CHAPTER 1
METRIC SPACES

Functional analysis is an abstract branch of mathematics that originated from classical analysis. Its development started about eighty years ago, and nowadays functional analytic methods and results are important in various fields of mathematics and its applications. The impetus came from linear algebra, linear ordinary and partial differential equations, calculus of variations, approximation theory and, in particular, linear integral equations, whose theory had the greatest effect on the development and promotion of the modern ideas. Mathematicians observed that problems from different fields often enjoy related features and properties. This fact was used for an effective unifying approach towards such problems, the unification being obtained by the omission of unessential details. Hence the advantage of such an *abstract approach* is that it concentrates on the essential facts, so that these facts become clearly visible since the investigator's attention is not disturbed by unimportant details. In this respect the abstract method is the simplest and most economical method for treating mathematical systems. Since any such abstract system will, in general, have various concrete realizations (concrete *models*), we see that the abstract method is quite versatile in its application to concrete situations. It helps to free the problem from isolation and creates relations and transitions between fields which have at first no contact with one another.

In the abstract approach, one usually starts from a set of elements satisfying certain axioms. The nature of the elements is left unspecified. This is done on purpose. The theory then consists of logical consequences which result from the axioms and are derived as theorems once and for all. This means that in this axiomatic fashion one obtains a mathematical structure whose theory is developed in an abstract way. Those general theorems can then later be applied to various special sets satisfying those axioms.

For example, in algebra this approach is used in connection with fields, rings and groups. In functional analysis we use it in connection with *abstract spaces;* these are of basic importance, and we shall consider some of them (Banach spaces, Hilbert spaces) in great detail. We shall see that in this connection the concept of a "space" is used in a very wide and surprisingly general sense. An *abstract space* will be a set of (unspecified) elements satisfying certain axioms. And by choosing different sets of axioms we shall obtain different types of abstract spaces.

The idea of using abstract spaces in a systematic fashion goes back to M. Fréchet (1906)[1] and is justified by its great success.

In this chapter we consider metric spaces. These are fundamental in functional analysis because they play a role similar to that of the real line **R** in calculus. In fact, they generalize **R** and have been created in order to provide a basis for a unified treatment of important problems from various branches of analysis.

We first define metric spaces and related concepts and illustrate them with typical examples. Special spaces of practical importance are discussed in detail. Much attention is paid to the concept of completeness, a property which a metric space may or may not have. Completeness will play a key role throughout the book.

Important concepts, brief orientation about main content

A *metric space* (cf. 1.1-1) is a set X with a *metric* on it. The metric associates with any pair of elements (*points*) of X a *distance.* The metric is defined axiomatically, the axioms being suggested by certain simple properties of the familiar distance between points on the real line **R** and the complex plane **C**. Basic examples (1.1-2 to 1.2-3) show that the concept of a metric space is remarkably general. A very important additional property which a metric space may have is *completeness* (cf. 1.4-3), which is discussed in detail in Secs. 1.5 and 1.6. Another concept of theoretical and practical interest is *separability* of a metric space (cf. 1.3-5). Separable metric spaces are simpler than nonseparable ones.

1.1 Metric Space

In calculus we study functions defined on the real line **R**. A little reflection shows that in limit processes and many other considerations we use the fact that on **R** we have available a distance function, call it d, which associates a *distance* $d(x, y) = |x - y|$ with every pair of points

[1] References are given in Appendix 3, and we shall refer to books and papers listed in Appendix 3 as is shown here.

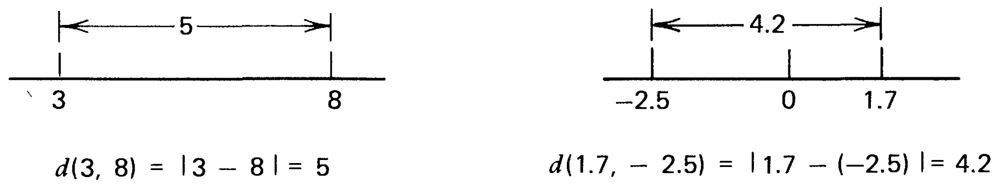

Fig. 2. Distance on **R**

$x, y \in \mathbf{R}$. Figure 2 illustrates the notation. In the plane and in "ordinary" three-dimensional space the situation is similar.

In functional analysis we shall study more general "spaces" and "functions" defined on them. We arrive at a sufficiently general and flexible concept of a "space" as follows. We replace the set of real numbers underlying **R** by an *abstract* set X (set of elements whose nature is left unspecified) and introduce on X a "distance function" which has only a few of the most fundamental properties of the distance function on **R**. But what do we mean by "most fundamental"? This question is far from being trivial. In fact, the choice and formulation of axioms in a definition always needs experience, familiarity with practical problems and a clear idea of the goal to be reached. In the present case, a development of over sixty years has led to the following concept which is basic and very useful in functional analysis and its applications.

1.1-1 Definition (Metric space, metric). A *metric space* is a pair (X, d), where X is a set and d is a *metric on* X (or *distance function on* X), that is, a function defined[2] on $X \times X$ such that for all $x, y, z \in X$ we have:

(M1) d is real-valued, finite and nonnegative.

(M2) $d(x, y) = 0$ if and only if $x = y$.

(M3) $d(x, y) = d(y, x)$ (*Symmetry*).

(M4) $d(x, y) \leqq d(x, z) + d(z, y)$ **(Triangle inequality).** ∎

[2] The symbol $\times$ denotes the *Cartesian product* of sets: $A \times B$ is the set of all ordered pairs (a, b), where $a \in A$ and $b \in B$. Hence $X \times X$ is the set of all ordered pairs of elements of X.

A few related terms are as follows. X is usually called the *underlying set* of (X, d). Its elements are called *points*. For fixed x, y we call the nonnegative number $d(x, y)$ the *distance* from x to y. Properties (M1) to (M4) are the *axioms of a metric*. The name "triangle inequality" is motivated by elementary geometry as shown in Fig. 3.

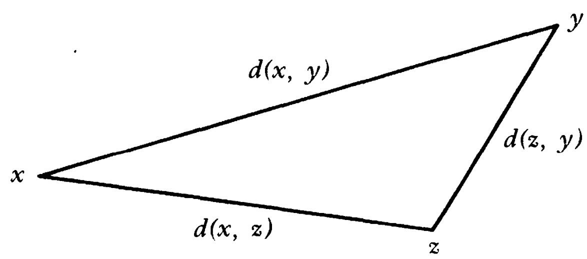

Fig. 3. Triangle inequality in the plane

From (M4) we obtain by induction the *generalized triangle inequality*

$$(1) \qquad d(x_1, x_n) \leqq d(x_1, x_2) + d(x_2, x_3) + \cdots + d(x_{n-1}, x_n).$$

Instead of (X, d) we may simply write X if there is no danger of confusion.

A **subspace** $(Y, \tilde{d})$ of (X, d) is obtained if we take a subset $Y \subset X$ and restrict d to $Y \times Y$; thus the metric on Y is the restriction[3]

$$\tilde{d} = d|_{Y \times Y}.$$

$\tilde{d}$ is called the metric **induced** on Y by d.

We shall now list examples of metric spaces, some of which are already familiar to the reader. To prove that these are metric spaces, we must verify in each case that the axioms (M1) to (M4) are satisfied. Ordinarily, for (M4) this requires more work than for (M1) to (M3). However, in our present examples this will not be difficult, so that we can leave it to the reader (cf. the problem set). More sophisticated

[3] Appendix 1 contains a review on mappings which also includes the concept of a restriction.

metric spaces for which (M4) is not so easily verified are included in the next section.

Examples

1.1-2 Real line R. This is the set of all real numbers, taken with the usual metric defined by

(2) $$d(x, y) = |x - y|.$$

1.1-3 Euclidean plane $\mathbf{R}^2$. The metric space $\mathbf{R}^2$, called the *Euclidean plane,* is obtained if we take the set of ordered pairs of real numbers, written[4] $x = (\xi_1, \xi_2)$, $y = (\eta_1, \eta_2)$, etc., and the *Euclidean metric* defined by

(3) $$d(x, y) = \sqrt{(\xi_1 - \eta_1)^2 + (\xi_2 - \eta_2)^2} \qquad (\geqq 0).$$

See Fig. 4.

Another metric space is obtained if we choose the same set as before but another metric d_1 defined by

(4) $$d_1(x, y) = |\xi_1 - \eta_1| + |\xi_2 - \eta_2|.$$

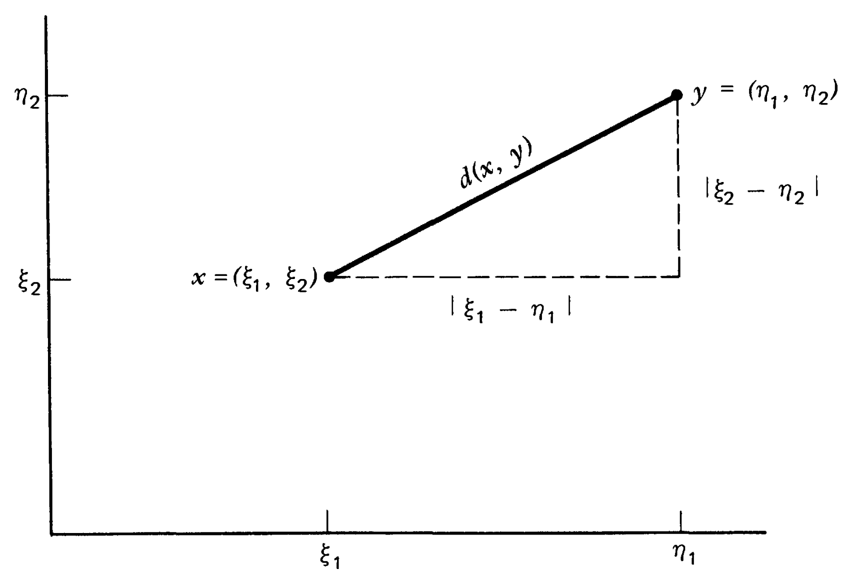

Fig. 4. Euclidean metric on the plane

[4] We do not write $x = (x_1, x_2)$ since $x_1, x_2, \cdots$ are needed later in connection with sequences (starting in Sec. 1.4).

This illustrates the important fact that from a given set (having more than one element) we can obtain various metric spaces by choosing different metrics. (The metric space with metric d_1 does not have a standard name. d_1 is sometimes called the *taxicab metric.* Why? $\mathbf{R}^2$ is sometimes denoted by E^2.)

1.1-4 Three-dimensional Euclidean space $\mathbf{R}^3$. This metric space consists of the set of ordered triples of real numbers $x=(\xi_1, \xi_2, \xi_3)$, $y=(\eta_1, \eta_2, \eta_3)$, etc., and the *Euclidean metric* defined by

$$d(x, y)=\sqrt{(\xi_1-\eta_1)^2+(\xi_2-\eta_2)^2+(\xi_3-\eta_3)^2} \qquad (\geqq 0). \tag{5}$$

1.1-5 Euclidean space $\mathbf{R}^n$, unitary space $\mathbf{C}^n$, complex plane $\mathbf{C}$. The previous examples are special cases of *n-dimensional Euclidean space* $\mathbf{R}^n$. This space is obtained if we take the set of all ordered n-tuples of real numbers, written

$$x=(\xi_1, \cdots, \xi_n), \qquad y=(\eta_1, \cdots, \eta_n)$$

etc., and the *Euclidean metric* defined by

$$d(x, y)=\sqrt{(\xi_1-\eta_1)^2+\cdots+(\xi_n-\eta_n)^2} \qquad (\geqq 0). \tag{6}$$

n-dimensional unitary space $\mathbf{C}^n$ is the space of all ordered n-tuples of *complex* numbers with metric defined by

$$d(x, y)=\sqrt{|\xi_1-\eta_1|^2+\cdots+|\xi_n-\eta_n|^2} \qquad (\geqq 0). \tag{7}$$

When $n=1$ this is the *complex plane* $\mathbf{C}$ with the usual metric defined by

$$d(x, y)=|x-y|. \tag{8}$$

($\mathbf{C}^n$ is sometimes called *complex Euclidean n-space.*)

1.1-6 Sequence space l^∞. This example and the next one give a first impression of how surprisingly general the concept of a metric space is.

As a set X we take the set of all bounded sequences of complex numbers; that is, every element of X is a complex sequence

$$x = (\xi_1, \xi_2, \cdots) \qquad \text{briefly} \qquad x = (\xi_j)$$

such that for all $j = 1, 2, \cdots$ we have

$$|\xi_j| \leqq c_x$$

where c_x is a real number which may depend on x, but does not depend on j. We choose the metric defined by

$$d(x, y) = \sup_{j \in \mathbf{N}} |\xi_j - \eta_j| \tag{9}$$

where $y = (\eta_j) \in X$ and $\mathbf{N} = \{1, 2, \cdots\}$, and sup denotes the supremum (least upper bound).[5] The metric space thus obtained is generally denoted by l^∞. (This somewhat strange notation will be motivated by 1.2-3 in the next section.) l^∞ is a *sequence space* because each element of X (each point of X) is a sequence.

1.1-7 Function space $C[a, b]$. As a set X we take the set of all real-valued functions $x, y, \cdots$ which are functions of an independent real variable t and are defined and continuous on a given closed interval $J = [a, b]$. Choosing the metric defined by

$$d(x, y) = \max_{t \in J} |x(t) - y(t)|, \tag{10}$$

where max denotes the maximum, we obtain a metric space which is denoted by $C[a, b]$. (The letter C suggests "continuous.") This is a *function space* because every point of $C[a, b]$ is a function.

The reader should realize the great difference between calculus, where one ordinarily considers a single function or a few functions at a time, and the present approach where a function becomes merely a single point in a large space.

[5] The reader may wish to look at the review of sup and inf given in A1.6; cf. Appendix 1.

1.1-8 Discrete metric space. We take any set X and on it the so-called *discrete metric* for X, defined by

$$d(x, x) = 0, \qquad d(x, y) = 1 \qquad (x \neq y).$$

This space (X, d) is called a *discrete metric space.* It rarely occurs in applications. However, we shall use it in examples for illustrating certain concepts (and traps for the unwary). ∎

From 1.1-1 we see that a metric is defined in terms of axioms, and we want to mention that axiomatic definitions are nowadays used in many branches of mathematics. Their usefulness was generally recognized after the publication of Hilbert's work about the foundations of geometry, and it is interesting to note that an investigation of one of the *oldest* and simplest parts of mathematics had one of the most important impacts on *modern* mathematics.

Problems

1. Show that the real line is a metric space.

2. Does $d(x, y) = (x-y)^2$ define a metric on the set of all real numbers?

3. Show that $d(x, y) = \sqrt{|x-y|}$ defines a metric on the set of all real numbers.

4. Find all metrics on a set X consisting of two points. Consisting of one point.

5. Let d be a metric on X. Determine all constants k such that (*i*) kd, (*ii*) $d+k$ is a metric on X.

6. Show that d in 1.1-6 satisfies the triangle inequality.

7. If A is the subspace of l^∞ consisting of all sequences of zeros and ones, what is the induced metric on A?

8. Show that another metric $\tilde{d}$ on the set X in 1.1-7 is defined by

$$\tilde{d}(x, y) = \int_a^b |x(t) - y(t)| \, dt.$$

9. Show that d in 1.1-8 is a metric.

10. (Hamming distance) Let X be the set of all ordered triples of zeros and ones. Show that X consists of eight elements and a metric d on X is defined by $d(x, y)$ = number of places where x and y have different entries. (This space and similar spaces of n-tuples play a role in switching and automata theory and coding. $d(x, y)$ is called the *Hamming distance* between x and y; cf. the paper by R. W. Hamming (1950) listed in Appendix 3.)

11. Prove (1).

12. (Triangle inequality) The triangle inequality has several useful consequences. For instance, using (1), show that

$$|d(x, y) - d(z, w)| \leqq d(x, z) + d(y, w).$$

13. Using the triangle inequality, show that

$$|d(x, z) - d(y, z)| \leqq d(x, y).$$

14. (Axioms of a metric) (M1) to (M4) could be replaced by other axioms (without changing the definition). For instance, show that (M3) and (M4) could be obtained from (M2) and

$$d(x, y) \leqq d(z, x) + d(z, y).$$

15. Show that nonnegativity of a metric follows from (M2) to (M4).

1.2 Further Examples of Metric Spaces

To illustrate the concept of a metric space and the process of verifying the axioms of a metric, in particular the triangle inequality (M4), we give three more examples. The last example (space l^p) is the most important one of them in applications.

1.2-1 Sequence space *s*. This space consists of the set of all (bounded or unbounded) sequences of complex numbers and the metric d

defined by

$$d(x, y)=\sum_{j=1}^{\infty}\frac{1}{2^j}\frac{|\xi_j-\eta_j|}{1+|\xi_j-\eta_j|}$$

where $x=(\xi_j)$ and $y=(\eta_j)$. Note that the metric in Example 1.1-6 would not be suitable in the present case. (Why?)

Axioms (M1) to (M3) are satisfied, as we readily see. Let us verify (M4). For this purpose we use the auxiliary function f defined on $\mathbf{R}$ by

$$f(t)=\frac{t}{1+t}.$$

Differentiation gives $f'(t)=1/(1+t)^2$, which is positive. Hence f is monotone increasing. Consequently,

$$|a+b|\leqq|a|+|b|$$

implies

$$f(|a+b|)\leqq f(|a|+|b|).$$

Writing this out and applying the triangle inequality for numbers, we have

$$\begin{aligned}\frac{|a+b|}{1+|a+b|}&\leqq\frac{|a|+|b|}{1+|a|+|b|}\\&=\frac{|a|}{1+|a|+|b|}+\frac{|b|}{1+|a|+|b|}\\&\leqq\frac{|a|}{1+|a|}+\frac{|b|}{1+|b|}.\end{aligned}$$

In this inequality we let $a=\xi_j-\zeta_j$ and $b=\zeta_j-\eta_j$, where $z=(\zeta_j)$. Then $a+b=\xi_j-\eta_j$ and we have

$$\frac{|\xi_j-\eta_j|}{1+|\xi_j-\eta_j|}\leqq\frac{|\xi_j-\zeta_j|}{1+|\xi_j-\zeta_j|}+\frac{|\zeta_j-\eta_j|}{1+|\zeta_j-\eta_j|}.$$