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.

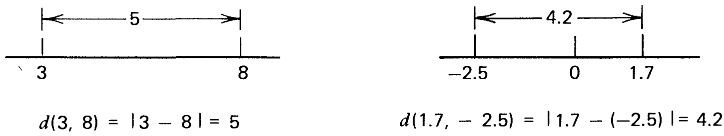

Fig. 2. Distance on **R**

$x, y \in \mathbf{R}$. 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.

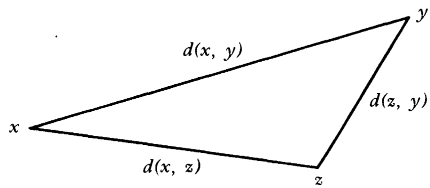

Fig. 3. Triangle inequality in the plane

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

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

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

$$d(x, y) = |x - y|. \tag{2}$$

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

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

See Fig. 4.

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

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

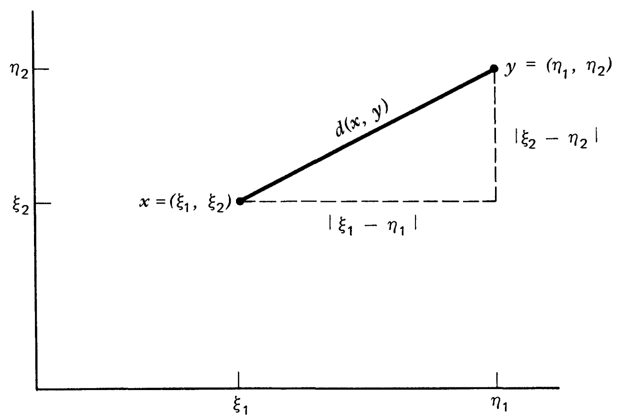

Fig. 4. Euclidean metric on the plane

[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 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* **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

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

If we multiply both sides by $1/2^j$ and sum over j from 1 to ∞, we obtain $d(x, y)$ on the left and the sum of $d(x, z)$ and $d(z, y)$ on the right:

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

This establishes (M4) and proves that s is a metric space.

1.2-2 Space $B(A)$ of bounded functions. By definition, each element $x \in B(A)$ is a function defined and bounded on a given set A, and the metric is defined by

$$d(x, y) = \sup_{t \in A} |x(t) - y(t)|,$$

where sup denotes the supremum (cf. the footnote in 1.1-6). We write $B[a, b]$ for $B(A)$ in the case of an interval $A = [a, b] \subset \mathbf{R}$.

Let us show that $B(A)$ is a metric space. Clearly, (M1) and (M3) hold. Also, $d(x, x) = 0$ is obvious. Conversely, $d(x, y) = 0$ implies $x(t) - y(t) = 0$ for all $t \in A$, so that $x = y$. This gives (M2). Furthermore, for every $t \in A$ we have

$$\begin{aligned} |x(t) - y(t)| &\leqq |x(t) - z(t)| + |z(t) - y(t)| \\ &\leqq \sup_{t \in A} |x(t) - z(t)| + \sup_{t \in A} |z(t) - y(t)|. \end{aligned}$$

This shows that $x - y$ is bounded on A. Since the bound given by the expression in the second line does not depend on t, we may take the supremum on the left and obtain (M4).

1.2-3 Space l^p, Hilbert sequence space l^2, Hölder and Minkowski inequalities for sums. Let $p \geqq 1$ be a fixed real number. By definition, each element in the space l^p is a sequence $x = (\xi_j) = (\xi_1, \xi_2, \cdots)$ of numbers such that $|\xi_1|^p + |\xi_2|^p + \cdots$ converges; thus

$$\sum_{j=1}^{\infty} |\xi_j|^p < \infty \qquad (p \geqq 1, \text{ fixed}) \tag{1}$$

and the metric is defined by

$$d(x, y) = \left(\sum_{j=1}^{\infty} |\xi_j - \eta_j|^p \right)^{1/p} \tag{2}$$

where $y=(\eta_j)$ and $\sum|\eta_j|^p<\infty$. If we take only real sequences [satisfying (1)], we get the *real space* l^p, and if we take complex sequences [satisfying (1)], we get the *complex space* l^p. (Whenever the distinction is essential, we can indicate it by a subscript **R** or **C**, respectively.)

In the case $p=2$ we have the famous *Hilbert sequence space* l^2 with metric defined by

$$d(x, y)=\sqrt{\sum_{j=1}^{\infty}|\xi_j-\eta_j|^2}. \tag{3}$$

This space was introduced and studied by D. Hilbert (1912) in connection with integral equations and is the earliest example of what is now called a *Hilbert space*. (We shall consider Hilbert spaces in great detail, starting in Chap. 3.)

We prove that l^p is a metric space. Clearly, (2) satisfies (M1) to (M3) provided the series on the right converges. We shall prove that it does converge and that (M4) is satisfied. Proceeding stepwise, we shall derive

(*a*) an auxiliary inequality,
(*b*) the Hölder inequality from (*a*),
(*c*) the Minkowski inequality from (*b*),
(*d*) the triangle inequality (M4) from (*c*).

The details are as follows.

(a) Let $p>1$ and define q by

$$\frac{1}{p}+\frac{1}{q}=1. \tag{4}$$

p and q are then called **conjugate exponents.** This is a standard term. From (4) we have

$$1=\frac{p+q}{pq}, \qquad pq=p+q, \qquad (p-1)(q-1)=1. \tag{5}$$

Hence $1/(p-1)=q-1$, so that

$$u=t^{p-1} \qquad \text{implies} \qquad t=u^{q-1}.$$

Let α and β be any positive numbers. Since $\alpha\beta$ is the area of the rectangle in Fig. 5, we thus obtain by integration the inequality

$$\alpha\beta \leqq \int_0^\alpha t^{p-1}\,dt + \int_0^\beta u^{q-1}\,du = \frac{\alpha^p}{p} + \frac{\beta^q}{q}. \tag{6}$$

Note that this inequality is trivially true if $\alpha = 0$ or $\beta = 0$.

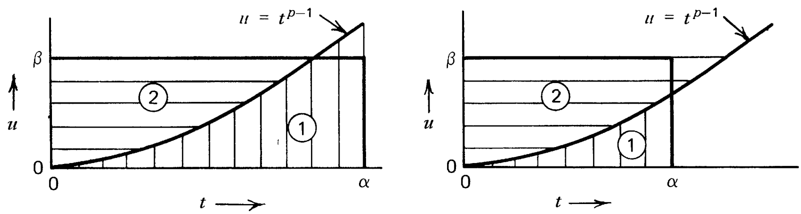

Fig. 5. Inequality (6), where ① corresponds to the first integral in (6) and ② to the second

(b) Let $(\tilde{\xi}_j)$ and $(\tilde{\eta}_j)$ be such that

$$\sum |\tilde{\xi}_j|^p = 1, \qquad \sum |\tilde{\eta}_j|^q = 1. \tag{7}$$

Setting $\alpha = |\tilde{\xi}_j|$ and $\beta = |\tilde{\eta}_j|$, we have from (6) the inequality

$$|\tilde{\xi}_j \tilde{\eta}_j| \leqq \frac{1}{p}|\tilde{\xi}_j|^p + \frac{1}{q}|\tilde{\eta}_j|^q.$$

If we sum over j and use (7) and (4), we obtain

$$\sum |\tilde{\xi}_j \tilde{\eta}_j| \leqq \frac{1}{p} + \frac{1}{q} = 1. \tag{8}$$

We now take any nonzero $x = (\xi_j) \in l^p$ and $y = (\eta_j) \in l^q$ and set

$$\tilde{\xi}_j = \frac{\xi_j}{\left(\sum |\xi_k|^p\right)^{1/p}}, \qquad \tilde{\eta}_j = \frac{\eta_j}{\left(\sum |\eta_m|^q\right)^{1/q}}. \tag{9}$$

Then (7) is satisfied, so that we may apply (8). Substituting (9) into (8) and multiplying the resulting inequality by the product of the denominators in (9), we arrive at the **Hölder inequality** *for sums*

$$\sum_{j=1}^{\infty} |\xi_j \eta_j| \leqq \left(\sum_{k=1}^{\infty} |\xi_k|^p\right)^{1/p} \left(\sum_{m=1}^{\infty} |\eta_m|^q\right)^{1/q} \tag{10}$$

where $p>1$ and $1/p+1/q=1$. This inequality was given by O. Hölder (1889).

If $p=2$, then $q=2$ and (10) yields the **Cauchy-Schwarz inequality** *for sums*

$$\sum_{j=1}^{\infty} |\xi_j \eta_j| \leqq \sqrt{\sum_{k=1}^{\infty} |\xi_k|^2} \sqrt{\sum_{m=1}^{\infty} |\eta_m|^2}. \tag{11}$$

It is too early to say much about this case $p=q=2$ in which p equals its conjugate q, but we want to make at least the brief remark that this case will play a particular role in some of our later chapters and lead to a space (a Hilbert space) which is "nicer" than spaces with $p \neq 2$.

(c) We now prove the **Minkowski inequality** *for sums*

$$\left(\sum_{j=1}^{\infty} |\xi_j + \eta_j|^p\right)^{1/p} \leqq \left(\sum_{k=1}^{\infty} |\xi_k|^p\right)^{1/p} + \left(\sum_{m=1}^{\infty} |\eta_m|^p\right)^{1/p} \tag{12}$$

where $x=(\xi_j) \in l^p$ and $y=(\eta_j) \in l^p$, and $p \geqq 1$. For finite sums this inequality was given by H. Minkowski (1896).

For $p=1$ the inequality follows readily from the triangle inequality for numbers. Let $p>1$. To simplify the formulas we shall write $\xi_j + \eta_j = \omega_j$. The triangle inequality for numbers gives

$$\begin{aligned} |\omega_j|^p &= |\xi_j + \eta_j| \, |\omega_j|^{p-1} \\ &\leqq (|\xi_j| + |\eta_j|)|\omega_j|^{p-1}. \end{aligned}$$

Summing over j from 1 to any fixed n, we obtain

$$\sum |\omega_j|^p \leqq \sum |\xi_j| \, |\omega_j|^{p-1} + \sum |\eta_j| \, |\omega_j|^{p-1}. \tag{13}$$

To the first sum on the right we apply the Hölder inequality, finding

$$\sum |\xi_j| \, |\omega_j|^{p-1} \leqq \left[\sum |\xi_k|^p\right]^{1/p} \left[\sum (|\omega_m|^{p-1})^q\right]^{1/q}.$$

On the right we simply have

$$(p-1)q=p$$

because $pq=p+q$; see (5). Treating the last sum in (13) in a similar fashion, we obtain

$$\sum |\eta_j| |\omega_j|^{p-1} \leqq \left[\sum |\eta_k|^p\right]^{1/p} \left[\sum |\omega_m|^p\right]^{1/q}.$$

Together,

$$\sum |\omega_j|^p \leqq \left\{\left[\sum |\xi_k|^p\right]^{1/p} + \left[\sum |\eta_k|^p\right]^{1/p}\right\}\left(\sum |\omega_m|^p\right)^{1/q}.$$

Dividing by the last factor on the right and noting that $1-1/q=1/p$, we obtain (12) with n instead of ∞. We now let $n \longrightarrow \infty$. On the right this yields two series which converge because $x, y \in l^p$. Hence the series on the left also converges, and (12) is proved.

(d) From (12) it follows that for x and y in l^p the series in (2) converges. (12) also yields the triangle inequality. In fact, taking any $x, y, z \in l^p$, writing $z=(\zeta_j)$ and using the triangle inequality for numbers and then (12), we obtain

$$\begin{aligned} d(x, y) &= \left(\sum |\xi_j - \eta_j|^p\right)^{1/p} \\ &\leqq \left(\sum [|\xi_j - \zeta_j| + |\zeta_j - \eta_j|]^p\right)^{1/p} \\ &\leqq \left(\sum |\xi_j - \zeta_j|^p\right)^{1/p} + \left(\sum |\zeta_j - \eta_j|^p\right)^{1/p} \\ &= d(x, z) + d(z, y). \end{aligned}$$

This completes the proof that l^p is a metric space. ∎

The inequalities (10) to (12) obtained in this proof are of general importance as indispensable tools in various theoretical and practical problems, and we shall apply them a number of times in our further work.

Problems

1. Show that in 1.2-1 we can obtain another metric by replacing $1/2^j$ with $\mu_j > 0$ such that $\sum \mu_j$ converges.

2. Using (6), show that the geometric mean of two positive numbers does not exceed the arithmetic mean.

3. Show that the Cauchy-Schwarz inequality (11) implies

$$(|\xi_1| + \cdots + |\xi_n|)^2 \leqq n(|\xi_1|^2 + \cdots + |\xi_n|^2).$$

4. (Space l^p) Find a sequence which converges to 0, but is not in any space l^p, where $1 \leqq p < +\infty$.

5. Find a sequence x which is in l^p with $p > 1$ but $x \notin l^1$.

6. (Diameter, bounded set) The *diameter* $\delta(A)$ of a nonempty set A in a metric space (X, d) is defined to be

$$\delta(A) = \sup_{x,y \in A} d(x, y).$$

A is said to be *bounded* if $\delta(A) < \infty$. Show that $A \subset B$ implies $\delta(A) \leqq \delta(B)$.

7. Show that $\delta(A) = 0$ (cf. Prob. 6) if and only if A consists of a single point.

8. (Distance between sets) The *distance* $D(A, B)$ between two nonempty subsets A and B of a metric space (X, d) is defined to be

$$D(A, B) = \inf_{\substack{a \in A \\ b \in B}} d(a, b).$$

Show that D does *not* define a metric on the power set of X. (For this reason we use another symbol, D, but one that still reminds us of d.)

9. If $A \cap B \neq \phi$, show that $D(A, B) = 0$ in Prob. 8. What about the converse?

10. The *distance* $D(x, B)$ from a point x to a non-empty subset B of (X, d) is defined to be

$$D(x, B) = \inf_{b \in B} d(x, b),$$

in agreement with Prob. 8. Show that for any $x, y \in X$,

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

11. If (X, d) is any metric space, show that another metric on X is defined by

$$\tilde{d}(x, y) = \frac{d(x, y)}{1 + d(x, y)}$$

and X is bounded in the metric $\tilde{d}$.

12. Show that the union of two bounded sets A and B in a metric space is a bounded set. (Definition in Prob. 6.)

13. (Product of metric spaces) The Cartesian product $X = X_1 \times X_2$ of two metric spaces (X_1, d_1) and (X_2, d_2) can be made into a metric space (X, d) in many ways. For instance, show that a metric d is defined by

$$d(x, y) = d_1(x_1, y_1) + d_2(x_2, y_2),$$

where $x = (x_1, x_2)$, $y = (y_1, y_2)$.

14. Show that another metric on X in Prob. 13 is defined by

$$\tilde{d}(x, y) = \sqrt{d_1(x_1, y_1)^2 + d_2(x_2, y_2)^2}.$$

15. Show that a third metric on X in Prob. 13 is defined by

$$\tilde{\tilde{d}}(x, y) = \max\,[d_1(x_1, y_1), d_2(x_2, y_2)].$$

(The metrics in Probs. 13 to 15 are of practical importance, and other metrics on X are possible.)

1.3 Open Set, Closed Set, Neighborhood

There is a considerable number of auxiliary concepts which play a role in connection with metric spaces. Those which we shall need are included in this section. Hence the section contains many concepts (more than any other section of the book), but the reader will notice that several of them become quite familiar when applied to Euclidean space. Of course this is a great convenience and shows the advantage of the terminology which is inspired by classical geometry.

We first consider important types of subsets of a given metric space $X=(X, d)$.

1.3-1 Definition (Ball and sphere). Given a point $x_0 \in X$ and a real number $r>0$, we define[6] three types of sets:

$$\begin{array}{llll} & (a) & B(x_0; r)=\{x \in X \mid d(x, x_0)<r\} & \textbf{(Open ball)} \\ \textbf{(1)} & (b) & \tilde{B}(x_0; r)=\{x \in X \mid d(x, x_0)\leq r\} & \textbf{(Closed ball)} \\ & (c) & S(x_0; r)=\{x \in X \mid d(x, x_0)=r\} & \textbf{(Sphere)} \end{array}$$

In all three cases, x_0 is called the *center* and r the *radius*. ▮

We see that an open ball of radius r is the set of all points in X whose distance from the center of the ball is less than r. Furthermore, the definition immediately implies that

$$S(x_0; r)=\tilde{B}(x_0; r)-B(x_0; r). \tag{2}$$

Warning. In working with metric spaces, it is a great advantage that we use a terminology which is analogous to that of Euclidean geometry. However, we should beware of a danger, namely, of assuming that balls and spheres in an arbitrary metric space enjoy the same properties as balls and spheres in $\mathbf{R}^3$. This is not so. An unusual property is that a sphere can be empty. For example, in a discrete metric space 1.1-8 we have $S(x_0; r)=\varnothing$ if $r\neq 1$. (What about spheres of radius 1 in this case?) Another unusual property will be mentioned later.

Let us proceed to the next two concepts, which are related.

1.3-2 Definition (Open set, closed set). A subset M of a metric space X is said to be *open* if it contains a ball about each of its points. A subset K of X is said to be *closed* if its complement (in X) is open, that is, $K^C=X-K$ is open. ▮

The reader will easily see from this definition that an open ball is an open set and a closed ball is a closed set.

[6] Some familiarity with the usual set-theoretic notations is assumed, but a review is included in Appendix 1.

An open ball $B(x_0; \varepsilon)$ of radius ε is often called an ε-*neighborhood* of x_0. (Here, $\varepsilon > 0$, by Def. 1.3-1.) By a **neighborhood**[7] of x_0 we mean any subset of X which contains an ε-neighborhood of x_0.

We see directly from the definition that every neighborhood of x_0 contains x_0; in other words, x_0 is a point of each of its neighborhoods. And if N is a neighborhood of x_0 and $N \subset M$, then M is also a neighborhood of x_0.

We call x_0 an **interior point** of a set $M \subset X$ if M is a neighborhood of x_0. The **interior** of M is the set of all interior points of M and may be denoted by M^0 or Int(M), but there is no generally accepted notation. Int(M) is open and is the largest open set contained in M.

It is not difficult to show that the collection of all open subsets of X, call it $\mathcal{T}$, has the following properties:

(T1) $\varnothing \in \mathcal{T}$, $X \in \mathcal{T}$.

(T2) The union of any members of $\mathcal{T}$ is a member of $\mathcal{T}$.

(T3) The intersection of finitely many members of $\mathcal{T}$ is a member of $\mathcal{T}$.

Proof. (T1) follows by noting that $\varnothing$ is open since $\varnothing$ has no elements and, obviously, X is open. We prove (T2). Any point x of the union U of open sets belongs to (at least) one of these sets, call it M, and M contains a ball B about x since M is open. Then $B \subset U$, by the definition of a union. This proves (T2). Finally, if y is any point of the intersection of open sets $M_1, \cdots, M_n$, then each M_j contains a ball about y and a smallest of these balls is contained in that intersection. This proves (T3). ∎

We mention that the properties (T1) to (T3) are so fundamental that one wants to retain them in a more general setting. Accordingly, one defines a **topological space** $(X, \mathcal{T})$ to be a set X and a collection $\mathcal{T}$ of subsets of X such that $\mathcal{T}$ satisfies the *axioms* (T1) to (T3). The set $\mathcal{T}$ is called *a topology for X*. From this definition we have:

A metric space is a topological space.

[7] In the older literature, neighborhoods used to be open sets, but this requirement has been dropped from the definition.

Open sets also play a role in connection with continuous mappings, where continuity is a natural generalization of the continuity known from calculus and is defined as follows.

1.3-3 Definition (Continuous mapping). Let $X=(X, d)$ and $Y=(Y, \tilde{d})$ be metric spaces. A mapping $T\colon X \longrightarrow Y$ is said to be *continuous at a point* $x_0 \in X$ if for every $\varepsilon>0$ there is a $\delta>0$ such that[8] (see Fig. 6)

$$\tilde{d}(Tx, Tx_0)<\varepsilon \qquad \text{for all } x \text{ satisfying} \qquad d(x, x_0)<\delta.$$

T is said to be *continuous* if it is continuous at every point of X. ∎

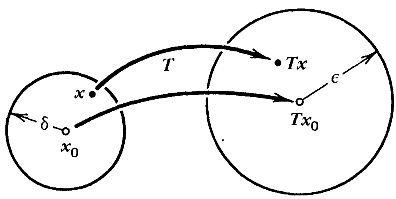

Fig. 6. Inequalities in Def. 1.3-3 illustrated in the case of Euclidean planes $X=\mathbf{R}^2$ and $Y=\mathbf{R}^2$

It is important and interesting that continuous mappings can be characterized in terms of open sets as follows.

1.3-4 Theorem (Continuous mapping). *A mapping T of a metric space X into a metric space Y is continuous if and only if the inverse image of any open subset of Y is an open subset of X.*

Proof. (*a*) Suppose that T is continuous. Let $S \subset Y$ be open and S_0 the inverse image of S. If $S_0=\varnothing$, it is open. Let $S_0 \neq \varnothing$. For any $x_0 \in S_0$ let $y_0=Tx_0$. Since S is open, it contains an ε-neighborhood N of y_0; see Fig. 7. Since T is continuous, x_0 has a δ-neighborhood N_0 which is mapped into N. Since $N \subset S$, we have $N_0 \subset S_0$, so that S_0 is open because $x_0 \in S_0$ was arbitrary.

(*b*) Conversely, assume that the inverse image of every open set in Y is an open set in X. Then for every $x_0 \in X$ and any

[8] In calculus we usually write $y=f(x)$. A corresponding notation for the image of x under T would be $T(x)$. However, to simplify formulas in functional analysis, it is customary to omit the parentheses and write Tx. A review of the definition of a mapping is included in A1.2; cf. Appendix 1.

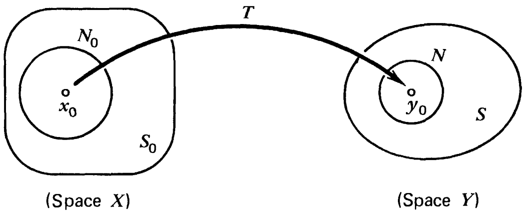

Fig. 7. Notation in part (a) of the proof of Theorem 1.3-4

ε-neighborhood N of Tx_0, the inverse image N_0 of N is open, since N is open, and N_0 contains x_0. Hence N_0 also contains a δ-neighborhood of x_0, which is mapped into N because N_0 is mapped into N. Consequently, by the definition, T is continuous at x_0. Since $x_0 \in X$ was arbitrary, T is continuous. ∎

We shall now introduce two more concepts, which are related. Let M be a subset of a metric space X. Then a point x_0 of X (which may or may not be a point of M) is called an **accumulation point** *of* M (or *limit point of* M) if every neighborhood of x_0 contains at least one point $y \in M$ distinct from x_0. The set consisting of the points of M and the accumulation points of M is called the **closure** *of* M and is denoted by

$$\bar{M}.$$

It is the smallest closed set containing M.

Before we go on, we mention another unusual property of balls in a metric space. Whereas in $\mathbf{R}^3$ the closure $\overline{B(x_0; r)}$ of an open ball $B(x_0; r)$ is the closed ball $\tilde{B}(x_0; r)$, this may not hold in a general metric space. We invite the reader to illustrate this with an example.

Using the concept of the closure, let us give a definition which will be of particular importance in our further work:

1.3-5 Definition (Dense set, separable space). A subset M of a metric space X is said to be *dense in* X if

$$\bar{M} = X.$$

X is said to be *separable* if it has a countable subset which is dense in X. (For the definition of a countable set, see A1.1 in Appendix 1 if necessary.) ∎

Hence if M is dense in X, then every ball in X, no matter how small, will contain points of M; or, in other words, in this case there is no point $x \in X$ which has a neighborhood that does not contain points of M.

We shall see later that separable metric spaces are somewhat simpler than nonseparable ones. For the time being, let us consider some important examples of separable and nonseparable spaces, so that we may become familiar with these basic concepts.

Examples

1.3-6 Real line R. *The real line* **R** *is separable.*

Proof. The set **Q** of all rational numbers is countable and is dense in **R**.

1.3-7 Complex plane C. *The complex plane* **C** *is separable.*

Proof. A countable dense subset of **C** is the set of all complex numbers whose real and imaginary parts are both rational.

1.3-8 Discrete metric space. *A discrete metric space X is separable if and only if X is countable.* (Cf. 1.1-8.)

Proof. The kind of metric implies that no proper subset of X can be dense in X. Hence the only dense set in X is X itself, and the statement follows.

1.3-9 Space l^∞. *The space l^∞ is not separable.* (Cf. 1.1-6.)

Proof. Let $y=(\eta_1, \eta_2, \eta_3, \cdots)$ be a sequence of zeros and ones. Then $y \in l^\infty$. With y we associate the real number $\hat{y}$ whose binary representation is

$$\frac{\eta_1}{2^1}+\frac{\eta_2}{2^2}+\frac{\eta_3}{2^3}+\cdots.$$

We now use the facts that the set of points in the interval $[0, 1]$ is uncountable, each $\hat{y} \in [0, 1]$ has a binary representation, and different $\hat{y}$'s have different binary representations. Hence there are uncountably many sequences of zeros and ones. The metric on l^∞ shows that any two of them which are not equal must be of distance 1 apart. If we let each of these sequences be the center of a small ball, say, of radius 1/3, these balls do not intersect and we have uncountably many of them. If M is any dense set in l^∞, each of these nonintersecting balls must contain an element of M. Hence M cannot be countable. Since M was an arbitrary dense set, this shows that l^∞ cannot have dense subsets which are countable. Consequently, l^∞ is not separable.

1.3-10 Space l^p. *The space l^p with $1 \leqq p < +\infty$ is separable.* (Cf. 1.2-3.)

Proof. Let M be the set of all sequences y of the form

$$y = (\eta_1, \eta_2, \cdots, \eta_n, 0, 0, \cdots)$$

where n is any positive integer and the η_j's are rational. M is countable. We show that M is dense in l^p. Let $x = (\xi_j) \in l^p$ be arbitrary. Then for every $\varepsilon > 0$ there is an n (depending on ε) such that

$$\sum_{j=n+1}^{\infty} |\xi_j|^p < \frac{\varepsilon^p}{2}$$

because on the left we have the remainder of a converging series. Since the rationals are dense in $\mathbf{R}$, for each ξ_j there is a rational η_j close to it. Hence we can find a $y \in M$ satisfying

$$\sum_{j=1}^{n} |\xi_j - \eta_j|^p < \frac{\varepsilon^p}{2}.$$

It follows that

$$[d(x, y)]^p = \sum_{j=1}^{n} |\xi_j - \eta_j|^p + \sum_{j=n+1}^{\infty} |\xi_j|^p < \varepsilon^p.$$

We thus have $d(x, y) < \varepsilon$ and see that M is dense in l^p.

Problems

1. Justify the terms "open ball" and "closed ball" by proving that (*a*) any open ball is an open set, (*b*) any closed ball is a closed set.

2. What is an open ball $B(x_0; 1)$ on $\mathbf{R}$? In $\mathbf{C}$? (Cf. 1.1-5.) In $C[a, b]$? (Cf. 1.1-7.) Explain Fig. 8.

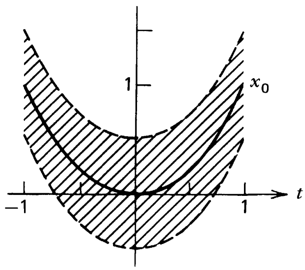

Fig. 8. Region containing the graphs of all $x \in C[-1, 1]$ which constitute the ε-neighborhood, with $\varepsilon = 1/2$, of $x_0 \in C[-1, 1]$ given by $x_0(t) = t^2$

3. Consider $C[0, 2\pi]$ and determine the smallest r such that $y \in \tilde{B}(x; r)$, where $x(t) = \sin t$ and $y(t) = \cos t$.

4. Show that any nonempty set $A \subset (X, d)$ is open if and only if it is a union of open balls.

5. It is important to realize that certain sets may be open and closed at the same time. (*a*) Show that this is always the case for X and $\varnothing$. (*b*) Show that in a discrete metric space X (cf. 1.1-8), every subset is open and closed.

6. If x_0 is an accumulation point of a set $A \subset (X, d)$, show that any neighborhood of x_0 contains infinitely many points of A.

7. Describe the closure of each of the following subsets. (*a*) The integers on $\mathbf{R}$, (*b*) the rational numbers on $\mathbf{R}$, (*c*) the complex numbers with rational real and imaginary parts in $\mathbf{C}$, (*d*) the disk $\{z \mid |z| < 1\} \subset \mathbf{C}$.

8. Show that the closure $\overline{B(x_0; r)}$ of an open ball $B(x_0; r)$ in a metric space can differ from the closed ball $\tilde{B}(x_0; r)$.

9. Show that $A \subset \bar{A}$, $\bar{\bar{A}} = \bar{A}$, $\overline{A \cup B} = \bar{A} \cup \bar{B}$, $\overline{A \cap B} \subset \bar{A} \cap \bar{B}$.

10. A point x not belonging to a *closed* set $M \subset (X, d)$ always has a nonzero distance from M. To prove this, show that $x \in \bar{A}$ if and only if $D(x, A) = 0$ (cf. Prob. 10, Sec. 1.2); here A is any nonempty subset of X.

11. (Boundary) A *boundary point* x of a set $A \subset (X, d)$ is a point of X (which may or may not belong to A) such that every neighborhood of x contains points of A as well as points not belonging to A; and the *boundary* (or *frontier*) of A is the set of all boundary points of A. Describe the boundary of (*a*) the intervals $(-1, 1)$, $[-1, 1)$, $[-1, 1]$ on

$\mathbf{R}$; (*b*) the set of all rational numbers on $\mathbf{R}$; (*c*) the disks $\{z \mid |z|<1\} \subset \mathbf{C}$ and $\{z \mid |z| \leqq 1\} \subset \mathbf{C}$.

12. **(Space $B[a, b]$)** Show that $B[a, b]$, $a<b$, is not separable. (Cf. 1.2-2.)

13. Show that a metric space X is separable if and only if X has a countable subset Y with the following property. For every $\varepsilon>0$ and every $x \in X$ there is a $y \in Y$ such that $d(x, y)<\varepsilon$.

14. **(Continuous mapping)** Show that a mapping $T: X \longrightarrow Y$ is continuous if and only if the inverse image of any closed set $M \subset Y$ is a closed set in X.

15. Show that the image of an open set under a continuous mapping need not be open.

1.4 Convergence, Cauchy Sequence, Completeness

We know that sequences of real numbers play an important role in calculus, and it is the metric on $\mathbf{R}$ which enables us to define the basic concept of convergence of such a sequence. The same holds for sequences of complex numbers; in this case we have to use the metric on the complex plane. In an arbitrary metric space $X=(X, d)$ the situation is quite similar, that is, we may consider a sequence (x_n) of elements $x_1, x_2, \cdots$ of X and use the metric d to define convergence in a fashion analogous to that in calculus:

1.4-1 Definition (Convergence of a sequence, limit). A sequence (x_n) in a metric space $X=(X, d)$ is said to *converge* or to *be convergent* if there is an $x \in X$ such that

$$\lim_{n \to \infty} d(x_n, x)=0.$$

x is called the *limit* of (x_n) and we write

$$\lim_{n \to \infty} x_n = x$$

or, simply,

$$x_n \longrightarrow x.$$

We say that (x_n) *converges to* x or *has the limit* x. If (x_n) is not convergent, it is said to be *divergent.* ∎

How is the metric d being used in this definition? We see that d yields the sequence of real numbers $a_n = d(x_n, x)$ whose convergence defines that of (x_n). Hence if $x_n \longrightarrow x$, an $\varepsilon > 0$ being given, there is an $N = N(\varepsilon)$ such that all x_n with $n > N$ lie in the ε-neighborhood $B(x; \varepsilon)$ of x.

To avoid trivial misunderstandings, we note that the limit of a convergent sequence must be a point of the space X in 1.4-1. For instance, let X be the open interval $(0, 1)$ on $\mathbf{R}$ with the usual metric defined by $d(x, y) = |x - y|$. Then the sequence $(\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots)$ is not convergent since 0, the point to which the sequence "wants to converge," is not in X. We shall return to this and similar situations later in the present section.

Let us first show that two familiar properties of a convergent sequence (uniqueness of the limit and boundedness) carry over from calculus to our present much more general setting.

We call a nonempty subset $M \subset X$ a *bounded set* if its *diameter*

$$\delta(M) = \sup_{x,y \in M} d(x, y)$$

is finite. And we call a sequence (x_n) in X a **bounded sequence** if the corresponding point set is a bounded subset of X.

Obviously, if M is bounded, then $M \subset B(x_0; r)$, where $x_0 \in X$ is any point and r is a (sufficiently large) real number, and conversely.

Our assertion is now as follows.

1.4-2 Lemma (Boundedness, limit). *Let $X = (X, d)$ be a metric space. Then:*

(a) *A convergent sequence in X is bounded and its limit is unique.*

(b) *If $x_n \longrightarrow x$ and $y_n \longrightarrow y$ in X, then $d(x_n, y_n) \longrightarrow d(x, y)$.*

Proof. **(a)** Suppose that $x_n \longrightarrow x$. Then, taking $\varepsilon = 1$, we can find an N such that $d(x_n, x) < 1$ for all $n > N$. Hence by the triangle inequality (M4), Sec. 1.1, for all n we have $d(x_n, x) < 1 + a$ where

$$a = \max\{d(x_1, x), \cdots, d(x_N, x)\}.$$

This shows that (x_n) is bounded. Assuming that $x_n \longrightarrow x$ and $x_n \longrightarrow z$, we obtain from (M4)

$$0 \leqq d(x, z) \leqq d(x, x_n) + d(x_n, z) \longrightarrow 0+0$$

and the uniqueness $x = z$ of the limit follows from (M2).

(b) By (1), Sec. 1.1, we have

$$d(x_n, y_n) \leqq d(x_n, x) + d(x, y) + d(y, y_n).$$

Hence we obtain

$$d(x_n, y_n) - d(x, y) \leqq d(x_n, x) + d(y_n, y)$$

and a similar inequality by interchanging x_n and x as well as y_n and y and multiplying by -1. Together,

$$|d(x_n, y_n) - d(x, y)| \leqq d(x_n, x) + d(y_n, y) \longrightarrow 0$$

as $n \longrightarrow \infty$. ∎

We shall now define the concept of completeness of a metric space, which will be basic in our further work. We shall see that completeness does *not* follow from (M1) to (M4) in Sec. 1.1, since there are *incomplete* (not complete) metric spaces. In other words, completeness is an additional property which a metric space may or may not have. It has various consequences which make complete metric spaces "much nicer and simpler" than incomplete ones—what this means will become clearer and clearer as we proceed.

Let us first remember from calculus that a sequence (x_n) of real or complex numbers converges on the real line $\mathbf{R}$ or in the complex plane $\mathbf{C}$, respectively, if and only if it satisfies the *Cauchy convergence criterion*, that is, if and only if for every given $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that

$$|x_m - x_n| < \varepsilon \qquad \text{for all } m, n > N.$$

(A proof is included in A1.7; cf. Appendix 1.) Here $|x_m - x_n|$ is the distance $d(x_m, x_n)$ from x_m to x_n on the real line $\mathbf{R}$ or in the complex plane **C**. Hence we can write the inequality of the Cauchy criterion in the form

$$d(x_m, x_n) < \varepsilon \qquad (m, n > N).$$

And if a sequence (x_n) satisfies the condition of the Cauchy criterion, we may call it a *Cauchy sequence.* Then the Cauchy criterion simply says that a sequence of real or complex numbers converges on **R** or in **C** if and only if it is a Cauchy sequence. This refers to the situation in **R** or **C**. Unfortunately, in more general spaces the situation may be more complicated, and there may be Cauchy sequences which do not converge. Such a space is then lacking a property which is so important that it deserves a name, namely, completeness. This consideration motivates the following definition, which was first given by M. Fréchet (1906).

1.4-3 Definition (Cauchy sequence, completeness). A sequence (x_n) in a metric space $X=(X, d)$ is said to be *Cauchy* (or *fundamental*) if for every $\varepsilon > 0$ there is an $N=N(\varepsilon)$ such that

$$d(x_m, x_n) < \varepsilon \qquad \text{for every } m, n > N. \tag{1}$$

The space X is said to be *complete* if every Cauchy sequence in X converges (that is, has a limit which is an element of X). ∎

Expressed in terms of completeness, the Cauchy convergence criterion implies the following.

1.4-4 Theorem (Real line, complex plane). *The real line and the complex plane are complete metric spaces.*

More generally, we now see directly from the definition that complete metric spaces are precisely those in which the Cauchy condition (1) continues to be necessary and sufficient for convergence.

Complete and incomplete metric spaces that are important in applications will be considered in the next section in a systematic fashion.

For the time being let us mention a few simple incomplete spaces which we can readily obtain. Omission of a point a from the real line yields the incomplete space $\mathbf{R}-\{a\}$. More drastically, by the omission of all irrational numbers we have the *rational line* **Q**, which is incomplete. An open interval (a, b) with the metric induced from **R** is another incomplete metric space, and so on.

It is clear from the definition that in an arbitrary metric space, condition (1) may no longer be sufficient for convergence since the space may be incomplete. A good understanding of the whole situation is important; so let us consider a simple example. We take $X=(0, 1]$, with the usual metric defined by $d(x, y)=|x-y|$, and the sequence (x_n), where $x_n=1/n$ and $n=1, 2, \cdots$. This is a Cauchy sequence, but it does not converge, because the point 0 (to which it "wants to converge") is not a point of X. This also illustrates that the concept of convergence is not an intrinsic property of the sequence itself but also depends on the space in which the sequence lies. In other words, a convergent sequence is not convergent "on its own" but it must converge to some point in the space.

Although condition (1) is no longer sufficient for convergence, it is worth noting that it continues to be necessary for convergence. In fact, we readily obtain the following result.

1.4-5 Theorem (Convergent sequence). *Every convergent sequence in a metric space is a Cauchy sequence.*

Proof. If $x_n \longrightarrow x$, then for every $\varepsilon>0$ there is an $N=N(\varepsilon)$ such that

$$d(x_n, x)<\frac{\varepsilon}{2} \qquad \text{for all } n>N.$$

Hence by the triangle inequality we obtain for $m, n>N$

$$d(x_m, x_n)\leqq d(x_m, x)+d(x, x_n)<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$

This shows that (x_n) is Cauchy. ∎

We shall see that quite a number of basic results, for instance in the theory of linear operators, will depend on the completeness of the corresponding spaces. Completeness of the real line **R** is also the main reason why in calculus we use **R** rather than the *rational line* **Q** (the set of all rational numbers with the metric induced from **R**).

Let us continue and finish this section with three theorems that are related to convergence and completeness and will be needed later.

1.4-6 Theorem (Closure, closed set). *Let M be a nonempty subset of a metric space (X, d) and $\bar{M}$ its closure as defined in the previous section. Then:*

(a) *$x \in \bar{M}$ if and only if there is a sequence (x_n) in M such that $x_n \longrightarrow x$.*

(b) *M is closed if and only if the situation $x_n \in M$, $x_n \longrightarrow x$ implies that $x \in M$.*

Proof. **(a)** Let $x \in \bar{M}$. If $x \in M$, a sequence of that type is $(x, x, \cdots)$. If $x \notin M$, it is a point of accumulation of M. Hence for each $n = 1, 2, \cdots$ the ball $B(x; 1/n)$ contains an $x_n \in M$, and $x_n \longrightarrow x$ because $1/n \longrightarrow 0$ as $n \longrightarrow \infty$.

Conversely, if (x_n) is in M and $x_n \longrightarrow x$, then $x \in M$ or every neighborhood of x contains points $x_n \neq x$, so that x is a point of accumulation of M. Hence $x \in \bar{M}$, by the definition of the closure.

(b) M is closed if and only if $M = \bar{M}$, so that (b) follows readily from (a). ∎

1.4-7 Theorem (Complete subspace). *A subspace M of a complete metric space X is itself complete if and only if the set M is closed in X.*

Proof. Let M be complete. By 1.4-6(a), for every $x \in \bar{M}$ there is a sequence (x_n) in M which converges to x. Since (x_n) is Cauchy by 1.4-5 and M is complete, (x_n) converges in M, the limit being unique by 1.4-2. Hence $x \in M$. This proves that M is closed because $x \in \bar{M}$ was arbitrary.

Conversely, let M be closed and (x_n) Cauchy in M. Then $x_n \longrightarrow x \in X$, which implies $x \in \bar{M}$ by 1.4-6(a), and $x \in M$ since $M = \bar{M}$ by assumption. Hence the arbitrary Cauchy sequence (x_n) converges in M, which proves completeness of M. ∎

This theorem is very useful, and we shall need it quite often. Example 1.5-3 in the next section includes the first application, which is typical.

The last of our present three theorems shows the importance of convergence of sequences in connection with the continuity of a mapping.

1.4-8 Theorem (Continuous mapping). *A mapping $T: X \longrightarrow Y$ of a metric space (X, d) into a metric space $(Y, \tilde{d})$ is continuous at a point*

$x_0 \in X$ *if and only if*

$$x_n \longrightarrow x_0 \qquad \text{implies} \qquad Tx_n \longrightarrow Tx_0.$$

Proof. Assume T to be continuous at x_0; cf. Def. 1.3-3. Then for a given $\varepsilon > 0$ there is a $\delta > 0$ such that

$$d(x, x_0) < \delta \qquad \text{implies} \qquad \tilde{d}(Tx, Tx_0) < \varepsilon.$$

Let $x_n \longrightarrow x_0$. Then there is an N such that for all $n > N$ we have

$$d(x_n, x_0) < \delta.$$

Hence for all $n > N$,

$$\tilde{d}(Tx_n, Tx_0) < \varepsilon.$$

By definition this means that $Tx_n \longrightarrow Tx_0$.

Conversely, we assume that

$$x_n \longrightarrow x_0 \qquad \text{implies} \qquad Tx_n \longrightarrow Tx_0$$

and prove that then T is continuous at x_0. Suppose this is false. Then there is an $\varepsilon > 0$ such that for every $\delta > 0$ there is an $x \neq x_0$ satisfying

$$d(x, x_0) < \delta \qquad \text{but} \qquad \tilde{d}(Tx, Tx_0) \geqq \varepsilon.$$

In particular, for $\delta = 1/n$ there is an x_n satisfying

$$d(x_n, x_0) < \frac{1}{n} \qquad \text{but} \qquad \tilde{d}(Tx_n, Tx_0) \geqq \varepsilon.$$

Clearly $x_n \longrightarrow x_0$ but (Tx_n) does not converge to Tx_0. This contradicts $Tx_n \longrightarrow Tx_0$ and proves the theorem. ∎

Problems

1. **(Subsequence)** If a sequence (x_n) in a metric space X is convergent and has limit x, show that every subsequence (x_{n_k}) of (x_n) is convergent and has the same limit x.

2. If (x_n) is Cauchy and has a convergent subsequence, say, $x_{n_k} \longrightarrow x$, show that (x_n) is convergent with the limit x.

3. Show that $x_n \longrightarrow x$ if and only if for every neighborhood V of x there is an integer n_0 such that $x_n \in V$ for all $n > n_0$.

4. (Boundedness) Show that a Cauchy sequence is bounded.

5. Is boundedness of a sequence in a metric space sufficient for the sequence to be Cauchy? Convergent?

6. If (x_n) and (y_n) are Cauchy sequences in a metric space (X, d), show that (a_n), where $a_n = d(x_n, y_n)$, converges. Give illustrative examples.

7. Give an indirect proof of Lemma 1.4-2(b).

8. If d_1 and d_2 are metrics on the same set X and there are positive numbers a and b such that for all $x, y \in X$,

$$ad_1(x, y) \leqq d_2(x, y) \leqq bd_1(x, y),$$

show that the Cauchy sequences in (X, d_1) and (X, d_2) are the same.

9. Using Prob. 8, show that the metric spaces in Probs. 13 to 15, Sec. 1.2, have the same Cauchy sequences.

10. Using the completeness of $\mathbf{R}$, prove completeness of $\mathbf{C}$.

1.5 Examples. Completeness Proofs

In various applications a set X is given (for instance, a set of sequences or a set of functions), and X is made into a metric space. This we do by choosing a metric d on X. The remaining task is then to find out whether (X, d) has the desirable property of being complete. To prove completeness, we take an arbitrary Cauchy sequence (x_n) in X and show that it converges in X. For different spaces, such proofs may vary in complexity, but they have approximately the same general pattern:

(i) Construct an element x (to be used as a limit).
(ii) Prove that x is in the space considered.
(iii) Prove convergence $x_n \longrightarrow x$ (in the sense of the metric).

We shall present completeness proofs for some metric spaces which occur quite frequently in theoretical and practical investigations.

The reader will notice that in these cases (Examples 1.5-1 to 1.5-5) we get help from the completeness of the real line or the complex plane (Theorem 1.4-4). This is typical.

Examples

1.5-1 Completeness of $\mathbf{R}^n$ and $\mathbf{C}^n$. *Euclidean space $\mathbf{R}^n$ and unitary space $\mathbf{C}^n$ are complete.* (Cf. 1.1–5.)

Proof. We first consider $\mathbf{R}^n$. We remember that the metric on $\mathbf{R}^n$ (the Euclidean metric) is defined by

$$d(x, y) = \left(\sum_{j=1}^{n} (\xi_j - \eta_j)^2\right)^{1/2}$$

where $x = (\xi_j)$ and $y = (\eta_j)$; cf. (6) in Sec. 1.1. We consider any Cauchy sequence (x_m) in $\mathbf{R}^n$, writing $x_m = (\xi_1^{(m)}, \cdots, \xi_n^{(m)})$. Since (x_m) is Cauchy, for every $\varepsilon > 0$ there is an N such that

$$(1) \qquad d(x_m, x_r) = \left(\sum_{j=1}^{n} (\xi_j^{(m)} - \xi_j^{(r)})^2\right)^{1/2} < \varepsilon \qquad (m, r > N).$$

Squaring, we have for $m, r > N$ and $j = 1, \cdots, n$

$$(\xi_j^{(m)} - \xi_j^{(r)})^2 < \varepsilon^2 \qquad \text{and} \qquad |\xi_j^{(m)} - \xi_j^{(r)}| < \varepsilon.$$

This shows that for each fixed j, $(1 \leqq j \leqq n)$, the sequence $(\xi_j^{(1)}, \xi_j^{(2)}, \cdots)$ is a Cauchy sequence of real numbers. It converges by Theorem 1.4-4, say, $\xi_j^{(m)} \longrightarrow \xi_j$ as $m \longrightarrow \infty$. Using these n limits, we define $x = (\xi_1, \cdots, \xi_n)$. Clearly, $x \in \mathbf{R}^n$. From (1), with $r \longrightarrow \infty$,

$$d(x_m, x) \leqq \varepsilon \qquad (m > N).$$

This shows that x is the limit of (x_m) and proves completeness of $\mathbf{R}^n$ because (x_m) was an arbitrary Cauchy sequence. Completeness of $\mathbf{C}^n$ follows from Theorem 1.4-4 by the same method of proof.

1.5-2 Completeness of l^∞. *The space l^∞ is complete.* (Cf. 1.1-6.)

Proof. Let (x_m) be any Cauchy sequence in the space l^∞, where $x_m = (\xi_1^{(m)}, \xi_2^{(m)}, \cdots)$. Since the metric on l^∞ is given by

$$d(x, y) = \sup_j |\xi_j - \eta_j|$$

[where $x = (\xi_j)$ and $y = (\eta_j)$] and (x_m) is Cauchy, for any $\varepsilon > 0$ there is an N such that for all $m, n > N$,

$$d(x_m, x_n) = \sup_j |\xi_j^{(m)} - \xi_j^{(n)}| < \varepsilon.$$

A fortiori, for every fixed j,

$$|\xi_j^{(m)} - \xi_j^{(n)}| < \varepsilon \qquad (m, n > N). \tag{2}$$

Hence for every fixed j, the sequence $(\xi_j^{(1)}, \xi_j^{(2)}, \cdots)$ is a Cauchy sequence of numbers. It converges by Theorem 1.4-4, say, $\xi_j^{(m)} \longrightarrow \xi_j$ as $m \longrightarrow \infty$. Using these infinitely many limits $\xi_1, \xi_2, \cdots$, we define $x = (\xi_1, \xi_2, \cdots)$ and show that $x \in l^\infty$ and $x_m \longrightarrow x$. From (2) with $n \longrightarrow \infty$ we have

$$|\xi_j^{(m)} - \xi_j| \leqq \varepsilon \qquad (m > N). \tag{2*}$$

Since $x_m = (\xi_j^{(m)}) \in l^\infty$, there is a real number k_m such that $|\xi_j^{(m)}| \leqq k_m$ for all j. Hence by the triangle inequality

$$|\xi_j| \leqq |\xi_j - \xi_j^{(m)}| + |\xi_j^{(m)}| \leqq \varepsilon + k_m \qquad (m > N).$$

This inequality holds for every j, and the right-hand side does not involve j. Hence (ξ_j) is a bounded sequence of numbers. This implies that $x = (\xi_j) \in l^\infty$. Also, from (2*) we obtain

$$d(x_m, x) = \sup_j |\xi_j^{(m)} - \xi_j| \leqq \varepsilon \qquad (m > N).$$

This shows that $x_m \longrightarrow x$. Since (x_m) was an arbitrary Cauchy sequence, l^∞ is complete.

1.5-3 Completeness of c. The space c consists of all convergent sequences $x = (\xi_j)$ of complex numbers, with the metric induced from the space l^∞.

The space c is complete.

Proof. c is a subspace of l^∞ and we show that c is closed in l^∞, so that completeness then follows from Theorem 1.4-7.

We consider any $x = (\xi_j) \in \bar{c}$, the closure of c. By 1.4-6(*a*) there are $x_n = (\xi_j^{(n)}) \in c$ such that $x_n \longrightarrow x$. Hence, given any $\varepsilon > 0$, there is an N such that for $n \geqq N$ and all j we have

$$|\xi_j^{(n)} - \xi_j| \leqq d(x_n, x) < \frac{\varepsilon}{3},$$

in particular, for $n = N$ and all j. Since $x_N \in c$, its terms $\xi_j^{(N)}$ form a convergent sequence. Such a sequence is Cauchy. Hence there is an N_1 such that

$$|\xi_j^{(N)} - \xi_k^{(N)}| < \frac{\varepsilon}{3} \qquad (j, k \geqq N_1).$$

The triangle inequality now yields for all j, $k \geqq N_1$ the following inequality:

$$|\xi_j - \xi_k| \leqq |\xi_j - \xi_j^{(N)}| + |\xi_j^{(N)} - \xi_k^{(N)}| + |\xi_k^{(N)} - \xi_k| < \varepsilon.$$

This shows that the sequence $x = (\xi_j)$ is convergent. Hence $x \in c$. Since $x \in \bar{c}$ was arbitrary, this proves closedness of c in l^∞, and completeness of c follows from 1.4-7. ∎

1.5-4 Completeness of l^p. *The space l^p is complete; here p is fixed and* $1 \leqq p < +\infty$. (Cf. 1.2-3.)

Proof. Let (x_n) be any Cauchy sequence in the space l^p, where $x_m = (\xi_1^{(m)}, \xi_2^{(m)}, \cdots)$. Then for every $\varepsilon > 0$ there is an N such that for all $m, n > N$,

$$d(x_m, x_n) = \left(\sum_{j=1}^{\infty} |\xi_j^{(m)} - \xi_j^{(n)}|^p \right)^{1/p} < \varepsilon. \tag{3}$$

It follows that for every $j = 1, 2, \cdots$ we have

$$|\xi_j^{(m)} - \xi_j^{(n)}| < \varepsilon \qquad (m, n > N). \tag{4}$$

We choose a fixed j. From (4) we see that $(\xi_j^{(1)}, \xi_j^{(2)}, \cdots)$ is a Cauchy sequence of numbers. It converges since $\mathbf{R}$ and $\mathbf{C}$ are complete (cf.

1.4-4), say, $\xi_j^{(m)} \longrightarrow \xi_j$ as $m \longrightarrow \infty$. Using these limits, we define $x = (\xi_1, \xi_2, \cdots)$ and show that $x \in l^p$ and $x_m \longrightarrow x$.

From (3) we have for all $m, n > N$

$$\sum_{j=1}^{k} |\xi_j^{(m)} - \xi_j^{(n)}|^p < \varepsilon^p \qquad (k = 1, 2, \cdots).$$

Letting $n \longrightarrow \infty$, we obtain for $m > N$

$$\sum_{j=1}^{k} |\xi_j^{(m)} - \xi_j|^p \leqq \varepsilon^p \qquad (k = 1, 2, \cdots).$$

We may now let $k \longrightarrow \infty$; then for $m > N$

$$\sum_{j=1}^{\infty} |\xi_j^{(m)} - \xi_j|^p \leqq \varepsilon^p. \tag{5}$$

This shows that $x_m - x = (\xi_j^{(m)} - \xi_j) \in l^p$. Since $x_m \in l^p$, it follows by means of the Minkowski inequality (12), Sec. 1.2, that

$$x = x_m + (x - x_m) \in l^p.$$

Furthermore, the series in (5) represents $[d(x_m, x)]^p$, so that (5) implies that $x_m \longrightarrow x$. Since (x_m) was an arbitrary Cauchy sequence in l^p, this proves completeness of l^p, where $1 \leqq p < +\infty$. ∎

1.5-5 Completeness of $C[a, b]$. *The function space $C[a, b]$ is complete; here $[a, b]$ is any given closed interval on* **R**. (Cf. 1.1-7.)

Proof. Let (x_m) be any Cauchy sequence in $C[a, b]$. Then, given any $\varepsilon > 0$, there is an N such that for all $m, n > N$ we have

$$d(x_m, x_n) = \max_{t \in J} |x_m(t) - x_n(t)| < \varepsilon \tag{6}$$

where $J = [a, b]$. Hence for any fixed $t = t_0 \in J$,

$$|x_m(t_0) - x_n(t_0)| < \varepsilon \qquad (m, n > N).$$

This shows that $(x_1(t_0), x_2(t_0), \cdots)$ is a Cauchy sequence of real numbers. Since **R** is complete (cf. 1.4-4), the sequence converges, say,

$x_m(t_0) \longrightarrow x(t_0)$ as $m \longrightarrow \infty$. In this way we can associate with each $t \in J$ a unique real number $x(t)$. This defines (pointwise) a function x on J, and we show that $x \in C[a, b]$ and $x_m \longrightarrow x$.

From (6) with $n \longrightarrow \infty$ we have

$$\max_{t \in J} |x_m(t) - x(t)| \leqq \varepsilon \qquad (m > N).$$

Hence for every $t \in J$,

$$|x_m(t) - x(t)| \leqq \varepsilon \qquad (m > N).$$

This shows that $(x_m(t))$ converges to $x(t)$ uniformly on J. Since the x_m's are continuous on J and the convergence is uniform, the limit function x is continuous on J, as is well known from calculus (cf. also Prob. 9). Hence $x \in C[a, b]$. Also $x_m \longrightarrow x$. This proves completeness of $C[a, b]$. ∎

In 1.1-7 as well as here we assumed the functions x to be real-valued, for simplicity. We may call this space the *real* $C[a, b]$. Similarly, we obtain the *complex* $C[a, b]$ if we take complex-valued continuous functions defined on $[a, b] \subset \mathbf{R}$. This space is complete, too. The proof is almost the same as before.

Furthermore, that proof also shows the following fact.

1.5-6 Theorem (Uniform convergence). *Convergence $x_m \longrightarrow x$ in the space $C[a, b]$ is uniform convergence, that is, (x_m) converges uniformly on $[a, b]$ to x.*

Hence the metric on $C[a, b]$ describes uniform convergence on $[a, b]$ and, for this reason, is sometimes called the *uniform metric.*

To gain a good understanding of completeness and related concepts, let us finally look at some

Examples of Incomplete Metric Spaces

1.5-7 Space Q. This is the set of all rational numbers with the usual metric given by $d(x, y) = |x - y|$, where $x, y \in \mathbf{Q}$, and is called the *rational line.* $\mathbf{Q}$ is not complete. (Proof?)

1.5-8 Polynomials. Let X be the set of all polynomials considered as functions of t on some finite closed interval $J = [a, b]$ and define a metric d on X by

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

This metric space (X, d) is not complete. In fact, an example of a Cauchy sequence without limit in X is given by any sequence of polynomials which converges uniformly on J to a continuous function, not a polynomial.

1.5-9 Continuous functions. Let X be the set of all continuous real-valued functions on $J = [0, 1]$, and let

$$d(x, y) = \int_0^1 |x(t) - y(t)| \, dt.$$

This metric space (X, d) is not complete.

Proof. The functions x_m in Fig. 9 form a Cauchy sequence because $d(x_m, x_n)$ is the area of the triangle in Fig. 10, and for every given $\varepsilon > 0$,

$$d(x_m; x_n) < \varepsilon \qquad \text{when} \qquad m, n > 1/\varepsilon.$$

Let us show that this Cauchy sequence does not converge. We have

$$x_m(t) = 0 \text{ if } t \in [0, \tfrac{1}{2}], \qquad x_m(t) = 1 \text{ if } t \in [a_m, 1]$$

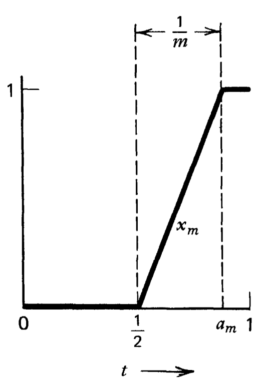

Fig. 9. Example 1.5-9

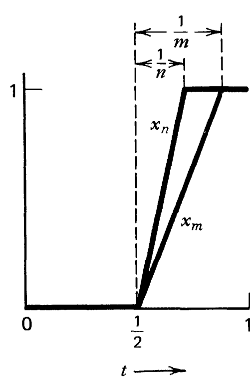

Fig. 10. Example 1.5-9

where $a_m = 1/2 + 1/m$. Hence for every $x \in X$,

$$d(x_m, x) = \int_0^1 |x_m(t) - x(t)| \, dt$$

$$= \int_0^{1/2} |x(t)| \, dt + \int_{1/2}^{a_m} |x_m(t) - x(t)| \, dt + \int_{a_m}^1 |1 - x(t)| \, dt.$$

Since the integrands are nonnegative, so is each integral on the right. Hence $d(x_m, x) \longrightarrow 0$ would imply that each integral approaches zero and, since x is continuous, we should have

$$x(t) = 0 \text{ if } t \in [0, \tfrac{1}{2}), \qquad\qquad x(t) = 1 \text{ if } t \in (\tfrac{1}{2}, 1].$$

But this is impossible for a continuous function. Hence (x_m) does not converge, that is, does not have a limit in X. This proves that X is not complete. ∎

Problems

1. Let $a, b \in \mathbf{R}$ and $a < b$. Show that the open interval (a, b) is an incomplete subspace of $\mathbf{R}$, whereas the closed interval $[a, b]$ is complete.

2. Let X be the space of all ordered n-tuples $x = (\xi_1, \cdots, \xi_n)$ of real numbers and

$$d(x, y) = \max_j |\xi_j - \eta_j|$$

where $y = (\eta_j)$. Show that (X, d) is complete.

3. Let $M \subset l^\infty$ be the subspace consisting of all sequences $x = (\xi_j)$ with at most finitely many nonzero terms. Find a Cauchy sequence in M which does not converge in M, so that M is not complete.

4. Show that M in Prob. 3 is not complete by applying Theorem 1.4-7.

5. Show that the set X of all integers with metric d defined by $d(m, n) = |m - n|$ is a complete metric space.

6. Show that the set of all real numbers constitutes an incomplete metric space if we choose

$$d(x, y) = |\arctan x - \arctan y|.$$

7. Let X be the set of all positive integers and $d(m, n) = |m^{-1} - n^{-1}|$. Show that (X, d) is not complete.

8. (Space $C[a, b]$) Show that the subspace $Y \subset C[a, b]$ consisting of all $x \in C[a, b]$ such that $x(a) = x(b)$ is complete.

9. In 1.5-5 we referred to the following theorem of calculus. If a sequence (x_m) of continuous functions on $[a, b]$ converges on $[a, b]$ and the convergence is uniform on $[a, b]$, then the limit function x is continuous on $[a, b]$. Prove this theorem.

10. (Discrete metric) Show that a discrete metric space (cf. 1.1-8) is complete.

11. (Space s) Show that in the space s (cf. 1.2-1) we have $x_n \longrightarrow x$ if and only if $\xi_j^{(n)} \longrightarrow \xi_j$ for all $j = 1, 2, \cdots$, where $x_n = (\xi_j^{(n)})$ and $x = (\xi_j)$.

12. Using Prob. 11, show that the sequence space s in 1.2-1 is complete.

13. Show that in 1.5-9, another Cauchy sequence is (x_n), where

$$x_n(t) = n \qquad \text{if } 0 \leqq t \leqq n^{-2}, \qquad\qquad x_n(t) = t^{-\frac{1}{2}} \qquad \text{if } n^{-2} \leqq t \leqq 1.$$

14. Show that the Cauchy sequence in Prob. 13 does not converge.

15. Let X be the metric space of all real sequences $x = (\xi_j)$ each of which has only finitely many nonzero terms, and $d(x, y) = \sum |\xi_j - \eta_j|$, where $y = (\eta_j)$. Note that this is a finite sum but the number of terms depends on x and y. Show that (x_n) with $x_n = (\xi_j^{(n)})$,

$$\xi_j^{(n)} = j^{-2} \qquad \text{for } j = 1, \cdots, n \qquad \text{and} \qquad \xi_j^{(n)} = 0 \qquad \text{for } j > n$$

is Cauchy but does not converge.

1.6 Completion of Metric Spaces

We know that the rational line $\mathbf{Q}$ is not complete (cf. 1.5-7) but can be "enlarged" to the real line $\mathbf{R}$ which is complete. And this "completion" $\mathbf{R}$ of $\mathbf{Q}$ is such that $\mathbf{Q}$ is dense (cf. 1.3-5) in $\mathbf{R}$. It is quite important that an arbitrary incomplete metric space can be "completed" in a similar fashion, as we shall see. For a convenient precise formulation we use the following two related concepts, which also have various other applications.

1.6-1 Definition (Isometric mapping, isometric spaces). Let $X=(X, d)$ and $\tilde{X}=(\tilde{X}, \tilde{d})$ be metric spaces. Then:

(a) A mapping T of X into $\tilde{X}$ is said to be *isometric* or an *isometry* if T preserves distances, that is, if for all $x, y \in X$,

$$\tilde{d}(Tx, Ty)=d(x, y),$$

where Tx and Ty are the images of x and y, respectively.

(b) The space X is said to be *isometric* with the space $\tilde{X}$ if there exists a bijective[9] isometry of X onto $\tilde{X}$. The spaces X and $\tilde{X}$ are then called *isometric spaces.* ∎

Hence isometric spaces may differ at most by the nature of their points but are indistinguishable from the viewpoint of metric. And in any study in which the nature of the points does not matter, we may regard the two spaces as identical—as two copies of the same "abstract" space.

We can now state and prove the theorem that every metric space can be completed. The space $\hat{X}$ occurring in this theorem is called the **completion** of the given space X.

1.6-2 Theorem (Completion). *For a metric space $X=(X, d)$ there exists a complete metric space $\hat{X}=(\hat{X}, \hat{d})$ which has a subspace W that is isometric with X and is dense in $\hat{X}$. This space $\hat{X}$ is unique except for isometries, that is, if $\tilde{X}$ is any complete metric space having a dense subspace $\tilde{W}$ isometric with X, then $\tilde{X}$ and $\hat{X}$ are isometric.*

[9] One-to-one and onto. For a review of some elementary concepts related to mappings, see A1.2 in Appendix 1. Note that an isometric mapping is always injective. (Why?)

Proof. The proof is somewhat lengthy but straightforward. We subdivide it into four steps (*a*) to (*d*). We construct:

(*a*) $\hat{X} = (\hat{X}, \hat{d})$

(*b*) an isometry T of X onto W, where $\bar{W} = \hat{X}$.

Then we prove:

(*c*) completeness of $\hat{X}$,

(*d*) uniqueness of $\hat{X}$, except for isometries.

Roughly speaking, our task will be the assignment of suitable limits to Cauchy sequences in X that do not converge. However, we should not introduce "too many" limits, but take into account that certain sequences "may want to converge with the same limit" since the terms of those sequences "ultimately come arbitrarily close to each other." This intuitive idea can be expressed mathematically in terms of a suitable equivalence relation [see (1), below]. This is not artificial but is suggested by the process of completion of the rational line mentioned at the beginning of the section. The details of the proof are as follows.

(a) *Construction of* $\hat{X} = (\hat{X}, \hat{d})$. Let (x_n) and (x_n') be Cauchy sequences in X. Define (x_n) to be *equivalent*[10] to (x_n'), written $(x_n) \sim (x_n')$, if

$$\lim_{n \to \infty} d(x_n, x_n') = 0. \tag{1}$$

Let $\hat{X}$ be the set of all equivalence classes $\hat{x}, \hat{y}, \cdots$ of Cauchy sequences thus obtained. We write $(x_n) \in \hat{x}$ to mean that (x_n) is a member of $\hat{x}$ (a *representative* of the class $\hat{x}$). We now set

$$\hat{d}(\hat{x}, \hat{y}) = \lim_{n \to \infty} d(x_n, y_n) \tag{2}$$

where $(x_n) \in \hat{x}$ and $(y_n) \in \hat{y}$. We show that this limit exists. We have

$$d(x_n, y_n) \leqq d(x_n, x_m) + d(x_m, y_m) + d(y_m, y_n);$$

hence we obtain

$$d(x_n, y_n) - d(x_m, y_m) \leqq d(x_n, x_m) + d(y_m, y_n)$$

and a similar inequality with m and n interchanged. Together,

$$|d(x_n, y_n) - d(x_m, y_m)| \leqq d(x_n, x_m) + d(y_m, y_n). \tag{3}$$

[10] For a review of the concept of equivalence, see A1.4 in Appendix 1.

Since (x_n) and (y_n) are Cauchy, we can make the right side as small as we please. This implies that the limit in (2) exists because $\mathbf{R}$ is complete.

We must also show that the limit in (2) is independent of the particular choice of representatives. In fact, if $(x_n) \sim (x_n')$ and $(y_n) \sim (y_n')$, then by (1),

$$|d(x_n, y_n) - d(x_n', y_n')| \leqq d(x_n, x_n') + d(y_n, y_n') \longrightarrow 0$$

as $n \longrightarrow \infty$, which implies the assertion

$$\lim_{n \to \infty} d(x_n, y_n) = \lim_{n \to \infty} d(x_n', y_n').$$

We prove that $\hat{d}$ in (2) is a metric on $\hat{X}$. Obviously, $\hat{d}$ satisfies (M1) in Sec. 1.1 as well as $\hat{d}(\hat{x}, \hat{x}) = 0$ and (M3). Furthermore,

$$\hat{d}(\hat{x}, \hat{y}) = 0 \qquad \Longrightarrow \qquad (x_n) \sim (y_n) \qquad \Longrightarrow \qquad \hat{x} = \hat{y}$$

gives (M2), and (M4) for $\hat{d}$ follows from

$$d(x_n, y_n) \leqq d(x_n, z_n) + d(z_n, y_n)$$

by letting $n \longmapsto \infty$.

(b) *Construction of an isometry* $T\colon X \longrightarrow W \subset \hat{X}$. With each $b \in X$ we associate the class $\hat{b} \in \hat{X}$ which contains the constant Cauchy sequence $(b, b, \cdots)$. This defines a mapping $T\colon X \longrightarrow W$ onto the subspace $W = T(X) \subset \hat{X}$. The mapping T is given by $b \longmapsto \hat{b} = Tb$, where $(b, b, \cdots) \in \hat{b}$. We see that T is an isometry since (2) becomes simply

$$\hat{d}(\hat{b}, \hat{c}) = d(b, c);$$

here $\hat{c}$ is the class of (y_n) where $y_n = c$ for all n. Any isometry is injective, and $T\colon X \longrightarrow W$ is surjective since $T(X) = W$. Hence W and X are isometric; cf. Def. 1.6-1(b).

We show that W is dense in $\hat{X}$. We consider any $\hat{x} \in \hat{X}$. Let $(x_n) \in \hat{x}$. For every $\varepsilon > 0$ there is an N such that

$$d(x_n, x_N) < \frac{\varepsilon}{2} \qquad (n > N).$$

Let $(x_N, x_N, \cdots) \in \hat{x}_N$. Then $\hat{x}_N \in W$. By (2),

$$\hat{d}(\hat{x}, \hat{x}_N) = \lim_{n\to\infty} d(x_n, x_N) \leqq \frac{\varepsilon}{2} < \varepsilon.$$

This shows that every ε-neighborhood of the arbitrary $\hat{x} \in \hat{X}$ contains an element of W. Hence W is dense in $\hat{X}$.

(c) *Completeness of* $\hat{X}$. Let $(\hat{x}_n)$ be any Cauchy sequence in $\hat{X}$. Since W is dense in $\hat{X}$, for every $\hat{x}_n$ there is a $\hat{z}_n \in W$ such that

$$\hat{d}(\hat{x}_n, \hat{z}_n) < \frac{1}{n}. \tag{4}$$

Hence by the triangle inequality,

$$\begin{aligned}\hat{d}(\hat{z}_m, \hat{z}_n) &\leqq \hat{d}(\hat{z}_m, \hat{x}_m) + \hat{d}(\hat{x}_m, \hat{x}_n) + \hat{d}(\hat{x}_n, \hat{z}_n)\\ &< \frac{1}{m} + \hat{d}(\hat{x}_m, \hat{x}_n) + \frac{1}{n}\end{aligned}$$

and this is less than any given $\varepsilon > 0$ for sufficiently large m and n because $(\hat{x}_m)$ is Cauchy. Hence $(\hat{z}_m)$ is Cauchy. Since $T\colon X \longrightarrow W$ is isometric and $\hat{z}_m \in W$, the sequence (z_m), where $z_m = T^{-1}\hat{z}_m$, is Cauchy in X. Let $\hat{x} \in \hat{X}$ be the class to which (z_m) belongs. We show that $\hat{x}$ is the limit of $(\hat{x}_n)$. By (4),

$$\begin{aligned}\hat{d}(\hat{x}_n, \hat{x}) &\leqq \hat{d}(\hat{x}_n, \hat{z}_n) + \hat{d}(\hat{z}_n, \hat{x})\\ &< \frac{1}{n} + \hat{d}(\hat{z}_n, \hat{x}).\end{aligned} \tag{5}$$

Since $(z_m) \in \hat{x}$ (see right before) and $\hat{z}_n \in W$, so that $(z_n, z_n, z_n, \cdots) \in \hat{z}_n$, the inequality (5) becomes

$$\hat{d}(\hat{x}_n, \hat{x}) < \frac{1}{n} + \lim_{m\to\infty} d(z_n, z_m)$$

and the right side is smaller than any given $\varepsilon > 0$ for sufficiently large n. Hence the arbitrary Cauchy sequence $(\hat{x}_n)$ in $\hat{X}$ has the limit $\hat{x} \in \hat{X}$, and $\hat{X}$ is complete.

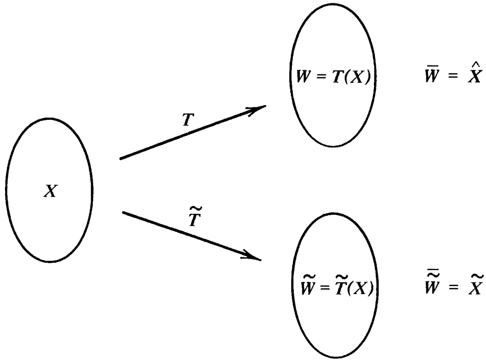

Fig. 11. Notations in part (d) of the proof of Theorem 1.6-2

(d) *Uniqueness of $\hat{X}$ except for isometries.* If $(\tilde{X}, \tilde{d})$ is another complete metric space with a subspace $\tilde{W}$ dense in $\tilde{X}$ and isometric with X, then for any $\tilde{x}, \tilde{y} \in \tilde{X}$ we have sequences $(\tilde{x}_n)$, $(\tilde{y}_n)$ in $\tilde{W}$ such that $\tilde{x}_n \longrightarrow \tilde{x}$ and $\tilde{y}_n \longrightarrow \tilde{y}$; hence

$$\tilde{d}(\tilde{x}, \tilde{y}) = \lim_{n \to \infty} \tilde{d}(\tilde{x}_n, \tilde{y}_n)$$

follows from

$$|\tilde{d}(\tilde{x}, \tilde{y}) - \tilde{d}(\tilde{x}_n, \tilde{y}_n)| \leqq \tilde{d}(\tilde{x}, \tilde{x}_n) + \tilde{d}(\tilde{y}, \tilde{y}_n) \qquad \longrightarrow \qquad 0$$

[the inequality being similar to (3)]. Since $\tilde{W}$ is isometric with $W \subset \hat{X}$ and $\overline{W} = \hat{X}$, the distances on $\tilde{X}$ and $\hat{X}$ must be the same. Hence $\tilde{X}$ and $\hat{X}$ are isometric. ∎

We shall see in the next two chapters (in particular in 2.3-2, 3.1-5 and 3.2-3) that this theorem has basic applications to individual incomplete spaces as well as to whole classes of such spaces.

Problems

1. Show that if a subspace Y of a metric space consists of finitely many points, then Y is complete.

2. What is the completion of (X, d), where X is the set of all rational numbers and $d(x, y) = |x - y|$?

3. What is the completion of a discrete metric space X? (Cf. 1.1-8.)

4. If X_1 and X_2 are isometric and X_1 is complete, show that X_2 is complete.

5. (Homeomorphism) A *homeomorphism* is a continuous bijective mapping $T: X \longrightarrow Y$ whose inverse is continuous; the metric spaces X and Y are then said to be *homeomorphic*. (a) Show that if X and Y are isometric, they are homeomorphic. (b) Illustrate with an example that a complete and an incomplete metric space may be homeomorphic.

6. Show that $C[0, 1]$ and $C[a, b]$ are isometric.

7. If (X, d) is complete, show that $(X, \tilde{d})$, where $\tilde{d} = d/(1+d)$, is complete.

8. Show that in Prob. 7, completeness of $(X, \tilde{d})$ implies completeness of (X, d).

9. If (x_n) and (x_n') in (X, d) are such that (1) holds and $x_n \longrightarrow l$, show that (x_n') converges and has the limit l.

10. If (x_n) and (x_n') are convergent sequences in a metric space (X, d) and have the same limit l, show that they satisfy (1).

11. Show that (1) defines an equivalence relation on the set of all Cauchy sequences of elements of X.

12. If (x_n) is Cauchy in (X, d) and (x_n') in X satisfies (1), show that (x_n') is Cauchy in X.

13. (Pseudometric) A *finite pseudometric* on a set X is a function $d: X \times X \longrightarrow \mathbf{R}$ satisfying (M1), (M3), (M4), Sec. 1.1, and

$$\text{(M2*)} \qquad d(x, x) = 0.$$

What is the difference between a metric and a pseudometric? Show that $d(x, y) = |\xi_1 - \eta_1|$ defines a pseudometric on the set of all ordered pairs of real numbers, where $x = (\xi_1, \xi_2)$, $y = (\eta_1, \eta_2)$. (We mention that some authors use the term *semimetric* instead of *pseudometric*.)

14. Does

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

define a metric or pseudometric on X if X is (*i*) the set of all real-valued continuous functions on $[a, b]$, (*ii*) the set of all real-valued Riemann integrable functions on $[a, b]$?

15. If (X, d) is a pseudometric space, we call a set

$$B(x_0; r) = \{x \in X \mid d(x, x_0) < r\} \qquad (r > 0)$$

an *open ball* in X with *center* x_0 and *radius* r. (Note that this is analogous to 1.3-1.) What are open balls of radius 1 in Prob. 13?