CHAPTER 1

First examples of metric spaces

1.1. The concept of a distance

In ancient times people were thinking about the space and its properties and such notions as proximity, distance, size etc. Historically, people came to the idea to use numbers as a measure of length and ancient geometry grew out of the need to measure distances and to plan urban constructions. Ancient astronomers were quite advanced, they were even able to measure distances to the Moon and the Sun. The earliest such accurate measurement was performed by a Greek astronomer and mathematician Hipparchus in the 2nd century BC.

FIGURE 1. Hipparchus.

In simple words one can describe the concept of a distance as follows. Given two points A and B, we connect them by a straight line segment and then count how many times the standard ruler fits into this segment, see Figure 2. Of course this elementary method will not work if we wish to measure the distance to the Moon, and for this purpose Hipparchus invented trigonometry.

FIGURE 2. The ruler for measuring distances.

Another practical way to measure distances is based on using a tape measure, see Figure 3. This method is used by builders, engineers and designers.

FIGURE 3. The tape measure

In this course we shall discuss mathematical meaning of the word “distance”, we shall analyse many interesting examples of different “distances” appearing in various situations in mathematics and in its applications in other sciences.

We also shall use distance to measure similarity between mathematical objects (such as geometric figures, functions, combinatorial structures etc.)

In Euclidean geometry one usually uses distances between points of the line by labelling its points by the real numbers $\mathbb{R}$, see Figure 4. For points on the plane we can use the theorem of

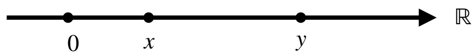

FIGURE 4. $d(x, y) = |x - y|$

Pythagoras: $d^2 = a^2 + b^2$ for a right triangle with sides a, b, d, see Figure 5. Using this theorem we

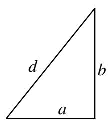

FIGURE 5. Right triangle

may define distance between points of the 2-dimensional plane as follows. The points of the plane are in one-to-one correspondence with pairs of real numbers and the distance between the points $A = (x_1, y_1)$ and $B = (x_2, y_2)$ is given by the formula

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \tag{1.1}$$

This is illustrated by Figure 6, where the sides of the triangle have length $a = |x_1 - x_2|$ and $b = |y_1 - y_2|$. Formula (1.1) defines *the Euclidean distance*, or distance corresponding to the Euclidean 2-dimensional geometry.

More generally, for points in $\mathbf{R}^m$ the Euclidean distance can be defined as follows. The points $v \in \mathbf{R}^m$ are represented by ordered sequences of m real numbers

$$v = (v_1, v_2, \dots, v_m), \quad v_i \in \mathbf{R},$$

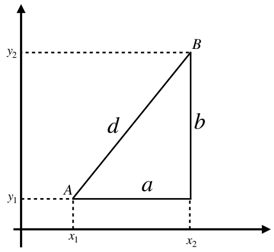

FIGURE 6. Euclidean distance between points of the plane

which are known as *the coordinates* of v. If $w = (w_1, w_2, \dots, w_m) \in \mathbf{R}^m$ is another point then the distance between v and w is defined as follows

$$d(v,w) = \sqrt{(v_1 - w_1)^2 + \cdots + (v_m - w_m)^2} = \left(\sum_{i=1}^{m} (v_i - w_i)^2\right)^{1/2}, \tag{1.2}$$

in analogy with (1.1). Formula (1.2) represents the basis of Euclidean geometry.

Here is a hypothetical picture of Euclid who lived around 300 BC:

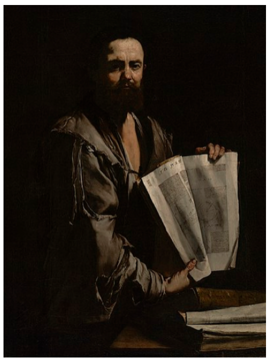

FIGURE 7. Euclid by Jusepe de Ribera

1.2. Hamming distance and the error correcting codes

The notion of distance is useful not only in geometry. Coding theory deals with methods of transmission of information via channels which are not perfect and sustain some accidental errors. *The error correcting codes* use the notion of distance to correct some errors, as we shall now explain.

Let Σ be a finite alphabet (i.e. a finite list of symbols). We are transmitting words of length n formed from symbols of the alphabet Σ. For example, Σ can be the English alphabet or $\Sigma = \{0, 1\}$ - the two element set consisting of 0 and 1. In the latter case words of length n are binary integers of length n such as (01011) or (11010), where $n = 5$. The set of all words of length n in Σ is denoted by Σ^n.

The Hamming distance $d(w_1, w_2)$ between two words $w_1, w_2 \in \Sigma^n$ is defined as the number of positions where these words have distinct entries. For example, the Hamming distance between the words (01011) and (11010) is 2 as they have different symbols on the first and on the last positions only.

Clearly, $d(w_1, w_2)$ is an integer, which can be equal $0, 1, \dots, n$, and $d(w_1, w_2) = 0$ if and only if the words w_1 and w_2 are identical.

When the word w_1 is transmitted by a communication channel, the obtained result w_2 might be different from the source w_1, and the Hamming distance $d(w_1, w_2)$ can be viewed as the measure of quality of transmission.

Let $C \subset \Sigma^n$ be a subset of words which we intend to transmit. This set C is *"the code"*, it is known in advance to both parties involved in information transmission. The number

$$\min\{d(w_1, w_2) | w_1, w_2 \in C,\ w_1 \neq w_2\} \ =\ \delta(C) \tag{1.3}$$

plays a crucial role; it is the minimal distance between distinct code words. The number $\delta(C)$ is *the code distance*, it characterises the ability of error correction in the communication channel.

The simplest method of error correction can work as follows. If on the receiving end a word $w \in \Sigma^n$ has appeared, one searches for a code word $w' \in C$, closest to w with respect to the Hamming distance. The decoder can correct x errors assuming that

$$x < \delta(C)/2.$$

Under this assumption the closest code word $w' \in C$ is unique. We refer the reader to the book [**1**] for further information about the error correcting codes.

1.3. Definition of a metric space

We start this section with a formal definition of a metric space.

Definition 1.1. A metric space is a pair (X, d) where X is a non-empty set and

$$d : X \times X \to \mathbf{R} \tag{1.4}$$

is a function (called *the metric*) satisfying the following three conditions:

(M1) for all $x, y \in X$ one has $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;
(M2) $d(x, y) = d(y, x)$ for $x, y \in X$;
(M3) for all $x, y, z \in X$ one has

$$d(x, z) \leq d(x, y) + d(y, z). \tag{1.5}$$

The latter inequality is known as *the triangle inequality.* This inequality states that a side of a triangle cannot be longer than the sum of the two other sides.

We now consider examples of metric spaces.

Example 1.2. Let $X = \mathbf{R}$ and $d(x, y) = |x - y|$. The axioms (M1), (M2) and (M3) are satisfied. In the case of (M3) we have

$$|x - z| \leq |x - y| + |y - z|$$

which is equivalent to

$$|a + b| \leq |a| + |b| \tag{1.6}$$

where $a = x - y,\ b = y - z$. The inequality (1.6) is an equality when the numbers a and b are both positive or both negative; otherwise one has a strict inequality in (1.6).

Example 1.3. Let X be an arbitrary set. Define the metric $d : X \times X \to \mathbf{R}$ by

$$d(x,y) = \begin{cases} 0, & \text{if} \quad x = y, \\ 1, & \text{if} \quad x \neq y. \end{cases}$$

For $|X|$ equal 3 and 4 this metric spaces are shown on Figure 8. They can be viewed as the sets

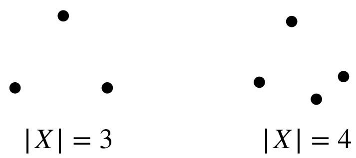

Figure 8. Equilateral metric space

of vertexes of a regular triangle or a regular tetrahedron in the 3-dimensional space. It is obvious that the axioms (M1) - (M3) are satisfied.

Example 1.4. Here is a generalisation of the previous example. Let Γ be a graph with the vertex set V. We may define a metric on V by setting $d_\Gamma(x,y) = 1$ if $x, y \in V$ are two distinct vertexes connected by an edge; besides, we shall set $d_\Gamma(x,y) = 2$ for any pair of distinct vertexes not connected by an edge. The axioms (M1) - (M3) are satisfied.

This example shows that the graphs are in 1-1 correspondence with metric spaces such that their metric takes the values $0, 1, 2$.

Example 1.5. Consider the space $X = \mathbf{R}^m$ with the Euclidean metric (1.2). The properties (M1) and (M2) are obvious. (M3) in this case reads

$$\sqrt{\sum_{i=1}^{m}(x_i - z_i)^2} \leq \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} + \sqrt{\sum_{i=1}^{m}(y_i - z_i)^2}.$$

Denoting $a_i = x_i - y_i$ and $b_i = y_i - z_i$ the above inequality can be wiritten as

$$\sqrt{\sum_{i=1}^{m}(a_i + b_i)^2} \leq \sqrt{\sum_{i=1}^{m} a_i^2} + \sqrt{\sum_{i=1}^{m} b_i^2}.$$

To prove the latter inequality we note that

$$\begin{aligned} \sum_{i=1}^{m}(a_i + b_i)^2 &= \sum_{i=1}^{m} a_i^2 + 2\sum_{i=1}^{m} a_i b_i + \sum_{i=1}^{m} b_i^2 \\ &\leq \sum_{i=1}^{m} a_i^2 + 2\sqrt{\sum_{i=1}^{m} a_i^2} \cdot \sqrt{\sum_{i=1}^{m} b_i^2} + \sum_{i=1}^{m} b_i^2 \\ &= \left(\sqrt{\sum_{i=1}^{m} a_i^2} + \sqrt{\sum_{i=1}^{m} b_i^2}\right)^2 \end{aligned}$$

Here we used the Cauchy inequality

$$\sum_{i=1}^{m} a_i b_i \leq \sqrt{\sum_{i=1}^{m} a_i^2} \cdot \sqrt{\sum_{i=1}^{m} b_i^2}, \tag{1.7}$$

which follows from the following identity

$$\left(\sum_{i=1}^{m} a_i b_i\right)^2 = \sum_{i=1}^{m} a_i^2 \cdot \sum_{i=1}^{m} b_i^2 - \frac{1}{2}\sum_{i,j}(a_i b_j - b_i a_j)^2. \tag{1.8}$$

1.4. Spaces $\mathbf{R}_p^m$

Let $p \in [1, \infty)$ be a fixed real number. In analogy with the Euclidean metric (1.2) we shall define the following metric

$$d_p : \mathbf{R}^m \times \mathbf{R}^m \to \mathbf{R}, \tag{1.9}$$

where

$$d_p(v, v') = \left(\sum_{i=1}^{m} |x_i - x_i'|^p\right)^{1/p} \tag{1.10}$$

for $v = (x_1, \ldots, x_m)$ and $v' = (x_1', \ldots, x_m')$. Clearly d_p coincides with the Euclidean metric for $p = 2$.

Note also that for $m = 1$ (the real line $\mathbf{R}$) the metrics d_p are equal to each other and coincide with the metric of Example 1.2.

The properties (M1) and (M2) are obvious. The property (M3) amounts to the inequality

$$\left(\sum_{i=1}^{m} |x_i - z_i|^p\right)^{1/p} \leq \left(\sum_{i=1}^{m} |x_i - y_i|^p\right)^{1/p} + \left(\sum_{i=1}^{m} |y_i - z_i|^p\right)^{1/p} \tag{1.11}$$

known as the Minkowski inequality. It will be proven in the following section.

The case $p = 1$ is easy since for any $i = 1, \ldots, m$ we have

$$|x_i - z_i| \leq |x_i - y_i| + |y_i - z_i|.$$

We may also include the case $p = \infty$ by defining

$$d_\infty(v, v') = \sup_{i=1,\ldots,m} |x_i - x_i'|. \tag{1.12}$$

The triangle inequality for d_∞ looks as follows:

$$\sup_i |x_i - z_i| \leq \sup_i |x_i - y_i| + \sup_i |y_i - z_i|$$

which can equivalently be written as

$$\sup_i |a_i + b_i| \leq \sup_i |a_i| + \sup_i |b_i| \tag{1.13}$$

where $a_i = x_i - y_i$ and $b_i = y_i - z_i$. Inequality (1.13) follows from $|a_i + b_i| \leq |a_i| + |b_i|$ by taking the supremum on both sides.

Our next task is to prove the Minkowski inequality (1.11) in the case $p \in (1, \infty)$.

1.5. Young's inequality

THEOREM 1.6. *Let $p, q > 1$ be such that*

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{1.14}$$

Then for any $a \geq 0$ and $b \geq 0$ one has

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}. \tag{1.15}$$

Moreover, the equality in (1.15) holds if and only if $a^p = b^q$.

PROOF. Consider the graph of the function $\beta = \alpha^{p-1}$ and the areas S_1 and S_2 indicated on Figure 9. Geometrically it is obvious (see Figure 10) that

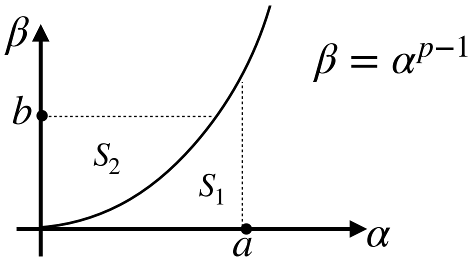

FIGURE 9

$$S_1 + S_2 \geq ab \tag{1.16}$$

and the equality holds iff $b = a^{p-1}$. Note that the product ab is the area of the rectangle built on

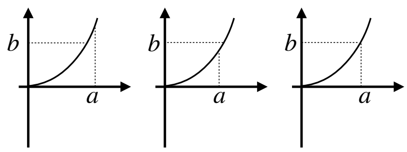

FIGURE 10

a and b, see Figure 10. Computing S_1 we obtain

$$S_1 = \int_0^a \alpha^{p-1} d\alpha = \frac{a^p}{p}.$$

To compute S_2 we express α as a function of β. Since $\beta = \alpha^{p-1}$ one has

$$\alpha = \beta^{\frac{1}{p-1}} = \beta^{q-1}$$

where we used the equality $q - 1 = \frac{1}{p-1}$ which follows from (1.14). Therefore,

$$S_2 = \int_0^b \beta^{q-1} d\beta = \frac{b^q}{q}.$$

Our statement now follows from (1.16). □

1.6. Hölder's inequality

For $p \in (1, \infty)$ define the norm

$$|| \cdot ||_p : \mathbf{R}^m \to \mathbf{R}$$

as follows

$$||v||_p = \left(\sum_{i=1}^{m} |x_i|^p \right)^{1/p}, \quad v = (x_1, \dots, x_m) \in \mathbf{R}^m. \tag{1.17}$$

THEOREM 1.7. *For $p, q \in (1, \infty)$ satisfying*

$$\frac{1}{p} + \frac{1}{q} = 1$$

and for any two vectors $v, w \in \mathbf{R}^m$ one has

$$v \cdot w \leq ||v||_p \cdot ||w||_q, \tag{1.18}$$

where $v \cdot w$ denotes the scalar product, i.e.

$$v \cdot w = \sum_{i=1}^{m} x_i y_i, \quad v = (x_1, \dots, x_m), \ w = (y_1, \dots, y_m).$$

Note that Hölder's inequality (1.18) in the special case $p = 2 = q$ turns into Cauchy's inequality (3.2).

PROOF. Since each side of the inequality (1.18) is homogeneous with respect to v and w, we may assume without loss of generality that $||v||_p = 1$ and $||w||_q = 1$.

For any $i = 1, \dots, m$, Young's inequality (1.15) gives

$$|x_i y_i| \leq \frac{|x_i|^p}{p} + \frac{|y_i|^q}{q}.$$

Summing up we obtain

$$v \cdot w = \sum_{i=1}^{m} x_i y_i \leq \sum_{i=1}^{m} |x_i y_i| \leq \frac{1}{p} \sum_{i=1}^{m} |x_i|^p + \frac{1}{q} \sum_{i=1}^{m} |y_i|^q = \frac{1}{p} + \frac{1}{q} = 1.$$

□

1.7. Minkowski's inequality

THEOREM 1.8. *For every real $p \in [1, \infty]$ and every vectors $v, w \in \mathbf{R}^m$ one has the inequality*

$$||v + w||_p \leq ||v||_p + ||w||_p. \tag{1.19}$$

PROOF. The cases $p = 1$ and $p = \infty$ were considered earlier in §1.4; we shall assume below that $p \in (1, \infty)$.

For $v = (x_1, \dots, x_m)$ and $w = (y_1, \dots, y_m)$ one has

$$|x_i + y_i|^p \leq |x_i + y_i|^{p-1} \cdot |x_i| + |x_i + y_i|^{p-1} \cdot |y_i| \tag{1.20}$$

and applying Hölder's inequality twice we obtain

$$(||v + w||_p)^p = \sum_{i=1}^{m} |x_i + y_i|^p \leq \left(\sum_{i=1}^{m} |x_i + y_i|^{(p-1)q} \right)^{1/q} \cdot (||v||_p + ||w||_p) \tag{1.21}$$

where $p^{-1} + q^{-1} = 1$ and hence $(p-1) \cdot q = p$. Therefore

$$\left(\sum_{i=1}^{m} |x_i + y_i|^{(p-1)q} \right)^{1/q} = (||v+w||_p)^{p/q} \,.$$

Thus, (1.21) reads

$$(||v+w||_p)^p \leq (||v+w||_p)^{p/q} \cdot (||v||_p + ||w||_p) \,.$$

Dividing both sides of this inequality by $(||v+w||_p)^{p/q}$ and taking into account that $p - \frac{p}{q} = 1$ we obtain (1.19). □

EXERCISE 1.9. Show that the Hamming distance $d : \Sigma^n \times \Sigma^n \to \mathbf{R}$ (defined in §1.2) satisfies the axioms (M1), (M2), (M3) of §1.3.

EXERCISE 1.10. For $p \in (0,1)$ consider the function $|| \cdot ||_p : \mathbf{R}^m \to \mathbf{R}$ where for $v \in \mathbf{R}^m$ the "norm" $||v||_p$ is defined by formula (1.17). Show that the Minkowski inequality

$$||v+w||_p \leq ||v||_p + ||w||_p,$$

is not satisfied for some $v, w \in \mathbf{R}^m$ if $m > 1$.

EXERCISE 1.11. Show that for any two fixed vectors $v, v' \in \mathbf{R}^m$ and for $p \in (1, \infty)$ the distance $d_p(v,v')$ tends to $d_\infty(v,v')$ when $p \to \infty$.