)x, \ \forall x \in X$ and $\lambda, \ \mu \in F$,

8. $1x = x, \ \forall x \in X$.

The elements of a vectored space are called **vectors**.

Definition 8 Let X be a vector space over a field of numbers F. The set $A \subseteq X$ is called **convex**, if for every pair $x, y \in X$ and every $t \in [0, 1]$, the element $tx + (1 - t)y$ belongs to the set A as well.

Definition 9 A real function $f : A \to \mathbb{R}$ defined over a convex subset of a linear space X is called **convex**, if for every $x, y \in A$ and $t \in [0, 1]$,

$$f\big(tx + (1 - t)y\big) \leq tf(x) + (1 - t)f(y).$$

Definition 10 A real function $f : A \to \mathbb{R}$ defined over a convex subset of a linear space X is called **concave**, if for every $x, y \in A$ and $t \in [0, 1]$,

$$f\big(tx + (1 - t)y\big) \geq tf(x) + (1 - t)f(y).$$

Let $\mathbb{R}$ be the vector space. Thus, the absolute value obeys the following properties:

1. $|x| \geq 0, \ x \in \mathbb{R}$ and $|x| = 0 \Leftrightarrow x = 0$

2. $|\lambda x| = |\lambda||x|, \ x \in \mathbb{R}, \ \lambda \in \mathbb{R}$, (positive homogeneous)

3. $|x + y| \leq |x| + |y|, \ x, \ y \in \mathbb{R}$, (triangular inequality)

Therefore the function $f(x) = |x|$ is positive homogeneous, convex and $f(x) > 0$ for $x \neq 0$. A norm is the generalization of the absolute value in higher-dimensional vector spaces.

Definition 11 Let $(X, +, \cdot)$ be a real vector space. The map $\| \cdot \| : X \longrightarrow \mathbb{R}$ is called **norm** if it obeys the following properties:

1. $\|x\| \geq 0, \ \forall x \in X$ and $\|x\| = 0 \Leftrightarrow x = 0$,

2. $\|\lambda x\| = |\lambda|\|x\|, \ \forall x \in X$ and $\lambda \in \mathbb{R}$, (positive homogeneous)

3. $\|x + y\| \leq \|x\| + \|y\|, \ \forall x, y \in X$, (triangular inequality)

The pair $(X, \| \cdot \|)$ is called a **normed space**.

It follows that $f(x) = \|x\|$ is also a positive homogeneous, convex function with $f(x) > 0$ for $x \neq 0$. It is not difficult to see that if $\|x - y\| = f(x - y) = d(x, y)$ for $x, y \in X$, then d is a metric in X with $d(x, 0) = \|x\|$. Moreover, if ρ is a metric in a vector space X satisfying the additional properties $\rho(x + z, y + z) = \rho(x, y)$, $x, y \in X$ and $\rho(\lambda x, 0) = |\lambda| \rho(x, 0)$ with $x \in X$, $\lambda \in \mathbb{R}$ (positive homogeneous), then $\rho(x, 0) = \|x\|$ is a norm. However, we will see that some of the metrics defined do not derive from norms.

Likewise metric spaces $X = X_1 \times X_2 \times ... \times X_n$, the $p-$norms in $(X_i, \| \cdot \|_i)$ are defined.

Examples of normed spaces for $X = \mathbb{R}^n$:

- The **p$-$norms** for $1 < p < +\infty$:

$$\|\mathbf{x}\|_p = \Big(\sum_{i=1}^{n} |x_i|^p \Big)^{\frac{1}{p}}.$$

- The **Euclidean norm $(p = 2)$:**

$$\|\mathbf{x}\|_2 = \Big(\sum_{i=1}^{n} x_i^2 \Big)^{\frac{1}{2}}.$$

- The **addition norm ($1-$norm)** $\| \; \|_1$ and the **norm** $\| \; \|_\infty$ respectively:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|,$$

$$\|\mathbf{x}\|_\infty = max\{|x_i| : i = 1, 2, ..., n\}.$$

At this point we should emphasise that $1 \leq p \leq +\infty$, so that $\| \cdot \|_p$ is a norm in $\mathbb{R}^n$, which could be easily proved using the Minkowski inequality.

Because of the demand for an optimization function, we set

$$f_p(\mathbf{x}) = \Big(\sum_{i=1}^{n} |x_i|^p \Big)^{\frac{1}{p}}, \; p > 0 \tag{13}$$

which is positive homogeneous for every $p > 0$, whilst convex only for $p \geq 1$. For $0 < p < 1$ the function is partially concave, hence the triangular inequality is not satisfied (Fig. 9).

Metric $\sigma_s(x, y) = |x - y|^s$, $x, y \in \mathbb{R}$ is transposition invariant, though not positive homogeneous for $s \neq 1$. Hence, $d_{(\sigma_s; \; p)}(\mathbf{x}, \mathbf{0})$ for $0 < s < 1$ in $\mathbb{R}^n$ are not norms (Fig. 10 for $p = 1$).

From another point of view, the geometric interpretation gives us a clear image of all above. For $p \geq 1$ the closed balls are convex sets, unlike for $0 < p < 1$. Suppose $(X, \| \cdot \|)$ is a normed vector space, the balls $\widetilde{S}_{\|\cdot\|}(x, r)$ are always convex sets. Indeed, if $y, z \in \widetilde{S}_{\|\cdot\|}(x, r)$ then $\|y - x\|$, $\|z - x\| \leq r$, thus for $\lambda \in (0, 1)$, $\|\lambda y + (1 - \lambda z) - x\| =$

$\|\lambda(y - x) + (1 - \lambda)(z - x)\| \leq \lambda r + (1 - \lambda)r \leq r$, i.e. $\lambda y + (1 - \lambda)z \in \widetilde{S}_{\|\cdot\|}(x, r)$, hence the set $\widetilde{S}_{\|\cdot\|}(x, r)$ is convex.

Remark: The property of convexity is of great importance. Suppose that $(X, \|\cdot\|)$ is a normed vector space and let $K \subseteq X$ be a convex and symmetric $(K = -K)$ open set, such that $R, r > 0$ exist and $S_{\|\cdot\|}(0, r) \subseteq K \subseteq S_{\|\cdot\|}(0, R)$. Thus K constitutes the unit ball of another norm, i.e. $\|x\|_K = \inf\{\lambda > 0 : x \in \lambda K\}$ (Minkowski functional) which is topological equivalent to the initial.
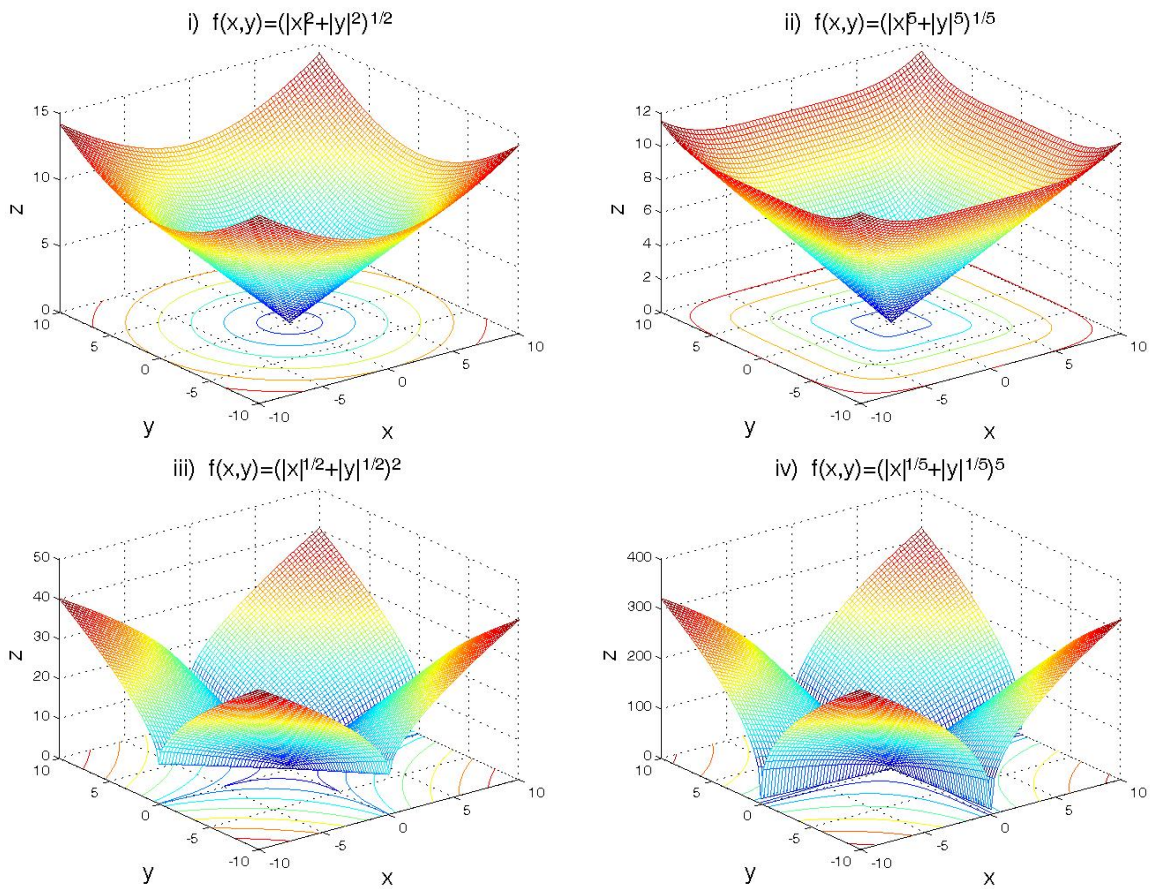

Figure 9: Function $f(x, y) = \left(|x|^p + |y|^p\right)^{\frac{1}{p}}$ for $p > 0$: i) - ii) For $p \geq 1$ those functions are convex. iii) - iv) For $0 < p < 1$ functions are not convex. i) - iv) All functions are positive homogeneous for all $p > 0$.
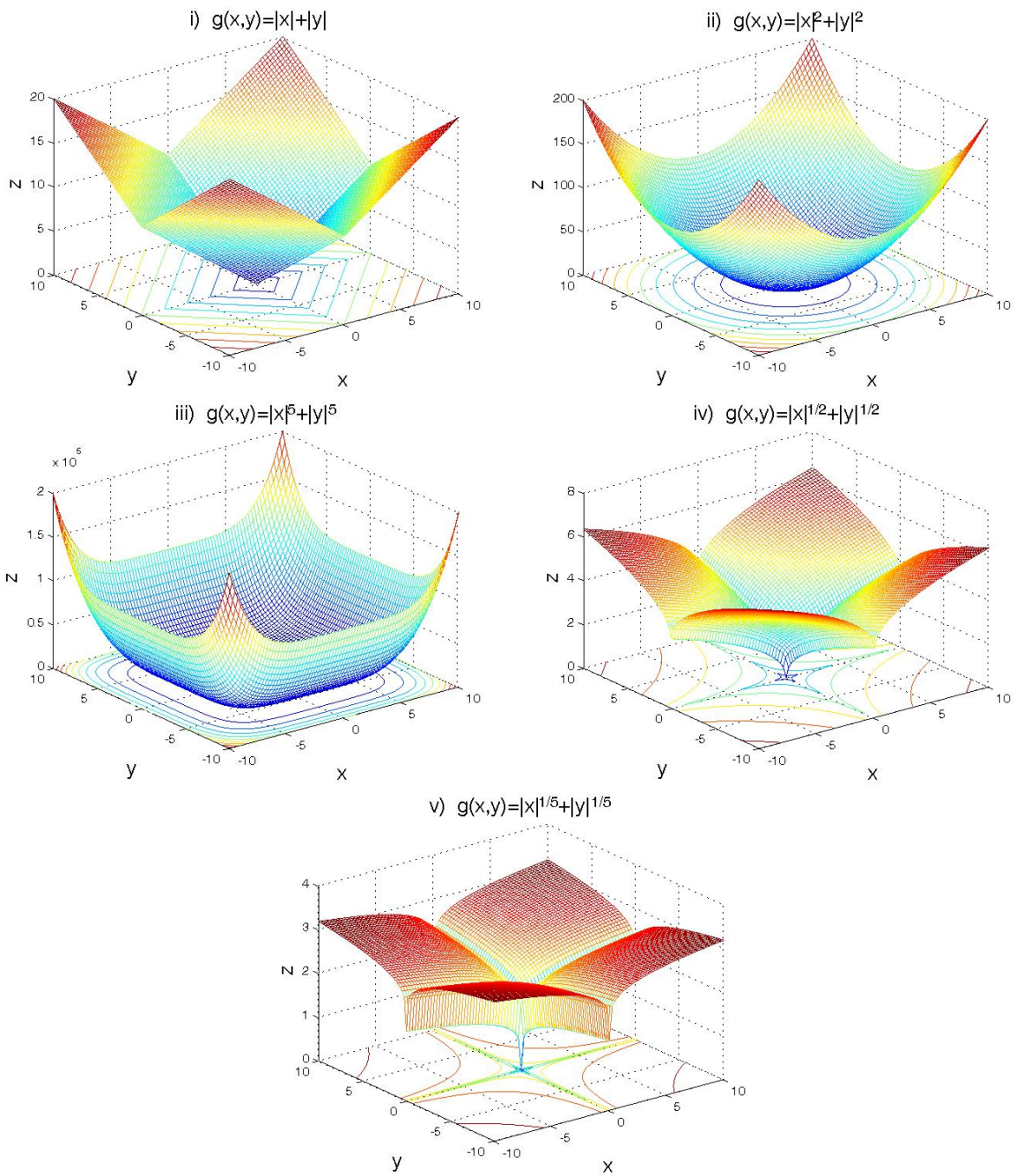
Figure 10: Function $g(x, y) = |x|^s + |y|^s$ for $s > 0$: i) - ii) For $s \geq 1$ the functions are convex. iv) - v) For $0 < s < 1$ functions are not convex. ii) - v) For $s \neq 1$ the functions are not positive homogeneous.

References

[1] Herbert Amann, Joachim Escher, *"Analysis I"*, Birkhäuser Verlag, 1998.

[2] Alfred M. Bruckstein, David L. Donoho, Micheal Elad, *"From Sparse Solutions of Equations to Sparse Modeling of signals and Images"*, SIAM Review Vol.51, No. 1, 2009.

[3] David L. Donoho, Michael Elad, *"Optimally Sparse Representation in General (non-Orthogonal) Dictionaries via l^1 minimization"*, Proceedings, National Academy of Sciences, Vol. 100, pp. 2197-2202, 2003.

[4] Stephen Boyd, Lieven Vandenberghe, *"Convex Optimization"*, Cambridge University Press, 2004.

The University of Athens
Department of Mathematics
Panepistemiopolis 15784
Athens
Greece
Email:
ldalla@math.uoa.gr
ge99210@hotmail.com