CMMS05 BASIC ANALYSIS

Yu. Safarov

1. Metrics, norms and inner products

Definition 1.1. Let X be a non-empty set. A function $\rho : X \times X \to \mathbb{R}$ is called a *metric* on X if it satisfies

(1) $\rho(x, y) > 0$ if $x \neq y$ and $\rho(x, x) = 0$,
(2) $\rho(x, y) = \rho(y, x)$,
(3) $\rho(x, z) \leqslant \rho(x, y) + \rho(y, z)$ (this is called the *triangle inequality*),

where x, y and z are arbitrary elements of X.

The pair (X, ρ) is said to be a metric space. If $X_0 \subset X$ then ρ is also a metric on X_0. The metric space (X_0, ρ) is called a subspace of (X, ρ). The function $\rho(x, y)$ can (and should) be interpreted as the distance between the elements x and y in the space (X, ρ).

If the X is a linear space, it is often possible to express the metric ρ in terms of a function of one variable that can be thought of as the length of each element (i.e., its distance from 0).

Definition 1.2. Let X be a complex or real linear space, that is, a vector space over $\mathbb{C}$ or $\mathbb{R}$. A function $\|\cdot\| : X \to \mathbb{R}$ is called a *norm* on X if it satisfies

(1) $\|x\| > 0$ if $x \neq 0$ and $\|0\| = 0$,
(2) $\|\lambda x\| = |\lambda|\,\|x\|$ for all $x \in X$ and $\lambda \in \mathbb{C}$ (or $\lambda \in \mathbb{R}$),
(3) $\|x+y\| \leqslant \|x\| + \|y\|$ for all $x, y \in X$ (this is a particular case of the triangle inequality).

A linear space provided with a norm is called a *normed* space.

Given a norm on X, the function $\rho(x, y) = \|x - y\|$ is a metric on X (one can easily check that each property of the norm $\|\cdot\|$ implies the corresponding property of the metric $\rho(x, y) = \|x - y\|$). However, not every metric arises in this way (see Example 1.6 below).

Definition 1.3. Let X be a complex or real linear space. A function $(\cdot, \cdot) : X \times X \to \mathbb{C}$ or, respectively, $(\cdot, \cdot) : X \times X \to \mathbb{C}$ is called an *inner product* on X if it satisfies

(1) $(x, x) \geqslant 0$ for all $x \in X$ and $(x, x) = 0$ if and only if $x = 0$;
(2) $(x, y) = \overline{(y, x)}$ for all $x, y \in X$, where the bar denotes the complex conjugation, in the real case we have $(x, y) = (y, x)$;
(3) $(\alpha x_1 + \beta x_2, y) = \alpha(x_1, y) + \beta(x_2, y)$ for all $x_1, x_2, y \in X$ and all $\alpha, \beta \in \mathbb{C}$ or, if X is a real space, for all $\alpha, \beta \in \mathbb{R}$.

The inner product generates the norm $\|\cdot\| := (\cdot, \cdot)$. Therefore an inner product space is a normed space (and is therefore a metric space).

Example 1.4. $\mathbb{R}^n$ and $\mathbb{C}^n$ are inner product spaces. The standard (Euclidean) inner product and metric are defined by the equalities $(x,y) = \sum_{i=1}^n x_i\,\bar{y}_i$ and $\rho(x,y) = \left(\sum_{i=1}^n |x_i - y_i|^2\right)^{1/2}$, where x_i and y_i are coordinates of x and y.

Example 1.5. $\mathbb{R}^n$ and $\mathbb{C}^n$ can also be provided with the norm $\|x\| = \sum_{i=1}^n |x_i|$. This norm generates the metric $\rho(x,y) = \sum_{i=1}^n |x_i - y_i|$ but is not associated with an inner product.

Example 1.6 (discrete metric). For any set X, define the metric ρ by

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

The discrete metric is not generated by a norm even if X is a linear space. Indeed, the function $d(x,0)$ on X does not satisfy conditions of Definition 1.2.

Example 1.7. Let X be the linear space of all infinite complex sequences $x = (x_1, x_2, \dots)$ such that $\sum_{i=1}^\infty |x_i|^p < \infty$ where q is a positive number from the interval $[1,\infty)$. Then $\|x\|_p := (\sum_{i=1}^\infty |x_i|^p)^{1/p}$ is a norm on X (in this particular case the triangle inequality is called the Minkowski inequality). The space X provided with the norm $\|\cdot\|_p$ is usually denoted l^p.

Example 1.8. Let l^∞ be the linear space of all infinite complex sequences $x = (x_1, x_2, \dots)$ such that $\sup_i |x_i| < \infty$. Then $\|x\|_\infty = \sup_i |x_i|$ is a norm on l_∞.

Example 1.9. The norm $\|\cdot|_2$ on l^2 is generated by the inner product $(x,y) = \sum_{i=1}^\infty x_i\,\bar{y}_i$. However, if $p \neq 2$ then $\|\cdot\|_p$ cannot be associated with an inner product.

Example 1.10. $C[a,b]$ usually denotes the linear space of all continuous (real or complex-valued) functions on the interval $[a,b]$. The standard norm on $C[a,b]$ is defined by the equality $\|f\| = \sup_{x\in[a,b]} |f(x)|$. It is not generated by an inner product. The corresponding metric is $\rho(f,g) = \sup_{x\in[a,b]} |f(x) - g(x)|$.

Example 1.11. The linear space $B(S)$ of all bounded (real or complex-valued) functions on a nonempty set S can be provided with the norm $\|f\| = \sup_{x\in S} |f(x)|$.

If $S = [a,b]$ then $C[a,b] \subset B[a,b]$ and the metric introduced in Example 1.11 is the same as in Example 1.10. However, the space $C[a,b]$ is strictly smaller than $B[a,b]$ (every continuous function on $[a,b]$ is bounded but there are bounded functions which are not continuous). Note that Example 1.8 is a particular case of Example 1.11 where the set S is countable.

Example 1.12. Let $L_2[a,b]$ be the linear space of integrable functions f on the interval $[a,b]$ such that $\int_a^b |f(t)|^2\,\mathrm{d}t < \infty$. The standard inner product on $L_2[a,b]$ is $(f,g) = \int_a^b f(t)\,\bar{g}(t)\,\mathrm{d}t$, and the corresponding norm is $\|f\| = \left(\int_a^b |f(t)|^2\,\mathrm{d}t\right)^{1/2}$.

Note that the notion "an integrable function" depends on our definition of integration. In order to define the space L_2 rigorously, one has to explain for which functions the integral is well defined.

2. Convergence in metric spaces

Definition 2.1. A sequence x_n of elements of a metric space (X,ρ) is said to *converge* to $x \in X$ if for any $\varepsilon > 0$ there exists an integer n_ε such that $\rho(x_n, x) < \varepsilon$ for all $n > n_\varepsilon$.

Lemma 2.2 (another definition of convergence). *$x_n \to x$ in (X,ρ) if and only if $\rho(x_n,x) \to 0$ in $\mathbb{R}$.*

Proof. By definition, the sequence of non-negative numbers $\rho(x_n,x)$ converges to zero if and only if for any $\varepsilon > 0$ there exists an integer n_ε such that $\rho(x_n,x) \leqslant \varepsilon$ for all $n > n_\varepsilon$. $\square$

Definition 2.3. We say that the metrics ρ_1 and ρ_2 defined on the same set X are equivalent if $x_n \to x$ in (X,ρ_1) if and only if $x_n \to x$ in (X,ρ_2). Two norms on the same linear space X are said to be equivalent when the corresponding metrics are equivalent.

One can easily check the metric introduced in Example 1.5 is equivalent to the Euclidean metric (Example 1.4), whereas the discrete metric is not.

Definition 2.4. Convergence in the metric space $B(S)$ (or any subspace of $B(S)$) is called *uniform convergence* on S. Convergence with respect to the norm introduced in Example 1.12 is called *mean square convergence*.

Remark 2.5. The standard norm on $C[a,b]$ (see Example 1.10) is not equivalent to the norm introduced in Example 1.12.

Definition 2.6. A sequence of functions $f_n \in B(S)$ converges to $f \in B(S)$ *pointwise* if for any $x \in S$ and $\varepsilon > 0$ there exists an integer $n_{\varepsilon,x}$ such that $|f(x) - f_n(x)| \leqslant \varepsilon$ for all $n > n_{\varepsilon,x}$.

In Definition 2.6 the integer $n_{\varepsilon,x}$ may depend on x. If for any ε the set $\{n_{\varepsilon,x}\}_{x\in S}$ is bounded from above, that is $n_{\varepsilon,x} \leqslant n_\varepsilon$ for all $x \in S$, then $f_n \to f$ uniformly. Note that pointwise convergence is not defined in terms of a metric. We shall see later that this convergence cannot be associated with any metric unless the set S is countable.

3. Open Sets and Closed Sets

Let (X,ρ) be a metric space and r be a strictly positive number.

Definition 3.1. If $\alpha \in X$ then the set $B_r(\alpha) = \{x \in X : \rho(x,\alpha) < r\}$ is called the *open ball*, and the set $B_r[\alpha] = \{x \in X : \rho(x,\alpha) \leqslant r\}$ is called the *closed ball* centre α radius r.

If there is a need to emphasize the metric, we write $B^\rho_r(\alpha)$ and $B^\rho_r[\alpha]$. Clearly,

$$\alpha \in B_{r-\varepsilon}(\alpha) \subset B_r(\alpha) \subset B_r[\alpha] \subset B_{r+\varepsilon}(\alpha)\,, \qquad \forall r > \varepsilon > 0\,. \tag{3.1}$$

Definition 3.2. A set $A \subset X$ is said to be a *neighbourhood* of $\alpha \in X$ if A contains an open ball $B_r(\alpha)$.

In view of (3.1) the balls $B_r(\alpha)$ and $B_r[\alpha]$ are neighbourhoods of the point α. Now we can rephrase Definition 2.1 as follows.

Definition 2.1'. A sequence x_n in a metric space converges to x if for any ball $B_r(x)$ (or any neighbourhood A of x) there exists an integer n' such that for all $n > n'$ we have $x_n \in B_r(x)$ (or $x_n \in A$).

Exercise 3.3. Prove that two metrics ρ and σ are equivalent if and only if every open ball $B_r^\rho(x)$ contains an open ball $B_s^\sigma(x)$ and every open ball $B_s^\sigma(x)$ contains an open ball $B_r^\rho(x)$.

By the above, if A is a neighbourhood of x in (X, ρ) then it is a neighbourhood of x with respect to every metric which is equivalent to ρ.

Definition 3.4. A set is *open* if if it contains a ball about each of its points. (Equivalently, a set is open if it contains a neighbourhood of each of its points.)

Exercise 3.5. Prove that an open ball in a metric space (X, ρ) is open.

Theorem 3.6. *If (X, ρ) is a metric space then*

(1) *the whole space X and the empty set $\varnothing$ are both open,*
(2) *the union of any collection of open subsets of X is open,*
(3) *the intersection of any finite collection of open subsets of X is open.*

Proof.
(1) The whole space is open because it contains all open balls and the empty set is open because it does not contain any points.
(2) If x belongs to the union of open sets A_ν then x belongs to at least one of the sets A_ν. Since this set is open, it also contains an open ball about x. This ball lies in the union of A_ν, so the union is an open set.
(3) If $A_1, A_2, \dots, A_k$ are open sets and $x \in \cap_{n=1}^k A_n$ then $x \in A_n$ for every $n = 1, \dots, k$. Since A_n are open, for each n there exists r_n such that $B_{r_n}(x) \subset A_n$. Let $r = \min\{r_1, r_2, \dots, r_k\}$. Then, in view of (3.1), $B_r(x) \subset B_{r_n}(x) \subset A_n$ for all $n = 1, \dots, k$, so $B_r(x) \subset \left(\cap_{n=1}^k A_n\right)$. $\square$

Exercise 3.5. Show that the intersection of an infinite collection open sets is not necessarily open.

Lemma 3.8. *A set is open if and only if it coincides with the union of a collection of open balls.*

Proof. According to Theorem 3.6 the union of any collection of open balls is open. On the other hand, if A is open then for every point $x \in A$ there exists a ball $B(x)$ about x lying in A. We have $A = \cup_{x \in A} B(x)$. Indeed, the union $\cup_{x \in A} B(x)$ is a subset of A because every ball $B(x)$ is a subset of A, and the union contains every point $x \in A$ because $x \in B(x)$. $\square$

Definition 3.9. A point $x \in A$ is said to be an *interior* point of the set A if there exists an open ball $B_r(x)$ lying in A. The *interior* of a set A is the union of all open sets contained in A, that is, the maximal open set contained in A. The interior of A is denoted by $int(A)$.

Clearly, the interior of A coincides with set of interior points of A. Indeed, if x is an interior point then there exists an open ball $B_r(x)$ lying in A. This ball is an open set lying in A and therefore is a subset of the maximal open set $int(A) \subset A$. Conversely, if $x \in int(A)$ then (since $int(A)$ is open) there exists a ball $B_r(x) \subset int(A) \subset A$, so x is an interior point of A.

Definition 3.10. A point $x \in X$ is called a *limit* point of a set A if every ball about x contains a point of A distinct from x. Other terms for "limit point" are point of accumulation or cluster point. The set of limit points of A is denoted A'.

Lemma 3.11. *A point x is a limit point of a set A if and only if there is a sequence x_n of elements of A distinct from x which converges to x.*

Proof. If $x_n \to x$ then every ball about x contains a point x_n (see Definition 2.1'). If every ball about x contains a point of A distinct from x then there exists a sequence of points $x_n \in A$ distinct from x and lying in the balls $B_{1/n}(x)$. Obviously, this sequence converges to x. □

Definition 3.12. A set is *closed* if it contains all its limit points.

Lemma 3.13 (another definition of closed sets). *A set A is closed if and only if the limit of any convergent sequence of elements of A lies in A.*

Proof. If a sequence of elements of A has a limit then either this limit coincides with one of the elements of the sequence (and then it lies in A) or it is a limit point of A. Therefore a closed set A contains the limits of all convergent sequences $\{x_n\} \subset A$.

Conversely, every limit point is a limit of some sequence $\{x_n\} \subset A$. Therefore A contains all its limit points, provided that the limit of any convergent sequence of elements of A lies in A. □

Exercise 3.14. Prove that a closed ball in a metric space (X, ρ) is closed.

Definition 3.15. If $A \subset X$ then $\mathcal{C}(A)$ denotes the complement of the set A in X, that is, the set of all points $x \in X$ which do not belong to A.

Theorem 3.16. *If A is open then $\mathcal{C}(A)$ is closed. If A is closed then $\mathcal{C}(A)$ is open.*

Proof. If A is open then for every point of A there exists a ball about this point lying in A. Clearly, such a ball does not contain any points from $\mathcal{C}(A)$. This means that every point of A is not a limit point of $\mathcal{C}(A)$, that is, $\mathcal{C}(A)$ contains all its limit points.

If A is closed then it contains all its limit points, so any point $x \in \mathcal{C}(A)$ is not a limit point of A. This means that there exists a ball $B_r(x)$ which lies in $\mathcal{C}(A)$, that is, $\mathcal{C}(A)$ is open. □

Exercise 3.17. Prove that, in a metric space (X, ρ),

(1) the whole space X and the empty set $\varnothing$ are both closed,
(2) the intersection of any collection of closed sets is closed,
(3) the union of any finite collection of closed sets is closed.

Hint. This follows from Theorems 3.6 and Theorem 3.16.

Definition 3.18. The *closure* of a set A is the intersection of all closed sets containing A, that is, the minimal closed set containing A. The closure is denoted by $cl(A)$ or $\bar{A}$.

Theorem 3.19. *$x \in \bar{A}$ if and only if there exists a sequence $\{x_n\} \subset A$ which converges to x.*

Proof. Denote by $\widetilde{A}$ the set of limits of all convergent sequences $\{x_n\} \subset A$. Note that $A \subset \widetilde{A}$ because for every $x \in A$ the sequence $\{x, x, x, \dots\} \subset A$ converges to x. We have to prove that $\widetilde{A} = \bar{A}$.

Let $x \notin \widetilde{A}$. If for every $\varepsilon > 0$ we can find $y(\varepsilon) \in A$ satisfying $\rho(x, y) \leqslant \varepsilon$ then, taking $\varepsilon = 1/n$, we obtain a sequence $y_n = y(1/n)$ which converges to x. Thus, there exists a positive ε such that $\rho(x, y) > \varepsilon$ for all $y \in A$ or, in other words, there exists a ball $B_\varepsilon(x) \in \mathcal{C}(\widetilde{A})$. Therefore the complement $\mathcal{C}(\widetilde{A})$ is open and, consequently, $\widetilde{A}$ is closed.

It remains to prove that $\widetilde{A}$ is the minimal closed set which contains A. Let $K \subset \widetilde{A}$ and $K \neq \widetilde{A}$. Then K does not contain at least one limit x of a convergent sequence $\{x_n\} \subset A$. If $x = x_n$ for some n then $x \in A$ and therefore $A \not\subset K$. If $x \neq x_n$ all n then x is a limit point. Therefore K is not the closure of A. $\square$

Example 3.20. Let (X, ρ) be a metric space with the discrete metric

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

Then

$$B_r[a] = \begin{cases} a, & \text{if } r < 1, \\ X, & \text{if } r \geqslant 1. \end{cases} \qquad B_r(a) = \begin{cases} a, & \text{if } r \leqslant 1, \\ X, & \text{if } r > 1, \end{cases}$$

Since the open ball is open, this implies that any point is an open set. Since any set coincides with the union of its elements, Theorem 3.6 implies that any subset of X is open. Therefore, by Theorem 3.16, any subset of X is closed.

The closure of the open ball $B_r(a)$ does not necessarily coincide with the closed ball $B_r[a]$. In particular, in the Example 3.21 $\overline{B}_1(a) = B_1(a) = a$ (since $B_1(a)$ is closed) but $B_1[a] = X$.

Exercise 3.21. Proved that in a normed linear space $\overline{B_r(\alpha)} = B_r[\alpha]$.

4. Completeness

Definition 4.1. A sequence x_n of elements of a metric space (X, ρ) is called a *Cauchy sequence* if, given any $\varepsilon > 0$, there exists n_ε such that $\rho(x_n, x_m) < \varepsilon$ for all $n, m > n_\varepsilon$.

Exercise 4.2. Show that every convergent sequence is a Cauchy sequence.

Exercise 4.3. Prove the following statement: if a Cauchy sequence has a convergent subsequence then it is convergent with the same limit.

Definition 4.4. A metric space (X, ρ) is said to be *complete* if any Cauchy sequence $\{x_n\} \subset X$ converges to a limit $x \in X$.

There are incomplete metric spaces. If a metric space (X, ρ) is not complete then it has Cauchy sequences which do not converge. This means, in a sense, that there are gaps (or missing elements) in X. Every incomplete metric space can be made complete by adding new elements, which can be thought of as the missing limits of non-convergent Cauchy sequences. More precisely, we have the following theorem.

Theorem 4.5. *Let (X, ρ) be an arbitrary metric space. Then there exists a complete metric space $(\widetilde{X}, \widetilde{\rho})$ such that*

(1) *$X \subset \widetilde{X}$ and $\widetilde{\rho}(x, y) = \rho(x, y)$ whenever $x, y \in X$;*
(2) *for every $\widetilde{x} \in \widetilde{X}$ there exists a sequence of elements $x_n \in X$ such that $x_n \to \widetilde{x}$ as $n \to \infty$ in the space $(\widetilde{X}, \widetilde{\rho})$.*

The metric space $(\widetilde{X}, \widetilde{\rho})$ is said to be the *completion* of (X, ρ). If (X, ρ) is already complete then necessarily $X = \widetilde{X}$ and $\rho = \widetilde{\rho}$.

Theorem 4.6. *Let (A, ρ) be a subspace of a complete metric space (X, ρ) and $\bar{A}$ be the closure of A in (X, ρ). Then $(\bar{A}, \rho)$ is the completion of (A, ρ).*

Proof. Let $\{x_n\}$ be a Cauchy sequence in $\bar{A}$. Since (X, ρ) is complete and $\bar{A} \subset X$, this sequence converges to some element $x \in X$. Since $\bar{A}$ is closed, by Lemma 3.13, we have $x \in \bar{A}$. Therefore the space $(\bar{A}, \rho)$ is complete. Now the theorem follows from Theorem 3.19. □

Example 4.7. Let X be the set of rational numbers with the standard metric $\rho(x, y) = |x - y|$. This metric space is not complete because any sequence x_n which converges to an irrational number is a Cauchy sequence but does not have a limit in X. The completion of this space is the set of all real numbers $\mathbb{R}$ with the same metric $\rho(x, y) = |x-y|$. Any irrational number can be written as an infinite decimal fraction $0.a_1a_2\ldots$ or, in other words, can be identified with the Cauchy sequence 0, $0.a_1$, $0.a_1a_2$, ... of rational numbers which does not converge to a rational limit.

The space of real numbers $\mathbb{R}$ is defined as the completion of the space of rational numbers and therefore, by definition, is complete.

Example 4.8. Since $\mathbb{R}$ is complete, the space of complex numbers $\mathbb{C}$ with the standard metric $\rho(x, y) = |x - y|$ is also complete. Indeed, if $\{c_n\}$ is a sequence of complex numbers and $c_n = a_n + ib_n$, where $a_n = \operatorname{Re} c_n$ and $b_n = \operatorname{Im} c_n$, then

$\{c_n\}$ is a Cauchy sequnce if and only if $\{a_n\}$ and $\{b_n\}$ are Cauchy sequnces of real numbers;

the sequence $\{c_n\}$ converges if and only if the sequences $\{a_n\}$ and $\{b_n\}$ converge.

Theorem 4.9. *$B(S)$ is complete.*

Proof. Let $f_1, f_2, \ldots$ be a Cauchy sequence in $B(S)$. Then for any $\varepsilon > 0$ there exists n_ε such that

$$\sup_{x \in S} |f_n(x) - f_m(x)| \leqslant \varepsilon/2\,, \qquad \forall n, m > n_\varepsilon\,.$$

This implies that for each fixed $x \in S$ the numbers $f_n(x)$ form a Cauchy sequence of real (or complex, if f_n are complex-valued functions) numbers. Since the space of real (or complex) numbers is complete, this sequence has a limit. Let us denote this limit by $f(x)$. Then $f_n(x) \to f(x)$ for each fixed $x \in S$, that is, for any $\varepsilon > 0$ there exists an integer $n_{\varepsilon,x}$ (which may depend on x) such that

$$|f_n(x) - f(x)| \leqslant \varepsilon/2\,, \qquad \forall n > n_{\varepsilon,x}\,.$$

We have

$$|f_n(x) - f(x)| \leqslant |f_n(x) - f_m(x)| + |f_m(x) - f(x)|$$

If $n, m > n_\varepsilon$ and $m > n_{\varepsilon,x}$ then the right hand side is estimated by ε. Therefore the left hand side is estimated by ε for all $x \in S$ (indeed, given x we can always choose m in the right hand side to be greater than n_ε and $n_{\varepsilon,x}$). This implies that $\sup_{x \in S} |f_n(x) - f(x)| \leqslant \varepsilon$ for all $n > n_\varepsilon$, which means that $f_n \to f$ uniformly.

It remains to prove that f is bounded. Choosing $n > n_\varepsilon$ we obtain

$$\begin{aligned}\sup_{x \in S} |f(x)| \leqslant \sup_{x \in S} |f_n(x) - f_n(x) + f(x)| &\leqslant \sup_{x \in S} \left(|f_n(x)| + |f_n(x) - f(x)|\right) \\ &\leqslant \sup_{x \in S} |f_n(x)| + \sup_{x \in S} |f_n(x) - f(x)| \leqslant \sup_{x \in S} |f_n(x)| + \varepsilon\,.\end{aligned}$$

Since f_n is bounded, this estimate implies that f is also bounded. □

Corollary 4.10. *$C[a,b]$ is complete.*

Proof. Since continuous functions on $[a,b]$ are bounded, Theorem 4.9 implies that any Cauchy sequence of continuous functions f_k uniformly converges to a bounded function f on $[a,b]$, and we only need to prove that the function f is continuous.

In order to prove that we have to show that for any $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| \leqslant \varepsilon$ whenever $|x-y| \leqslant \delta$. We have

$$\begin{aligned}|f(x) - f(y)| &= |f(x) - f_n(x) + f_n(x) - f_n(y) + f_n(y) - f(y)| \\ &\leqslant |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)|\,.\end{aligned}$$

Since $f_n \to f$ in $B(S)$, we can choose n such that $|f(x) - f_n(x)| \leqslant \varepsilon/3$ and $|f(y) - f_n(y)| \leqslant \varepsilon/3$. Since the function f_n is continuous, there exists $\delta > 0$ such that $|f_n(x) - f_n(y)| \leqslant \varepsilon/3$ whenever $|x-y| \leqslant \delta$. Therefore

$$|f(x) - f(y)| \leqslant \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon$$

whenever $|x-y| \leqslant \delta$. $\square$

Definition 4.11. A complete normed linear space is called a *Banach space*.

Example 4.12. In view of Theorem 4.9 and Corollary 4.10, $B(S)$ and $C[a,b]$ are Banach spaces.

Let x_n be a sequence of elements of a normed linear space X.

Definition 4.13. The series $\sum_{n=1}^{\infty} x_n$ is said to be *convergent* if the sequence σ_k defined by $\sigma_k = \sum_{n=1}^{k} x_n$ is convergent in X. If $\sigma_k \to x \in X$ as $k \to \infty$ then we write $\sum_{n=1}^{\infty} x_n = x$.

Definition 4.14. The series $\sum_{n=1}^{\infty} x_n$ is said to be *absolutely convergent* if $\sum_{n=1}^{\infty} \|x_n\| < \infty$.

Theorem 4.15. *In a Banach space every absolutely convergent series is convergent.*

Proof. Let $s_k = \sum_{n=1}^{k} \|x_n\|$. Since the series $\sum_{n=1}^{\infty} \|x_n\|$ is convergent, the sequence of positive numbers $\{s_k\}$ converges and therefore it is a Cauchy sequence. If $m > k$, $\sigma_m = \sum_{n=1}^{m} x_n$ and $\sigma_k = \sum_{n=1}^{k} x_n$ then

$$\rho(\sigma_m, \sigma_k) = \|\sigma_m - \sigma_k\| = \|\sum_{n=k+1}^{m} x_n\| \leqslant \sum_{n=k+1}^{m} \|x_n\| = |s_m - s_k|\,.$$

This implies that $\{\sigma_n\}_{n=1,2,\dots}$ is a Cauchy sequence in our Banach space, and therefore it converges. $\square$

Corollary 4.16. *Let f_n be bounded functions defined on a set S. If $\sum_{n=1}^{\infty} \sup_{x\in S} |f_n(x)| < \infty$ then there exists a bounded function f on S such that*

$$\sup_{x \in S} |f(x) - \sum_{n=1}^{k} f_n(x)| \underset{k\to\infty}{\to} 0\,.$$

If $S = [a,b]$ is a bounded interval and the functions f_n are continuous on $[a,b]$ then f is also continuous.

Proof. The corollary immediately follows from Theorem 4.15 and the fact that $B(S)$ and $C[a,b]$ are Banach spaces (see Theorem 4.9 and Corollary 4.10). $\square$

5. Baire Category Theorem

Definition 5.1. A subset A of a metric space (X, ρ) is said to be dense if its closure $\bar{A}$ coincides with X.

Definition 5.2. A subset A of a metric space is said to be nowhere dense if its closure $\bar{A}$ has empty interior.

The following results is called the *Baire category theorem.*

Theorem 5.3. *The complement of any countable union of nowhere dense subsets of a complete metric space (X, ρ) is dense in X.*

Proof. Suppose that A_n, $n = 1, 2, \ldots$, is a countable collection of nowhere dense subsets of X. Set $X_0 := X \setminus \bigcup_{n=1}^{\infty} A_n$. We wish to show that X_0 is dense in X. Now, $X \setminus \bigcup_{n=1}^{\infty} \bar{A}_n \subset X_0$, and a set is nowhere dense if and only if its closure is. Hence, by taking closures if necessary, we may assume that each A_n is closed. Suppose then, by way of contradiction, that X_0 is not dense in X. Then $X \setminus \bar{X}_0 \neq \varnothing$. Since $X \setminus \bar{X}_0$ is open, and non-empty, there exists $x_0 \in X \setminus \bar{X}_0$ and r_0 such that $B_{r_0}(x_0) \subset X \setminus \bar{X}_0 \subset \bigcup_{n=1}^{\infty} A_n$. The idea of the proof is to construct a sequence of points in X with a limit $x \in B_{r_0}(x_0)$ which does not belong to any of the sets $A_1, A_2, \ldots$, which will contradict to the inclusion $B_{r_0}(x_0) \subset \bigcup_{n=1}^{\infty} A_n$.

We start by noticing that, since A_1 is nowhere dense, the open ball $B_{r_0}(x_0)$ is not contained in A_1. This means that there is some point $x_1 \in B_{r_0}(x_0) \setminus A_1$. Furthermore, the set $B_{r_0}(x_0) \setminus A_1$ coincides with the intersection of open sets $B_{r_0}(x_0)$ and $\mathcal{C}(A_1)$ (see Theorem 3.16). Therefore it is open and contains a closed ball $B_{r_1}[x_1]$ with $r_1 < 1$. We have $B_{r_1}[x_1] \subset B_{r_0}(x_0)$ and $B_{r_1}[x_1] \bigcap A_1 = \varnothing$.

Now, since A_2 is nowhere dense, the open ball $B_{r_1}(x_1)$ is not contained in A_2. Thus, there is some point $x_1 \in B_{r_1}(x_1) \setminus A_2$. The set $B_{r_1}(x_1) \setminus A_2$ is open because it coincides with the intersection of open sets $B_{r_1}(x_1)$ and $\mathcal{C}(A_2)$. Therefore $B_{r_1}(x_1) \setminus A_2$ contains a closed ball $B_{r_2}[x_2]$ with $r_1 < 1/2$. We have $B_{r_2}[x_2] \subset B_{r_1}(x_1)$ and $B_{r_2}[x_2] \bigcap A_2 = \varnothing$.

Recursively, we obtain a sequence $x_0, x_1, x_2, \ldots$ in X and positive real numbers $r_0, r_1, r_2, \ldots$ satisfying $r_n < 1/n$, such that $B_{r_n}[x_n] \subset B_{r_{n-1}}(x_{n-1})$ and $B_{r_n}[x_n] \bigcap A_n = \varnothing$. If $m, n > N$ then both points x_m and x_n belong to $B_{r_N}[x_N]$ and

$$\rho(x_m, x_n) \leqslant \rho(x_m, x_N) + \rho(x_N, x_m) \leqslant 2/N.$$

Hence $\rho(x_m, x_n) \to 0$ as $n, m \to \infty$, which means that x_n form a Cauchy sequence. Therefore there is some $x \in X$ such that $x_n \to x$. Since $x_n \in B_{r_n}[x_n] \subset B_{r_N}[x_N]$ for all $n \geqslant N$, it follows that $x \in B_{r_N}[x_N]$ for all $N = 1, 2, \ldots$. However, by our construction, $B_{r_{N+1}}[x_{N+1}] \subset B_{r_N}(x_N)$ and $B_{r_N}(x_N) \bigcap A_N = \varnothing$. Hence $x \notin A_N$ for any N. This is our required contradiction and the result follows. $\square$

The Baire category theorem implies, in particular, that a complete metric space cannot be given as a countable union of nowhere dense sets. In other words, if a complete metric space is equal to a countable union of sets, then not all of these can be nowhere dense; that is, at least one of them has a closure with non-empty interior. Another corollary to the theorem is that if a metric space can be expressed as a countable union of nowhere dense sets, then it is not complete.

6. Linear operators in normed spaces

1. Continuity.

Given a map $T : X \to Y$ and a subset $A \subset Y$, the set $\{x \in X : Tx \in A\}$ is denoted $T^{-1}(A)$ and called the *inverse image* of A. Note that $T^{-1}(A)$ is a well-defined set irrespective of whether T has any inverse.

Definition 6.1. Let (X, ρ) and (Y, d) be metric spaces. A map $T : X \to Y$ is said to be *continuous at* $\alpha \in X$ if for any open ball $B_\varepsilon(T\alpha)$ about $T\alpha$ there exists a ball $B_\delta(\alpha)$ about α such that $B_\delta(\alpha) \subset T^{-1}(B_\varepsilon(T\alpha))$.

This definition can be reformulated as follows: the map $T : (X, \rho) \to (Y, d)$ is continuous at $\alpha \in X$ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $d(Tx, T\alpha) < \varepsilon$ whenever $\rho(x, \alpha) < \delta$.

Exercise 6.2. Show that the map $T : (X, \rho) \to (Y, d)$ is continuous at $\alpha \in X$ if and only if for every sequence x_n converging to α in (X, ρ), the sequence Tx_n converges to $T\alpha$ in (Y, d).

Theorem 6.3. *Let (X, ρ) and (Y, σ) be metric space and $T : X \to Y$ be a map from X to Y. Then the following statements are equivalent:*

(1) *T is continuous,*
(2) *the inverse image of every open subset of Y is an open subset of X,*
(3) *the inverse image of every closed subset of Y is a closed subset of X.*

Proof. The inverse image of the complement of a set A coincides with the complement of the inverse image $T^{-1}(A)$. Therefore Theorem 3.16 implies that (2) is equivalent to (3). Let us prove that (2) is equivalent to (1).

Assume first that T is continuous. Let A be an open subset of Y and $x \in T^{-1}(A) \subset X$. Since A is open, there exists a ball $B_\varepsilon(Tx)$ about the point Tx such that $B_\varepsilon(Tx) \subset A$. Since T is continuous, there exists a ball $B_\delta(x)$ about x such that $B_\delta(x) \subset T^{-1}(B_\varepsilon(Tx)) \subset T^{-1}(A)$ (Definition 6.1′). Therefore for every point $x \in T^{-1}(A)$ there exists a ball $B_\delta(x)$ lying in $T^{-1}(A)$, which means that $T^{-1}(A)$ is open.

Assume now that the inverse image of any open set is open. Let $x \in X$ and $B_\varepsilon(Tx)$ is a ball about $Tx \in Y$. The inverse image $T^{-1}(B_\varepsilon(Tx))$ is an open set which contains the point x. Therefore there exists a ball $B_\delta(x)$ about x such that $B_\delta(x) \subset T^{-1}(B_\varepsilon(Tx))$. This implies that T is continuous. □

In the rest of this section we shall be assuming that X and Y are normed linear spaces and denote by $\|\cdot\|_X$, $\|\cdot\|_Y$ the the corresponding norms.

Recall that a map $T : X \to Y$ is called a linear operator if $T(x + y) = Tx + Ty$ and $T(\lambda x) = \lambda Tx$ for all $x, y \in X$ and $\lambda \in \mathbb{R}$ (or $\lambda \in \mathbb{C}$). In particular, if T is a linear operator then $T0 = 0$. Linear operators form a linear space which is usually denoted $\mathcal{B}(X, Y)$ (we define $(T_1 + T_2)x = T_1x + T_2x$ and $(\lambda T)x = \lambda Tx$).

Definition 6.4. A linear map $T : X \to Y$ is said to be *bounded* if there exists a positive constant C such that $\|Tx\|_Y \leqslant C\|x\|_X$ for all $x \in X$.

Theorem 6.5. *Let X and Y be normed linear spaces and $T : X \to Y$ be a linear map. Then T is continuous if and only if it is bounded.*

Proof. Let $\rho(x_1, x_2) = \|x_1 - x_2\|_X$ and $d(y_1, y_2) = \|y_1 - y_2\|_Y$ be the metrics on X and Y generated by the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$.

Assume first that T is bounded, that is, $\|Tx\|_Y \leqslant C\|x\|_X$. Let $\alpha \in X$ and $\varepsilon > 0$. Take $\delta = C^{-1}\varepsilon$. Then, for all $x \in X$ satisfying $\rho(x,\alpha) = \|x-\alpha\|_X \leqslant \delta$, we have

$$d(Tx, T\alpha) = \|Tx - T\alpha\|_Y = \|T(x-\alpha)\|_Y \leqslant C\|x-\alpha\|_X \leqslant \varepsilon.$$

This implies that T is continuous.

Assume now that T is continuous. Then T is continuous at 0, and therefore there exists $\delta > 0$ such that

$$d(0, Tx_0) = \|Tx_0\|_Y \leqslant 1 \quad \text{whenever} \quad \rho(0, x_0) = \|x_0\|_X \leqslant \delta\,. \tag{6.1}$$

If $x \in X$, let us denote $c = \delta\|x\|_X^{-1}$ and $x_0 = c\,x$. Then $\|x_0\|_X = \delta$. Since T is a linear map, (6.1) implies $\|Tx\|_Y = c^{-1}\|Tx_0\|_Y \leqslant c^{-1} = \delta^{-1}\|x\|_X$, which means that T is bounded. $\square$

If T is bounded, then there is the minimal constant C such that $\|Tx\|_y \leqslant C\,\|x\|_X$, which is denoted $\|T\|$.

Exercise 6.6. Prove that the function $\|T\|$ defined on $\mathcal{B}(X,Y)$ satisfies conditions of Definition 1.2 and is therefore a norm on the linear space $\mathcal{B}(X,Y)$.

2. The Banach-Steinhaus theorem.

The following result is called the Banach-Steinhaus theorem or the Principle of Uniform Boundedness (for obvious reasons).

Theorem 6.7. *Let X be a Banach space and let $\mathcal{F}$ be a family of bounded linear operators from X into a normed space Y such that the set $\{Tx : T \in \mathcal{F}\}$ is bounded for each $x \in X$. Then the set of norms $\{\|T\| : T \in \mathcal{F}\}$ is bounded.*

Proof. Let A_n be the set of all $x \in X$ such that $\|Tx\|_Y \leqslant n$ for all $T \in \mathcal{F}$, where $n = 1, 2, \ldots$. Then each set A_n coincides with the intersection $\bigcap_{T\in\mathcal{F}} T^{-1}(B_n[0])$ of inverse images $T^{-1}(B_n[0])$ of the closed balls $B_n[0] \subset Y$. Since the maps T are continuous, these inverse images are closed, and so is their intersection. Thus, A_n are closed subsets of X. Moreover, by hypothesis, each $x \in X$ lies in some A_n. Thus, we may write $X = \bigcup_{n=1}^{\infty} A_n$.

By the Baire Theorem, together with the fact that the sets A_n are closed, it follows that there is some positive integer m such that A_m has non-empty interior. If x_0 is an interior point of A_m then A_m contains a closed ball $B_r[x_0]$. By the definition of A_m, this means that $\|Tx\|_Y \leqslant m$ whenever $\|x - x_0\|_X \leqslant r$ and $T \in \mathcal{F}$. But then for any $x \in X$ with $\|x\|_X \leqslant r$ we have $\|(x+x_0) - x_0\|_X \leqslant r$ and therefore

$$\|Tx\|_Y = \|T(x+x_0) - Tx_0\|_Y \leqslant \|T(x+x_0)\|_Y + \|Tx_0\|_Y \leqslant 2m$$

for all $T \in \mathcal{F}$. If $x \in X$, $x \neq 0$ and $c_r = r/\|x\|_X$ then the norm of the vector $c_r x$ is equal to r. Therefore $2m \geqslant \|T(c_r x)\|_Y = c_r\,\|Tx\|_Y$, which implies that $\|Tx\|_Y \leqslant 2mr^{-1}\|x\|_X$ and, consequently, $\|T\| \leqslant 2mr^{-1}$. $\square$

3. The Open Mapping Theorem.

Definition 6.8. A map $T : X \to Y$ is said to be open if, for every open set $A \in X$, its image $T(A)$ is open in Y.

Definition 6.9. A subset A of a linear space X is said to be *symmetric* if $-x \in A$ whenever $x \in A$. The subset A is said to be convex if $tx_1 + (1-t)x_2 \in A$ whenever $x_1, x_2 \in A$ and $t \in [0,1]$ (in other words, A is convex if any line segment joining two points from A lies in A).

Exercise 6.10. Prove that, in a normed linear space X, the closure of a symmetric set is symmetric and the closure of a convex set is convex.

If $A \in X$ and $\delta > 0$, we shall denote by δA the set of points x such that $\delta^{-1}x \in A$.

Exercise 6.11. Let X be a normed linear space. Prove that $\delta A_0 \subset \delta A$ whenever $A_0 \subset A$, and that $\overline{\delta A} = \delta \bar{A}$.

The following result, called the Open Mapping Theorem, is also a consequence of Theorem 5.3.

Theorem 6.12. *Let X and Y be Banach spaces and $T : X \to Y$ be a bounded linear operator mapping X onto Y. Then T is an open map.*

In order to prove the theorem, we shall need the following auxiliary results.

Lemma 6.13. *Let X be a Banach space, Y be a normed space and $T : X \to Y$ be a bounded linear operator. Assume that $B_r(0) \subset \overline{T(B_R(0))}$. Then $B_r(0) \subset T(B_R(0))$.*

Proof. Let $\varepsilon \in (0, \frac{1}{2})$, and let $y \in B_r(0)$. Then $y \in \overline{T(B_R(0))}$ which implies that there is $y_1 \in T(B_R(0))$ such that $\|y - y_1\|_Y < \varepsilon r$. That is, there exists $x_1 \in B_R(0)$ such that $y_1 = Tx_1$ satisfies $\|y - Tx_1\| < \varepsilon r$. In other words, $y - Tx_1 \in B_{\varepsilon r}(0)$. The inclusion $B_r(0) \subset \overline{T(B_R(0))}$ implies that $B_{\varepsilon r}(0) \subset \overline{T(B_{\varepsilon R}(0))}$ (see Exercise 6.11), that is, $y - Tx_1$ belongs to $B_{\varepsilon r}(0) \subset \overline{T(B_{\varepsilon R}(0))}$.

Similarly, we can find a point $x_2 \in B_{\varepsilon r}(0)$ such that $\|(y - Tx_1) - Tx_2\|_Y < \varepsilon^2 r$. Continuing in this way (i.e., by recursion) we obtain a sequence $x_1, x_2, x_3, \dots$ such that $x_n \in B_{\varepsilon^{n-1}R}(0)$ and

$$y - Tx_1 - Tx_2 - \cdots - Tx_n \in B_{\varepsilon^n r}(0) \subset \overline{T(B_{\varepsilon^n R}(0))}\,.$$

The above inclusion implies that $y = \sum_{n=1}^{\infty} Tx_n$. On the other hand, $\|x_n\|_X < \varepsilon^{n-1}R$ and, consequently, $\sum_n \|x_n\|_X < \infty$. By Theorem 4.15, the series $\sum_n x_n$ converges to some vector $x \in X$. Since T is continuous, we have $\sum_{n=1}^{k} Tx_n = T\sum_{n=1}^{n} x_k \to Tx$ as $k \to \infty$. Therefore $y = Tx$. Furthermore,

$$\|\sum_{n=1}^{k} x_n\|_X \leqslant \sum_{n=1}^{k} \|x_n\|_X \leqslant \sum_{n=1}^{k} \varepsilon^{k-1}R < R/(1-\varepsilon),$$

so that $\|x\|_X \leqslant R/(1-\varepsilon) < R/(1-2\varepsilon)$. Hence $x \in B_{R_\varepsilon}(0)$ and $y = Tx \in T(B_{R_\varepsilon}(0))$, where $R_\varepsilon := R/(1-2\varepsilon)$.

Thus we have proved that $B_r(0) \subset T(B_{R_\varepsilon}(0))$ for all $\varepsilon \in (0, \frac{1}{2})$. Therefore $B_d(0) \subset T(B_{dR_\varepsilon/r}(0))$ for all $d > 0$ (Exercise 6.11). Let $y \in B_r(0)$ and $d > 0$ be such that $\|y\|_Y < d < r$. Then $y \in B_d(0) \subset T(B_{dR_\varepsilon/r}(0))$. Since $d/r < 1$ and $r_\varepsilon \to R$ as $\varepsilon \to 0$, for all sufficiently small ε we have $dR_\varepsilon/r < R$. It follows that $y \in T(B_R(0))$. Hence $B_r(0) \subset T(B_R(0))$. $\square$

Proof of Theorem 6.11. Let G be an open subset of X. We want to prove that $T(G)$ is open. If $G = \varnothing$ then $T(G) = \varnothing$ and is open. Assume that $G \neq \varnothing$ and take

$y_0 \in T(G)$. Then there exists a vector $x_0 \in G$ such that $Tx_0 = y_0$. Since G is open, we can find a ball $B_r(x_0) \subset G$. In order to show that $T(G)$ is open, it is sufficient to construct a ball $B_{r'}(y_0) \subset T(B_r(x_0))$. Note that $B_{r'}(y_0)$ coincides with the set of vectors $y \in Y$ such that $y - y_0 \in B_{r'}(0)$, and $B_r(x_0)$ is the set of vectors $x \in X$ such that $x - x_0 \in B_r(0)$. Since $Tx_0 = y_0$, we have $B_{r'}(y_0) \subset T(B_r(x_0))$ if and only if $B_{r'}(0) \subset T(B_r(0))$. Thus, we need to show that for every $r > 0$ there exists $r' > 0$ such that $B_{r'}(0) \subset T(B_r(0))$.

The collection of balls $\{B_n(0)\}_{n=1,2,\dots}$ covers X. Since $T(X) = Y$, it follows that $Y = \bigcup_{n=1}^{\infty} T(B_n(0))$. By Baire Theorem, there is a positive integer m such that the closure $\overline{T(B_m(0))}$ has a nonempty interior. Thus there exists $y \in Y$ and $\varepsilon > 0$ such that $B_\varepsilon(y) \subset \overline{T(B_m(0))}$.

Since the ball $B_m(0)$ is symmetric and convex, so is its image $T(B_m(0))$. Therefore the closure $\overline{T(B_m(0))}$ is also symmetric and convex. The former implies that $B_\varepsilon(-y) \subset \overline{T(B_m(0))}$. If $\|z\|_Y < \varepsilon$ then $z + y \in B_\varepsilon(y) \subset \overline{T(B_m(0))}$ and $z - y \in B_\varepsilon(-y) = -B_\varepsilon(y) \subset \overline{T(B_m(0))}$. Since $\overline{T(B_m(0))}$ is convex, we see that

$$z = \frac{1}{2}(z + y) + \frac{1}{2}(z - y) \in \overline{T(B_m(0))}\,.$$

Thus we have $z \in \overline{T(B_m(0))}$ whenever $\|z\|_Y < \varepsilon$ or, in other words, $B_\varepsilon(0) \subset \overline{T(B_m(0))}$. But then $B_\varepsilon(0) \subset T(B_m(0))$ (Lemma 6.13). Multiplying ε and m by r/m, we see that $B_{r'}(0) \subset T(B_r(0))$ for $r' = r\varepsilon/m$. $\square$

The following corollary of the Open Mapping Theorem is known as the Inverse Mapping Theorem.

Theorem 6.14. *Any one-to-one and onto bounded linear mapping between Banach spaces has a bounded inverse.*

Proof. Any one-to-one map T from X onto Y has an inverse $T^{-1} : Y \to X$. The inverse image of a set $A \in X$ by the map T^{-1} coincides with $T(A)$. Therefore Theorem 6.14 is an immediate consequence of the Open Mapping Theorem and Theorem 6.3. $\square$

4. The Closed Graph Theorem.

The direct product $X \times Y$ of two normed linear spaces x and Y is a linear space (by definition, $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$ and $(x, y) = (\lambda x, \lambda y)$). Given $(x, y) \in X \times Y$, we define $\|(x, y)\| := \sqrt{\|x\|_X^2 + \|y\|_Y^2}$.

Exercise 6.15. Check that $\|(x, y)\|$ defined on $X \times Y$ satisfies conditions of Definition 1.2 and is therefore a norm on the linear space $X \times Y$. Show that $X \times Y$ is complete with respect to this norm if and only if both spaces x and Y are complete.

Definition 6.16. The graph of a map $T : X \to Y$, denoted $\Gamma(T)$, is the linear subspace of $X \times Y$ given by $\Gamma(T) = \{(x, y) \in X \times Y : y = Tx\}$. In other words, $\Gamma(T) = \{(x, Tx) : x \in X\}$.

Theorem 6.17. *Let X and Y be Banach spaces and $T : X \to Y$ be a linear operator. Then T is bounded if and only if its graph $\Gamma(T)$ is closed in $X \times Y$.*

Proof. Assume first that T is bounded and consider a sequence $(x_n, y_n) \in G(T)$ such that $(x_n, y_n) \to (x, y)$ as $n \to \infty$ in the Banach space $X \times Y$. Then $x_n \to x$ in X and $y_n \to y$ in Y. Since T is continuous, we also have $Tx_n \to Tx$ in Y.

However, $Tx_n = y_n$. Therefore $y = Tx$ which means that $(x, y) \in G(T)$. Thus, $G(T)$ is closed.

Conversely, suppose that $G(T)$ is closed in the Banach space $X \times Y$. Then, by Theorem 4.6, $G(T)$ is itself a Banach space with respect to the norm inherited from $X \times Y$. Consider the maps $\pi_1 : G(T) \to X$ and $\pi_2 : G(T) \to Y$ defined by $\pi_1(x, Tx) = x$ and $\pi_2(x, Tx) = Tx$. Evidently, both π_1 and π_2 are norm decreasing, and so are bounded linear operators. Moreover, it is clear that π_1 is both injective and surjective. It follows, by the Inverse Mapping Theorem, that $\pi_1^{-1} : X \to G(T)$ is bounded. But then

$$x \overset{\pi_1^{-1}}{\mapsto} (x, Tx) \overset{\pi_2}{\mapsto} Tx \,,$$

that is, $T = \pi_2 \circ \pi_1^{-1}$. Since π_1^{-1} and π_2 are continuous, T is also continuous. $\square$

Exercise 6.18. Let (X_k, ρ_k) be metric spaces and $T_1 : X_1 \to X_2$, $T_2 : X_2 \to X_3$ be continuous maps. Prove that the composition $T_2 \circ T_1$ is a continuous map from X_1 to X_3.

The closed graph theorem can be a great help in establishing the boundedness of linear operators between Banach spaces. Indeed, in order to show that a linear operator $T : X \to Y$ is bounded, one must establish essentially two things: firstly, that if $x_n \to x$ in X, then Tx_n converges in Y and, secondly, that this limit is Tx. The closed graph theorem says that to prove that T is bounded it is enough to prove that its graph is closed (provided, of course, that X and Y are Banach spaces). This means that we may assume that $x_n \to x$ and $Tx_n \to y$, for some $y \in Y$, and then need only show that $y = Tx$. In other words, thanks to the closed graph theorem, the convergence of Tx_n can be taken as part of the hypothesis rather than forming part of the proof itself.

Definition 6.19. A complete inner product space is called a *Hilbert space.*

The following corollary of Theorem 6.17 is known as Hellinger-Toeplitz theorem.

Corollary 6.20. *Let $A : H \to H$ be a linear operator on a real or complex Hilbert space H (defined on the whole space H) such that $(Ax, y) = (x, Ay)$ for all $x, y \in H$. Then A is bounded.*

Proof. In view of Theorem 6.17, it is sufficient to show that $G(A)$ is closed. Assume that $x_n \to x$ and $Tx_n \to y$. We need to show that $(x, y) \in G(A)$ or, in other words, that $y = Ax$. For every $z \in H$,

$$(z, y) = \lim_{n\to\infty} (z, Ax_n) = \lim_{n\to\infty} (Az, x_n) = (Az, x) = (z, Ax) \,.$$

Therefore $(z, y - Ax) = 0$ for all $z \in H$. Taking $z = y - Ax$, we see that $y = Ax$. $\square$

7. Compactness in metric spaces

Intervals which are bounded and closed figure prominently in analysis on the real line. The appropriate generalization of their essential properties that are relevant to analysis in more general spaces is compactness. There are two definitions of compactness in metric spaces which can be shown to be equivalent.

Definition 7.1. A subset K of a metric space (X, ρ) is said to be *(sequentially) compact* if any sequence of elements of K has a subsequence which converges to a limit in K.

It is clear from the definition that K is compact in (X, ρ) if and only if it is compact in (X, σ) for any metric σ equivalent to ρ.

The second definition needs a little terminology. If $\hat{S}$ is a family of subsets of X and $K \subseteq \cup_{\hat{S}} S$ then $\hat{S}$ is called a *cover* of K. If each member of $\hat{S}$ is open, it is called an *open cover* of K. If $\hat{S}$ is a cover of K and a subset $\hat{S}_0$ of $\hat{S}$ also covers K then $\hat{S}_0$ is called a *subcover* of $\hat{S}$. A cover (or subcover) is said to be finite if it has a finite number of members.

Definition 7.2. A subset K of a metric space (X, ρ) is said to be *compact* if any open cover of K has a finite subcover.

Definition 7.3. A metric space (X, ρ) is said to be (sequentially) compact if the set X is (sequentially) compact.

Theorem 7.4. *A set is compact if and only if it is sequentially compact.*

In order to prove this theorem we need the following auxiliary lemmas.

Lemma 7.5. *A sequentially compact metric space (X, ρ) is complete.*

Proof. Let $\{x_n\} \subset X$ be an arbitrary Cauchy sequence. Since X is compact, this sequence has a subsequence which converges to a limit in X. By Lemma 4.3, the whole sequence $\{x_n\}$ converges to the same limit. $\square$

Lemma 7.6. *A closed subset of a sequentially compact set is sequentially compact.*

Proof. Let K be compact, K_0 be a closed subset of K and $\{x_n\}$ be a sequence of elements of K_0. Since $\{x_n\} \subset K$, this sequence has a convergent subsequence. Since K_0 is closed, the limit of this subsequence lies in K_0. Therefore any sequence of elements of K_0 has a subsequence which converges to a limit in K_0 which means that K_0 is compact. $\square$

Lemma 7.7. *If K is sequentially compact then, for any positive r, K can be represented as the union of a finite collection of closed subsets whose diameters are not greater than r.*

Proof. Let $x_1 \in K$ and $A_1 = K \cap B_{r/2}[x_1]$. Since $A_1 \subset B_{r/2}[x_1]$, the diameter of A_1 is not greater than r. Indeed, for all $x, y \in A_1$ we have

$$\rho(x, y) \leqslant \rho(x, x_1) + \rho(y, x_1) \leqslant r/2 + r/2 = r\,.$$

If A_1 coincides with K, we have obtained the required representation of K. If not, we take $x_2 \in \mathcal{C}(A_1) \cap K$ and define $A_2 = K \cap B_{r/2}[x_2]$. If $A_1 \cup A_2 \neq K$ we take $x_3 \in \mathcal{C}(A_1 \cup A_2) \cap K$, define $A_3 = K \cap B_{r/2}[x_3]$ and so on. If this procedure stops after finite number of steps then $A_1 \cup A_2 \ldots \cup A_k = K$. If not, we obtain a sequence x_k such that $\rho(x_k, x_j) > r/2$ for all $j \neq k$ (because x_j lies outside the ball $B_{r/2}[x_k]$). Clearly, such a sequence does not have any convergent subsequences, which cannot be true since K is compact. $\square$

Proof of Theorem 7.8. Let K be compact and suppose it is not sequentially compact. Then there is a sequence x_n of elements of K with no subsequence converging to an element of K.

We make the following observation: for any element x, if every open ball with centre x contains x_n for an infinite number of values of n then there is a subsequence of x_n converging to x. (To see this, simply choose an increasing sequence of integers n_k such that x_{n_k} is in the ball centre x radius $\frac{1}{k}$.) Hence every element of K must fail this condition, or in other words, each element x of K is contained in some open ball $B_{r_x}(x)$ which contains x_n for only a finite number (possibly none) of values of n. Now $\{B_{r_x}(x) : x \in K\}$ is an open cover of K and since K is compact, it has a finite subcover. But then the union of these sets, and hence K would contain x_n for only a finite number of values of n which contradicts that $x_n \in K$ for all positive integers n. Thus no such sequence x_n exists and K is sequentially compact.

Let K be sequentially compact and suppose that K is not compact. Then there exists an open cover $\hat{S}$ of K which does not have any finite subcover.

Let us represent K as the union of a finite collection of closed subsets whose diameters are not greater than 1. At least one of these subsets cannot be covered by a finite number of open sets $S \in \hat{S}$. Denote this subset by A_1. By Lemma 7.6 A_1 is compact. Now we represent A_1 as the union of a finite collection of closed subsets whose diameters are not greater than $1/2$, denote the subset which does not admit a finite subcover by A_2, and so on. Then we obtain a sequence of closed sets $A_1 \supset A_2 \supset A_3 \ldots$ such that $\operatorname{diam} A_k \leqslant k^{-1}$.

Let $x_n \in A_n$ be arbitrary elements of A_n. Then $\{x_n\}$ is a Cauchy sequence. Indeed, since $x_n \in A_{n_0}$ for all $n \geqslant n_0$ and $\operatorname{diam} A_{n_0} \leqslant n_0^{-1}$, we have $\rho(x_n, x_k) \leqslant n_0^{-1}$ for all $k, n > n_0$ and n_0^{-1} can be made arbitrarily small by choosing large n_0. By Lemma 7.5 the sequence x_k converges to a limit $x \in K$. Since $x_n \in A_{n_0}$ for all $n \geqslant n_0$ and A_k are closed, the limit belongs to A_k for every k, that is, $x \in \cap_{k=1}^{\infty} A_k$.

We have proved that $\cap_{k=1}^{\infty} A_k$ is not empty. Now let us note that $\cap_{k=1}^{\infty} A_k$ cannot contain more than one point. Indeed, two points x and y cannot lie in a set whose diameter is less than $\rho(x, y)$, but $\operatorname{diam} A_k \to 0$ as $k \to 0$. Thus the intersection $\cap_{k=1}^{\infty} A_k$ consists of one point x.

This point x belongs to at least one open set $S_x \in \hat{S}$. Since S_x is open, it also contains an open ball $B_r(x)$ for some positive r. If $k^{-1} < r$ then

$$\rho(y, x) \leqslant \operatorname{diam} A_k \leqslant k^{-1} < r$$

for all $y \in A_k$, which means that $A_k \subset B_r(x)$. Thus A_k is covered by the set $S_x \supset B_r(x)$ which contradicts to the fact that A_k does not admit a finite subcover. $\square$

Definition 7.8. A subset K of a metric space (X, ρ) is *bounded* if, for some $x \in X$ and $r > 0$, we have $K \subset B_r(x)$.

Theorem 7.9. *A compact set K is bounded and closed.*

Proof. If K is not bounded then, for every $x \in X$, the family of balls $B_n(x)$, $n = 1, 2, \ldots$, is an open cover of K which does not have a finite subcover.

If K is not closed, it does not contain at least one of its limit points. Consider a sequence of elements of K which converges to this limit point. Every subsequence of this sequence converges to the same limit point. Therefore such a sequence does not have a subsequence which converges to a limit in K. $\square$

Lemma 7.10. *If K and L are compact subsets of metric spaces (X, ρ) and (Y, σ) respectively then $K \times L$ as a subset of $X \times Y$ with the metric*

$$d\left((x_1, y_1), (x_2, y_2)\right) = \sqrt{\rho(x_1, x_2)^2 + \sigma(y_1, y_2)^2}\,.$$

is compact.

Proof. Let (x_n, y_n) be an arbitrary sequence in $K \times L$. Since K is compact, there is a subsequence x_{n_k} which converges to a limit $x \in K$ as $k \to \infty$. Since L is compact, the sequence y_{n_k} has a subsequence $y_{n_{k\,i}}$ which converges to a limit $y \in L$ as $i \to \infty$. Since $x_{n_k} \to x$ as $k \to \infty$, we also have $x_{n_{k\,i}} \to x$ as $i \to \infty$. By definition of convergence, $\rho(x_{n_{k\,i}}, x) \to 0$ and $\sigma(y_{n_{k\,i}}, y) \to 0$ as $i \to \infty$. This implies that $d\big((x_{n_{k\,i}}, y_{n_{k\,i}}), (x, y)\big) \to 0$ as $i \to \infty$, that is, $(x_{n_{k\,i}}, y_{n_{k\,i}}) \to (x, y) \in K \times L$. Therefore any sequence (x_n, y_n) of elements of $K \times L$ has a subsequence which converges to a limit in $K \times L$. □

Theorem 7.11. *A bounded and closed subset of* $\mathbb{R}^n$ *is compact.*

Proof. Since any bounded subset lies in a closed cube Q^n, in view of Lemma 7.6 it is sufficient to prove that the closed cube is compact. The closed cube Q^n is a direct product of a one dimensional closed cube $\mathbb{Q}^1$ (a closed interval) and a closed cube $Q^{n-1} \subset \mathbb{R}^{n-1}$. If $\mathbb{Q}^1$ and $\mathbb{Q}^{n-1}$ are compact then, by Lemma 7.10, $\mathbb{Q}^n$ is also compact. Therefore it is sufficient to prove that a closed interval is compact (after that the required result is obtained by induction in n).

Let x_n be an arbitrary sequence of numbers lying in a closed interval $[a, b]$. Let us split $[a, b]$ into the union of two intervals of length $\delta/2$, where $\delta = b - a$. At least one of these intervals contains infinitely many elements x_n of our sequence. Let us choose one of these elements and denote it by y_1. Now we split the interval of length $\delta/2$ which contains infinitely many elements x_n into the union of two intervals of length $\delta/4$. Again, at least one of these intervals contains infinitely many elements x_n. We choose one of these elements (distinct from y_1) and denote it by y_2. Repeating this procedure, we obtain a subsequence $\{y_k\}$ of the sequence $\{x_n\}$ such that y_k lie in an interval of length 2^{-k_0} for all $k \geqslant k_0$. Clearly, $\{y_k\}$ is a Cauchy sequence. Since $\mathbb{R}$ is a complete metric space, $\{y_k\}$ converges to a limit. Since a closed interval is a closed set, this limit belongs to $[a, b]$. Thus, any sequence of elements of $[a, b]$ has a subsequence which converges to a limit in $[a, b]$, which means that the closed interval is compact. □

Theorem 7.12. *The image of a compact set by a continuous map is compact.*

Proof. Let K be a compact set and T be a continuous map. Let y_n be an arbitrary sequence of elements of $T(K)$. Then $y_n = Tx_n$ where $x_n \in K$. Since K is compact, the sequence $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ which converges to a limit $x \in K$. Then by Theorem 4.2 the subsequence $y_{n_k} = Tx_{n_k}$ converges to the limit $Tx \in T(K)$. This proves that $T(K)$ is compact. □

Theorem 7.13. *Let* (X, ρ)*,* (Y, σ) *be compact metric spaces and* $T : (X, \rho) \to (Y, \sigma)$ *be a continuous bijection. Then the inverse mapping* T^{-1} *is continuous.*

Proof. Applying Theorem 4.8 to T^{-1}, we see that it is sufficient to prove that the inverse image $(T^{-1})^{-1}(B) = T(B) \subset Y$ is closed whenever the set $B \subset X$ is closed.

If B is closed then, by Lemma 7.6 it is compact. By Theorem 7.12 $T(B)$ is also compact and therefore is closed (Theorem 7.9). □

Example 7.14. Let X be the space of continuously differentiable functions on a closed interval $[a,b]$ and ρ, σ be the metrics on X defined as follows:

$$\rho(f,g) = \sup_{x\in[a,b]} |f(x)-g(x)| + \sup_{x\in[a,b]} |f'(x)-g'(x)|\,,$$
$$\sigma(f,g) = \sup_{x\in[a,b]} |f(x)-g(x)|\,.$$

The identical map $(X,\rho)\to(X,\sigma)$ is a bijection and is continuous because $f_n \xrightarrow{\rho} f$ implies $f_n \xrightarrow{\sigma} f$. However, the inverse mapping is not continuous. Indeed, the sequence $f_n(x) = n^{-1}\sin(n^2x)$ converges to the zero function with respect to the metric σ but does not converge with respect to the metric ρ.

Definition 7.15. We say that a (real or complex-valued) function f defined on a metric space $f:(X,\rho)$ is *uniformly continuous* if for any $\varepsilon>0$ there exists $\delta>0$ such that $|f(x)-f(y)|\leqslant\varepsilon$ whenever $\rho(x,y)\leqslant\delta$.

Obviously, a uniformly continuous function is continuous.

Theorem 7.16. *If (X,ρ) is a compact metric space then any continuous function f on (X,ρ) is uniformly continuous.*

Proof. Let $\varepsilon>0$. Since f is continuous, for every point $x\in X$ there exists $\delta_x>0$ such that

$$|f(y)-f(x)|\leqslant\varepsilon/2 \quad\text{whenever}\quad \rho(y,x)\leqslant\delta_x. \tag{7.1}$$

Let $J_x = B_{\delta_x/2}(x)$. Since $x\in J_x$, the collection of open balls $\{J_x\}_{x\in X}$, is an open cover of X. Since X is compact, it has a finite subcover, that is, there exists a finite collection of points $x_1,x_2,\dots,x_k$ such that $X=\cup_{n=1}^k J_{x_k}$. Denote $\delta=\frac{1}{2}\min\{\delta_{x_1},\dots\delta_{x_k}\}$. Since the number of points x_n is finite, we have $\delta>0$.

Let $x,y\in X$ and $\rho(x,y)\leqslant\delta$. Since $X=\cup_{n=1}^k J_{x_k}$, there exists n such that $x\in J_{x_n}$, that is, $\rho(x,x_n)\leqslant\delta_{x_n}/2$. By the triangle inequality

$$\rho(y,x_n)\leqslant\rho(x,x_n)+\rho(x,y)\leqslant\delta_{x_n}/2+\delta\leqslant\delta_{x_n}$$

and, in view of (7.1), $|f(y)-f(x)|\leqslant|f(y)-f(x_n)|+|f(x_n)-f(x)|\leqslant\varepsilon/2+\varepsilon/2=\varepsilon$. Thus we have proved that for any $\varepsilon>0$ there exists $\delta>0$ such that $|f(x)-f(y)|\leqslant\varepsilon$ whenever $\rho(x,y)\leqslant\delta$. $\square$

Further on in this section, for the sake of simplicity, we shall deal only with real-valued functions. Theorems 7.18 and 7.20 can be easily generalized to complex-valued functions by considering their real and imaginary parts separately.

Definition 7.17. If (K,ρ) is a metric space then $C(K)$ denotes the linear space of continuous functions $f:K\to\mathbb{R}$ provided with the norm $\|f\|=\sup_{x\in K}|f(x)|\,.$

Theorem 7.18 (Weierstrass' approximation theorem). *Let $\mathcal{P}$ be the set of polynomials in $C[a,b]$, where $[a,b]$ is a finite closed interval. Then $\overline{\mathcal{P}}=C[a,b]$.*

Proof. The theorem follows from Theorem 7.20 (see below).

There are many equivalent ways to state the above theorem. For example:

(1) Any continuous function on a closed bounded interval can be uniformly approximated by polynomials.
(2) Let f be continuous on $[a,b]$. Given any $\varepsilon>0$ there exists a polynomial p such that $|p(x)-f(x)|<\varepsilon$ for all $x\in[a,b]$.
(3) The polynomials are dense in $C[a,b]$.

Clearly, if $f, g \in C(K)$ then $fg \in C(K)$. A linear space with this property is called an algebra.

Definition 7.19. A set of functions $\mathcal{P} \in C(K)$ is said to be a subalgebra of $C(K)$ if $\mathcal{P}$ is a linear space and $fg \in \mathcal{P}$ whenever $f, g \in \mathcal{P}$.

Theorem 7.20 (Stone–Weierstrass). *Let K be a compact metric space and $\mathcal{P}$ be a subalgebra of $C(K)$. If*

1. *$\mathcal{P}$ contains the constant functions and*
2. *for every pair of points $x, y \in K$ there exists a function $f \in \mathcal{P}$ such that $f(x) \neq f(y)$ (in other words, $\mathcal{P}$ separates the points of K)*

then $\overline{\mathcal{P}} = C(K)$.

Example 7.21. Finite sums of the form

$$c + \sum_{i=1}^{k} a_i \sin(ix) + \sum_{j=1}^{l} b_j \cos(jx)\,,$$

where c, a_i, b_j are some constants, are called trigonometric polynomials. The Stone–Weierstrass theorem implies that the trigonometric polynomials are dense in $C[a, b]$ for any closed interval $[a, b]$, provided that $b - a < 2\pi$. Indeed,

(1) from the equalities

$$\begin{aligned} 2\sin(nx)\cos(mx) &= \sin((n+m)x) + \sin((n-m)x)\,,\\ 2\sin(nx)\sin(mx) &= \cos((n-m)x) - \cos((n+m)x)\,,\\ 2\cos(nx)\cos(mx) &= \cos((n-m)x) + \cos((n+m)x) \end{aligned}$$

it follows that the set of trigonometric polynomials is a subalgebra;

(2) if $\sin(nx) = \sin(ny)$ and $\cos(nx) = \cos(ny)$ for all n then the equalities

$$\begin{aligned} 0 = \sin(nx) - \sin(ny) &= 2\cos\frac{n(x+y)}{2}\,\sin\frac{n(x-y)}{2}\,,\\ 0 = \cos(ny) - \cos(nx) &= 2\sin\frac{n(x+y)}{2}\,\sin\frac{n(x-y)}{2} \end{aligned}$$

imply that $\sin\frac{n(x-y)}{2} = 0$ for all n. This is only possible if $\frac{(x-y)}{2} = k\pi$ for some integer k, which implies that $x = y$ (since $b - a < 2\pi$).

Proof of Theorem 7.20. The proof of Theorem 7.20 proceeds in a number of steps.

Step 1. If $b > 0$ and $\varepsilon > 0$ then the function $f(y) = \sqrt{y+\varepsilon}$ can be uniformly approximated by polynomials on the interval $[0, b]$.

Proof 1. The function $\sqrt{z}$ is analytic on the open half-line $(0, +\infty)$. Therefore its Taylor series $\sum_{k=1}^{\infty} c_k\,(z-c)^k$ converges to $\sqrt{z}$ uniformly on every closed interval $[c-L, c+L] \subset (0, +\infty)$ (it is a standard result from complex analysis). Taking $c = \frac{b}{2}+\varepsilon$, $L = \frac{b}{2}$ and $z = y+\varepsilon$ we see that the sequence of polynomials $\sum_{k=1}^{n} c_k\,(y-\frac{b}{2})^k$ converges to $\sqrt{y+\varepsilon}$ uniformly on $[0, b]$ as $n \to \infty$.

Step 2. Given $a > 0$ there exists a sequence of polynomials $P_n(t)$ converging uniformly to the function $f(t) = |t|$ on $[-a, a]$.

Proof 2. Let $f_n(t) = \sqrt{t^2 + n^{-1}}$. Then

$$\sup_{t\in[-a,a]} |f(t) - f_n(t)| = \sup_{t\in[-a,a]} \left(\sqrt{t^2 + n^{-1}} - |t|\right) \leqslant n^{-1/2}.$$

By the above, the function $\sqrt{y + n^{-1}}$ can be uniformly approximated by polynomials on the interval $[0, a^2]$. Let us choose a polynomial $Q_n(y)$ such that

$$\sup_{y\in[0,a^2]} |\, Q_n(y) - \sqrt{y + n^{-1}}\, | \leqslant n^{-1}.$$

Then

$$\sup_{t\in[-a,a]} |f(t) - Q_n(t^2)| \leqslant \sup_{t\in[-a,a]} |f(t) - f_n(t)| + \sup_{t\in[-a,a]} |f_n(t) - Q_n(t^2)|$$
$$= \sup_{t\in[-a,a]} |f(t) - f_n(t)| + \sup_{y\in[0,a^2]} |Q_n(y) - \sqrt{y + n^{-1}}| \leqslant n^{-1} + n^{-1/2} \to 0$$

as $n \to \infty$. Thus, we can take $P_n(t) := Q_n(t^2)$. $\square$

Step 3. If $\varphi \in \overline{\mathcal{P}}$ then $|\varphi| \in \overline{\mathcal{P}}$.

Proof 3. First note that $\overline{\mathcal{P}}$ is an algebra. For this we need to show that sums, products and scalar multiples of elements of $\overline{\mathcal{P}}$ are in $\overline{\mathcal{P}}$. If $f, g \in \overline{\mathcal{P}}$ then we have $p_n, q_n \in \mathcal{P}$ such that $p_n \to f$ and $q_n \to g$ in $C(K)$. But then since $\mathcal{P}$ is an algebra, $p_n q_n \in \mathcal{P}$ and from the relation

$$\|p_n q_n - fg\| \leqslant \|p_n\| \, \|q_n - g\| + \|p_n - f\| \, \|g\|$$

it is clear that $p_n q_n \to fg$, showing that $fg \in \overline{\mathcal{P}}$. Similar (even easier) arguments show that $f + g \in \overline{\mathcal{P}}$ and $\lambda f \in \overline{\mathcal{P}}$.

Now suppose $\varphi \in \overline{\mathcal{P}}$ and let $a = \|\varphi\| = \sup_{x\in K} |\varphi(x)|$. For $x \in K$, define $P_{\varphi,n}(x) = P_n[\varphi(x)]$, where P_n are the same polynomials as in Step 2. Then $P_{\varphi,n} \in \overline{\mathcal{P}}$. Also, if

$$|\, P_n(t) - |t|\, | < \varepsilon, \qquad \forall t \in [-a, a],$$

then

$$|\, P_{\varphi,n}(x) - |\varphi(x)|\, | < \varepsilon, \qquad \forall x \in K.$$

Thus it follows from Step 2 that $P_{\varphi,n} \to |\varphi|$ in $C(K)$ and so $|\varphi| \in \overline{\mathcal{P}}$. $\square$

Step 4. If $\varphi, \psi \in \overline{\mathcal{P}}$ then $\varphi \vee \psi$ and $\varphi \wedge \psi$ defined by

$$(\varphi \vee \psi)(x) = \max\{\varphi(x), \psi(x)\}, \qquad (\varphi \wedge \psi)(x) = \min\{\varphi(x), \psi(x)\}$$

are both in $\overline{\mathcal{P}}$.

Proof 4. This is obvious from Step 3 and the relations

$$(\varphi \vee \psi)(x) = \frac{1}{2}\left[\, \varphi(x) + \psi(x) + |\varphi(x) - \psi(x)|\, \right]$$
$$(\varphi \wedge \psi)(x) = \frac{1}{2}\left[\, \varphi(x) + \psi(x) - |\varphi(x) - \psi(x)|\, \right].$$

$\square$

Step 5. Given any $f \in C(K)$ and any $y \in K$, for any $\varepsilon > 0$ there exists $\varphi_y \in \overline{\mathcal{P}}$ such that $f(y) = \varphi_y(y)$ and $\varphi_y(x) > f(x) - \varepsilon$ for all $x \in K$.

Proof 5. We first find, for each $z \in K$, an element $\psi_{y,z}$ of $\overline{\mathcal{P}}$ such that $\psi_{y,z}(y) = f(y)$ and $\psi_{y,z}(z) = f(z)$. This is easy since there is a function $p \in \mathcal{P}$ such that $p(y) \neq p(z)$. We now solve

$$f(z) = \lambda + \mu\, p(z)\,,$$
$$f(y) = \lambda + \mu\, p(y)$$

for λ and μ and take $\psi_{y,z}(x) = \lambda + \mu\, p(x)$ (using that the constant function λ is in $\mathcal{P}$).

Since f and $\psi_{y,z}$ are continuous, we can find δ_z such that $|\psi_{y,z}(x) - f(x)| < \varepsilon$ whenever $x \in B_{\delta_z}(z)$. Then we have $\psi_{y,z}(x) > f(x) - \varepsilon$ for all $x \in B_{\delta_z}(z)$.

The family $\{B_{\delta_z}(z) : z \in K\}$ covers K and, since K is compact, we can find a finite subcover $\{B_{\delta_i}(z_i) : 1 \leqslant i \leqslant n\}$ (we write δ_i for δ_{z_i}). Now put

$$\varphi_y = \psi_{y,z_1} \vee \psi_{y,z_2} \vee \ldots \vee \psi_{y,z_n} = \max\{\psi_{y,z_1}, \psi_{y,z_2}, \ldots , \psi_{y,z_n}\}\,.$$

Since $\psi_{y,z_i}(y) = f(y)$ for each i it is clear that $\varphi_y(y) = f(y)$. Also, for any $x \in K$ we have $x \in B_{\delta_j}(z_j)$ for some j and so

$$\varphi_y(x) \;\geqslant\; \psi_{y,z_j}(x) \;>\; f(x) - \varepsilon.$$

□

Step 6. Given any $f \in C(K)$, for any $\varepsilon > 0$ there exists $\varphi \in \overline{\mathcal{P}}$ such that $f(x)+\varepsilon > \varphi(x) > f(x) - \varepsilon$ for all $x \in K$.

Proof 6. We use the same idea as in Step 5 but work exclusively with the functions $\{\varphi_y\}$ which all satisfy $\varphi_y(x) > f(x) - \varepsilon$ for all $x \in K$.

Since φ_y is continuous and $\varphi_y(y) = f(y)$, we can find δ_y such that $|\varphi_y(x) - f(x)| < \varepsilon$ whenever $x \in B_{\delta_y}(y)$. Then we have $\varphi_y(x) < f(x) + \varepsilon$ for all $x \in B_{\delta_y}(y)$.

The family $\{B_{\delta_y}(y) : y \in K\}$ covers K and, since K is compact, we can find a finite subcover $\{B_{\delta_i}(y_i) : 1 \leqslant i \leqslant m\}$ (we write δ_i for δ_{y_i}). Now put

$$\varphi = \varphi_{y_1} \wedge \varphi_{y_1} \wedge \ldots \wedge \varphi_{y_1} = \min\{\varphi_{y_1}, \varphi_{y_2}, \ldots , \varphi_{y_n}\}\,.$$

Clearly $\varphi(x) > f(x) - \varepsilon$ for all $x \in K$ since it is the minimum of a finite set of functions each having this property. Also for any $x \in K$ we have $x \in B_{\delta_j}(y_j)$ for some j and so

$$\varphi(x) \;\leqslant\; \varphi_{y_j}(x) \;<\; f(x) + \varepsilon\,.$$

Since ε can be chosen arbitrarily small, this completes the proof of the Stone-Weierstrass Theorem. □

8. Linear continuous functionals and seminorms

1. Locally convex linear spaces.

Let X be a linear space. We shall be assuming, for the sake of definiteness, that X is a real vector space, in other words, $\lambda x \in X$ and $x_1 + x_2 \in X$ whenever $x, x_1, x_2 \in X$ and $\lambda \in \mathbb{R}$. Similar results remain valid for complex vectors space (where we take $\lambda \in \mathbb{C}$).

Definition 8.1. A function $p : X \to \mathbb{R}$ is said to be a seminorm on X if

(1) $p(x) \geqslant 0$ for all $x \in X$,
(2) $p(\lambda x) = |\lambda|\, p(x)$ for all $x \in X$ and $\lambda \in \mathbb{R}$,
(3) $p(x_1 + x_2) \leqslant p(x_1) + p(x_2)$ for all $x_1, x_2 \in X$ (it is called the *triangle inequality*).

Recall that a norm $\|\cdot\|$ on X satisfies all the same conditions and, in addition, vanishes only on the zero vector. If p is a seminorm then, in view of (2), we have $p(0) = 0$ but we do NOT assume that $p(x) = 0$ only if $x = 0$.

Consider a family of seminorms $P = \{p_\theta\}_{\theta\in\Theta}$ defined on the same space X; here Θ is an arbitrary non-empty index set. If

(S) for every $x \neq 0$ there exists $p_\theta \in P$ such that $p_\theta(x) \neq 0$

then one can define convergence in the space X as follows.

Definition 8.2. Let the condition (S) be fulfilled. We say that a sequence $\{x_n\}$ converges to x in X if $p_\theta(x - x_n) \to 0$ as $n \to \infty$ for all $p_\theta \in P$.

Remark. If P includes only one seminorm p and p is a norm then Definition 8.2 coincides with the usual definition of convergence in a normed linear space. If the condition S is not fulfilled then Definition 8.2 does not make much sense. Indeed, if $x \neq 0$ but $p_\theta(x) = 0$ for all $p_\theta \in P$ then the sequence $x_n := n\,x$ converges to the zero vector which is nonsense.

Obviously, the notion of convergence in X does not change if we add to the family P a new seminorm p such that

$$p(x) \;\leqslant\; C\, \max\{p_{\theta_1}(x), p_{\theta_2}(x), \dots, p_{\theta_m}(x)\}\,, \qquad \forall x \in X\,, \tag{8.1}$$

where C is a positive constant and $p_{\theta_1}, p_{\theta_2}, \dots, p_{\theta_m}$ is a **finite** family of seminorms $p_{\theta_j} \in P$. Let $P_{\max}$ be the family of seminorms which is obtained by adding to P all possible seminorms p on X satisfying (8.1). Clearly, $P \subset P_{\max}$ and a seminorm p belongs to $P_{\max}$ whenever p satisfies (8.1) with some $p_{\theta_j} \in P_{\max}$.

Example 8.3. Consider the linear space X of real-valued functions f on a non-empty set Θ. If we define the seminorms p_θ by $p(f) := |f(\theta)|$ then convergence in X is equivalent to pointwise convergence. A seminorm p belongs to $P_{\max}$ if and only if there exist a constant $C > 0$ and a finite family of points $\theta_1, \theta_2, \dots, \theta_n \in \Theta$ such that $p(f) \;\leqslant\; C\, \max\{|f(\theta_1)|, |f(\theta_2)|, \dots, |f(\theta_n)|\}$ for all $f \in X$.

Theorem 8.4. *If P satisfies* (S) *and is countable then there exists a metric on X such that $x_n \to x$ if and only if $\rho(x_n, x) \to 0$.*

Proof. If $P = \{p_1, p_2, \dots\}$ then we can take

$$\rho(x, y) := \sum_{k=1}^{\infty} 2^{-k}\, \frac{p_k(x-y)}{1 + p_k(x-y)}\,. \tag{8.2}$$

Indeed this function is a metric and $\rho(x_n, x) \to 0$ if and only if $p_k(x - x_n) \to 0$ for all k. $\square$

Given a family of seminorms P satisfying (S), one can define open and closed subsets of X as follows.

Definition 8.5. Consider the class of subsets of X which consists of

(1) all inverse images $p^{-1}(D)$, where $p \in P_{\max}$ and D is an open subset of $\mathbb{R}$;
(2) all sets of the form $\{x_0 + p^{-1}(D)\}$ with $x_0 \in X$, that is, all sets obtained from the inverse images introduced in (1) by adding a fixed vector $x_0 \in X$ to each element;
(3) all possible unions of the sets introduced in (2).

These sets are said to be open. The closed sets are defined as complements of open sets.

Definition 8.6. The pair (X, P), where X is linear space and P is a family of seminorms satisfying the condition (S), is called a *locally convex space*. The open and closed sets and convergence in the locally convex space are defined as above.

Remark. Every normed linear space is a locally convex space. A general metric space (X, ρ) may not be locally convex even if X is a linear space.

If $p \in P_{\max}$ and $x \in X$, let us denote

$$B_{r,p}(x) \;:=\; \{y \in X : p(x-y) < r\}\,.$$

Obviously, $B_{r,p}(x) = B_{1,p'}(x)$ where $p'(x) := r^{-1}p(x) \in P_{\max}$.

Lemma 8.7. *A set $\Omega \subset X$ is open if and only if for every $x \in \Omega$ there exists a seminorm $p \in P_{\max}$ such that $B_{1,p}(x) \subset \Omega$.*

Proof. Assume first that Ω is open and $x \in \Omega$. By Definition 8.5 (3), Ω is the union of a collection of sets of the form $\{x_0 + p^{-1}(D)\}$. Therefore there exists $x_0 \in X$ and an open set $D \subset \mathbb{R}$ such that

$$x \;\in\; \{x_0 + p^{-1}(D)\} \;\subset\; \Omega\,.$$

Denote $p(x - x_0) := a$. The inclusion $x \in \{x_0 + p^{-1}(D)\}$ means that $a \in D$. Since D is open, there exists a constant $r > 0$ such that $(a - r, a + r) \subset D$. If $y \in B_{r,p}(x)$ then, by the triangle inequality,

$$|p(y - x_0) - a| = |p(y - x_0) - p(x - x_0)| \leqslant p(y - x) < r\,.$$

Therefore $p(y - x_0) \in (a - r, a + r) \subset D$, that is, $y \in \{x_0 + p^{-1}(D)\} \subset \Omega$. This proves that $B_{r,p}(x) \subset \Omega$ or, equivalently, $B_{1,p'}(x) \subset \Omega$ where $p'(x) := r^{-1}p(x)$.

On the other hand, we have $x \in B_{1,p}(x) = \{x + p^{-1}(-1, 1)\}$. If for every $x \in \Omega$ there exists $p_x \in P_{\max}$ such that $\{x + p_x^{-1}(-1, 1)\} \subset \Omega$ then $\bigcup_{x\in\Omega}\{x + p_x^{-1}(-1, 1)\} = \Omega$ and Ω is open by Definition 8.5 (3). □

Now we can give another definition of convergence in terms of open sets.

Definition 8.2′. We say that $x_n \to x$ if for every open set Ω containing x there exists n_Ω such that $x_n \in \Omega$ for all $n > n_\Omega$.

One can easily prove that Definition 8.2′ is equivalent to Definition 8.2.

Remark. If there exists a metric on X which generates the same open set as the collection of seminorms P then the locally convex space (X, P) is called *metrizable*. Lemma 8.7 implies that, under conditions of Theorem 8.4, a set $\Omega \in X$ is open if and only if it is open with respect to the metric (8.2), that is, the space (X, P) is metrizable. Note that the standard definitions of closed and compact sets, continuous functions and so on, which involve convergent sequences (see CM321A Lecture Notes) do not work in a locally convex space if it is not metrizable. In particular, the fact that $q(x_n) \to q(x)$ whenever $x_n \to x$ does NOT necessarily imply that the function q is continuous in the sense of Definition 8.9 (see below).

Example 8.8. If X is the locally convex space introduced in Example 8.3 and Ω is a subset of X then the minimal closed set which contains Ω does NOT necessarily coincide with the set obtained by adding to Ω the limits of all convergent sequences $\{f_n\} \subset \Omega$. Indeed, assume that Θ is not countable and consider the set Ω of all functions f on Θ each of which takes the value 1 on a finite subset of Ω and is equal to 0 outside this subset. Let f_0 be the function identically equal to 1 on Θ. Then a sequence $\{f_n\} \subset \Omega$ cannot converge to f_0 because all the function f_n vanish at some point $\theta \in \Theta$. On the other hand, every open 'ball' $B_{r,p}(f_0)$ contains at least one element of Ω. This implies that f_0 belongs to the intersection of all closed subsets which contain Ω.

Definition 8.9. Let (X, P) and (Y, Q) be locally convex spaces. We say that a map $T : X \to Y$ is continuous if the inverse image of every open subset of Y is open in X. In particular, a function $q : X \to \mathbb{R}$ is said to be continuous if the inverse image of every open subset of $\mathbb{R}$ is open in X.

Note that all seminorms $p \in P_{\max}$ are continuous. Indeed, the inverse images $p^{-1}(D)$ of open sets $D \in \mathbb{R}$ are open by Definition 8.5 (1).

Theorem 8.10. *A linear map $T : (X, P) \to (Y, Q)$ is continuous if and only if for every $q \in Q_{\max}$ there exists $p \in P_{\max}$ such that $q(Tx) \leqslant p(x)$ for all $x \in X$.*

Proof. Assume first that T is continuous. Consider the inverse image $T^{-1}(B_{1,q}(0))$ of the open set $B_{1,q}(0) = q^{-1}(-1, 1) \subset Y$. By Lemma 8.7 there exists a seminorm $p \in P_{\max}$ such that $B_{1,p}(0) \subset T^{-1}(B_{1,q}(0))$, that is, $q(Tx) < 1$ whenever $p(x) < 1$. This implies that $q(Tx) \leqslant 2p(x)$ for all $x \in X$ such that $p(x) = 1/2$. Since $q(T\lambda x) = \lambda q(Tx)$ and $p(\lambda x) = \lambda p(x)$ for all $\lambda > 0$, the same estimate holds true for all $x \in X$. Therefore we have $q(Tx) \leqslant p'(x)$ for $p'(x) := 2p(x) \in P_{\max}$.

Assume now that for every $q \in Q_{\max}$ there exists $p \in P_{\max}$ such that $q(Tx) \leqslant p(x)$. Let $\Omega \subset Y$ be an open set and $x \in T^{-1}(\Omega)$ be an arbitrary element of $T^{-1}(\Omega)$. Then $Tx \in \Omega$ and, by Lemma 8.7, there exists a seminorm $q \in Q_{\max}$ such that $B_{1,q}(Tx) \subset \Omega$. Let p be the corresponding seminorm on X. If $y \in B_{1,p}(x)$ then $q(Ty - Tx) \leqslant p(y - x) < 1$. This implies that $Ty \in B_{1,q}(Tx) \subset \Omega$ and, consequently, $y \in T^{-1}(\Omega)$. Therefore $B_{1,p}(x) \subset T^{-1}(\Omega)$ and, by Lemma 8.7, the set $T^{-1}(\Omega)$ is open. $\square$

Definition 8.11. A linear map $q : X \to \mathbb{R}$ is said to be a *linear functional* on X.

Let X^* be the space of linear continuous functionals on X. We shall denote elements of X^* by x^* and the value of the functional $x^* \in X^*$ on the vector $x \in X$ by $\langle x, x^* \rangle$. If x^* is a linear functional on X then the function p_{x^*} defined by $p_{x^*}(x) := |\langle x, x^* \rangle|$ is a seminorm on X as it satisfies the conditions (1)–(3) of Definition 8.1. By Theorem 8.10, a linear functional x^* is continuous if and only if $p_{x^*} \in P_{\max}$. This immediately implies that X^* is a linear space.

The following result (given without proof) is the celebrated Hahn–Banach theorem.

Theorem 8.12. *Let $p \in P_{\max}$ and Y be an arbitrary subspace of X. Then any linear functional x^* on Y such that $|\langle x, x^* \rangle| \leqslant p(x)$ for all $x \in Y$ can be extended to a functional on the whole space X which satisfies the same estimate for all $x \in X$.*

2. Linear continuous functionals on a Hilbert space. In this subsection we shall consider a Hilbert space H, that is, a complete inner product space H with an inner product $(\cdot,\cdot)$ and the norm defined by $\|x\| := \sqrt{(x,x)}$.

The following theorem is easily seen to be false in some normed linear spaces.

Theorem 8.13. *Let H be a Hilbert space, H_0 be a closed subspace of H and x be an arbitrary element of H. Then there exists a unique vector $y_0 \in H_0$ such that $\|x-y_0\| \leqslant \|x-y\|$ for all $y \in H_0$.*

Proof. Let $\delta = \inf\{\|x-y\| : y \in H_0\}$; we seek $y_0 \in H_0$ with $\delta = \|x-y_0\|$. Choose any sequence $y_n \in H_0$ such that $\|x-y_n\| \to \delta$ We wish to replace the distance-approximating sequence y_n by a distance-achieving vector y. We will show that the sequence y_n. converges to a suitable vector y_0; to this end it will suffice to show that y_n is a Cauchy sequence. If m and n are large, we know that $\|x-y_m\|$ and $\|x-y_n\|$ are near δ; we need to show that y_m and y_n are near each other.

Consider $y, z \in H_0$. We are interested in estimating $\|z-y\|$ in terms of $\|x-y\|$ and $\|z-x\|$. This suggests looking at the equation $z-y = (x-y)-(x-z)$. By the parallelogram law,

$$\|(x-y)-(x-z)\|^2 + \|(x-y)+(x-z)\|^2 = 2\|x-y\|^2 + 2\|z-x\|^2,$$

thus

$$\|z-y\|^2 = 2\|x-y\|^2 + 2\|z-x\|^2 - \|2x-y-z\|^2.$$

Since $z+y \in H_0$, we have $\|2x-y-z\|^2 \geqslant 4\delta^2$. Therefore

$$\|z-y\|^2 \leqslant 2\|x-y\|^2 + 2\|z-x\|^2 - 4\delta^2. \tag{8.3}$$

In particular,

$$\|y_m-y_n\|^2 \leqslant 2\|x-y_n\|^2 + 2\|y_m-x\|^2 - 4\delta^2.$$

Since the right side tends to 0 as $m,n \to \infty$, we have $\|y_m-y_n\| \to 0$, which means that y_n is a Cauchy sequence. Since H is complete, the sequence y_n converges to a limit y_0, for which we have $\|x-y_0\| = \delta$. This proves existence of the minimizing vector.

To see that y_0, is unique, suppose that $z_0 \in H_0$ also satisfies $\|x-z_0\| = \delta$. Then, by (8.3), $\|z-y\|^2 \leqslant 0$, which implies $z_0 = y_0$. □

If $H_0 \subset H$, denote by $H_0^\perp$ the set of vectors $z \in H$ such that $(y,z) = 0$ for all $y \in H_0$. The set $H_0^\perp$ is called the *annihilator* of H_0. Obviously, H_0 is a linear subspace of H.

Theorem 8.14. *Let H_0 be a closed linear subspace of a Hilbert space H. Then for each $x \in H$ there exists a unique representation $x = y+z$ with $y \in H_0$ and $z \in H_0^\perp$.*

Proof. Uniqueness is easy: if $x = y_1+z_1 = y_2+z_2$ with $y_1, y_2 \in H_0$ and $z_1, z_2 \in H_0^\perp$ then the vector $y_1-y_2 = z_2-z_1$ belongs both to H_0 and to $H_0^\perp$, which implies that $\|y_1-y_2\| = 0$.

In order to prove existence, let us consider the vector $y_0 \in H_0$ such that $\|x-y_0\| \leqslant \|x-\widetilde{y}\|$ for all $\widetilde{y} \in H_0$ (Theorem 8.13). Set $z = x-y_0$; it will suffice to show that $z \in H_0^\perp$. Given $y \in H_0$, we must show that $(z,y) = 0$; we can suppose $\|y\| = 1$. Direct calculation shows that

$$\|z-(z,y)y\|^2 = \|z\|^2 - |(z,y)|^2\,.$$

Since $z - (z,y)y = x - (y_0 + (z,y)y)$ and $\widetilde{y} := (y_0 + (z,y)y) \in H_0$, it follows from the definition of y_0 that

$$\|z\|^2 = \|x - y_0\|^2 \leqslant \|x - \widetilde{y}\|^2 = \|z - (z,y)y\|^2 = \|z\|^2 - |(z,y)|^2.$$

Thus $(z,y) = 0$. $\square$

Lemma 8.15. *If H_0 is a closed linear subspace of H then $H_0^{\perp\perp} = H_0$.*

proof. If $y \in H_0$ then $(z,y) = 0$ for all $z \in H_0^{\perp}$, which implies that $y \in H_0^{\perp\perp}$. On the other hand, if $x \in H_0^{\perp\perp}$ then , by Theorem 8.14 we have $x = y + z$ with $y \in H_0$ and $z \in H_0^{\perp}$. Since $y \in H_0^{\perp\perp}$, we also have $z = x - y \in H_0^{\perp\perp}$. Therefore $(z,z) = 0$, that is, $z = 0$. This implies that $x \in H_0$. $\square$

The following result is often called the Riesz representation theorem or Riesz lemma.

Theorem 8.16. *If $x' \in H$ then $x \to (x,x')$ is a linear continuous functional on H. Conversely, for every linear continuous functional x^* on a Hilbert space H there exists a unique vector $x' \in H$ such that $\langle x, x^* \rangle = (x,x')$ for all $x \in H$.*

Proof. Since $|(x,x')| \leqslant c\,\|x\|$ with $c = \|x'\|$, the first statement follows from Theorem 8.10.

Let us consider an arbitrary linear continuous functional x^* on H. If $\langle x, x^* \rangle = 0$ for all $x \in H$ then we can take $x' = 0$. If x^* is not identically equal to zero then the set $H_0 = \{x \in H : \langle x, x^* \rangle = 0\}$ is a linear subspace of H and $H_0 \neq H$. The subspace H_0 is closed because it is the inverse image of the closed subset of the real line which consists of one point 0. By Theorem 8.14 there exists a nonzero $z \in H_0^{\perp}$; we can suppose that $\langle z, x^* \rangle = 1$ (otherwise we multiply z by a suitable constant). Then $x - \langle x, x^* \rangle z \in H_0$ for every $x \in H$. This implies that $(x - \langle x, x^* \rangle z, z) = 0$, that is, $(x,z) = \langle x, x^* \rangle \|z\|^2$. Thus, we obtain $\langle x, x^* \rangle = (x,x')$ with $x' = \|z\|^{-2} z$. If $\langle x, x^* \rangle = (x, \widetilde{x}')$ for some other vector $\widetilde{x}'$ then $(x, x' - \widetilde{x}') = 0$ for all $x \in H$. Taking $x = x' - \widetilde{x}'$ we see that $x' = \widetilde{x}'$. This proves the uniqueness of x'. $\square$

9. Tempered distributions

Definition 9.1. Denote by $\mathcal{S}(\mathbb{R})$ the space of infinitely differentiable functions φ on $\mathbb{R}$ such that

$$p_{m,n}(\varphi) \ := \ \sup_{x \in \mathbb{R}1} |x^m \, \varphi^{(n)}(x)| \ < \ \infty$$

for all $n, m = 0, 1, \dots$, where $\varphi^{(n)}$ denotes the nth derivative of φ and $\varphi^{(0)} := \varphi$.

If P is the family of seminorms $\{p_{m,n}\}_{m,n=0,1,\dots}$ then $\{S(\mathbb{R}1), P\}$ is a locally convex space which is called the space of test functions.

A map $f : \mathcal{S}(\mathbb{R}1) \to \mathbb{C}$ is said to be a functional on $\mathcal{S}(\mathbb{R}1)$. The value of functional f on the function $\varphi \in \mathcal{S}(\mathbb{R}1)$ is denoted by $\langle f, \varphi \rangle$. We say that the functional f is linear if

$$\langle f, \alpha\varphi + \beta\psi \rangle \ = \ \alpha \langle f, \varphi \rangle \ + \ \beta \langle f, \psi \rangle \, , \qquad \forall \varphi, \psi \in \mathcal{S}(\mathbb{R}1) \, , \quad \forall\, \alpha, \beta \in \mathbb{C} \, .$$

Definition 9.2. Linear continuous functionals on $\mathcal{S}(\mathbb{R}1)$ are said to be the (tempered) distributions.

The distributions form a linear space. If $\varphi, \psi \in \mathcal{S}(\mathbb{R})$ and $\alpha, \beta \in \mathbb{C}$, we define the distribution $\alpha f + \beta g$ by

$$\langle \alpha f + \beta g, \varphi \rangle \;=\; \alpha \langle f, \varphi \rangle \;+\; \beta \langle g, \varphi \rangle\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R})\,.$$

The linear space of distributions is denoted by $\mathcal{S}'(\mathbb{R})$.

Example 9.3. If $f \in \mathcal{S}(\mathbb{R})$ then the functional on $\mathcal{S}(\mathbb{R})$ defined by $\langle f, \varphi \rangle = \int f\, \varphi\, dx$ is a distribution.

Example 9.4. Let $x \in \mathbb{R}$ be a fixed point. The distribution δ_x defined by

$$\langle \delta_x, \varphi \rangle \;=\; \varphi(x)\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}),$$

is said to be the δ-function at x. The δ-function is one of the simplest distributions; its value $\langle \delta_x, \varphi \rangle$ depends only on the value of φ at one fixed point x.

One can do with the distributions almost all the same things as with the functions from $\mathcal{S}(\mathbb{R})$. The basic idea is as follows. Assume that we are going to extend a linear operator T in the space $\mathcal{S}(\mathbb{R})$ to the space $\mathcal{S}'(\mathbb{R})$. First, we take $\psi \in \mathcal{S}(\mathbb{R})$ and write

$$\langle T\psi, \varphi \rangle \;=\; \int T\psi\, \varphi\, dx\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R})$$

(here we consider the function $T\psi$ as a distribution). Then we try to find a linear operator T' in $\mathcal{S}(\mathbb{R})$ such that

$$\langle T\psi, \varphi \rangle \;=\; \int \psi\, T'\varphi\, dx\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R})\,.$$

Finally, for $f \in \mathcal{S}(\mathbb{R}1)$ we define Tf by

$$\langle Tf, \varphi \rangle \;=\; \langle f, T'\varphi \rangle\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}1)\,. \tag{9.1}$$

Obviously, (9.1) defines a linear functional Tf on $\mathcal{S}(\mathbb{R}1)$. If T' is continuous in $\mathcal{S}(\mathbb{R}1)$ then $\varphi_k \to \varphi$ implies $\langle Tf, \varphi_k \rangle \to \langle Tf, \varphi \rangle$. In this case the functional Tf is continuous, so $Tf \in \mathcal{S}'(\mathbb{R}1)$.

Definition 9.5. Let h be an infinitely smooth function on $\mathbb{R}1$ such that

$$|d^k h/dx^k| \;\leqslant\; c_k\,(1 + x^2)^{m_k}\,, \qquad \forall\, k = 0, 1, \dots,$$

with some constants $c_k > 0$ and $m_k > 0$. Then $\varphi \to h\,\varphi$ is a continuous operator in $\mathcal{S}(\mathbb{R}1)$, and for $f \in \mathcal{S}'(\mathbb{R}1)$ we define hf by

$$\langle hf, \varphi \rangle \;=\; \langle f, h\,\varphi \rangle\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R})\,.$$

Definition 9.6. Differentiation is a continuous operator in $\mathcal{S}(\mathbb{R})$, so for $f \in \mathcal{S}'(\mathbb{R})$ we define f' by $\langle f', \varphi \rangle := -\langle f, \varphi' \rangle\,,\ \varphi \in \mathcal{S}(\mathbb{R})\,.$

Definitions 9.5 and 9.6 allow one to consider differential operators on the space of distributions. It is very important in applications as many differential equations do not have classical solutions but have solutions in the class $\mathcal{S}'(\mathbb{R})$ (so-called generalised solutions).

Example 9.7. Let

$$f(x) \;=\; \begin{cases} 0, & x < 0, \\ 1, & x \geqslant 0\,. \end{cases}$$

Then for all $\varphi \in \mathcal{S}(\mathbb{R})$ we have

$$-\langle f, \varphi' \rangle \;=\; -\int f\,\varphi'\,dx \;=\; -\int_0^\infty \varphi'\,dx \;=\; \varphi(0)\,.$$

Therefore f' coincides with the δ-function at $x = 0$.

References

CM321A Lecture Notes
http://www.mth.kcl.ac.uk/∼ ysafarov/Lectures/CM321A/

CMMS11 Lecture Notes
http://www.mth.kcl.ac.uk/∼ ysafarov/Lectures/Fourier/