Some basic topics in analysis

Stephen Semmes
Rice University

Preface

These informal notes deal with some possible topics for a second course in analysis. In particular, the reader is assumed to be familiar with metric spaces, sequences and series, and continuous functions. Some topics may be mentioned in a first course, with some review or elaboration here.

Of course, there are many textbooks in analysis, a few of which are mentioned in the bibliography. The aim here is to complement these textbooks, while trying to look ahead a bit to more advanced courses.

Although some basic notions are used frequently throughout the text, there is a fair amout of independence between the various sections and chapters. Thus the reader may wish to focus more on some parts, at least initially.

Some aspects of history related to topics like those considered here may be found in [3, 15, 17, 18, 19, 20, 32, 33, 41, 42, 45, 66, 67, 68, 71, 72, 73, 74, 75, 76, 77, 79, 80, 96, 113, 152, 164], for instance. Some songs related to some topics like those considered here may be found in [118, 119, 120]. Some remarks concerning the clarity of explanations in mathematics may be found in [78].

Contents

Chapter 1

Some mappings, metrics, and norms

1.1 Lipschitz mappings

Let $(X, d_X(\cdot,\cdot))$ and $(Y, d_Y(\cdot,\cdot))$ be metric spaces. A mapping f from X into Y is said to be *Lipschitz* if there is a nonnegative real number C such that

$$d_Y(f(x), f(w)) \le C\, d_X(x, w) \tag{1.1.1}$$

for every $x, w \in X$. It is easy to see that Lipschitz mappings are uniformly continuous. Note that (1.1.1) holds with $C = 0$ if and only if f is constant on X.

Let X be a set, and let $d(\cdot,\cdot)$ and $d'(\cdot,\cdot)$ be metrics on X. Consider the condition that there be a nonnegative real number C such that

$$d'(x, w) \le C\, d(x, w) \tag{1.1.2}$$

for every $x, w \in X$. This is the same as (1.1.1), with $d_X = d$, $Y = X$, $d_Y = d'$, and f taken to be the identity mapping on X.

Let $(X, d(\cdot,\cdot))$ be a metric space, and put

$$\rho(x, w) = \sqrt{d(x, w)} \tag{1.1.3}$$

for every $x, w \in X$. One can check that this defines a metric on X as well. This corresponds to the second part of Exercise 11 at the end of Chapter 2 in [155], when we start with the standard Euclidean metric on the real line.

One can check that the identity mapping on X is uniformly continuous as a mapping from X equipped with $d(\cdot,\cdot)$ into X equipped with $\rho(x, w)$. Similarly, one can check that the identity mapping on X in uniformly continuous as a mapping from X equipped with $\rho(\cdot,\cdot)$ into X equipped with $d(\cdot,\cdot)$.

1.2 Lipschitz conditions on R

Let E be a nonempty subset of the real line $\mathbf{R}$, and let f be a real-valued function on E. Note that f is Lipschitz on E with respect to the standard Euclidean metric on $\mathbf{R}$ and its restriction to E if and only if there is a nonnegative real number C such that

$$|f(x) - f(w)| \le C\,|x - w| \tag{1.2.1}$$

for every $x, w \in E$. Here $|t|$ denotes the usual absolute value of a real number t.

Of course, (1.2.1) holds automatically when $x = w$. If $x \ne w$, then (1.2.1) is the same as saying that

$$\frac{|f(x) - f(w)|}{|x - w|} \le C. \tag{1.2.2}$$

Suppose that $x \in E$ is a limit point of E. The *derivative* of f at x is defined as usual by

$$f'(x) = \lim_{\substack{w \in E \\ w \to x}} \frac{f(w) - f(x)}{w - x}, \tag{1.2.3}$$

when the limit on the right exists. In this case, f is said to be *differentiable* at x, as a function on E. If f is differentiable at x, and (1.2.1) holds for all $w \in E$, or at least when $w \in E$ is sufficiently close to x, then one can check that

$$|f'(x)| \le C. \tag{1.2.4}$$

If f is differentiable at x, then f is continuous at x, as a function on E, by a standard argument. More precisely, if C is a real number such that

$$|f'(x)| < C, \tag{1.2.5}$$

then one can verify that (1.2.1) holds for all $w \in E$ that are sufficiently close to x.

Let a and b be real numbers with $a < b$, and suppose for the moment that E is the corresponding open interval (a, b) in $\mathbf{R}$. We may also allow $a = -\infty$ or $b = +\infty$ here, so that E could be the real line, or an open half-line in $\mathbf{R}$. Suppose that f is differentiable at every point in E, and that there is a nonnegative real number C such that (1.2.4) holds for every $x \in E$. Under these conditions, the mean value theorem implies that (1.2.1) holds for every $x, w \in E$.

Let a and b be real numbers with $a < b$ again, and suppose now that E is the corresponding closed interval $[a, b]$ in $\mathbf{R}$. Suppose that f is continuous on $[a, b]$, and differentiable at every point in (a, b). If there is a nonnegative real number C such that (1.2.4) holds for every $x \in (a, b)$, then the mean-value theorem implies that (1.2.1) holds for every $x, w \in E$.

Of course, there are analogous statements when E is a half-open, half-closed interval in $\mathbf{R}$, or a closed half-line in $\mathbf{R}$.

See [13, 34, 39] for some related perspectives on the mean value theorem. Some additional results related to the mean value theorem can be found in [99, 181]. Some aspects of calculus on the rationals are discussed in [100, 102]. Some more connections between Lipschitz conditions and derivatives will be considered in Chapter 8.

1.3 Norms on $\mathbf{R}^n$

Let n be a positive integer. Remember that $\mathbf{R}^n$ is the space of ordered n-tuples $x = (x_1, \ldots, x_n)$ such that $x_j \in \mathbf{R}$ for each $j = 1, \ldots, n$. Addition can be defined on $\mathbf{R}^n$ coordinatewise, so that

$$x + y = (x_1 + y_1, \ldots, x_n + y_n) \tag{1.3.1}$$

for every $x, y \in \mathbf{R}^n$. Similarly, if $t \in \mathbf{R}$ and $x \in \mathbf{R}^n$, then

$$t\,x = (t\,x_1, \ldots, t\,x_n) \tag{1.3.2}$$

defines another element of $\mathbf{R}^n$. Using these definitions of addition and scalar multiplication, $\mathbf{R}^n$ becomes a vector space over the real numbers. Although we shall not discuss the formal definition of a vector space here, the relevant notions will hopefully be clear in the examples. Note that we shall use 0 to refer to the element of $\mathbf{R}^n$ whose coordinates are equal to the real number 0, which will hopefully also be clear from the context.

A nonnegative real-valued function N on $\mathbf{R}^n$ is said to define a *norm* on $\mathbf{R}^n$ if it satisfies the following three conditions. First, $N(x) = 0$ if and only if $x = 0$. Second,

$$N(t\,x) = |t|\,N(x) \tag{1.3.3}$$

for every $t \in \mathbf{R}$ and $x \in \mathbf{R}^n$. Third,

$$N(x + y) \le N(x) + N(y) \tag{1.3.4}$$

for every $x, y \in \mathbf{R}^n$, which is the *triangle inequality* for norms.

The standard *Euclidean norm* is defined by

$$\|x\|_2 = \Big(\sum_{j=1}^{n} x_j^2\Big)^{1/2} \tag{1.3.5}$$

for every $x \in \mathbf{R}^n$. Of course, this uses the nonnegative square root on the right side of the equation. It is easy to see that this satisfies the first two requirements of a norm mentioned in the preceding paragraph. The triangle inequality is more complicated, and can be obtained from the Cauchy–Schwarz inequality. See Theorem 1.37 on p16 of [155].

One can check directly that

$$\|x\|_1 = \sum_{j=1}^{n} |x_j| \tag{1.3.6}$$

defines a norm on $\mathbf{R}^n$. (Exercise.) Similarly,

$$\|x\|_\infty = \max(|x_1|, \ldots, |x_n|) \tag{1.3.7}$$

defines a norm on $\mathbf{R}^n$, where more precisely the right side is the maximum of $|x_1|, \ldots, |x_n|$. In particular, if $x, y \in \mathbf{R}^n$, then

$$\begin{aligned} \|x+y\|_\infty &= \max(|x_1+y_1|, \ldots, |x_n+y_n|) \\ &\le \max(|x_1|+|y_1|, \ldots, |x_n|+|y_n|) \le \|x\|_\infty + \|y\|_\infty, \end{aligned} \tag{1.3.8}$$

using the triangle inequality for the absolute value function on $\mathbf{R}$ in the second step.

If N is any norm on $\mathbf{R}^n$, then

$$d_N(x, y) = N(x - y) \tag{1.3.9}$$

defines a metric on $\mathbf{R}^n$. Indeed, the first requirement of a norm ensures that (1.3.9) is equal to 0 if and only if $x = y$. The homogeneity condition (1.3.3) implies that (1.3.9) is symmetric in x and y, by taking $t = -1$ in (1.3.3). The triangle inequality for (1.3.9) as a metric on $\mathbf{R}^n$ follows from the triangle inequality (1.3.4) for N as a norm on $\mathbf{R}^n$.

The metric

$$d_2(x, y) = \|x - y\|_2 \tag{1.3.10}$$

associated to the standard Euclidean norm (1.3.5) is the standard *Euclidean metric* on $\mathbf{R}^n$. Let

$$d_1(x, y) = \|x - y\|_1 \tag{1.3.11}$$

and

$$d_\infty(x, y) = \|x - y\|_\infty \tag{1.3.12}$$

be the metrics on $\mathbf{R}^n$ corresponding to the norms (1.3.6) and (1.3.7), respectively. If $n = 1$, then the norms (1.3.5), (1.3.6), and (1.3.7) reduce to the absolute value function on $\mathbf{R}$, and the corresponding metrics are the same as the standard Euclidean metric on $\mathbf{R}$.

If $a_1, \ldots, a_n, b_1, \ldots, b_n$ are nonnegative real numbers, then

$$\sum_{j=1}^n a_j\, b_j \le \Big(\sum_{j=1}^n a_j^2\Big)^{1/2} \Big(\sum_{j=1}^n b_j^2\Big)^{1/2}. \tag{1.3.13}$$

This is a version of the *Cauchy–Schwarz inequality.* This is often formulated a bit differently for arbitrary real or complex numbers, as in Theorem 1.35 on p15 of [155]. This formulation is included in the other one, by restricting one's attention to nonnegative real numbers. The other formulation can also be obtained from this one, by applying (1.3.13) to the absolute values of the given real or complex numbers.

1.4 Norms on $\mathbf{C}^n$

Let $\mathbf{C}$ be the complex plane, as usual, and let n be a positive integer again. As before, $\mathbf{C}^n$ is the set of ordered n-tuples of complex numbers. Addition can be defined on $\mathbf{C}^n$ coordinatewise, as in (1.3.1). One can also multiply an element of $\mathbf{C}^n$ by a complex number coordinatewise, as in (1.3.2). In this way, $\mathbf{C}^n$ becomes a vector space over the complex numbers.

A nonnegative real-valued function N on $\mathbf{C}^n$ is said to be a *norm* on $\mathbf{C}^n$ if it satisfies the same type of conditions as in the previous section. More precisely, the first and third conditions conditions are the same as before. In this situation, the homogeneity condition (1.3.3) should hold for all complex numbers t and elements of $\mathbf{C}^n$, where $|t|$ is the usual absolute value function on $\mathbf{C}$.

The standard *Euclidean norm* is defined on $\mathbf{C}^n$ by

$$\|z\|_2 = \Big(\sum_{j=1}^{n} |z_j|^2\Big)^{1/2}, \tag{1.4.1}$$

where $|z_j|$ is the absolute value of $z_j \in \mathbf{C}$ for each $j = 1, \ldots, n$. The triangle inequality for (1.4.1) can be reduced to the real case, by taking the absolute values of the coordinates of elements of $\mathbf{C}^n$ to get n-tuples of nonnegative real numbers. This argument uses the triangle inequality for the absolute value function on $\mathbf{C}$, which is the same as the $n = 1$ case. The triangle inequality for (1.4.1) on $\mathbf{C}^n$ can also be obtained from the Cauchy–Schwarz inequality, using an argument analogous to the one in the real case. As before, it is easy to verify the other two requirements for (1.4.1) to be a norm on $\mathbf{C}^n$ directly from the definition.

Alternatively, the complex plane can be identified with $\mathbf{R}^2$, using the real and imaginary parts of a complex number. Using this identification, the absolute value of a complex number corresponds to the standard Euclidean norm on $\mathbf{R}^2$. Similarly, $\mathbf{C}^n$ can be identified with $\mathbf{R}^{2n}$, using the real and imaginary parts of the n coordinates of an element of $\mathbf{C}^n$. Using this identification, the standard Euclidean norm (1.4.1) on $\mathbf{C}^n$ corresponds exactly to the standard Euclidean norm on $\mathbf{R}^{2n}$. This permits one to get the triangle inequality for (1.4.1) on $\mathbf{C}^n$ from the triangle inequality for the standard Euclidean norm on $\mathbf{R}^{2n}$, because addition on $\mathbf{C}^n$ corresponds exactly to addition on $\mathbf{R}^{2n}$ with respect to this identification.

As before, one can verify directly that

$$\|z\|_1 = \sum_{j=1}^{n} |z_j| \tag{1.4.2}$$

and

$$\|z\|_\infty = \max(|z_1|, \ldots, |z_n|) \tag{1.4.3}$$

define norms on $\mathbf{C}^n$ as well. Of course, $\mathbf{R}^n$ may be considered as a subset of $\mathbf{C}^n$, because $\mathbf{R}$ is contained in $\mathbf{C}$. The restrictions of (1.4.2) and (1.4.3) to $z \in \mathbf{R}^n$ are the same as the corresponding norms defined on $\mathbf{R}^n$ in the previous section. Similarly, the restriction of (1.4.1) to $z \in \mathbf{R}^n$ is the same as the standard Euclidean norm on $\mathbf{R}^n$.

If N is any norm on $\mathbf{C}^n$, then

$$d_N(z, w) = N(z - w) \tag{1.4.4}$$

defines a metric on $\mathbf{C}^n$, for the same reasons as in the real case. The standard *Euclidean metric* on $\mathbf{C}^n$ is the metric

$$d_2(z,w) = \|z-w\|_2 \tag{1.4.5}$$

associated to the standard Euclidean norm (1.4.1). Similarly, let

$$d_1(z,w) = \|z-w\|_1 \tag{1.4.6}$$

and

$$d_\infty(z,w) = \|z-w\|_\infty \tag{1.4.7}$$

be the metrics on $\mathbf{C}^n$ associated to the norms (1.4.2) and (1.4.3), respectively. If $n=1$, then the norms (1.4.1), (1.4.2), and (1.4.3) reduce to the absolute value function on $\mathbf{C}$, so that the corresponding metrics are the same as the standard Euclidean metric on $\mathbf{C}$.

The restriction of any norm N on $\mathbf{C}^n$ to $\mathbf{R}^n$ defines a norm on $\mathbf{R}^n$. In this case, the restriction of (1.4.4) to $z,w \in \mathbf{R}^n$ is the same as the metric on $\mathbf{R}^n$ associated to the restriction of N to $\mathbf{R}^n$. In particular, the restrictions of (1.4.5), (1.4.6), and (1.4.7) to $z,w \in \mathbf{R}^n$ are the same as the corresponding metrics defined on $\mathbf{R}^n$ in the previous section.

Let p be a positive real number, and put

$$\|z\|_p = \Big(\sum_{j=1}^n |z_j|^p\Big)^{1/p} \tag{1.4.8}$$

for every $z \in \mathbf{C}^n$. It is easy to see that this satisfies that positivity and homogeneity requirements of a norm. If $p \geq 1$, then it is well known that (1.4.8) satisfies the triangle inequality, and hence defines a norm on $\mathbf{C}^n$. This is a version of *Minkowski's inequality* for sums. Of course, (1.4.8) is the same as (1.4.1) when $p=2$, and it is the same as (1.4.2) when $p=1$. If $n=1$, then (1.4.8) reduces to the absolute value function on $\mathbf{C}$. If $0<p<1$ and $n \geq 2$, then one can show that (1.4.8) does not satisfy the triangle inequality, because the corresponding balls in $\mathbf{C}^n$ are not convex. There are analogous statements for the restriction of (1.4.8) to $\mathbf{R}^n$.

1.5 Some basic inequalities

Let n be a positive integer, and let $z \in \mathbf{C}^n$ be given. It is easy to see that

$$\|z\|_\infty \leq \|z\|_2,\ \|z\|_1, \tag{1.5.1}$$

directly from the definitions of these norms in the previous section. Similarly,

$$\|z\|_2^2 = \sum_{j=1}^n |z_j|^2 \leq \|z\|_\infty \sum_{j=1}^n |z_j| = \|z\|_\infty \, \|z\|_1 \leq \|z\|_1^2, \tag{1.5.2}$$

so that

$$\|z\|_2 \le \|z\|_1. \tag{1.5.3}$$

It follows that the corresponding metrics satisfy

$$d_\infty(z,w) \le d_2(z,w) \le d_1(z,w) \tag{1.5.4}$$

for every $z, w \in \mathbf{C}^n$.

In the other direction, we have that

$$\|z\|_2 \le n^{1/2}\,\|z\|_\infty \tag{1.5.5}$$

and

$$\|z\|_1 \le n\,\|z\|_\infty \tag{1.5.6}$$

for every $z \in \mathbf{C}^n$. One can also check that

$$\|z\|_1 \le n^{1/2}\,\|z\|_2 \tag{1.5.7}$$

for every $z \in \mathbf{C}^n$, using the Cauchy-Schwarz inequality (1.3.13). This implies that

$$d_2(z,w) \le n^{1/2}\,d_\infty(z,w), \tag{1.5.8}$$

$$d_1(z,w) \le n\,d_\infty(z,w), \tag{1.5.9}$$

and

$$d_1(z,w) \le n^{1/2}\,d_2(z,w) \tag{1.5.10}$$

for every $z, w \in \mathbf{C}^n$.

Using these simple relationships, we get that $d_1(z,w)$, $d_2(z,w)$, and $d_\infty(z,w)$ have many of the same properties on $\mathbf{C}^n$. They determine the same collections of open sets, closed sets, compact sets, and bounded sets, for instance. They also determine the same limit points of subsets of $\mathbf{C}^n$, convergent sequences in $\mathbf{C}^n$, and Cauchy sequences. Using (1.5.4), it is easy to see that the identity mapping on $\mathbf{C}^n$ is Lipschitz as a mapping from $\mathbf{C}^n$ equipped with $d_1(z,w)$ into $\mathbf{C}^n$ equipped with $d_2(z,w)$, and from $\mathbf{C}^n$ equipped with $d_2(z,w)$ into $\mathbf{C}^n$ equipped with $d_\infty(z,w)$. Similarly, the identity mapping on $\mathbf{C}^n$ is Lipschitz as a mapping from $\mathbf{C}^n$ equipped with $d_1(z,w)$ into $\mathbf{C}^n$ equipped with $d_\infty(z,w)$. We can use (1.5.8) to get that the identity mapping on $\mathbf{C}^n$ is Lipschitz as a mapping from $\mathbf{C}^n$ equipped with $d_\infty(z,w)$ into $\mathbf{C}^n$ equipped with $d_2(z,w)$, and (1.5.10) implies that the identity mapping on $\mathbf{C}^n$ is Lipschitz as a mapping from $\mathbf{C}^n$ equipped with $d_2(z,w)$ into $\mathbf{C}^n$ equipped with $d_1(z,w)$. One can use (1.5.9) to get that the identity mapping on $\mathbf{C}^n$ is Lipschitz as a mapping from $\mathbf{C}^n$ equipped with $d_\infty(z,w)$ into $\mathbf{C}^n$ equipped with $d_1(z,w)$. Of course, there are analogous statements for the restrictions of these metrics to $\mathbf{R}^n$.

If N is any norm on $\mathbf{R}^n$ or $\mathbf{C}^n$, then one can show that N is bounded by a constant times the standard Euclidean norm, or equivalently by a constant times either of the norms $\|\cdot\|_1$, $\|\cdot\|_\infty$. To see this, one can express any element of $\mathbf{R}^n$ or $\mathbf{C}^n$ as a linear combination of the standard basis vectors, to estimate

N in terms of the absolute values of the coodinates of the given vector. One can use this to show that N is continuous as a real-valued function on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, with respect to the standard Euclidean metric. It follows that N attains its minimum on the unit sphere in $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, with respect to the standard Euclidean metric again. Because the infimum is positive, by definition of a norm, one can verify that the standard Euclidean norm is bounded by a constant times N on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate.

1.6 Functions with finite support

Let X be a nonempty set, and let f be a real or complex-valued function on X. The *support* of f is defined to be the subset of X given by

$$\operatorname{supp} f = \{x \in X : f(x) \neq 0\}. \tag{1.6.1}$$

Let $c_{00}(X, \mathbf{R})$ be the space of real-valued functions on X whose support has only finitely many elements, and let $c_{00}(X, \mathbf{C})$ be the space of complex-valued functions on X with finite support. If f and g are real or complex-valued functions on X with finite support, then their sum $f(x) + g(x)$ also defines a real or complex-valued function on X with finite support. Similarly, if f is a real or complex-valued function on X with finite support, and t is a real or complex number, as appropriate, then $t\, f(x)$ has finite support in X. More precisely, $c_{00}(X, \mathbf{R})$ and $c_{00}(X, \mathbf{C})$ are vector spaces over the real and complex numbers, respectively, with respect to pointwise addition and scalar multiplication of functions. These may be considered as linear subspaces of the spaces of all real or complex-valued functions on X, respectively.

Of course, if X has only finitely many elements, then every real or complex-valued function on X automatically has finite support. Let n be a positive integer, and suppose for the moment that

$$X = \{1, \ldots, n\} \tag{1.6.2}$$

is the set of positive integers from 1 to n. In this case, $c_{00}(X, \mathbf{R})$ and $c_{00}(X, \mathbf{C})$ can be identified with $\mathbf{R}^n$ and $\mathbf{C}^n$, respectively. Similarly, if X is the set $\mathbf{Z}_+$ of positive integers, then a real or complex-valued function on X corresponds to an infinite sequence of real or complex numbers. Thus $c_{00}(\mathbf{Z}_+, \mathbf{R})$ and $c_{00}(\mathbf{Z}_+, \mathbf{C})$ can be identified with the spaces of infinite sequences of real or complex numbers for which all but finitely many terms are equal to 0, respectively.

As usual, a nonnegative real-valued function N on $c_{00}(X, \mathbf{R})$ or $c_{00}(X, \mathbf{C})$ is said to be a *norm* if it satisfies the following three conditions. First, $N(f) = 0$ if and only if $f = 0$. Second, if $f \in c_{00}(X, \mathbf{R})$ or $c_{00}(X, \mathbf{C})$ and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then

$$N(t\, f) = |t|\, N(f). \tag{1.6.3}$$

Third,

$$N(f + g) \leq N(f) + N(g) \tag{1.6.4}$$

for every $f, g \in c_{00}(X, \mathbf{R})$ or $c_{00}(X, \mathbf{C})$, as appropriate. In this case,

$$d_N(f, g) = N(f - g) \tag{1.6.5}$$

defines a metric on $c_{00}(X, \mathbf{R})$ or $c_{00}(X, \mathbf{C})$, as appropriate.

Let f be a real or complex-valued function on X. If A is a nonempty finite subset of X, then

$$\sum_{x \in A} f(x) \tag{1.6.6}$$

can be defined as a real or complex number, as appropriate. Suppose that f has finite support in X, and observe that the finite sums (1.6.6) are all the same when $\operatorname{supp} f \subseteq A$. This permits us to define the sum

$$\sum_{x \in X} f(x) \tag{1.6.7}$$

as a real or complex number, as appropriate, as the value of (1.6.6) when A is a nonempty finite subset of X that contains the support of f.

If f is a real or complex-valued function on X with finite support, then

$$\|f\|_1 = \sum_{x \in X} |f(x)| \tag{1.6.8}$$

is defined as a nonnegative real number, as in the previous paragraph. Similarly,

$$\|f\|_2 = \Big(\sum_{x \in X} |f(x)|^2 \Big)^{1/2} \tag{1.6.9}$$

is defined as a nonnegative real number, where the sum on the right is defined as before. We can also put

$$\|f\|_\infty = \max_{x \in X} |f(x)|, \tag{1.6.10}$$

where the maximum of $|f(x)|$ over $x \in X$ is clearly attained in this situation. One can check that these define norms on $c_{00}(X, \mathbf{R})$ and $c_{00}(X, \mathbf{C})$. In particular, the triangle inequality for (1.6.9) reduces to the analogous statement for $\mathbf{R}^n$ and $\mathbf{C}^n$, mentioned in Sections 1.3 and 1.4.

Using these norms, we get metrics

$$d_1(f, g) = \|f - g\|_1, \tag{1.6.11}$$

$$d_2(f, g) = \|f - g\|_2, \tag{1.6.12}$$

and

$$d_\infty(f, g) = \|f - g\|_\infty \tag{1.6.13}$$

on $c_{00}(X, \mathbf{R})$ and $c_{00}(X, \mathbf{C})$. If X is as in (1.6.2) for some positive integer n, then the norms mentioned in the previous paragraph correspond to the analogous norms defined on $\mathbf{R}^n$ and $\mathbf{C}^n$ in Sections 1.3 and 1.4. Similarly, these metrics correspond to the analogous metrics defined earlier on $\mathbf{R}^n$ and $\mathbf{C}^n$ in this case.

If f is a real or complex-valued function on X with finite support, then

$$\|f\|_\infty \le \|f\|_2 \le \|f\|_1. \tag{1.6.14}$$

This follows from (1.5.1) and (1.5.3), since it is enough to look at the finitely many elements of X in the support of f. This implies that

$$d_\infty(f,g) \le d_2(f,g) \le d_1(f,g) \tag{1.6.15}$$

for all real and complex-valued functions f and g on X with finite support.

It follows that the identity mappings on $c_{00}(X,\mathbf{R})$ and $c_{00}(X,\mathbf{C})$ are Lipschitz with respect to $d_2(f,g)$ on the domain and $d_\infty(f,g)$ on the range, as before. Similarly, the identity mappings on $c_{00}(X,\mathbf{R})$ and $c_{00}(X,\mathbf{C})$ are Lipschitz with respect to $d_1(f,g)$ on the domain and $d_2(f,g)$ on the range. The identity mappings on $c_{00}(X,\mathbf{R})$ and $c_{00}(X,\mathbf{C})$ are also Lipschitz with respect to $d_1(f,g)$ on the domain and $d_\infty(f,g)$ on the range.

1.7 Cauchy sequences and completeness

Let $(X, d_X(\cdot,\cdot))$ be a metric space. Remember that a sequence $\{x_j\}_{j=1}^\infty$ of elements of X is said to be a *Cauchy sequence* with respect to $d_X(\cdot,\cdot)$ if for every $\epsilon > 0$ there is a positive integer L such that

$$d_X(x_j, x_l) < \epsilon \tag{1.7.1}$$

for every $j, l \ge L$. It is not difficult to verify that

(1.7.2) convergent sequences in X are Cauchy sequences.

If

(1.7.3) every Cauchy sequence in X converges to an element of X,

then X is said to be *complete* with respect to $d_X(\cdot,\cdot)$. It is well known that $\mathbf{R}$ and $\mathbf{C}$ are complete with respect to their standard Euclidean metrics.

Let E be a subset of X, and remember that the restriction of $d_X(x,w)$ to $x, w \in E$ defines a metric on E. If $\{x_j\}_{j=1}^\infty$ is a sequence of elements of E, then it is easy to see that

(1.7.4) $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence as a sequence of elements of E

if and only if

(1.7.5) $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence as a sequence of elements of X,

with respect to $d_X(\cdot,\cdot)$ and its restriction to E. If X is complete with respect to $d_X(\cdot,\cdot)$, and if E is a closed set in X, then

(1.7.6) E is complete with respect to the restriction of $d_X(\cdot,\cdot)$ to E.

More precisely, if $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence of elements of E, then $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence in X, which converges to some $x \in X$, because X is complete. We also have that $x \in E$, because E is a closed set in X, so that $\{x_j\}_{j=1}^\infty$ converges to x in E, with respect to the restriction of $d_X(\cdot,\cdot)$ to E.

Suppose now that E is complete with respect to the restriction of $d_X(\cdot,\cdot)$ to E, and let us check that

$$E \text{ is a closed set in X.} \tag{1.7.7}$$

Let $x \in X$ be a limit point of E, which implies that there is a sequence $\{x_j\}_{j=1}^\infty$ of elements of E that converges to x in X. Note that $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence as a sequence of elements of X, and hence as a sequence of elements of E. Because E is complete, there is an $x' \in E$ such that $\{x_j\}_{j=1}^\infty$ converges to x' with respect to the restriction of $d_X(\cdot,\cdot)$ to E. Of course, $\{x_j\}_{j=1}^\infty$ converges to x' in X as well, so that $x = x'$, and thus $x \in E$.

Let $(Y, d_Y(\cdot,\cdot))$ be another metric space, and let f be a uniformly continuous mapping from X into Y. If $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence of elements of X, then one can check that

$$\{f(x_j)\}_{j=1}^\infty \text{ is a Cauchy sequence in } Y. \tag{1.7.8}$$

This is the first part of Exercise 11 at the end of Chapter 4 in [155]. If Y is complete, then it follows that $\{f(x_j)\}_{j=1}^\infty$ converges in Y. If $\{x_j\}_{j=1}^\infty$ converges to an element x of X, then continuity of f at x implies that $\{f(x_j)\}_{j=1}^\infty$ converges to $f(x)$ in Y.

1.8 Pointwise and uniform convergence

Let X be a set, and let $(Y, d_Y(\cdot,\cdot))$ be a metric space. Also let $\{f_j\}_{j=1}^\infty$ be a sequence of mappings from X into Y, and let f be another mapping from X into Y. We say that $\{f_j\}_{j=1}^\infty$ converges to f *pointwise* on X if for every $x \in X$, $\{f_j(x)\}_{j=1}^\infty$ converges to $f(x)$ in Y. This means that for every $x \in X$ and $\epsilon > 0$ there is a positive integer L such that

$$d_Y(f_j(x), f(x)) < \epsilon \tag{1.8.1}$$

for every $j \ge L$. We say that $\{f_j\}_{j=1}^\infty$ converges to f *uniformly* on X if for every $\epsilon > 0$ there is a positive integer L such that (1.8.1) holds for every $x \in X$ and $j \ge L$. Note that uniform convergence implies pointwise convergence. If X has only finitely many elements, and $\{f_j\}_{j=1}^\infty$ converges to f pointwise on X, then one can check that $\{f_j\}_{j=1}^\infty$ converges to f uniformly on X.

As an example, let us take X to be the closed unit interval $[0,1]$ in the real line, and $Y = \mathbf{R}$ with the standard metric. Put

$$f_j(x) = x^j \tag{1.8.2}$$

for each positive integer j and $0 \le x \le 1$. In this case,

$$\lim_{j \to \infty} f_j(x) = 0 \quad \text{when } 0 \le x < 1 \tag{1.8.3}$$
$$= 1 \quad \text{when } x = 1.$$

However, $\{f_j\}_{j=1}^\infty$ does not converge uniformly on $[0, 1]$, because for each positive integer j we have that x^j is as close to 1 as we want when x is sufficiently close to 1. If r is a positive real number with $r < 1$, then $\{f_j\}_{j=1}^\infty$ does converge to 0 uniformly on $[0, r]$.

Now let (X, d_X) be a metric space, and let (Y, d_Y) be a metric space again too. Also let

(1.8.4) $\{f_j\}_{j=1}^\infty$ be a sequence of mappings from X into Y that converges uniformly to a mapping f from X into Y,

and let $x \in X$ be given. If

(1.8.5) f_j is continuous at x for every $j \ge 1$,

then

(1.8.6) f is continuous at x

as well. To see this, let $\epsilon > 0$ be given. Because $\{f_j\}_{j=1}^\infty$ converges uniformly to f on X, there is an $L \in \mathbf{Z}_+$ such that

$$d_Y(f_j(w), f(w)) < \epsilon/3 \tag{1.8.7}$$

for every $j \ge L$ and $w \in X$. In particular, this holds at x, so that

$$d_Y(f_j(x), f(x)) < \epsilon/3 \tag{1.8.8}$$

for every $j \ge L$. Because f_L is continuous at x, there is a $\delta_L > 0$ such that

$$d_Y(f_L(x), f_L(w)) < \epsilon/3 \tag{1.8.9}$$

for every $w \in X$ with $d_X(x, w) < \delta_L$. Observe that

$$d_Y(f(x), f(w)) \le d_Y(f(x), f_L(x)) + d_Y(f_L(x), f_L(w)) + d_Y(f_L(w), f(w)) \tag{1.8.10}$$

for every $w \in X$, by the triangle inequality. It follows that

$$d_Y(f(x), f(w)) < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon \tag{1.8.11}$$

for every $w \in X$ with $d_X(x, w) < \delta_L$, as desired.

Similarly, if $\{f_j\}_{j=1}^\infty$ is a sequence of uniformly continuous mappings from X into Y that converges uniformly to a mapping f from X into Y, then

(1.8.12) f is uniformly continuous on X.

As before, we let $\epsilon > 0$ be given, and let L be a positive integer such that (1.8.8) holds for every $j \geq L$ and $w \in X$. In this case, the uniform continuity of f_L implies that there is a $\delta_L > 0$ such that (1.8.9) holds for every $x, w \in X$ with $d_X(x, w) < \delta_L$. This implies that (1.8.11) holds for every $x, w \in X$ with $d_X(x, w) < \delta_L$, as before.

Note that there are related statements about limits of functions at a given point, instead of continuity at a point.

1.9 Bounded sets

Let (X, d_X) be a metric space. If $x \in X$ and r is a positive real number, then the *open ball* in X centered at x with radius r is defined as usual by

$$B(x, r) = B_X(x, r) = \{w \in X : d_X(x, w) < r\}. \tag{1.9.1}$$

If x' is another element of X, then it is easy to see that

$$B(x, r) \subseteq B(x', r + d_X(x, x')), \tag{1.9.2}$$

using the triangle inequality. It is well known that open balls in X are open sets.

A subset E of X is said to be *bounded* in X if there is an $x \in X$ and an $r > 0$ such that

$$E \subseteq B(x, r). \tag{1.9.3}$$

This implies that for every $x' \in X$ there is an $r' > 0$ such that

$$E \subseteq B(x', r'), \tag{1.9.4}$$

because of (1.9.2). Of course, this condition implies the previous one when $X \neq \emptyset$. To avoid minor technicalities, the empty set will be considered as a bounded set even when $X = \emptyset$.

If K is a compact subset of X, then K is bounded in X. This is trivial when $X = \emptyset$, because the empty set is automatically considered to be a bounded set, and so we may suppose that $X \neq \emptyset$. If x is any element of X, then the collection of open balls $B(x, j)$ with $j \in \mathbf{Z}_+$ is an open covering of K, because

$$\bigcup_{j=1}^{\infty} B(x, j) = X. \tag{1.9.5}$$

If K is compact, then there are finitely many positive integers $j_1, \ldots, j_n$ such that

$$K \subseteq \bigcup_{l=1}^{n} B(x, j_l). \tag{1.9.6}$$

This implies that $K \subseteq B(x, r)$, with $r = \max(j_1, \ldots, j_n)$.

Note that subsets of bounded sets are bounded. Let $E_1, \ldots, E_n$ be finitely many bounded subsets of X, and let us check that their union $\bigcup_{j=1}^{n} E_j$ is bounded. As before, this is trivial when $X = \emptyset$, and so we may suppose that $X \neq \emptyset$. If x is any element of X, then for each $j = 1, \ldots, n$ there is a positive real number r_j such that

$$E_j \subseteq B(x, r_j). \tag{1.9.7}$$

This implies that

$$\bigcup_{j=1}^{n} E_j \subseteq B\big(x, \max_{1 \le j \le n} r_j\big), \tag{1.9.8}$$

as desired.

If $x \in X$ and r is a nonnegative real number, then the *closed ball* in X centered at x with radius r is defined by

$$\overline{B}(x, r) = \overline{B}_X(x, r) = \{w \in X : d_X(x, w) \le r\}. \tag{1.9.9}$$

If x' is another element of X, then

$$\overline{B}(x, r) \subseteq \overline{B}(x', r + d_X(x, x')), \tag{1.9.10}$$

as in (1.9.2). One can check that closed balls in X are closed sets.

If $X \neq \emptyset$, then a subset E of X is bounded if and only if it is contained in a closed ball in X. In this case, E is contained in a closed ball centered at any point in X, as before. In particular, if E is bounded, then the closure $\overline{E}$ of E in X is bounded too.

A sequence $\{x_j\}_{j=1}^{\infty}$ of elements of X is said to be *bounded* in X if the set of x_j's, $j \in \mathbf{Z}_+$, is bounded in X. One can check that

(1.9.11) convergent sequences in X are bounded.

Similarly, one can verify that

(1.9.12) Cauchy sequences in X are bounded.

Totally bounded subsets of metric spaces are discussed in Section 4.2. In particular, compact sets are totally bounded, and totally bounded sets are bounded.

1.10 Some remarks and examples

Let n be a positive integer, and let E be a subset of $\mathbf{R}^n$ or $\mathbf{C}^n$. If E is bounded with respect to any of the metrics d_1, d_2, or d_∞ defined in Sections 1.3 or 1.4, as appropriate, then it is easy to see that E is bounded with respect to the other two metrics, using the inequalities in Section 1.5. Similarly, if a sequence of elements of $\mathbf{R}^n$ or $\mathbf{C}^n$ converges with respect to any of these three metrics, then it converges with respect to the other two metrics, and with the same limit.

Let X be a nonempty set, and let E be a subset of $c_{00}(X, \mathbf{R})$ or $c_{00}(X, \mathbf{C})$. If E is bounded with respect to the metric d_2 defined in Section 1.6, then E is bounded with respect to d_∞. Similarly, if E is bounded with respect to d_1, then

E is bounded with respect to d_2 and d_∞. If X has only finitely many elements, and E is bounded with respect to d_∞, then E is bounded with respect to d_1 and d_2.

Let $\{f_j\}_{j=1}^\infty$ be a sequence of elements of $c_{00}(X, \mathbf{R})$ or $c_{00}(X, \mathbf{C})$, and let f be another element of the same space. If $\{f_j\}_{j=1}^\infty$ converges to f with respect to d_2, then $\{f_j\}_{j=1}^\infty$ converges to f with respect to d_∞. If $\{f_j\}_{j=1}^\infty$ converges to f with respect to d_1, then $\{f_j\}_{j=1}^\infty$ converges to f with respect to d_2 and d_∞. If X has only finitely many elements, and $\{f_j\}_{j=1}^\infty$ converges to f with respect to d_∞, then $\{f_j\}_{j=1}^\infty$ converges to f with respect to d_1 and d_2.

One can check that $\{f_j\}_{j=1}^\infty$ converges to f with respect to d_∞ if and only if $\{f_j\}_{j=1}^\infty$ converges to f uniformly on X. This uses the standard metric on $\mathbf{R}$ or $\mathbf{C}$, as appropriate.

If $x \in X$, then let δ_x be the real-valued function on X equal to 1 at x, and to 0 at every other element of X. It is easy to see that the collection of δ_x's, $x \in X$, is a basis for each of $c_{00}(X, \mathbf{R})$ and $c_{00}(X, \mathbf{C})$, as vector spaces over $\mathbf{R}$ and $\mathbf{C}$, respectively, if one is familiar with these notions from linear algebra. Basically, this means that every element f of $c_{00}(X, \mathbf{R})$ or $c_{00}(X, \mathbf{C})$ can be expressed in a unique way as a linear combination of the δ_x's, $x \in X$, with coefficients in $\mathbf{R}$ or $\mathbf{C}$, as appropriate. In fact, the coefficient of δ_x is equal to $f(x)$ for each $x \in X$.

Observe that

$$\|\delta_x\|_1 = \|\delta_x\|_2 = \|\delta_x\|_\infty = 1 \tag{1.10.1}$$

for every $x \in X$. If $x, y \in X$ and $x \neq y$, then

$$\|\delta_x - \delta_y\|_\infty = 1, \tag{1.10.2}$$

$$\|\delta_x - \delta_y\|_2 = \sqrt{2}, \tag{1.10.3}$$

$$\|\delta_x - \delta_y\|_1 = 2. \tag{1.10.4}$$

Let us now take $X = \mathbf{Z}_+$, and let δ_j be as before for each positive integer j. It is easy to see that $\{\delta_j\}_{j=1}^\infty$ converges to 0 pointwise on $\mathbf{Z}_+$. Note that $\{\delta_j\}_{j=1}^\infty$ is bounded with respect to each of d_1, d_2, and d_∞, and that $\{\delta_j\}_{j=1}^\infty$ does not converge to 0 with respect to any of these three metrics.

Similarly, $\{j\,\delta_j\}_{j=1}^\infty$ converges to 0 pointwise on $\mathbf{Z}_+$. However,

$$\|j\,\delta_j\|_1 = \|j\,\delta_j\|_2 = \|j\,\delta_j\|_\infty = j \tag{1.10.5}$$

for each j, so that $\{j\,\delta_j\}_{j=1}^\infty$ is not bounded with respect to d_1, d_2, or d_∞.

If $j \in \mathbf{Z}_+$, then let f_j be the real-valued function on $\mathbf{Z}_+$ defined by

$$\begin{aligned} f_j(l) &= 1 \quad \text{when } l \le j \\ &= 0 \quad \text{when } l > j. \end{aligned} \tag{1.10.6}$$

Observe that $\{f_j\}_{j=1}^\infty$ converges pointwise to the function equal to 1 everywhere on $\mathbf{Z}_+$. One can check that $\{f_j\}_{j=1}^\infty$ does not converge uniformly on $\mathbf{Z}_+$. We also have that

$$\|f_j\|_\infty = 1, \tag{1.10.7}$$

$$\|f_j\|_2 = \sqrt{j}, \tag{1.10.8}$$

$$\|f_j\|_1 = j \tag{1.10.9}$$

for every j.

If $\alpha \in \mathbf{R}$, then

$$\|j^{-\alpha} f_j\|_\infty = j^{-\alpha} \tag{1.10.10}$$

for every j. This means that $\{j^{-\alpha} f_j\}_{j=1}^\infty$ is bounded with respect to d_∞ exactly when $\alpha \geq 0$, and that $\{j^{-\alpha} f_j\}_{j=1}^\infty$ converges to 0 with respect to d_∞ exactly when $\alpha > 0$. Similarly,

$$\|j^{-\alpha} f_j\|_2 = j^{(1/2)-\alpha} \tag{1.10.11}$$

for every j, so that $\{j^{-\alpha} f_j\}_{j=1}^\infty$ is bounded with respect to d_2 if and only if $\alpha \geq 1/2$, and $\{j^{-\alpha} f_j\}_{j=1}^\infty$ converges to 0 with respect to d_2 exactly when $\alpha > 1/2$. In the same way,

$$\|j^{-\alpha} f_j\|_1 = j^{1-\alpha} \tag{1.10.12}$$

for every j, so that $\{j^{-\alpha} f_j\}_{j=1}^\infty$ is bounded with respect to d_1 if and only if $\alpha \geq 1$, and $\{j^{-\alpha} f_j\}_{j=1}^\infty$ converges to 0 with respect to d_1 if and only if $\alpha > 1$.

1.11 Bounded functions

Let X be a set, and let (Y, d_Y) be a metric space. A mapping f from X into Y is said to be *bounded* if

(1.11.1) the image $f(X)$ of X under f is a bounded subset of Y.

Let $\mathcal{B}(X, Y)$ be the space of bounded mappings from X into Y.

Let $\{f_j\}_{j=1}^\infty$ be a sequence of bounded mappings from X into Y that converges uniformly to a mapping f from X into Y. This implies that there is an $L \in \mathbf{Z}_+$ such that

$$d_Y(f_j(x), f(x)) < 1 \tag{1.11.2}$$

for every $j \geq L$ and $x \in X$. One can use this to check that

(1.11.3) f is bounded,

because f_L is bounded.

Suppose that $X \neq \emptyset$, and let f, g be bounded mappings from X into Y. It is easy to see that

$$d_Y(f(x), g(x)) \tag{1.11.4}$$

is bounded as a nonnegative real-valued function of x on X, using the triangle inequality. Put

$$\theta(f, g) = \sup\{d_Y(f(x), g(x)) : x \in X\}. \tag{1.11.5}$$

If $f = g$, then $f(x) = g(x)$ for every $x \in X$, so that

$$d_Y(f(x), g(x)) = 0 \tag{1.11.6}$$

for every $x \in X$, and hence

$$\theta(f, g) = 0. \tag{1.11.7}$$

Conversely, if (1.11.7) holds, then (1.11.6) holds for every $x \in X$, so that $f(x) = g(x)$ for every $x \in X$, which means that $f = g$. We also have that

$$\theta(f,g) = \theta(g,f), \tag{1.11.8}$$

because $d_Y(f(x), g(x)) = d_Y(g(x), f(x))$ for every $x \in X$. If h is another bounded mapping from X into Y, then

$$\begin{aligned} d_Y(f(x), h(x)) &\le d_Y(f(x), g(x)) + d_Y(g(x), h(x)) \\ &\le \theta(f,g) + \theta(g,h) \end{aligned} \tag{1.11.9}$$

for every $x \in X$. This implies that

$$\theta(f,h) \le \theta(f,g) + \theta(g,h). \tag{1.11.10}$$

Thus (1.11.5) defines a metric on $\mathcal{B}(X,Y)$, which is known as the *supremum metric*.

Let $\{f_j\}_{j=1}^\infty$ be a sequence of bounded mappings from X into Y, and let f be another bounded mapping from X into Y. If $\{f_j\}_{j=1}^\infty$ converges to f with respect to the supremum metric, then for each $\epsilon > 0$ there is an $L(\epsilon) \in \mathbf{Z}_+$ such that

$$\theta(f_j, f) < \epsilon \tag{1.11.11}$$

for every $j \ge L(\epsilon)$. It follows that

$$d_Y(f_j(x), f(x)) < \epsilon \tag{1.11.12}$$

for every $j \ge L(\epsilon)$ and $x \in X$, so that $\{f_j\}_{j=1}^\infty$ converges to f uniformly on X.

Conversely, if $\{f_j\}_{j=1}^\infty$ converges to f uniformly on X, then for each $\epsilon > 0$ there is an $L'(\epsilon) \in \mathbf{Z}_+$ such that (1.11.12) holds for every $j \ge L'(\epsilon)$ and $x \in X$. This implies that

$$\theta(f_j, f) = \sup\{d_Y(f_j(x), f(x)) : x \in X\} \le \epsilon \tag{1.11.13}$$

for every $j \ge L'(\epsilon)$, and hence that $\{f_j\}_{j=1}^\infty$ converges to f with respect to the supremum metric.

1.12 Completeness of $\mathcal{B}(X,Y)$

Let us continue with the same notation and hypotheses as in the previous section. Suppose that

$$Y \text{ is complete with respect to } d_Y, \tag{1.12.1}$$

and let us check that

$$\mathcal{B}(X,Y) \text{ is complete with respect to the supremum metric.} \tag{1.12.2}$$

Let $\{f_j\}_{j=1}^\infty$ be a sequence of bounded mappings from X into Y that is a Cauchy sequence with respect to the supremum metric. This means that for each $\epsilon > 0$ there is an $L(\epsilon) \in \mathbf{Z}_+$ such that

$$\theta(f_j, f_l) < \epsilon \tag{1.12.3}$$

for every $j, l \geq L(\epsilon)$. Thus

$$d_Y(f_j(x), f_l(x)) < \epsilon \tag{1.12.4}$$

for every $j, l \geq L(\epsilon)$ and $x \in X$. In particular,

$$\{f_j(x)\}_{j=1}^\infty \text{ is a Cauchy sequence in } Y \tag{1.12.5}$$

for every $x \in X$. Because Y is complete,

$$\{f_j(x)\}_{j=1}^\infty \text{ converges in } Y \tag{1.12.6}$$

for every $x \in X$, and we put

$$f(x) = \lim_{j \to \infty} f_j(x). \tag{1.12.7}$$

This defines a mapping f from X into Y, and one can check that

$$d_Y(f_j(x), f(x)) \leq \epsilon \tag{1.12.8}$$

for every $j \geq L(\epsilon)$ and $x \in X$, using (1.12.4). Indeed,

$$\begin{aligned} d_Y(f_j(x), f(x)) &\leq d_Y(f_j(x), f_l(x)) + d_Y(f_l(x), f(x)) \\ &< \epsilon + d_Y(f_l(x), f(x)) \end{aligned} \tag{1.12.9}$$

for all $j, l \geq L(\epsilon)$ and $x \in X$, because of (1.12.4) and the triangle inequality. This implies (1.12.8), because $\{f_l(x)\}_{l=1}^\infty$ converges to $f(x)$ in Y.

It follows that $\{f_j\}_{j=1}^\infty$ converges to f uniformly on X, and hence that f is bounded on X, as in the previous section. This implies that $\{f_j\}_{j=1}^\infty$ converges to f with respect to the supremum metric, as before.

One can also verify that (1.12.2) implies (1.12.1). Indeed, every element of Y corresponds to a constant function on X with values in Y. It is easy to see that a Cauchy sequence in Y corresponds to a Cauchy sequence in $\mathcal{B}(X, Y)$ with respect to the supremum metric in this way. If a sequence of constant mappings from X to Y converges with respect to the supremum metric, then it converges uniformly on X, and thus pointwise on X, and it is easy to see that the limit is a constant mapping as well.

1.13 More on bounded functions

The spaces of bounded real and complex-valued functions on a empty set X are also denoted

$$\ell^\infty(X, \mathbf{R}) \text{ and } \ell^\infty(X, \mathbf{C}), \tag{1.13.1}$$

respectively. This implicitly uses the standard Euclidean metrics on $\mathbf{R}$ and $\mathbf{C}$. If f and g are bounded real or complex-valued functions on X, then it is easy to see that

$$f + g \text{ is bounded on } X \text{ as well.} \tag{1.13.2}$$

Similarly, if f is a bounded real or complex-valued function on X, and t is a real or complex number, as appropriate, then

$$t\,f \text{ is bounded on } X \tag{1.13.3}$$

too. Thus $\ell^\infty(X, \mathbf{R})$ and $\ell^\infty(X, \mathbf{C})$ are linear subspaces of the real and complex vector spaces of all real and complex-valued functions on X, respectively.

If f is a bounded real or complex-valued function on X, then put

$$\|f\|_\infty = \sup\{|f(x)| : x \in X\}. \tag{1.13.4}$$

Note that $\|f\|_\infty = 0$ if and only if $f = 0$ on X. If $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then one can check that

$$\|t\,f\|_\infty = |t|\,\|f\|_\infty. \tag{1.13.5}$$

If g is another bounded real or complex-valued function on X, then

$$|f(x) + g(x)| \le |f(x)| + |g(x)| \le \|f\|_\infty + \|g\|_\infty \tag{1.13.6}$$

for every $x \in X$. This implies that

$$\|f + g\|_\infty \le \|f\|_\infty + \|g\|_\infty. \tag{1.13.7}$$

It follows that (1.13.4) defines a norm on each of $\ell^\infty(X, \mathbf{R})$ and $\ell^\infty(X, \mathbf{C})$, which is known as the *supremum norm.* The corresponding metric

$$d_\infty(f, g) = \|f - g\|_\infty \tag{1.13.8}$$

is the same as the supremum metric on these spaces, associated to the standard Euclidean metric on $\mathbf{R}$ or $\mathbf{C}$, as appropriate.

1.14 Continuous functions

Let (X, d_X) and (Y, d_Y) be metric spaces, and let $C(X, Y)$ be the space of continuous mappings from X into Y. Also let

$$C_b(X, Y) = \mathcal{B}(X, Y) \cap C(X, Y) \tag{1.14.1}$$

be the space of bounded continuous mappings from X into Y. If f is a continuous mapping from X into Y and X is compact, then it is well known that $f(X)$ is compact in Y, so that $f(X)$ is bounded in Y in particular. Thus $C_b(X, Y)$ is the same as $C(X, Y)$ when X is compact.

Suppose that $X \ne \emptyset$, so that the supremum metric can be defined on $\mathcal{B}(X, Y)$ as in Section 1.11. Note that

$$C_b(X, Y) \text{ is a closed set in } \mathcal{B}(X, Y), \tag{1.14.2}$$

with respect to the supremum metric. More precisely, if $\{f_j\}_{j=1}^\infty$ is a sequence of bounded continuous mappings from X into Y that converges to a bounded mapping f from X into Y with respect to the supremum metric, then we have seen that $\{f_j\}_{j=1}^\infty$ converges to f uniformly on X, and hence that f is continuous on X. Alternatively, if $f \in \mathcal{B}(X, Y)$ is a limit point of $C_b(X, Y)$ with respect to the supremum metric, then one can show that f is continuous on X. This is very similar to the argument used to show that uniform limits of continuous mappings are continuous, as in Section 1.8.

Of course, $C_b(X, Y)$ may be considered as a metric space, using the restriction of the supremum metric on $\mathcal{B}(X, Y)$ to $C_b(X, Y)$. Suppose that Y is complete with respect to d_Y, so that $\mathcal{B}(X, Y)$ is complete with respect to the supremum metric, as in Section 1.12. Under these conditions, we get that

$$C_b(X, Y) \text{ is complete as a metric space} \tag{1.14.3}$$

with respect to the supremum metric, as in Section 1.7. This also uses (1.14.2).

Let us now take $Y = \mathbf{R}$ or $\mathbf{C}$, with their standard Euclidean metrics. If f and g are continuous real or complex-valued functions on X, then it is well known that their sum $f + g$ is continuous on X too. Similarly, if f is a continuous real or complex-valued function on X, and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then $t\,f$ is continuous on X. This means that $C(X, \mathbf{R})$ and $C(X, \mathbf{C})$ are linear subspaces of the real and complex vector spaces of all real and complex-valued functions on X, respectively. We may also consider

$$C_b(X, \mathbf{R}) = \ell^\infty(X, \mathbf{R}) \cap C(X, \mathbf{R}) \tag{1.14.4}$$

as a linear subspace of both $\ell^\infty(X, \mathbf{R})$ and $C(X, \mathbf{R})$, and

$$C_b(X, \mathbf{C}) = \ell^\infty(X, \mathbf{C}) \cap C(X, \mathbf{C}) \tag{1.14.5}$$

as a linear subspace of both $\ell^\infty(X, \mathbf{C})$ and $C(X, \mathbf{C})$.

1.15 Continuous functions on $[0, 1]$

In this section, we take X to be the closed unit interval $[0, 1]$ in the real line, equipped with the restriction of the standard Euclidean metric on $\mathbf{R}$ to $[0, 1]$. It is well known that $[0, 1]$ is compact as a subset of $\mathbf{R}$, and thus as a subset of itself. This means that every continuous real or complex-valued function f on $[0, 1]$ is bounded, as before.

A nonnegative real-valued function N on $C([0, 1], \mathbf{R})$ or $C([0, 1], \mathbf{C})$ is said to be a *norm* if it satisfies the usual three conditions, as follows. First, $N(f) = 0$ if and only if $f = 0$. Second, if f is a continuous real or complex-valued function on $[0, 1]$ and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then

$$N(t\,f) = |t|\,N(f). \tag{1.15.1}$$

Third, if f and g are continuous real or complex-valued functions on $[0, 1]$, then

$$N(f + g) \le N(f) + N(g). \tag{1.15.2}$$

In this case,

$$d_N(f,g) = N(f-g) \tag{1.15.3}$$

defines a metric on $C([0,1],\mathbf{R})$ or $C([0,1],\mathbf{C})$, as appropriate. The supremum norm (1.13.4) defines a norm on each of $C([0,1],\mathbf{R})$ and $C([0,1],\mathbf{C})$, for which the corresponding metric (1.13.8) is the supremum metric, as in Section 1.13.

If f is a continuous real or complex-valued function on $[0,1]$, then put

$$\|f\|_1 = \int_0^1 |f(x)|\,dx. \tag{1.15.4}$$

More precisely, it is well known and not difficult to verify that $|f(x)|$ is also continuous on $[0,1]$, so that the Riemann integral on the right side of (1.15.4) exists. If $f(x_0) \neq 0$ for some $0 \le x_0 \le 1$, then

$$|f(x)| \ge |f(x_0)|/2 > 0 \tag{1.15.5}$$

when $0 \le x \le 1$ is sufficiently close to x_0, because f is continuous at x_0. This implies that $\|f\|_1 > 0$, so that (1.15.4) satisfies the first condition in the definition of a norm. It is easy to see that (1.15.4) satisfies (1.15.1) and (1.15.2), so that (1.15.4) defines a norm on $C([0,1],\mathbf{R})$ and $C([0,1],\mathbf{C})$.

Similarly, put

$$\|f\|_2 = \Big(\int_0^1 |f(x)|^2\,dx\Big)^{1/2} \tag{1.15.6}$$

for every continuous real or complex-valued function f on $[0,1]$. One can check that (1.15.6) is equal to 0 exactly when $f=0$ on $[0,1]$, using the same type of argument as in the preceding paragraph. It is easy to see that (1.15.6) satisfies the homogeneity condition (1.15.1). If g is another continuous real or complex-valued function on $[0,1]$, then it is well known that

$$\|f\,g\|_1 \le \|f\|_2\,\|g\|_2, \tag{1.15.7}$$

which is an integral version of the Cauchy–Schwarz inequality. This can be used to show that (1.15.6) satisfies the triangle inequality (1.15.2), as usual, so that (1.15.6) defines a norm on $C([0,1],\mathbf{R})$ and $C([0,1],\mathbf{C})$.

Clearly

$$\|f\|_1, \|f\|_2 \le \|f\|_\infty \tag{1.15.8}$$

for every continuous real or complex-valued function f on $[0,1]$. One can can also get that

$$\|f\|_1 \le \|f\|_2, \tag{1.15.9}$$

using (1.15.7). Let

$$d_1(f,g) = \|f-g\|_1 \tag{1.15.10}$$

and

$$d_2(f,g) = \|f-g\|_2 \tag{1.15.11}$$

be the metrics on $C([0,1],\mathbf{R})$ and $C([0,1],\mathbf{C})$ associated to (1.15.4) and (1.15.6) as in (1.15.3), respectively. Using (1.15.8) and (1.15.9), we get that

$$d_1(f,g) \le d_2(f,g) \le d_\infty(f,g) \tag{1.15.12}$$

for all continuous real and complex-valued functions f and g on $[0,1]$.

It follows that the identity mappings on $C([0,1],\mathbf{R})$ and $C([0,1],\mathbf{C})$ are Lipschitz with respect to $d_2(f,g)$ on the domain and $d_1(f,g)$ on the range. Similarly, the identity mappings on $C([0,1],\mathbf{R})$ and $C([0,1],\mathbf{C})$ are Lipschitz with respect to $d_\infty(f,g)$ on the domain and $d_2(f,g)$ on the range. The identity mappings on $C([0,1],\mathbf{R})$ and $C([0,1],\mathbf{C})$ are also Lipschitz with respect to $d_\infty(f,g)$ on the domain and $d_1(f,g)$ on the range.

Chapter 2

Basic ℓ^1 and ℓ^2 spaces

This chapter deals with classical ℓ^1 and ℓ^2 spaces, of absolutely summable and square-summable sequences of real or complex numbers, respectively. We also consider c_0 spaces of real or complex-valued functions on arbitrary nonempty sets that vanish at infinity. Sums over arbitrary nonempty sets and corresponding ℓ^1, ℓ^2 spaces will be discussed in Chapter 11.

2.1 Infinite series

Remember that an infinite series

$$\sum_{j=1}^{\infty} a_j \tag{2.1.1}$$

of real or complex numbers is said to *converge* if the corresponding sequence of partial sums

$$\sum_{j=1}^{n} a_j \tag{2.1.2}$$

converges with respect to the standard Euclidean metric on $\mathbf{R}$ or $\mathbf{C}$, as appropriate. Of course, the value of the sum (2.1.1) is defined to be the limit of the sequence of partial sums (2.1.2) in this case. If (2.1.1) converges, and if $\sum_{j=1}^{\infty} b_j$ is another convergent series of real or complex numbers, as appropriate, then $\sum_{j=1}^{\infty}(a_j + b_j)$ converges, with

$$\sum_{j=1}^{\infty}(a_j + b_j) = \sum_{j=1}^{\infty} a_j + \sum_{j=1}^{\infty} b_j. \tag{2.1.3}$$

This reduces to the corresponding statement for sums of convergent sequences of real or complex numbers, applied to the partial sums of these series. Similarly, if (2.1.1) converges, and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then $\sum_{j=1}^\infty t\,a_j$ converges, with

$$\sum_{j=1}^\infty t\,a_j = t \sum_{j=1}^\infty a_j. \tag{2.1.4}$$

If a_j is a nonnegative real number for each $j \geq 1$, then the partial sums (2.1.2) increase monotonically. It is well known that a monotonically increasing sequence of real numbers converges with respect to the standard Euclidean metric on $\mathbf{R}$ if and only if the sequence has an upper bound in $\mathbf{R}$, in which case the sequence converges to the supremum of the set of its terms. If a_j is a nonnegative real number for each j, but the partial sums (2.1.2) do not have an upper bound in $\mathbf{R}$, then it is sometimes convenient to consider the value of the sum (2.1.1) to be $+\infty$. Note that the partial sums (2.1.2) tend to $+\infty$ as $n \to \infty$ in this situation. If a_j and b_j are nonnegative real numbers for each $j \geq 1$, then one can check that (2.1.3) holds, where the right side of (2.1.3) is considered to be $+\infty$ when either of the individual sums is $+\infty$. Similarly, (2.1.4) holds for every positive real number t, where the right side is considered to be $+\infty$ when (2.1.1) is $+\infty$. If a_j and b_j are nonnegative real numbers with

$$a_j \leq b_j \tag{2.1.5}$$

for every $j \geq 1$, then

$$\sum_{j=1}^\infty a_j \leq \sum_{j=1}^\infty b_j, \tag{2.1.6}$$

which is trivial when the right side if $+\infty$.

An infinite series (2.1.1) of real or complex numbers is said to converge *absolutely* if

$$\sum_{j=1}^\infty |a_j| \tag{2.1.7}$$

converges as an infinite series of nonnegative real numbers. This means that (2.1.7) is finite, in terms of the conventions for sums of nonnegative real numbers mentioned in the previous paragraph. If (2.1.1) converges absolutely, then it is well known that (2.1.1) converges in the usual sense. One can also check that

$$\left|\sum_{j=1}^\infty a_j\right| \leq \sum_{j=1}^\infty |a_j| \tag{2.1.8}$$

under these conditions. This uses the fact that

$$\left|\sum_{j=1}^n a_j\right| \leq \sum_{j=1}^n |a_j| \tag{2.1.9}$$

for every positive integer n, by the triangle inequality.

If (2.1.1) converges, then it is well known that $\{a_j\}_{j=1}^\infty$ converges to 0 as a sequence of real or complex numbers, as appropriate. In particular, this holds when (2.1.1) converges absolutely.

2.2 Basic ℓ^1 spaces

Let $\ell^1(\mathbf{Z}_+, \mathbf{R})$ be the space of real-valued functions f on the set $\mathbf{Z}_+$ of positive integers such that

$$\|f\|_1 = \sum_{j=1}^{\infty} |f(j)| \tag{2.2.1}$$

is finite, which is to say that the right side converges as an infinite series of nonnegative real numbers. Similarly, let $\ell^1(\mathbf{Z}_+, \mathbf{C})$ be the space of complex-valued functions f on $\mathbf{Z}_+$ such that (2.2.1) is finite. In both cases, the convergence of the series on the right side of (2.2.1) implies that

$$\lim_{j \to \infty} f(j) = 0, \tag{2.2.2}$$

as mentioned in the previous section.

If $f \in \ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, then (2.2.1) is a nonnegative real number, which is equal to 0 exactly when $f(j) = 0$ for every $j \in \mathbf{Z}_+$. If $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then $t\,f(j)$ defines another real or complex-valued function on $\mathbf{Z}_+$, as appropriate, and

$$\|tf\|_1 = \sum_{j=1}^{\infty} |t\,f(j)| = |t|\,\|f\|_1. \tag{2.2.3}$$

In particular, $t\,f \in \ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate.

Let g be another element of $\ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate. Thus $f + g$ is a real or complex-valued function on $\mathbf{Z}_+$, as appropriate, and

$$\|f + g\|_1 = \sum_{j=1}^{\infty} |f(j) + g(j)| \le \sum_{j=1}^{\infty} (|f(j)| + |g(j)|) = \|f\|_1 + \|g\|_1. \tag{2.2.4}$$

This implies that $f + g \in \ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate. Hence $\ell^1(\mathbf{Z}_+, \mathbf{R})$ and $\ell^1(\mathbf{Z}_+, \mathbf{C})$ are linear subspaces of the real and complex vector spaces of all real and complex-valued functions on $\mathbf{Z}_+$, respectively.

This means that $\ell^1(\mathbf{Z}_+, \mathbf{R})$ and $\ell^1(\mathbf{Z}_+, \mathbf{C})$ are vector spaces over the real and complex numbers, respectively, with respect to pointwise addition and scalar multiplication. We also get that (2.2.1) defines a norm on each of these spaces, by the remarks in the preceding paragraphs. Using (2.2.3) and (2.2.4), one can check that

$$d_1(f, g) = \|f - g\|_1 \tag{2.2.5}$$

defines a metric on each of $\ell^1(\mathbf{Z}_+, \mathbf{R})$ and $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as usual.

If $f \in \ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, then

$$\sum_{j=1}^{\infty} f(j) \tag{2.2.6}$$

converges as an infinite series of real or complex numbers, as in the previous section. If g is another element of $\ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate, then

$$\left|\sum_{j=1}^{\infty} f(j) - \sum_{j=1}^{\infty} g(j)\right| = \left|\sum_{j=1}^{\infty}(f(j) - g(j))\right| \le \sum_{j=1}^{\infty} |f(j) - g(j)| = \|f - g\|_1, \tag{2.2.7}$$

using (2.1.8) in the second step. This implies that the mapping from f to the sum (2.2.6) is uniformly continuous as a mapping from $\ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}, \mathbf{C})$ into $\mathbf{R}$ or $\mathbf{C}$, respectively, using the ℓ^1 metric (2.2.5) on $\ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, and the standard Euclidean metric on $\mathbf{R}$ or $\mathbf{C}$.

If $f \in \ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, then it is easy to see directly that f is bounded on $\mathbf{Z}_+$, with

$$\|f\|_\infty \le \|f\|_1. \tag{2.2.8}$$

Here $\|f\|_\infty$ is the supremum norm of f on $\mathbf{Z}_+$, as in (1.13.4). If g is another element of $\ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate, then we get that

$$d_\infty(f, g) \le d_1(f, g), \tag{2.2.9}$$

where $d_\infty(f, g)$ is the supremum metric for bounded real or complex-valued functions on $\mathbf{Z}_+$, as in (1.13.8).

Remember that $c_{00}(\mathbf{Z}_+, \mathbf{R})$, $c_{00}(\mathbf{Z}_+, \mathbf{C})$ are the spaces of real and complex-valued functions on $\mathbf{Z}_+$ with finite support, respectively, as in Section 1.6. Clearly

$$c_{00}(\mathbf{Z}_+, \mathbf{R}) \subseteq \ell^1(\mathbf{Z}_+, \mathbf{R}), \quad c_{00}(\mathbf{Z}_+, \mathbf{C}) \subseteq \ell^1(\mathbf{Z}_+, \mathbf{C}), \tag{2.2.10}$$

since an infinite series automatically converges when all but finitely many terms are equal to 0. One can also check that $c_{00}(\mathbf{Z}_+, \mathbf{R})$ and $c_{00}(\mathbf{Z}_+, \mathbf{C})$ are dense subsets of $\ell^1(\mathbf{Z}_+, \mathbf{R})$ and $\ell^1(\mathbf{Z}_+, \mathbf{C})$, respectively, with respect to the ℓ^1 metric (2.2.5).

2.3 Basic ℓ^2 spaces

Let $\ell^2(\mathbf{Z}_+, \mathbf{R})$ and $\ell^2(\mathbf{Z}_+, \mathbf{C})$ be the spaces of real and complex-valued functions f on $\mathbf{Z}_+$ such that

$$\sum_{j=1}^{\infty} |f(j)|^2 \tag{2.3.1}$$

converges as an infinite series of nonnegative real numbers, respectively. In both cases, we put

$$\|f\|_2 = \Big(\sum_{j=1}^{\infty} |f(j)|^2\Big)^{1/2}, \tag{2.3.2}$$

using the nonnegative square root on the right side. Note that this is equal to 0 if and only if $f(j) = 0$ for every $j \ge 1$. The convergence of the series (2.3.1) implies that

$$\lim_{j \to \infty} |f(j)|^2 = 0, \tag{2.3.3}$$

and hence that $f(j) \to 0$ as $j \to \infty$.

Let $f \in \ell^2(\mathbf{Z}_+, \mathbf{R})$ or $\ell^2(\mathbf{Z}_+, \mathbf{C})$ be given. If $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then

$$\sum_{j=1}^{\infty} |t\, f(j)|^2 = |t|^2 \sum_{j=1}^{\infty} |f(j)|^2, \tag{2.3.4}$$

and in particular the series on the left converges. This means that $t\, f$ is an element of $\ell^2(\mathbf{Z}_+, \mathbf{R})$ or $\ell^2(\mathbf{Z}_+, \mathbf{C})$, as appropriate, with

$$\|t\, f\|_2 = |t|\, \|f\|_2. \tag{2.3.5}$$

If a and b are nonnegative real numbers, then it is well known that

$$a\, b \le \frac{1}{2}\,(a^2 + b^2), \tag{2.3.6}$$

because $0 \le (a - b)^2 = a^2 - 2\, a\, b + b^2$. Suppose that g is another element of $\ell^2(\mathbf{Z}_+, \mathbf{R})$ or $\ell^2(\mathbf{Z}_+, \mathbf{C})$, as appropriate. Using (2.3.6), we get that

$$|f(j)|\, |g(j)| \le \frac{1}{2}\,(|f(j)|^2 + |g(j)|^2) \tag{2.3.7}$$

for every $j \in \mathbf{Z}_+$. Hence

$$\begin{aligned} \sum_{j=1}^{\infty} |f(j)|\, |g(j)| &\le \sum_{j=1}^{\infty} \frac{1}{2}\,(|f(j)|^2 + |g(j)|^2) \\ &= \frac{1}{2} \sum_{j=1}^{\infty} |f(j)|^2 + \frac{1}{2} \sum_{j=1}^{\infty} |g(j)|^2 = \frac{1}{2}\, \|f\|_2^2 + \frac{1}{2}\, \|g\|_2^2. \end{aligned} \tag{2.3.8}$$

In particular, the series on the left converges, so that $f\, g \in \ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate.

In fact, we have that

$$\sum_{j=1}^{\infty} |f(j)|\, |g(j)| \le \|f\|_2\, \|g\|_2 \tag{2.3.9}$$

under these conditions, which is another version of the Cauchy–Schwarz inequality. This follows from (2.3.8) when $\|f\|_2 = \|g\|_2 = 1$, and otherwise one can reduce to that case using (2.3.5).

Observe that

$$|f(j) + g(j)|^2 \le (|f(j)| + |g(j)|)^2 \le |f(j)|^2 + 2\, |f(j)|\, |g(j)| + |g(j)|^2 \tag{2.3.10}$$

for every $j \ge 1$, so that

$$\sum_{j=1}^{\infty} |f(j) + g(j)|^2 \le \sum_{j=1}^{\infty} |f(j)|^2 + 2 \sum_{j=1}^{\infty} |f(j)|\,|g(j)| + \sum_{j=1}^{\infty} |g(j)|^2. \tag{2.3.11}$$

This implies that the series on the left converges, so that $f + g \in \ell^2(\mathbf{Z}_+, \mathbf{R})$ or $\ell^2(\mathbf{Z}_+, \mathbf{C})$, as appropriate. Combining (2.3.9) and (2.3.11), we get that

$$\|f + g\|_2^2 \le \|f\|_2^2 + 2\,\|f\|_2\,\|g\|_2 + \|g\|_2^2 = (\|f\|_2 + \|g\|_2)^2, \tag{2.3.12}$$

so that

$$\|f + g\|_2 \le \|f\|_2 + \|g\|_2. \tag{2.3.13}$$

This shows that $\ell^2(\mathbf{Z}_+, \mathbf{R})$ and $\ell^2(\mathbf{Z}_+, \mathbf{C})$ are linear subspaces of the real and complex vector spaces of all real and complex-valued functions on $\mathbf{Z}_+$, respectively. We also get that (2.3.2) defines a norm on each of $\ell^2(\mathbf{Z}_+, \mathbf{R})$ and $\ell^2(\mathbf{Z}_+, \mathbf{C})$, by (2.3.5) and (2.3.13). It follows that

$$d_2(f, g) = \|f - g\|_2 \tag{2.3.14}$$

defines a metric on each of $\ell^2(\mathbf{Z}_+, \mathbf{R})$ and $\ell^2(\mathbf{Z}_+, \mathbf{C})$, as usual.

If $f \in \ell^2(\mathbf{Z}_+, \mathbf{R})$ or $\ell^2(\mathbf{Z}_+, \mathbf{C})$, then one can check directly that f is bounded on $\mathbf{Z}_+$, with

$$\|f\|_\infty \le \|f\|_2. \tag{2.3.15}$$

If g is another element of $\ell^2(\mathbf{Z}_+, \mathbf{R})$ or $\ell^2(\mathbf{Z}_+, \mathbf{C})$, as appropriate, then we get that

$$d_\infty(f, g) \le d_2(f, g). \tag{2.3.16}$$

Suppose now that $f \in \ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$. Observe that

$$\sum_{j=1}^{\infty} |f(j)|^2 \le \|f\|_\infty \sum_{j=1}^{\infty} |f(j)| = \|f\|_\infty\,\|f\|_1 \le \|f\|_1^2, \tag{2.3.17}$$

using (2.2.8) in the third step. This implies that $f \in \ell^2(\mathbf{Z}_+, \mathbf{R})$ or $\ell^2(\mathbf{Z}_+, \mathbf{C})$, as appropriate, with

$$\|f\|_2 \le \|f\|_1. \tag{2.3.18}$$

In particular, we get that

$$\ell^1(\mathbf{Z}_+, \mathbf{R}) \subseteq \ell^2(\mathbf{Z}_+, \mathbf{R}), \quad \ell^1(\mathbf{Z}_+, \mathbf{C}) \subseteq \ell^2(\mathbf{Z}_+, \mathbf{C}). \tag{2.3.19}$$

If g is another element of $\ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate, then it follows that

$$d_2(f, g) \le d_1(f, g). \tag{2.3.20}$$

If f is a real or complex-valued function on $\mathbf{Z}_+$ with finite support, then f is clearly an element of $\ell^2(\mathbf{Z}_+, \mathbf{R})$ or $\ell^2(\mathbf{Z}_+, \mathbf{C})$, as appropriate. One can verify that $c_{00}(\mathbf{Z}_+, \mathbf{R})$ and $c_{00}(\mathbf{Z}_+, \mathbf{C})$ are dense subsets of $\ell^2(\mathbf{Z}_+, \mathbf{R})$ and $\ell^2(\mathbf{Z}_+, \mathbf{C})$, respectively, with respect to the ℓ^2 metric (2.3.14).

2.4 Completeness of ℓ^1, ℓ^2

Let $\{f_l\}_{l=1}^\infty$ be a sequence of elements of $\ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, and suppose that the f_l's have bounded ℓ^1 norms, so that there is a nonnegative real number C_1 such that

$$\sum_{j=1}^\infty |f_l(j)| \le C_1 \tag{2.4.1}$$

for every $l \ge 1$. Suppose also that $\{f_l\}_{l=1}^\infty$ converges to a real or complex-valued function f pointwise on $\mathbf{Z}_+$, as appropriate. Let us verify that $f \in \ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate, with

$$\sum_{j=1}^\infty |f(j)| \le C_1. \tag{2.4.2}$$

If $n \in \mathbf{Z}_+$, then

$$\sum_{j=1}^n |f(j)| = \lim_{l\to\infty} \sum_{j=1}^n |f_l(j)| \le C_1, \tag{2.4.3}$$

using pointwise convergence in the first step, and (2.4.1) in the second step. This implies (2.4.2), since this estimate holds for all $n \ge 1$.

We would like to show that $\ell^1(\mathbf{Z}_+, \mathbf{R})$ and $\ell^1(\mathbf{Z}_+, \mathbf{C})$ are complete with respect to the ℓ^1 metric (2.2.5). Let $\{f_l\}_{l=1}^\infty$ be a sequence of elements of $\ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$ that is a Cauchy sequence with respect to (2.2.5). This implies that for every $\epsilon > 0$ there is a positive integer $L(\epsilon)$ such that

$$\sum_{j=1}^\infty |f_k(j) - f_l(j)| = \|f_k - f_l\|_1 < \epsilon \tag{2.4.4}$$

for all $k, l \ge L(\epsilon)$. Hence

$$|f_k(j) - f_l(j)| < \epsilon \tag{2.4.5}$$

for every $j \in \mathbf{Z}_+$ when $k, l \ge L(\epsilon)$. This means that $\{f_l(j)\}_{l=1}^\infty$ is a Cauchy sequence of real or complex numbers, as appropriate, for every $j \in \mathbf{Z}_+$, and with respect to the standard Euclidean metric on $\mathbf{R}$ or $\mathbf{C}$. Remember that $\mathbf{R}$ and $\mathbf{C}$ are complete as metric spaces with respect to their standard Euclidean metrics. It follows that $\{f_l(j)\}_{l=1}^\infty$ converges in $\mathbf{R}$ or $\mathbf{C}$, as appropriate, for every $j \in \mathbf{Z}_+$, with respect to the standard Euclidean metric. Thus

$$f(j) = \lim_{l\to\infty} f_l(j) \tag{2.4.6}$$

defines a real or complex-valued function on $\mathbf{Z}_+$, as appropriate. We would like to check that $\sum_{j=1}^\infty |f(j)|$ converges, so that f is an element of $\ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate. We would also like to verify that $\{f_l\}_{l=1}^\infty$ converges to f with respect to the ℓ^1 metric.

Let $\epsilon > 0$ and $l \ge L(\epsilon)$ be given. Note that $\{f_k - f_l\}_{k=L(\epsilon)}^\infty$ is a sequence of elements of $\ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate, that converges to $f - f_l$

pointwise on $\mathbf{Z}_+$. The remarks at the beginning of the section imply that $f - f_l$ is an element of $\ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate, with

$$\|f - f_l\|_1 = \sum_{j=1}^{\infty} |f(j) - f_l(j)| \le \epsilon, \tag{2.4.7}$$

because of (2.4.4). Hence $f \in \ell^1(\mathbf{Z}_+, \mathbf{R})$ or $\ell^1(\mathbf{Z}_+, \mathbf{C})$, as appropriate, because of the corresponding property of f_l. It is easy to see that $\{f_l\}_{l=1}^{\infty}$ converges to f with respect to the ℓ^1 metric, because (2.4.7) holds for every $l \ge L(\epsilon)$.

Now let $\{f_l\}_{l=1}^{\infty}$ be a sequence of elements of $\ell^2(\mathbf{Z}_+, \mathbf{R})$ or $\ell^2(\mathbf{Z}_+, \mathbf{C})$ with bounded ℓ^2 norms, so that

$$\|f_l\|_2 \le C_2 \tag{2.4.8}$$

for some $C_2 \ge 0$ and every $l \ge 1$. This is the same as saying that

$$\sum_{j=1}^{\infty} |f_l(j)|^2 \le C_2^2 \tag{2.4.9}$$

for every $l \ge 1$. Suppose that $\{f_l\}_{l=1}^{\infty}$ also converges pointwise to a real or complex-valued function f on $\mathbf{Z}_+$, which implies that $\{|f_l|^2\}_{l=1}^{\infty}$ converges to $|f|^2$ pointwise on $\mathbf{Z}_+$ too. It follows that

$$\sum_{j=1}^{\infty} |f(j)|^2 \le C_2^2, \tag{2.4.10}$$

by the remarks at the beginning of the section, applied to $\{|f_l|^2\}_{l=1}^{\infty}$. This means that $f \in \ell^2(\mathbf{Z}_+, \mathbf{R})$ or $\ell^2(\mathbf{Z}_+, \mathbf{C})$, as appropriate, with

$$\|f\|_2 \le C_2. \tag{2.4.11}$$

Using this, one can show that $\ell^2(\mathbf{Z}_+, \mathbf{R})$ and $\ell^2(\mathbf{Z}_+, \mathbf{C})$ are complete with respect to the ℓ^2 metric (2.3.14). The argument is similar to the previous one for $\ell^1(\mathbf{Z}_+, \mathbf{R})$, $\ell^1(\mathbf{Z}_+, \mathbf{C})$.

2.5 Vanishing at infinity

Let X be a nonempty set, and let f be a real or complex-valued function on X. We say that f *vanishes at infinity* on X if for every $\epsilon > 0$,

$$|f(x)| < \epsilon \tag{2.5.1}$$

for all but finitely many $x \in X$. Equivalently, this means that for each $\epsilon > 0$,

$$E_\epsilon(f) = \{x \in X : |f(x)| \ge \epsilon\} \tag{2.5.2}$$

has only finitely many elements. Let $c_0(X, \mathbf{R})$ and $c_0(X, \mathbf{C})$ be the spaces of real and complex-valued functions on X that vanish at infinity, respectively.

Remember that $c_{00}(X, \mathbf{R})$ and $c_{00}(X, \mathbf{C})$ denote the spaces of real and complex-valued functions f on X such that the support of f has only finitely many elements, respectively, as in Section 1.6. In this case, f automatically vanishes at infinity on X, so that

$$c_{00}(X, \mathbf{R}) \subseteq c_0(X, \mathbf{R}), \quad c_{00}(X, \mathbf{C}) \subseteq c_0(X, \mathbf{C}). \tag{2.5.3}$$

In particular, if X has only finitely many elements, then every real or complex-valued function on X vanishes at infinity.

If f is a real or complex-valued function on the set $\mathbf{Z}_+$ of positive integers, then f vanishes at infinity on $\mathbf{Z}_+$ if and only if

$$\lim_{j \to \infty} f(j) = 0. \tag{2.5.4}$$

Let X be any nonempty set again, and let $\{x_j\}_{j=1}^{\infty}$ be an infinite sequence of distinct elements of X. Also let f be a real or complex-valued function on X, and suppose that the support of f is contained in the set of x_j's, $j \in \mathbf{Z}_+$. Under these conditions, f vanishes at infinity on X if and only if

$$\lim_{j \to \infty} f(x_j) = 0. \tag{2.5.5}$$

Let f be any real or complex-valued function on X, and remember that the support of f is the set of $x \in X$ such that $f(x) \neq 0$. Equivalently,

$$\operatorname{supp} f = \bigcup_{n=1}^{\infty} E_{1/n}(f), \tag{2.5.6}$$

where $E_{1/n}(f)$ is as in (2.5.2). If f vanishes at infinity on X, then it follows that the support of f has only finitely or countably many elements.

If f is a real or complex-valued function on X that vanishes at infinity, then it is easy to see that f is bounded on X. In fact, we have that

$$\|f\|_\infty = \max_{x \in X} |f(x)|, \tag{2.5.7}$$

which is to say that the maximum on the right is attained. Of course, this is trivial when $f \equiv 0$ on X. Otherwise, if $f(x_0) \neq 0$ for some $x_0 \in X$, then there are only finitely many $x \in X$ such that $|f(x)| \geq |f(x_0)|$. In this case, it suffices to take the maximum of $|f(x)|$ over this finite set.

Thus

$$c_0(X, \mathbf{R}) \subseteq \ell^\infty(X, \mathbf{R}), \quad c_0(X, \mathbf{C}) \subseteq \ell^\infty(X, \mathbf{C}). \tag{2.5.8}$$

If g is another real or complex-valued function on X that vanishes at infinity, then one can check that $f + g$ vanishes at infinity on X as well. Similarly, if $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then $t\, f$ vanishes at infinity on X. This means that $c_0(X, \mathbf{R})$ and $c_0(X, \mathbf{C})$ are linear subspaces of $\ell^\infty(X, \mathbf{R})$ and $\ell^\infty(X, \mathbf{C})$, respectively.

Let f be a bounded real or complex-valued function on X. Suppose that f is a limit point of $c_0(X, \mathbf{R})$ or $c_0(X, \mathbf{C})$ in $\ell^\infty(X, \mathbf{R})$ or $\ell^\infty(X, \mathbf{C})$, respectively, with respect to the supremum metric. We would like to verify that f also vanishes at infinity on X in this case. Let $\epsilon > 0$ be given. By hypothesis, there is a real or complex-valued function g on X, as appropriate, such that g vanishes at infinity on X, and

$$\|f - g\|_\infty < \epsilon/2. \tag{2.5.9}$$

This implies that

$$|f(x)| \le |g(x)| + |f(x) - g(x)| \le |g(x)| + \|f - g\|_\infty < |g(x)| + \epsilon/2 \tag{2.5.10}$$

for every $x \in X$. Of course, $|g(x)| < \epsilon/2$ for all but finitely many $x \in X$, because g vanishes at infinity on X. It follows that

$$|f(x)| < \epsilon/2 + \epsilon/2 = \epsilon \tag{2.5.11}$$

for all but finitely many $x \in X$, as desired.

This shows that $c_0(X, \mathbf{R})$ and $c_0(X, \mathbf{C})$ are closed sets in $\ell^\infty(X, \mathbf{R})$ and $\ell^\infty(X, \mathbf{C})$, respectively, with respect to the supremum metric. As a slightly different version of this, let $\{f_j\}_{j=1}^\infty$ be a sequence of real or complex-valued functions on X that vanish at infinity and converge uniformly to a real or complex-valued function f on X, as appropriate. Under these conditions, f vanishes at infinity on X too. This can be obtained from the previous statement, or using the same argument as in the preceding paragraph.

Let f be a real or complex-valued function on X that vanishes at infinity, and let $\epsilon > 0$ be given. Let f_ϵ be the real or complex-valued function, as appropriate, defined on X by

$$\begin{aligned} f_\epsilon(x) &= f(x) \quad \text{when } |f(x)| \ge \epsilon \\ &= 0 \qquad \text{when } |f(x)| < \epsilon. \end{aligned} \tag{2.5.12}$$

Note that f_ϵ has finite support in X, because f vanishes at infinity on X. By construction,

$$\|f - f_\epsilon\|_\infty < \epsilon, \tag{2.5.13}$$

where the strict inequality uses the hypothesis that f vanish at infinity on X again. This shows that $c_{00}(X, \mathbf{R})$ and $c_{00}(X, \mathbf{C})$ are dense in $c_0(X, \mathbf{R})$ and $c_0(X, \mathbf{C})$, respectively, with respect to the supremum metric.

Chapter 3

Some more metric spaces

3.1 Functions on intervals

Let a, b be real numbers with $a < b$, and let

$$[a, b] = \{x \in \mathbf{R} : a \le x \le b\} \tag{3.1.1}$$

be the usual closed interval in the real line from a to b. It is well known that $[a, b]$ is compact with respect to the standard metric on $\mathbf{R}$. As in Section 1.14, $C([a, b], \mathbf{R})$ and $C([a, b], \mathbf{C})$ are the spaces of continuous real and complex-valued functions on $[a, b]$. This uses the restriction of the standard metric on $\mathbf{R}$ to $[a, b]$, and the standard Euclidean metrics on $\mathbf{R}$ and $\mathbf{C}$, as appropriate. As before, these spaces are linear subspaces of the real and complex vector spaces of all real and complex-valued functions on $[a, b]$, as appropriate.

As usual, a nonnegative real-valued function N on $C([a, b], \mathbf{R})$ or $C([a, b], \mathbf{C})$ is said to be a *norm* if it satisfies the following three conditions. First, $N(f) = 0$ if and only if $f = 0$. Second, if f is a continuous real or complex-valued function on $[a, b]$ and $t \in \mathbf{R}$ of $\mathbf{C}$, as appropriate, then

$$N(t\, f) = |t|\, N(f). \tag{3.1.2}$$

Third, if f and g are continuous real or complex-valued functions on $[a, b]$, as appropriate, then

$$N(f + g) \le N(f) + N(g). \tag{3.1.3}$$

In this case,

$$d_N(f, g) = N(f - g) \tag{3.1.4}$$

defines a metric on $C([a, b], \mathbf{R})$ or $C([a, b], \mathbf{C})$, as appropriate.

Remember that continuous real and complex-valued functions on $[a, b]$ are bounded, because $[a, b]$ is compact. The *supremum norm* is defined on each of $C([a, b], \mathbf{R})$ and $C([a, b], \mathbf{C})$ by

$$\|f\|_\infty = \sup_{a \le x \le b} |f(x)|, \tag{3.1.5}$$

as in (1.13.4). Note that the supremum is attained in this situation, because of the extreme value theorem. The corresponding metric

$$d_\infty(f, g) = \|f - g\|_\infty \tag{3.1.6}$$

is the same as the supremum metric, as before.

If f is a continuous real or complex-valued function on $[a, b]$, then put

$$\|f\|_1 = \int_a^b |f(x)| \, dx, \tag{3.1.7}$$

using the standard Riemann integral on the right side. This is the same as (1.15.4) in the case of the unit interval in $\mathbf{R}$. As before, one can check that (3.1.7) defines a norm on each of $C([a, b], \mathbf{R})$ and $C([a, b], \mathbf{C})$. Let

$$d_1(f, g) = \|f - g\|_1 = \int_a^b |f(x) - g(x)| \, dx \tag{3.1.8}$$

be the metric associated to (3.1.7) on each of $C([a, b], \mathbf{R})$ and $C([a, b], \mathbf{C})$.

If f is a continuous real or complex-valued function on $[a, b]$, then

$$\|f\|_1 = \int_a^b |f(x)| \, dx \le (b - a) \, \|f\|_\infty. \tag{3.1.9}$$

Hence

$$d_1(f, g) = \|f - g\|_1 \le (b - a) \, \|f - g\|_\infty = (b - a) \, d_\infty(f, g) \tag{3.1.10}$$

for all real or complex-valued continuous functions f and g on $[a, b]$. This implies that the identity mappings on $C([a, b], \mathbf{R})$ and $C([a, b], \mathbf{C})$ are Lipschitz with respect to $d_\infty(f, g)$ on the domain and $d_1(f, g)$ on the range.

3.2 The square norm

Let a, b be real numbers with $a < b$ again. Put

$$\|f\|_2 = \Big(\int_a^b |f(x)|^2 \, dx \Big)^{1/2} \tag{3.2.1}$$

for each continuous real or complex-valued continuous function f on $[a, b]$, using the nonnegative square root on the right side, as usual. Note that this is the same as (1.15.6) in the case of the unit interval. It is easy to see that this satisfies the homogeneity and positivity requirements of a norm, as before. To get the triangle inequality, one can use an integral version of the Cauchy–Schwarz inequality, as follows.

Let f, g be continuous real or complex-valued functions on $[a, b]$. Observe that

$$|f(x)| \, |g(x)| \le \frac{1}{2} \, |f(x)|^2 + \frac{1}{2} \, |g(x)|^2 \tag{3.2.2}$$

for every $a \le x \le b$, as in (2.3.6). This implies that

$$\int_a^b |f(x)|\,|g(x)|\,dx \le \frac{1}{2}\int_a^b |f(x)|^2\,dx + \frac{1}{2}\int_a^b |g(x)|^2\,dx = \frac{1}{2}\,\|f\|_2^2 + \frac{1}{2}\,\|g\|_2^2. \tag{3.2.3}$$

Using this, we can get that

$$\int_a^b |f(x)|\,|g(x)|\,dx \le \|f\|_2\,\|g\|_2, \tag{3.2.4}$$

which is the integral version of the Cauchy–Schwarz inequality mentioned in the preceding paragraph. More precisely, (3.2.4) follows from (3.2.3) when $\|f\|_2 = \|g\|_2 = 1$. If $\|f\|_2, \|g\|_2 > 0$, then one can reduce to the previous case, using scalar multiplication. Otherwise, (3.2.4) is trivial when $f = 0$ or $g = 0$ on $[a, b]$.

Clearly

$$\begin{aligned} \|f+g\|_2^2 &= \int_a^b |f(x)+g(x)|^2\,dx \\ &\le \int_a^b (|f(x)| + |g(x)|)^2\,dx \\ &= \int_a^b |f(x)|^2\,dx + 2\int_a^b |f(x)|\,|g(x)|\,dx + \int_a^b |g(x)|^2\,dx. \end{aligned} \tag{3.2.5}$$

It follows that

$$\|f+g\|_2^2 \le \|f\|_2^2 + 2\,\|f\|_2\,\|g\|_2 + \|g\|_2^2 = (\|f\|_2 + \|g\|_2)^2, \tag{3.2.6}$$

using (3.2.4) in the first step. Thus

$$\|f+g\|_2 \le \|f\|_2 + \|g\|_2, \tag{3.2.7}$$

so that (3.2.1) defines a norm on each of $C([a,b],\mathbf{R})$ and $C([a,b],\mathbf{C})$. The associated metric is given by

$$d_2(f,g) = \|f-g\|_2 = \Big(\int_a^b |f(x)-g(x)|^2\,dx\Big)^{1/2}. \tag{3.2.8}$$

If f is a continuous real or complex-valued function on $[a, b]$, then

$$\|f\|_2^2 = \int_a^b |f(x)|^2\,dx \le (b-a)\,\|f\|_\infty^2. \tag{3.2.9}$$

Equivalently,

$$\|f\|_2 \le (b-a)^{1/2}\,\|f\|_\infty. \tag{3.2.10}$$

This implies that

$$d_2(f,g) \le (b-a)^{1/2}\,d_\infty(f,g) \tag{3.2.11}$$

for all real or complex-valued continuous functions f and g on $[a, b]$. In particular, the identity mappings on $C([a, b], \mathbf{R})$ and $C([a, b], \mathbf{C})$ are Lipschitz with respect to $d_\infty(f, g)$ on the domain and $d_2(f, g)$ on the range.

Observe that

$$\|f\|_1 = \int_a^b |f(x)|\, dx \le (b-a)^{1/2}\, \|f\|_2 \tag{3.2.12}$$

for every continuous real or complex-valued function f on $[a, b]$, by (3.2.4). Hence

$$d_1(f, g) \le (b-a)^{1/2}\, d_2(f, g) \tag{3.2.13}$$

for all continuous real or complex-valued functions f and g on $[a, b]$. It follows that the identity mappings on $C([a, b], \mathbf{R})$ and $C([a, b], \mathbf{C})$ are Lipschitz with respect to $d_2(f, g)$ on the domain and $d_1(f, g)$ on the range.

3.3 Riemann–Stieltjes integrals

Let a, b be real numbers with $a < b$, and let α be a monotonically increasing real-valued function on $[a, b]$. If f is a continuous real-valued function on $[a, b]$, then the corresponding *Riemann–Stieltjes integral*

$$\int_a^b f\, d\alpha = \int_a^b f(x)\, d\alpha(x) \tag{3.3.1}$$

can be defined as a real number in a natural way. This reduces to the ordinary Riemann integral of f on $[a, b]$ when $\alpha(x) = x$ for every $x \in [a, b]$.

To be a bit more precise, let $\mathcal{P} = \{t_j\}_{j=0}^l$ be a *partition* of $[a, b]$. This is a finite sequence of real numbers with

$$a = t_0 < t_1 < \cdots < t_{l-1} < t_l = b, \tag{3.3.2}$$

as usual. In the definition of the Riemann integral, one considers certain sums associated to such a partition, involving the values of f on each interval $[t_{j-1}, t_j]$, and the length of the interval. In the definition of the Riemann–Stieltjes integral, one considers sums of a similar type, but using

$$\alpha(t_j) - \alpha(t_{j-1}) \tag{3.3.3}$$

instead of the length of $[t_{j-1}, t_j]$. Note that (3.3.3) is greater than or equal to 0 for each j, and that

$$\sum_{j=1}^{l} (\alpha(t_j) - \alpha(t_{j-1})) = \alpha(b) - \alpha(a). \tag{3.3.4}$$

Riemann–Stieltjes integrals are discussed in many textbooks, as well as the article [86]. See [28] for a geometric interpretation of the Riemann–Stieltjes integral, and [153, 177] for some variants of the Riemann–Stieltjes integral.

The Riemann–Stieltjes integral has many of the same properties as the Riemann integral. In particular, (3.3.1) is greater than or equal to 0 when $f \geq 0$ on $[a, b]$. In this case, if $f(x_0) > 0$ for some $x_0 \in [a, b]$, and if α is not constant near x_0, then (3.3.1) is positive. If α is strictly increasing on $[a, b]$, then this works for any $x_0 \in [a, b]$. If α is constant on $[a, b]$, then (3.3.1) is equal to 0 for any function f on $[a, b]$.

If α is continuously-differentiable on $[a, b]$, then it is well known that

$$\int_a^b f(x)\, d\alpha(x) = \int_a^b f(x)\, \alpha'(x)\, dx. \tag{3.3.5}$$

In fact, this works when α is differentiable at every point in $[a, b]$, and its derivative α' is Riemann integrable on $[a, b]$. Although monotonically increasing functions on $[a, b]$ are not necessarily continuous, it is well known that their only possible discontinuities are jump discontinuities. It is also well known that such a function can have only finitely or countably many discontinuities in $[a, b]$. It is interesting to consider the Riemann–Stieltjes integral when α is a monotonically increasing step function on $[a, b]$, for instance.

It is well known that the Riemann integral of a bounded real-valued function f on $[a, b]$ can be defined when f is Riemann integrable on $[a, b]$, in a suitable sense, which includes the case where f is continuous on $[a, b]$. Similarly, (3.3.1) can be defined when f is Riemann–Stieltjes integrable on $[a, b]$ with respect to α, in a suitable sense, which includes the case when f is continuous on $[a, b]$. If f is a complex-valued function on $[a, b]$, then one can consider the corresponding integrability properties of the real and imaginary parts of f. If the real and imaginary parts of f are Riemann–Stieltjes integrable on $[a, b]$ with respect to α, then one can define the Riemann–Stieltjes integral of f as a complex number, whose real and imaginary parts are the corresponding Riemann–Stieltjes integrals of the real and imaginary parts of f. In particular, this works when f is continuous on $[a, b]$.

3.4 Riemann–Stieltjes integrals and seminorms

Let a, b be real numbers with $a < b$, and let N be a nonnegative real-valued function on $C([a, b], \mathbf{R})$ or $C([a, b], \mathbf{C})$. As in Section A.6, N is said to be a *seminorm* if

$$N(t\, f) = |t|\, N(f) \tag{3.4.1}$$

and

$$N(f + g) \leq N(f) + N(g) \tag{3.4.2}$$

for every $f, g \in C([a, b], \mathbf{R})$ or $C([a, b], \mathbf{C})$ and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate. In this case, $N(0) = 0$, by taking $t = 0$ in (3.4.1). If we also have that $N(f) > 0$ when $f \not\equiv 0$ on $[a, b]$, then N is a norm on $C([a, b], \mathbf{R})$ or $C([a, b], \mathbf{C})$, as appropriate.

Let α be a monotonically increasing real-valued function on $[a, b]$. If f is a continuous real or complex-valued function on $[a, b]$, then

$$\|f\|_{1,\alpha} = \int_a^b |f(x)| \, d\alpha(x) \tag{3.4.3}$$

is defined as a nonnegative real number, as in the previous section. Of course, this is the same as (3.1.7) when $\alpha(x) = x$ on $[a, b]$. It is easy to see that

$$\|t\, f\|_{1,\alpha} = |t| \, \|f\|_{1,\alpha} \tag{3.4.4}$$

and

$$\|f + g\|_{1,\alpha} \le \|f\|_{1,\alpha} + \|g\|_{1,\alpha} \tag{3.4.5}$$

for all continuous real or complex-valued functions f and g on $[a, b]$ and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate. Thus (3.4.3) defines a seminorm on each of $C([a, b], \mathbf{R})$ and $C([a, b], \mathbf{C})$.

If α is strictly increasing on $[a, b]$, and if f is a continuous real or complex-valued function on $[a, b]$ such that $f(x_0) \ne 0$ for some $x_0 \in [a, b]$, then one can check that (3.4.3) is positive, as before. This implies that (3.4.3) defines a norm on each of $C([a, b], \mathbf{R})$ and $C([a, b], \mathbf{C})$ in this case. It follows that

$$d_{1,\alpha}(f, g) = \|f - g\|_{1,\alpha} \tag{3.4.6}$$

defines a metric on each of $C([a, b], \mathbf{R})$ and $C([a, b], \mathbf{C})$ under these conditions.

If f, g are continuous real or complex-valued function on $[a, b]$, then

$$\begin{aligned} \left| \int_a^b f \, d\alpha - \int_a^b g \, d\alpha \right| &= \left| \int_a^b (f - g) \, d\alpha \right| \\ &\le \int_a^b |f - g| \, d\alpha = \|f - g\|_{1,\alpha}, \end{aligned} \tag{3.4.7}$$

using basic properties of the Riemann–Stieltjes integral in the first two steps. Suppose for the moment that α is strictly increasing on $[a, b]$. It follows from (3.4.7) that

$$f \mapsto \int_a^b f \, d\alpha \tag{3.4.8}$$

is Lipschitz as a mapping from $C([a, b], \mathbf{R})$ or $C([a, b], \mathbf{C})$ into $\mathbf{R}$ or $\mathbf{C}$, as appropriate, with respect to (3.4.6) on the domain and the standard Euclidean metric on the range.

If f is a continuous real or complex-valued function on $[a, b]$, then

$$\|f\|_{1,\alpha} \le (\alpha(b) - \alpha(a)) \, \|f\|_\infty, \tag{3.4.9}$$

where $\|f\|_\infty$ is the supremum norm of f on $[a, b]$, as in (3.1.5). Combining (3.4.7) and (3.4.9), we get that

$$\left| \int_a^b f \, d\alpha - \int_a^b g \, d\alpha \right| \le (\alpha(b) - \alpha(a)) \, \|f - g\|_\infty, \tag{3.4.10}$$

for all continuous real or complex-valued functions f, g on $[a,b]$. This implies that (3.4.8) is Lipschitz as a mapping from $C([a,b],\mathbf{R})$ or $C([a,b],\mathbf{C})$ into $\mathbf{R}$ or $\mathbf{C}$, as appropriate, with respect to the supremum metric on the domain and the standard Euclidean metric on the range.

Let $\{f_j\}_{j=1}^{\infty}$ be a sequence of real or complex-valued functions on $[a,b]$ that converges uniformly to a real or complex-valued function f on $[a,b]$, as appropriate. If f_j is continuous on $[a,b]$ for each j, then f is continuous on $[a,b]$, and

$$\|f_j - f\|_{1,\alpha} \to 0 \text{ as } j \to \infty, \tag{3.4.11}$$

by (3.4.9). This means that $\{f_j\}_{j=1}^{\infty}$ converges to f with respect to (3.4.6) when α is strictly increasing on $[a,b]$. We also get that

$$\lim_{j\to\infty} \int_a^b f_j(x)\, d\alpha(x) = \int_a^b f(x)\, d\alpha(x), \tag{3.4.12}$$

by (3.4.10).

More precisely, if f_j is Riemann–Stieltjes integrable on $[a,b]$ with respect to α for each j, then it is well known that f is Riemann–Stieltjes integrable on $[a,b]$ with respect to α too. In this case, (3.4.12) holds for essentially the same reasons as before. Similarly, one can define (3.4.3) for Riemann–Stieltjes integrable functions on $[a,b]$ with respect to α, and (3.4.11) holds as well.

3.5 Square seminorms

Let a, b be real numbers with $a < b$, and let α be a monotonically increasing real-valued function on $[a,b]$. If f is a continuous real or complex-valued function on $[a,b]$, then

$$\|f\|_{2,\alpha} = \Big(\int_a^b |f(x)|^2\, d\alpha(x)\Big)^{1/2} \tag{3.5.1}$$

is defined as a nonnegative real number. This reduces to (3.2.1) when $\alpha(x) = x$ on $[a,b]$, as before. Note that

$$\|t\, f\|_{2,\alpha} = |t|\, \|f\|_{2,\alpha} \tag{3.5.2}$$

for every $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate.

Let g be another continuous real or complex-valued function on $[a,b]$. Observe that

$$\begin{aligned} \int_a^b |f(x)|\,|g(x)|\, d\alpha(x) &\le \frac{1}{2}\int_a^b |f(x)|^2\, d\alpha(x) + \frac{1}{2}\int_a^b |g(x)|^2\, d\alpha(x) \\ &= \frac{1}{2}\,\|f\|_{2,\alpha}^2 + \frac{1}{2}\,\|g\|_{2,\alpha}^2, \end{aligned} \tag{3.5.3}$$

using (3.2.2) in the first step. In particular,

$$\int_a^b |f(x)|\,|g(x)|\, d\alpha(x) \le 1 \tag{3.5.4}$$

when $\|f\|_{2,\alpha}, \|g\|_{2,\alpha} \le 1$.

Let f, g be arbitrary continuous real or complex-valued functions on $[a,b]$, and let r, t be positive real numbers such that

$$\|f\|_{2,\alpha} \le r, \quad \|g\|_{2,\alpha} \le t. \tag{3.5.5}$$

Thus $\|r^{-1}\, f\|_{2,\alpha}, \|t^{-1}\, g\|_{2,\alpha} \le 1$, so that

$$\int_a^b |r^{-1}\, f(x)|\, |t^{-1}\, g(x)|\, d\alpha(x) \le 1. \tag{3.5.6}$$

Equivalently, this means that

$$\int_a^b |f(x)|\, |g(x)|\, d\alpha(x) \le r\, t. \tag{3.5.7}$$

One can use this to get that

$$\int_a^b |f(x)|\, |g(x)|\, d\alpha(x) \le \|f\|_{2,\alpha}\, \|g\|_{2,\alpha}, \tag{3.5.8}$$

which is the integral version of the Cauchy–Schwarz inequality for Riemann–Stieltjes integrals on $[a,b]$.

As in Section 3.2, one can show that

$$\|f+g\|_{2,\alpha} \le \|f\|_{2,\alpha} + \|g\|_{2,\alpha}, \tag{3.5.9}$$

using (3.5.8). This implies that (3.5.1) defines a seminorm on each of $C([a,b],\mathbf{R})$ and $C([a,b],\mathbf{C})$. If α is strictly increasing on $[a,b]$, and if f is a continuous real or complex-valued function on $[a,b]$ that is not equal to 0 everywhere on $[a,b]$, then one can verify that (3.5.1) is positive, as usual. In this case, (3.5.1) defines a norm on each of $C([a,b],\mathbf{R})$ and $C([a,b],\mathbf{C})$. Thus

$$d_{2,\alpha}(f,g) = \|f-g\|_{2,\alpha} \tag{3.5.10}$$

defines a metric on each of these spaces in this situation.

Observe that

$$\|f\|_{2,\alpha} \le (\alpha(b)-\alpha(a))^{1/2}\, \|f\|_\infty \tag{3.5.11}$$

for every continuous real or complex-valued function f on $[a,b]$. Of course, this implies that

$$\|f-g\|_{2,\alpha} \le (\alpha(b)-\alpha(a))^{1/2}\, \|f-g\|_\infty \tag{3.5.12}$$

for all continuous real or complex-valued functions f, g on $[a,b]$. Similarly,

$$\|f\|_{1,\alpha} \le (\alpha(b)-\alpha(a))^{1/2}\, \|f\|_{2,\alpha}, \tag{3.5.13}$$

by (3.5.8). It follows that

$$\|f-g\|_{1,\alpha} \le (\alpha(b)-\alpha(a))^{1/2}\, \|f-g\|_{2,\alpha} \tag{3.5.14}$$

for every $f,g \in C([a,b],\mathbf{R})$ or $C([a,b],\mathbf{C})$.

3.6 Some more monotone functions

Let a, b be real numbers with $a < b$, and let α, β be monotonically increasing functions real-valued functions on $[a, b]$. Thus $\alpha+\beta$ is a monotonically increasing real-valued function on $[a, b]$ as well. If f is a continuous real-valued function on $[a, b]$, then the Riemann–Stieltjes integrals of f on $[a, b]$ with respect to α, β, $\alpha + \beta$ may be defined in the usual way. It is well known that

(3.6.1) $$\int_a^b f\, d(\alpha + \beta) = \int_a^b f\, d\alpha + \int_a^b f\, d\beta.$$

In particular, if $f \geq 0$ on $[a, b]$, then

(3.6.2) $$\int_a^b f\, d\alpha,\ \int_a^b f\, d\beta \leq \int_a^b f\, d(\alpha + \beta).$$

Suppose now that f is a continuous real or complex-valued function on $[a, b]$, so that $\|f\|_{1,\alpha}$ and $\|f\|_{2,\alpha}$ may be defined as in the previous two sections. Of course, the analogous quantities

(3.6.3) $$\|f\|_{1,\beta},\ \|f\|_{2,\beta},\ \|f\|_{1,\alpha+\beta},\ \|f\|_{2,\alpha+\beta}$$

associated to β and $\alpha + \beta$ may be defined in the same way. Observe that

(3.6.4) $$\|f\|_{1,\alpha+\beta} = \|f\|_{1,\alpha} + \|f\|_{1,\beta},$$

because of (3.6.1). Similarly,

(3.6.5) $$\|f\|_{2,\alpha+\beta}^2 = \|f\|_{2,\alpha}^2 + \|f\|_{2,\beta}^2.$$

It follows that

(3.6.6) $$\|f\|_{1,\alpha},\ \|f\|_{1,\beta} \leq \|f\|_{1,\alpha+\beta}$$

and

(3.6.7) $$\|f\|_{2,\alpha},\ \|f\|_{2,\beta} \leq \|f\|_{2,\alpha+\beta}$$

Let t be a nonnegative real number, so that $t\,\alpha$ is a monotonically increasing real-valued function on $[a, b]$ too. If f is a continuous real-valued function on $[a, b]$, then the Riemann–Stieltjes integral of f on $[a, b]$ with respect to $t\,\alpha$ may be defined in the usual way. It is well known that

(3.6.8) $$\int_a^b f\, d(t\,\alpha) = t \int_a^b f\, d\alpha.$$

If f is a continuous real or complex-valued function on $[a, b]$, then

(3.6.9) $$\|f\|_{1,t\,\alpha} \text{ and } \|f\|_{2,t\,\alpha}$$

may be defined as in the two previous sections again. It is easy to see that

(3.6.10) $$\|f\|_{1,t\,\alpha} = t\,\|f\|_{1,\alpha}$$

and

(3.6.11) $$\|f\|_{2,t\,\alpha} = \sqrt{t}\,\|f\|_{2,\alpha},$$

using (3.6.8).

3.7 Compact support

Let $(X, d(x, y))$ be a metric space, and let f be a real or complex-valued function on X. The *support* of f in X is defined by

$$\operatorname{supp} f = \overline{\{x \in X : f(x) \neq 0\}}, \tag{3.7.1}$$

which is to say the closure of the set of $x \in X$ with $f(x) \neq 0$ with respect to $d(\cdot, \cdot)$. Note that this is different from the definition of the support used in Section 1.6.

Suppose for the moment that $d(x, y)$ is the *discrete metric* on X, which is equal to 1 when $x \neq y$, and to 0 when $x = y$. In this case, it is easy to see that every subset of X is a closed set. Similarly, the closure of any subset E of X is the same as E. This means that (3.7.1) is the same as the previous definition of the support in Section 1.6 under these conditions.

If $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then

$$\begin{aligned} \{x \in X : t\, f(x) \neq 0\} &= \{x \in X : f(x) \neq 0\} \quad \text{when } t \neq 0 \\ &= \emptyset \quad \text{when } t = 0. \end{aligned} \tag{3.7.2}$$

This implies that

$$\begin{aligned} \operatorname{supp}(t\, f) &= \operatorname{supp} f \quad \text{when } t \neq 0 \\ &= \emptyset \quad \text{when } t = 0. \end{aligned} \tag{3.7.3}$$

If g is another real or complex-valued function on X, then it is easy to see that

$$\{x \in X : f(x) + g(x) \neq 0\} \subseteq \{x \in X : f(x) \neq 0\} \cup \{x \in X : g(x) \neq 0\}. \tag{3.7.4}$$

This implies that

$$\operatorname{supp}(f + g) \subseteq (\operatorname{supp} f) \cup (\operatorname{supp} g). \tag{3.7.5}$$

More precisely, this also uses the well-known fact that

$$\overline{(E_1 \cup E_2)} \subseteq \overline{E_1} \cup \overline{E_2} \tag{3.7.6}$$

for any two subsets E_1, E_2 of X.

Of course, f is said to have *compact support* in X if the support of f is a compact subset of X. Suppose that there is a compact subset K of X such that

$$\{x \in X : f(x) \neq 0\} \subseteq K. \tag{3.7.7}$$

This implies that

$$\operatorname{supp} f \subseteq K, \tag{3.7.8}$$

because compact subsets of metric spaces are closed sets. If E is a closed subset of X such that $E \subseteq K$, then it is well known that E is also compact in X. Thus (3.7.8) implies that $\operatorname{supp} f$ is compact, because $\operatorname{supp} f$ is a closed set in X, by construction.

If f has compact support in X, then $t\,f$ has compact support in X for every $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, because of (3.7.3). Let g be another real or complex-valued function on X, as appropriate, with compact support. It is well known and easy to see that the union of any two compact subsets of X is also compact, so that the union of $\operatorname{supp} f$ and $\operatorname{supp} g$ is compact. Thus (3.7.5) implies that the support of $f + g$ is contained in a compact subset of X. It follows that the support of $f + g$ is compact, as in the preceding paragraph.

Suppose that $X \neq \emptyset$, and let $C_{com}(X, \mathbf{R})$ and $C_{com}(X, \mathbf{C})$ be the spaces of continuous real and complex-valued functions on X with compact support, respectively. These are linear subspaces of the real and complex vector spaces $C(X, \mathbf{R})$ and $C(X, \mathbf{C})$ of continuous real and complex-valued functions on X, respectively, by the remarks in the previous paragraph. Of course, if X is compact, then every real or complex-valued function on X has compact support.

Suppose that f is a continuous real or complex-valued function on X with compact support. This implies that

$$f(\operatorname{supp} f) \tag{3.7.9}$$

is a compact subset of $\mathbf{R}$ or $\mathbf{C}$, as appropriate, because continuous mappings send compact sets to compact sets. In particular, this means that f is bounded on $\operatorname{supp} f$, because compact subsets of metric spaces are bounded. It follows that f is bounded on X, because f is equal to 0 on the complement of $\operatorname{supp} f$, by construction. Thus

$$C_{com}(X, \mathbf{R}) \subseteq C_b(X, \mathbf{R}), \quad C_{com}(X, \mathbf{C}) \subseteq C_b(X, \mathbf{C}), \tag{3.7.10}$$

where $C_b(X, Y)$ is as in Section 1.14.

Suppose for the moment again that X is equipped with the discrete metric. In this case, the only compact subsets of X are those with only finitely many elements. If Y is any metric space, then every mapping from X into Y is continuous. Hence $C_{com}(X, \mathbf{R})$ and $C_{com}(X, \mathbf{C})$ are the same as the spaces $c_{00}(X, \mathbf{R})$ and $c_{00}(X, \mathbf{C})$ defined in Section 1.6 in this situation.

3.8 Functions on R

In this section, we take the real line to be equipped with the standard Euclidean metric, as usual. Let f be a real or complex-valued function on $\mathbf{R}$. If f has compact support in $\mathbf{R}$, then there are real numbers a, b with $a < b$ and

$$\operatorname{supp} f \subseteq [a, b], \tag{3.8.1}$$

because compact subsets of $\mathbf{R}$ are bounded. Conversely, if (3.8.1) holds for some $a, b \in \mathbf{R}$ with $a \leq b$, then $\operatorname{supp} f$ is compact, because $[a, b]$ is a compact subset of $\mathbf{R}$. Remember that the spaces $C_{com}(\mathbf{R}, \mathbf{R})$ and $C_{com}(\mathbf{R}, \mathbf{C})$ of continuous real and complex-valued functions on $\mathbf{R}$ with compact support are linear subspaces of the real and complex vector spaces of all continuous real and complex-valued functions on $\mathbf{R}$, respectively, as in the previous section.

A nonnegative real-valued function N on $C_{com}(\mathbf{R}, \mathbf{R})$ or $C_{com}(\mathbf{R}, \mathbf{C})$ is said to be a *norm* if it satisfies the following three conditions, as usual. First, $N(f) = 0$ if and only if $f = 0$. Second, if f is a continuous real or complex-valued function on $\mathbf{R}$ with compact support and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then

$$N(t\,f) = |t|\,N(f). \tag{3.8.2}$$

Third, if f and g are continuous real or complex-valued functions on $\mathbf{R}$ with compact support, as appropriate, then

$$N(f+g) \le N(f) + N(g). \tag{3.8.3}$$

This implies that

$$d_N(f,g) = N(f-g) \tag{3.8.4}$$

defines a metric on $C_{com}(\mathbf{R}, \mathbf{R})$ or $C_{com}(\mathbf{R}, \mathbf{C})$, as appropriate.

Remember that continuous real and complex-valued functions on $\mathbf{R}$ with compact support are bounded, as in the previous section. The *supremum norm* is defined on each of $C_{com}(\mathbf{R}, \mathbf{R})$ and $C_{com}(\mathbf{R}, \mathbf{C})$ by

$$\|f\|_\infty = \sup_{x \in \mathbf{R}} |f(x)|, \tag{3.8.5}$$

as in (1.13.4). Equivalently,

$$\|f\|_\infty = \sup\{|f(x)| : x \in \operatorname{supp} f\} \tag{3.8.6}$$

when $\operatorname{supp} f \ne \emptyset$, in which case the supremum is attained, by the extreme value theorem. If (3.8.1) holds for some $a, b \in \mathbf{R}$ with $a \le b$, then

$$\|f\|_\infty = \sup_{a \le x \le b} |f(x)|. \tag{3.8.7}$$

As before, the corresponding metric

$$d_\infty(f,g) = \|f-g\|_\infty \tag{3.8.8}$$

is the same as the supremum metric.

Let f be a continuous real or complex-valued function on $\mathbf{R}$ with compact support, and let a, b be real numbers such that $a \le b$ and (3.8.1) holds. Under these conditions, one can define the integral

$$\int_{-\infty}^{\infty} f(x)\,dx \tag{3.8.9}$$

to be the Riemann integral

$$\int_a^b f(x)\,dx. \tag{3.8.10}$$

More precisely, one can check that (3.8.10) does not depend on the particular choices of a and b, as long as (3.8.1) holds. If g is another continuous real or complex-valued function on $\mathbf{R}$, as appropriate, with compact support, then

$$\int_{-\infty}^{\infty} (f(x) + g(x))\,dx = \int_{-\infty}^{\infty} f(x)\,dx + \int_{-\infty}^{\infty} g(x)\,dx. \tag{3.8.11}$$

To see this, one can choose $a, b \in \mathbf{R}$ such that $a < b$ and the supports of both f and g are contained in $[a, b]$, so that the support of $f + g$ is contained in $[a, b]$ too. In this case, (3.8.11) reduces to the well-known fact that

$$\int_{a}^{b} (f(x) + g(x))\,dx = \int_{a}^{b} f(x)\,dx + \int_{a}^{b} g(x)\,dx. \tag{3.8.12}$$

Of course, other basic properties of (3.8.9) follow from the corresponding statements for (3.8.10).

If f is a continuous real or complex-valued function on $\mathbf{R}$ with compact support, then put

$$\|f\|_1 = \int_{-\infty}^{\infty} |f(x)|\,dx, \tag{3.8.13}$$

where the integral on the right is defined as in the preceding paragraph. One can check that this defines a norm on each of $C_{com}(\mathbf{R}, \mathbf{R})$ and $C_{com}(\mathbf{R}, \mathbf{C})$, in essentially the same way as for functions on an interval. This leads to a metric

$$d_1(f, g) = \|f - g\|_1 = \int_{-\infty}^{\infty} |f(x) - g(x)|\,dx \tag{3.8.14}$$

on each of $C_{com}(\mathbf{R}, \mathbf{R})$ and $C_{com}(\mathbf{R}, \mathbf{C})$, as in (3.8.8).

Similarly, put

$$\|f\|_2 = \Big(\int_{-\infty}^{\infty} |f(x)|^2\,dx\Big)^{1/2} \tag{3.8.15}$$

for every continuous real or complex-valued function f on $\mathbf{R}$ with compact support, using the nonnegative square root on the right side. One can verify that this defines a norm on each of $C_{com}(\mathbf{R}, \mathbf{R})$ and $C_{com}(\mathbf{R}, \mathbf{C})$, using the analogous statements for functions on an interval. Thus

$$d_2(f, g) = \Big(\int_{-\infty}^{\infty} |f(x) - g(x)|^2\,dx\Big)^{1/2} \tag{3.8.16}$$

defines a metric on each of $C_{com}(\mathbf{R}, \mathbf{R})$ and $C_{com}(\mathbf{R}, \mathbf{C})$.

Let f be any continuous real or complex-valued function on $\mathbf{R}$ with compact support. Observe that

$$\|f\|_2^2 = \int_{-\infty}^{\infty} |f(x)|^2\,dx \le \|f\|_\infty \int_{-\infty}^{\infty} |f(x)|\,dx = \|f\|_1\,\|f\|_\infty. \tag{3.8.17}$$

3.9 Dense sets and Weierstrass' theorem

Let $(M, d(x, y))$ be a metric space, and let E be a subset of M. Remember that E is said to be *dense* in X if every element of M is an element of E, or a limit point of E, or both. This is the same as saying that the closure $\overline{E}$ of E in M is equal to M. Equivalently, it is easy to see that E is dense in M if and only if for every $x \in M$ and $\epsilon > 0$ there is a $y \in E$ such that

$$d(x, y) < \epsilon. \tag{3.9.1}$$

Alternatively, one can verify that E is dense in M if and only if for every $x \in M$ there is a sequence $\{x_j\}_{j=1}^\infty$ of elements of E that converges to x in M.

Let (X, d_X) and (Y, d_Y) be nonempty metric spaces, and let $\theta(\cdot, \cdot)$ be the corresponding supremum metric on the space $C_b(X, Y)$ of bounded continuous mappings from X into Y, as in Sections 1.11 and 1.14. A subset E of $C_b(X, Y)$ is dense in $C_b(X, Y)$ with respect to $\theta(\cdot, \cdot)$ if and only if for every $f \in C_b(X, Y)$ and $\epsilon > 0$ there is a $g \in E$ such that

$$\theta(f, g) < \epsilon, \tag{3.9.2}$$

as in the preceding paragraph. Note that (3.9.2) implies that

$$d_Y(f(x), g(x)) < \epsilon \text{ for every } x \in X. \tag{3.9.3}$$

If (3.9.3) holds, then

$$\theta(f, g) \le \epsilon, \tag{3.9.4}$$

by the definition of the supremum metric. Using this, we get that E is dense in $C_b(X, Y)$ with respect to $\theta(\cdot, \cdot)$ if and only if for every $f \in C_b(X, Y)$ and $\epsilon > 0$ there is a $g \in E$ such that (3.9.3) holds.

We also have that E is dense in $C_b(X, Y)$ with respect to $\theta(\cdot, \cdot)$ if and only if for every $f \in C_b(X, Y)$ there is a sequence $\{f_j\}_{j=1}^\infty$ of elements of E that converges to f uniformly on X. This uses the equivalence of uniform convergence and convergence with respect to $\theta(\cdot, \cdot)$ for sequences of bounded mappings from X into Y, as in Section 1.11.

Let a, b be real numbers with $a < b$, and let f be a continuous complex-valued function on $[a, b]$. Of course, this uses the restriction of the standard Euclidean metric on $\mathbf{R}$ to $[a, b]$, and the standard Euclidean metric on $\mathbf{C}$. A famous theorem of Weierstrass states that there is a sequence of polynomials on $\mathbf{R}$ with complex coefficients that converges to f uniformly on $[a, b]$. If f is a continuous real-valued function on $[a, b]$, then one can use polynomials on $\mathbf{R}$ with real coefficients. This means that the set of functions on $[a, b]$ corresponding to polynomials with complex coefficients is dense in the space $C([a, b], \mathbf{C})$ of continuous complex-valued functions on $[a, b]$, with respect to the supremum metric. Similarly, the set of functions on $[a, b]$ corresponding to polynomials with real coefficients is dense in the space $C([a, b], \mathbf{R})$ of continuous real-valued functions on $[a, b]$, with respect to the supremum metric.

Suppose now that (X, d_X) is a nonempty compact metric space. The Stone–Weierstrass theorem gives a criterion for a set of continuous real or complex-valued functions on X to be dense in $C(X, \mathbf{R})$ or $C(X, \mathbf{C})$, as appropriate, with respect to the supremum metric. This will be discussed in Section 3.11, after some preliminaries in the next section.

3.10 Functions on sets

Let X be a nonempty set, and let

$$c(X, \mathbf{R}),\ c(X, \mathbf{C}) \tag{3.10.1}$$

be the spaces of all real and complex-valued functions on X, respectively. Of course, if f and g are real or complex-valued functions on X, then $f + g$ is a real or complex-valued function on X, as appropriate. Similarly, if f is a real or complex-valued function on X, and t is a real or complex number, as appropriate, then $t\,f$ is a real or complex-valued function on X, as appropriate. These are basic (classes of) examples of vector spaces over the real or complex numbers.

A subset W of $c(X, \mathbf{R})$ or $c(X, \mathbf{C})$ is said to be a *linear subspace* of $c(X, \mathbf{R})$ or $c(X, \mathbf{C})$, as appropriate, if $0 \in W$, and W satisfies the following two additional conditions. First, if $f, g \in W$, then

$$f + g \in W. \tag{3.10.2}$$

Second, if $f \in W$ and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then

$$t\,f \in W. \tag{3.10.3}$$

If f and g are real or complex-valued functions on X, then their product $f(x)\,g(x)$ defines a real or complex-valued function on X, as appropriate, too. In fact, $c(X, \mathbf{R})$, $c(X, \mathbf{C})$ are basic (classes of) examples of *commutative rings*, although we shall not discuss this in detail here. More precisely, these are *commutative algebras* over the real and complex numbers, respectively.

Let $\mathcal{A}$ be a linear subspace of $c(X, \mathbf{R})$ or $c(X, \mathbf{C})$. Suppose that for every $f, g \in \mathcal{A}$, we have that

$$f\,g \in \mathcal{A}. \tag{3.10.4}$$

Under these conditions, $\mathcal{A}$ is said to be a *subalgebra* of $c(X, \mathbf{R})$ or $c(X, \mathbf{C})$, as appropriate.

If X is equipped with a metric, then the corresponding spaces $C(X, \mathbf{R})$ and $C(X, \mathbf{C})$ of continuous real and complex-valued functions on X are subalgebras of $c(X, \mathbf{R})$ and $c(X, \mathbf{C})$, respectively. If X is equipped with the discrete metric, then any mapping from X into another metric space is continuous. In particular, this means that

$$C(X, \mathbf{R}) = c(X, \mathbf{R}) \text{ and } C(X, \mathbf{C}) = c(X, \mathbf{C}) \tag{3.10.5}$$

in this case.

Let W be a subset of $c(X, \mathbf{R})$ or $c(X, \mathbf{C})$. We say that W is *nowhere vanishing* on X if for every $x \in X$ there is an $f \in W$ such that

$$f(x) \neq 0. \tag{3.10.6}$$

In particular, this holds when W contains a nonzero constant function on X. We say that W *separates points* in X if for every $x, w \in X$ with $x \neq w$ there is an $f \in W$ such that

$$f(x) \neq f(w). \tag{3.10.7}$$

Of course, at least one of $f(x)$ and $f(w)$ is not equal to 0 in this case.

Suppose now that W is a subset of $c(X, \mathbf{C})$. We shall sometimes be concerned with situations where W contains the complex-conjugate of each of its elements. If W is a linear subspace of $c(X, \mathbf{C})$, then this is equivalent to the condition that W contain the real and imaginary parts of all of its elements.

If W is a linear subspace of $c(X, \mathbf{R})$ or $c(X, \mathbf{C})$, then a subset of W may be called a linear subspace of W if it is a linear subspace of $c(X, \mathbf{R})$ or $c(X, \mathbf{C})$, as appropriate. Similarly, if $\mathcal{A}$ is a subalgebra of $c(X, \mathbf{R})$ or $c(X, \mathbf{C})$, then a subset of $\mathcal{A}$ may be called a subalgebra of $\mathcal{A}$ if it is a subalgebra of $c(X, \mathbf{R})$ or $c(X, \mathbf{C})$, as appropriate. In particular, if X is equipped with a metric, then the notions of linear subspaces and subalgebras of $C(X, \mathbf{R})$ and $C(X, \mathbf{C})$ may be defined in this way.

3.11 The Stone–Weierstrass theorem

Let $(X, d(x, y))$ be a metric space, and suppose that X is nonempty and compact. Also let $\mathcal{A}$ be a subalgebra of the algebra $C(X, \mathbf{R})$ of continuous real-valued functions on X, and suppose that

$$\mathcal{A} \text{ is nowhere vanishing on } X, \tag{3.11.1}$$

and that

$$\mathcal{A} \text{ separates points in } X. \tag{3.11.2}$$

Under these conditions, the *Stone–Weierstrass theorem* states that

$$\mathcal{A} \text{ is dense in } C(X, \mathbf{R}), \tag{3.11.3}$$

with respect to the supremum metric. It is easy to see that (3.11.1) and (3.11.2) are necessary for (3.11.3) to hold.

Let a, b be real numbers with $a < b$. Observe that the collection of functions on $[a, b]$ that can be expressed as a polynomial with real coefficients is a subalgebra of the algebra of continuous real-valued functions on $[a, b]$. It is easy to see that this subalgebra is nowhere vanishing on $[a, b]$ and that it separates points in $[a, b]$. Thus Weierstrass' original approximation theorem for approximating real-valued continuous functions on $[a, b]$ by polynomials with real coefficients uniformly on $[a, b]$ can be obtained from the Stone-Weierstrass theorem. However, Weierstrass' original theorem, or at least an interesting case of it, is typically used to show the Stone–Weierstrass theorem.

Let $(X, d(x, y))$ be any nonempty compact metric space again, and now let $\mathcal{A}$ be a subalgebra of the algebra $C(X, \mathbf{C})$ of continuous complex-valued functions on X. Suppose that $\mathcal{A}$ is nowhere vanishing on X, that $\mathcal{A}$ separates points in X, and that

(3.11.4) $\mathcal{A}$ contains the complex conjugate of each of its elements.

In this case, another version of the Stone–Weierstrass theorem states that

$$\mathcal{A} \text{ is dense in } C(X, \mathbf{C}), \tag{3.11.5}$$

with respect to the supremum metric.

This version of the Stone–Weierstrass theorem can be obtained from the previous one, as follows. Note that

(3.11.6) $\mathcal{A}$ contains the real and imaginary parts of all of its elements,

as in the previous section. One can check that

$$\text{the collection of real-valued functions in } \mathcal{A} \tag{3.11.7}$$

is a subalgebra of $C(X, \mathbf{R})$ that satisfies (3.11.1) and (3.11.2). Thus the previous version of the theorem implies that every continuous real-valued function on X can be approximated by elements of (3.11.7), uniformly on X. Using this, it is easy to see that every continuous complex-valued function on X can be approximated by elements of $\mathcal{A}$, uniformly on X.

It is well known that the version of the Stone–Weierstrass theorem for complex-valued functions does not always work without the hypothesis (3.11.6). One can get a counterexample by taking X to be the closed unit disk in the complex plane, and $\mathcal{A}$ to be the algebra of continuous complex-valued functions on the closed unit disk that are holomorphic or complex-analytic on the open unit disk. Alternatively, one can take X to be the unit circle in the complex plane, and $\mathcal{A}$ to be the algebra of continuous complex-valued functions on the circle that have a continuous extension to the closed unit disk that is holomorphic on the open unit disk. There are examples related to these that may be described without using complex analysis.

3.12 Algebras of bounded functions

Let X be a nonempty set, and remember that $\ell^\infty(X, \mathbf{R})$ and $\ell^\infty(X, \mathbf{C})$ denote the spaces of bounded real and complex-valued functions on X, respectively, as in Section 1.13. If f, g are bounded real or complex-valued functions on X, then it is easy to see that their product $f\,g$ is bounded on X as well. More precisely, one can check that the supremum norm of $f\,g$ on X satisfies

$$\|f\,g\|_\infty \le \|f\|_\infty\,\|g\|_\infty. \tag{3.12.1}$$

Thus $\ell^\infty(X, \mathbf{R})$ and $\ell^\infty(X, \mathbf{C})$ are subalgebras of the algebras $c(X, \mathbf{R})$ and $c(X, \mathbf{C})$ of all real and complex-valued functions on X, respectively, as in Section 3.10.

If W is a subset of $\ell^\infty(X, \mathbf{R})$ or $\ell^\infty(X, \mathbf{C})$, then let $\overline{W}$ be the closure of W in $\ell^\infty(X, \mathbf{R})$ or $\ell^\infty(X, \mathbf{C})$, as appropriate, with respect to the supremum metric. If W is a linear subspace of $\ell^\infty(X, \mathbf{R})$ or $\ell^\infty(X, \mathbf{C})$, then one can check that

(3.12.2) $$\overline{W} \text{ is a linear subspace of } \ell^\infty(X, \mathbf{R}) \text{ or } \ell^\infty(X, \mathbf{C}),$$

as appropriate. Similarly, if W is a subalgebra of $\ell^\infty(X, \mathbf{R})$ or $\ell^\infty(X, \mathbf{C})$, then one can verify that

(3.12.3) $$\overline{W} \text{ is a subalgebra of } \ell^\infty(X, \mathbf{R}) \text{ or } \ell^\infty(X, \mathbf{C}),$$

as appropriate.

Suppose now that $(X, d(x, y))$ is a nonempty metric space. The spaces $C_b(X, \mathbf{R})$ and $C_b(X, \mathbf{C})$ of bounded continuous real and complex-valued functions on X are subalgebras of $\ell^\infty(X, \mathbf{R})$ and $\ell^\infty(X, \mathbf{C})$, respectively. Remember that $C_b(X, \mathbf{R})$ and $C_b(X, \mathbf{C})$ are also closed sets in $\ell^\infty(X, \mathbf{R})$ and $\ell^\infty(X, \mathbf{C})$, respectively, with respect to the supremum metric, as in Section 1.14.

Let $\mathcal{A}$ be a subalgebra of $C_b(X, \mathbf{R})$ or $C_b(X, \mathbf{C})$, so that $\mathcal{A}$ may also be considered as a subalgebra of $\ell^\infty(X, \mathbf{R})$ or $\ell^\infty(X, \mathbf{C})$, as appropriate. The closure $\overline{\mathcal{A}}$ of $\mathcal{A}$ in $C_b(X, \mathbf{R})$ or $C_b(X, \mathbf{C})$, as appropriate, is the same as the closure of $\mathcal{A}$ in $\ell^\infty(X, \mathbf{R})$ or $\ell^\infty(X, \mathbf{C})$, as appropriate. In particular, $\overline{\mathcal{A}}$ is a subalgebra of $C_b(X, \mathbf{R})$ or $C_b(X, \mathbf{C})$, as appropriate, as before.

If X is compact, then one can use these remarks to reduce to the case of closed subalgebras of $C(X, \mathbf{R})$ or $C(X, \mathbf{C})$, with respect to the supremum metric, in the Stone-Weierstrass theorem.

3.13 Some remarks about closed subalgebras

Let a be a positive real number. It is well known that the absolute value function on $\mathbf{R}$ can be uniformly approximated on $[-a, a]$ by polynomials p with real coefficients such that

(3.13.1) $$p(0) = 0,$$

as in Corollary 7.27 on p161 of [155]. Of course, Weierstrass' approximation theorem implies that $|\cdot|$ may be approximated uniformly on $[-a, a]$ by polynomials q with real coefficients. The constant term of such an approximation should be small, because $|0| = 0$, and so one can replace q with $q - q(0)$ to get approximations with the additional condition (3.13.1). This type of approximation can also be obtained more directly, without using Weierstrass' theorem, as in Exercise 23 on p169 of [155].

Let X be a nonempty set, and let $\mathcal{A}$ be a subalgebra of the algebra $c(X, \mathbf{R})$ of all real-valued functions on X. If $f \in \mathcal{A}$, and if p is a polynomial with real coefficients that satisfies (3.13.1), then it is easy to see that

(3.13.2) $$p \circ f \in \mathcal{A}.$$

Now let $\mathcal{A}$ be a subalgebra of the algebra $\ell^\infty(X, \mathbf{R})$ of all bounded real-valued functions on X. Suppose that $\mathcal{A}$ is a closed set in $\ell^\infty(X, \mathbf{R})$, with respect to the supremum metric. If $f \in \mathcal{A}$, then it is well known that

$$|f| \in \mathcal{A} \tag{3.13.3}$$

too. This corresponds to Step 1 in the proof of Theorem 7.32 on p162 of [155], which is the Stone–Weierstrass theorem.

To see this, note that there is a nonnegative real number a such that $|f| \le a$ on X, which means that

$$f(X) \subseteq [-a, a]. \tag{3.13.4}$$

The absolute value function on $\mathbf{R}$ can be uniformly approximated on $[-a, a]$ by polynomials p with real coefficients that satisfy (3.13.1), as before, which implies that $|f|$ can be uniformly approximated on X by the corresponding functions $p \circ f$. Thus (3.13.3) can be obtained from (3.13.2), because $\mathcal{A}$ is a closed set in $\ell^\infty(X, \mathbf{R})$, by hypothesis.

If g is another element of $\mathcal{A}$, then it is well known that

$$\max(f, g),\ \min(f, g) \in \mathcal{A} \tag{3.13.5}$$

as well, as in Step 2 on p163 of [155]. This can be obtained by expressing the maximum and minimum of two real numbers in terms of the absolute value function, as in [155].

3.14 Local compactness

A metric space (X, d) is said to be *locally compact* at a point $x \in X$ if there is a positive real number r such that $\overline{B}(x, r)$ is a compact subset of X. Remember that $\overline{B}(x, r)$ is the closed ball in X centered at x with radius r, and that this is a closed set in X.

If there is a continuous real or complex-valued function f on X with compact support such that $f(x) \ne 0$, then X is locally compact at x. More precisely, if $f(x) \ne 0$, then there is an $r > 0$ such that f is nonzero at every point in $\overline{B}(x, r)$, because f is continuous at x. This implies that

$$\overline{B}(x, r) \subseteq \operatorname{supp} f, \tag{3.14.1}$$

and hence that $\overline{B}(x, r)$ is compact when $\operatorname{supp} f$ is compact, because $\overline{B}(x, r)$ is a closed set in X.

If (X, d) is any metric space and $x \in X$, then it is well known that the distance to x defines a continuous real-valued function on X. If r is a positive real number, then one can use this to get continuous nonnegative real-valued functions f on X such that $f(x) > 0$ and

$$\operatorname{supp} f \subseteq \overline{B}(x, r). \tag{3.14.2}$$

More precisely, one can obtain such functions f by composing the distance to x with suitable continuous functions on the real line. If X is locally compact at x, then one can choose $r > 0$ so that $\overline{B}(x, r)$ is compact in X. In this case, (3.14.2) implies that f has compact support in X, as before.

If X is locally compact at every $x \in X$, then X is said to be *locally compact* as a metric space. Of course, if X is compact, then X is automatically locally compact. If X is equipped with the discrete metric, then X is clearly locally compact. Note that $\mathbf{R}^n$ is locally compact with respect to the standard Euclidean metric for every positive integer n.

3.15 Another vanishing condition

Let $(X, d(x, y))$ be a metric space, and let f be a real or complex-valued function on X. We say that f *vanishes at infinity* on X if for every $\epsilon > 0$ there is a compact subset $K(\epsilon)$ of X such that

$$|f(x)| < \epsilon \tag{3.15.1}$$

for every $x \in X \setminus K(\epsilon)$. Note that this is not the same as the definition used in Section 2.5. However, if X is equipped with the discrete metric, then the two notions are equivalent, because the only compact subsets of X are those with only finitely many elements. If X is a compact metric space, then any real or complex-valued function on X automatically vanishes at infinity in this sense.

Equivalently, f vanishes at infinity on X if for every $\epsilon > 0$ there is a compact set $K(\epsilon) \subseteq X$ such that

$$E_\epsilon(f) = \{x \in X : |f(x)| \geq \epsilon\} \subseteq K(\epsilon). \tag{3.15.2}$$

This condition implies that

$$\overline{E_\epsilon(f)} \subseteq K(\epsilon), \tag{3.15.3}$$

where $\overline{E_\epsilon(f)}$ is the closure of $E_\epsilon(f)$ in X, because compact subsets of metric spaces are closed sets. It follows from (3.15.3) that $\overline{E_\epsilon(f)}$ is compact in X, because a closed set in a metric space is compact when it is contained in a compact set. Conversely, if $\overline{E_\epsilon(f)}$ is compact, then one might as well take it to be $K(\epsilon)$. Thus f vanishes at infinity on X if and only if $\overline{E_\epsilon(f)}$ is compact for every $\epsilon > 0$.

Note that

$$\overline{E_\epsilon(f)} \subseteq \operatorname{supp} f \tag{3.15.4}$$

for every $\epsilon > 0$, where $\operatorname{supp} f$ is as in (3.7.1). If f has compact support in X, then it follows that f vanishes at infinity on X.

If f vanishes at infinity on X, then it is easy to see that $t\,f$ vanishes at infinity for every $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate. Let g be another real or complex-valued function on X, as appropriate. If f and g both vanish at infinity on X, then one can check that $f + g$ vanishes at infinity on X too. This uses the fact that the union of two compact subsets of X is compact as well.

Suppose that $X \neq \emptyset$, and let $C_0(X, \mathbf{R})$ and $C_0(X, \mathbf{C})$ be the spaces of continuous real and complex-valued functions on X that vanish at infinity, respectively. These are linear subspaces of the real and complex vector spaces $C(X, \mathbf{R})$ and $C(X, \mathbf{C})$ of continuous real and complex-valued functions on X, respectively, by the remarks in the preceding paragraph. We also have that

$$C_{com}(X, \mathbf{R}) \subseteq C_0(X, \mathbf{R}), \quad C_{com}(X, \mathbf{C}) \subseteq C_0(X, \mathbf{C}) \tag{3.15.5}$$

because functions on X with compact support automatically vanish at infinity, as before. If X is equipped with the discrete metric, then $C_0(X, \mathbf{R})$ and $C_0(X, \mathbf{C})$ are the same as the spaces $c_{00}(X, \mathbf{R})$ and $c_{00}(X, \mathbf{C})$ defined in Section 2.5, respectively. This uses the fact that all functions on X are continuous in this case.

Let f be a continuous real or complex-valued function on X. Note that $E_\epsilon(f)$ is a closed set in X for each $\epsilon > 0$. Thus f vanishes at infinity on X if and only if $E_\epsilon(f)$ is compact for every $\epsilon > 0$. Of course, the restriction of f to any compact set $K \subseteq X$ is bounded, because $f(K)$ is compact, and hence bounded. If $E_\epsilon(f)$ is compact for any $\epsilon > 0$, then it follows that f is bounded on X. In particular, if f vanishes at infinity on X, then f is bounded on X. This means that

$$C_0(X, \mathbf{R}) \subseteq C_b(X, \mathbf{R}), \quad C_0(X, \mathbf{C}) \subseteq C_b(X, \mathbf{C}), \tag{3.15.6}$$

where $C_b(X, Y)$ is as in Section 1.14.

Let f be a continuous real or complex-valued function on X again, and let x be an element of X such that $f(x) \neq 0$. Thus $|f(x)| > \epsilon$ for some $\epsilon > 0$, and the continuity of f at x implies that there is an $r > 0$ such that $|f(y)| \geq \epsilon$ for every $y \in X$ with $d(x, y) \leq r$. Equivalently, this means that

$$\overline{B}(x, r) \subseteq E_\epsilon(f), \tag{3.15.7}$$

where $\overline{B}(x, r)$ is the closed ball in X centered at x with radius r, as before. If f vanishes at infinity on X, then it follows that $\overline{B}(x, r)$ is compact, because $\overline{B}(x, r)$ is a closed set in X. This implies that X is locally compact at x under these conditions.

Let $\{f_j\}_{j=1}^\infty$ be a sequence of real or complex-valued functions on X that converges uniformly to a real or complex-valued function f on X, as appropriate. Suppose that f_j vanishes at infinity on X for each j, and let us show that f vanishes at infinity on X too. Let $\epsilon > 0$ be given. Because $\{f_j\}_{j=1}^\infty$ converges to f uniformly on X, there is an $L \in \mathbf{Z}_+$ such that

$$|f_j(x) - f(x)| < \epsilon/2 \tag{3.15.8}$$

for every $x \in X$ and $j \geq L$. There is also a compact set $K \subseteq X$ such that

$$|f_L(x)| < \epsilon/2 \tag{3.15.9}$$

for every $x \in X \setminus K$, because f_L vanishes at infinity on X. Combining (3.15.8) and (3.15.9), we get that

$$|f(x)| \leq |f(x) - f_L(x)| + |f_L(x)| < \epsilon/2 + \epsilon/2 = \epsilon \tag{3.15.10}$$

for every $x \in X \setminus K$, as desired. In particular, this implies that $C_0(X, \mathbf{R})$ and $C_0(X, \mathbf{C})$ are closed sets in $C_b(X, \mathbf{R})$ and $C_b(X, \mathbf{C})$ with respect to the supremum metric.

Let f be a real or complex-valued function on X that vanishes at infinity. Also let ϕ be a mapping from $\mathbf{R}$ or $\mathbf{C}$ into itself such that $\phi(0) = 0$ and ϕ is continuous at 0. Under these conditions, one can check that $\phi \circ f$ vanishes at infinity on X too. If ϕ vanishes on a neighborhood of 0 in $\mathbf{R}$ or $\mathbf{C}$, as appropriate, then one can verify that $\phi \circ f$ has compact support in X. Of course, if f and ϕ are continuous, then $\phi \circ f$ is continuous on X too. It is easy to see that the identity mappings on $\mathbf{R}$ and $\mathbf{C}$ can be approximated uniformly by continuous functions that vanish on a neighborhood of 0. If f is a continuous function on X that vanishes at infinity, then one can use this to approximate f uniformly by continuous functions on X with compact support. This means that $C_{com}(X, \mathbf{R})$ and $C_{com}(X, \mathbf{C})$ are dense in $C_0(X, \mathbf{R})$ and $C_0(X, \mathbf{C})$, respectively, with respect to the supremum metric.

Chapter 4

Compactness and completeness

4.1 Diameters of sets

Let $(X, d(x,y))$ be a metric space, and let E be a subset of X. If $E \neq \emptyset$, then the *diameter* of E can be defined as a nonnegative extended real number by

$$\operatorname{diam} E = \sup\{d(x,y) : x, y \in E\}. \tag{4.1.1}$$

It is easy to see that this is finite exactly when E is bounded in X. It is sometimes convenient to define the diameter of the empty set to be 0. One can check that

$$\operatorname{diam} \overline{E} = \operatorname{diam} E \tag{4.1.2}$$

for every $E \subseteq X$, where $\overline{E}$ is the closure of E in X, as usual.

If $E_1 \subseteq E_2 \subseteq X$, then

$$\operatorname{diam} E_1 \le \operatorname{diam} E_2. \tag{4.1.3}$$

Observe that

$$\operatorname{diam} \overline{B}(x,r) \le 2\,r \tag{4.1.4}$$

for every $x \in X$ and nonnegative real number r, by the triangle inequality. Here $\overline{B}(x,r)$ is the closed ball in X centered at x with radius r, as before. If $E \subseteq X$ is a bounded set, then

$$E \subseteq \overline{B}(x, \operatorname{diam} E) \tag{4.1.5}$$

for every $x \in E$.

Let $E_1, E_2, E_3, \ldots$ be an infinite sequence of nonempty subsets of X such that

$$E_j \text{ is closed and bounded} \tag{4.1.6}$$

for each $j \ge 1$,

$$E_{j+1} \subseteq E_j \tag{4.1.7}$$

for every $j \geq 1$, and

$$\lim_{j \to \infty} \operatorname{diam} E_j = 0. \tag{4.1.8}$$

Let x_j be an element of E_j for each $j \geq 1$. If $1 \leq j \leq l$, then it follows that

$$x_l \in E_l \subseteq E_j. \tag{4.1.9}$$

It is easy to see that

$$\{x_j\}_{j=1}^\infty \text{ is a Cauchy sequence in } X. \tag{4.1.10}$$

Suppose for the moment that X is complete with respect to d, so that

$$\{x_j\}_{j=1}^\infty \text{ converges to an element } x \text{ of } X. \tag{4.1.11}$$

If l is any positive integer, then $\{x_j\}_{j=l}^\infty$ is a sequence of elements of E_l that converges to x. This implies that

$$x \in E_l \tag{4.1.12}$$

for every $l \geq 1$, because E_l is a closed set, by hypothesis. Thus

$$\bigcap_{j=1}^\infty E_j \neq \emptyset \tag{4.1.13}$$

under these conditions.

Now let $\{x_j\}_{j=1}^\infty$ be a Cauchy sequence in X, and put

$$E_j = \overline{\{x_l : l \geq j\}} \tag{4.1.14}$$

for each $j \geq 1$, which is the closure in X of the set of x_l with $j \geq l$. Clearly $E_j \neq \emptyset$ and (4.1.7) holds for every $j \geq 1$, and the E_j's are closed sets, by construction. It is well known that the E_j's are bounded in X, because $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence, and in fact we have that (4.1.8) holds. If

$$x \in \bigcap_{j=1}^\infty E_j, \tag{4.1.15}$$

then one can check that

$$\{x_j\}_{j=1}^\infty \text{ converges to } x \text{ in } X. \tag{4.1.16}$$

This means that completeness of X is necessary in order to get (4.1.13) under the conditions described in the preceding paragraph.

4.2 Totally bounded sets

Let $(X, d(x, y))$ be a metric space again. A subset E of X is said to be *totally bounded* if for every $r > 0$ there are finitely many elements $x_1, \ldots, x_l$ of X such that

$$E \subseteq \bigcup_{j=1}^{l} B(x_j, r). \tag{4.2.1}$$

If $E = \emptyset$, then this may be considered to hold trivially with $l = 0$ for each $r > 0$, even when $X = \emptyset$. If (4.2.1) holds for some $r > 0$ and finitely many points $x_1, \ldots, x_l$ in X, then E is a bounded set in X. Thus totally bounded sets are bounded in particular.

If $E \subseteq X$ is compact, then

$$E \text{ is totally bounded.} \tag{4.2.2}$$

More precisely, for each $r > 0$, E can be covered by open balls in X of radius r. Compactness implies that this open covering can be reduced to a finite subcovering, as desired.

Suppose that E is a bounded subset of $\mathbf{R}^n$ for some positive integer n, with respect to the standard Euclidean metric on $\mathbf{R}^n$. In this case, one can check directly that E is totally bounded.

Of course, finite subsets of X are totally bounded. If X is equipped with the discrete metric and $E \subseteq X$ is totally bounded, then

$$E \text{ has only finitely many elements.} \tag{4.2.3}$$

This can be seen by taking $r = 1$ in (4.2.1).

A subset E of X is totally bounded if and only if for every $t > 0$ there are finitely many elements $y_1, \ldots, y_l$ of X such that

$$E \subseteq \bigcup_{j=1}^{l} \overline{B}(y_j, t). \tag{4.2.4}$$

More precisely, the previous formulation implies this one, because $B(x, r)$ is contained in $\overline{B}(x, r)$ for every $x \in X$ and $r > 0$. To get the converse, one can use the fact that $\overline{B}(y, t)$ is contained in $B(y, r)$ when $t < r$. If E is totally bounded, then it follows from this reformulation that

$$\text{the closure } \overline{E} \text{ of } E \text{ in } X \text{ is totally bounded} \tag{4.2.5}$$

too.

As another reformulation, a subset E of X is totally bounded if and only if for every $r_1 > 0$,

$$\begin{aligned} &E \text{ is contained in the union of finitely many subsets of } X \\ &\text{with diameter strictly less than } r_1. \end{aligned} \tag{4.2.6}$$

More precisely, this formulation can be obtained from (4.2.1) using (4.1.4). Similarly, one can get (4.2.1) from this formulation using (4.1.5). Equivalently, $E \subseteq X$ is totally bounded if and only if for every $t_1 > 0$,

(4.2.7) E is contained in the union of finitely many subsets of X with diameter less than or equal to t_1.

Suppose that $E \subseteq X$ is not totally bounded in X. This means that there is an $r > 0$ such that E cannot be covered by finitely many open balls in X of radius r. In particular, $E \neq \emptyset$, and we let x_1 be an element of E. By hypothesis, E is not contained in $B(x_1, r)$, and so there is an element x_2 of E that is not in $B(x_1, r)$. If elements $x_1, \ldots, x_l$ of E have been chosen in this way for some positive integer l, then E is not contained in $\bigcup_{j=1}^{l} B(x_j, r)$, by hypothesis. This permits us to choose $x_{l+1} \in E$ such that x_{l+1} is not in $B(x_j, r)$ when $1 \leq j \leq l$. Continuing in this way, we get an infinite sequence $\{x_j\}_{j=1}^{\infty}$ of elements of E such that x_l is not in $B(x_j, r)$ when $j < l$. Equivalently,

$$d(x_j, x_l) \geq r \tag{4.2.8}$$

when $j < l$.

4.3 Separable metric spaces

A metric space $(X, d(x, y))$ is said to be *separable* if

(4.3.1) there is a dense subset of X with only finitely or countably many elements.

It is well known that

(4.3.2) $\mathbf{R}^n$ is separable

with respect to the standard Euclidean metric for each positive integer n, as in Exercise 22 at the end of Chapter 2 of [155]. More precisely, one can check that the set of points in $\mathbf{R}^n$ with rational coordinates is a countable dense set in $\mathbf{R}^n$, as in the hint given in [155].

If X is any set equipped with the discrete metric, then

(4.3.3) X is the only dense subset of itself.

Thus X is separable with respect to the discrete metric if and only if

(4.3.4) X has only finitely or countably many elements.

Let $(X, d(x, y))$ be any metric space again. One can check that X is separable if and only if for each $\epsilon > 0$ there is a subset $A(\epsilon)$ of X such that

(4.3.5) $A(\epsilon)$ has only finitely or countably many elements

and

$$\bigcup_{x \in A(\epsilon)} B(x, \epsilon) = X. \tag{4.3.6}$$

Indeed, if X is separable, then there is a dense set $A \subseteq X$ with only finitely or countably many elements, and one can take

$$A(\epsilon) = A \tag{4.3.7}$$

for every $\epsilon > 0$. Conversely, if $A(1/j) \subseteq X$ satisfies (4.3.6) with $\epsilon = 1/j$ for each positive integer j, then it is easy to see that

$$A = \bigcup_{j=1}^{\infty} A(1/j) \tag{4.3.8}$$

is dense in X. If $A(1/j)$ also has only finitely or countably many elements for each $j \geq 1$, then (4.3.8) has only finitely or countably many elements as well.

Note that X is totally bounded if and only if for every $\epsilon > 0$ there is a finite set $A(\epsilon) \subseteq X$ such that (4.3.6) holds. This implies that

$$X \text{ is separable,} \tag{4.3.9}$$

as before.

A collection $\mathcal{B}$ of open subsets of X is said to be a *base for the topology of* X if for every $x \in X$ and $r > 0$ there is a $U \in \mathcal{B}$ such that $x \in U$ and

$$U \subseteq B(x, r). \tag{4.3.10}$$

Equivalently, this means that every nonempty open subset of X can be expressed as a union of elements of $\mathcal{B}$. One may consider the empty set to be the union of the empty collection of sets.

Let E be a dense subset of X, and let $\mathcal{B}(E)$ be the collection of subsets of X of the form $B(y, 1/j)$, where $y \in E$ and $j \in \mathbf{Z}_+$. We would like to show that

$$\mathcal{B}(E) \text{ is a base for the topology of } X. \tag{4.3.11}$$

To do this, let $x \in X$ and $r > 0$ be given. Let j be a positive integer such that $2/j < r$. Because E is dense in X, there is a $y \in E$ such that $d(x, y) < 1/j$. This implies that $x \in B(y, 1/j)$, and one can also verify that

$$B(y, 1/j) \subseteq B(x, r), \tag{4.3.12}$$

using the triangle inequality, as desired.

If E has only finitely or countably many elements, then

$$\mathcal{B}(E) \text{ has only finitely or countably many elements.} \tag{4.3.13}$$

Conversely, if there is a base $\mathcal{B}$ for the topology of X with only finitely or countably many elements, then X is separable. More precisely, one can choose an element from every nonempty element of $\mathcal{B}$ to get a dense set in X with only finitely or countably many elements.

4.4 Lindelöf's theorem

Let $(X, d(x,y))$ be a metric space, and suppose that $\mathcal{B}$ is a base for the topology of X with only finitely or countably many elements. Let $\{U_\alpha\}_{\alpha \in A}$ be a family of open subsets of X, indexed by a nonempty set A. Under these conditions, a theorem of Lindelöf says that there is a subset A_1 of A such that

(4.4.1) A_1 has only finitely or countably many elements

and

$$\bigcup_{\alpha \in A_1} U_\alpha = \bigcup_{\alpha \in A} U_\alpha. \tag{4.4.2}$$

To see this, put

$$\mathcal{B}_\alpha = \{V \in \mathcal{B} : V \subseteq U_\alpha\} \tag{4.4.3}$$

for every $\alpha \in A$. Observe that

$$\bigcup_{V \in \mathcal{B}_\alpha} V = U_\alpha \tag{4.4.4}$$

for every $\alpha \in A$, because $\mathcal{B}$ is a base for the topology of X, and U_α is an open set. Put

$$\mathcal{B}' = \bigcup_{\alpha \in A} \mathcal{B}_\alpha, \tag{4.4.5}$$

so that

$$\bigcup_{V \in \mathcal{B}'} V = \bigcup_{\alpha \in A} \Big(\bigcup_{V \in \mathcal{B}_\alpha} V \Big) = \bigcup_{\alpha \in A} U_\alpha. \tag{4.4.6}$$

If $V \in \mathcal{B}'$, then $V \in \mathcal{B}_\alpha$ for some $\alpha \in A$, and we let $\alpha(V)$ be an element of A with this property. Let

$$A_1 = \{\alpha(V) : V \in \mathcal{B}'\} \tag{4.4.7}$$

be the collection of elements of A that have been chosen in this way.

Thus

$$\bigcup_{V \in \mathcal{B}'} V \subseteq \bigcup_{\alpha \in A_1} U_\alpha, \tag{4.4.8}$$

because $V \subseteq U_{\alpha(V)}$ for every $V \in \mathcal{B}'$, by construction. This implies that

$$\bigcup_{\alpha \in A} U_\alpha \subseteq \bigcup_{\alpha \in A_1} U_\alpha, \tag{4.4.9}$$

by (4.4.6). The opposite inclusion holds automatically, because $A_1 \subseteq A$, so that (4.4.2) holds. Clearly $\mathcal{B}'$ has only finitely or countably many elements, because $\mathcal{B}' \subseteq \mathcal{B}$, and $\mathcal{B}$ has only finitely or countably many elements, by hypothesis. It follows that A_1 has only finitely or countably many elements too, as desired.

A subset E of X is said to have the *Lindelöf property* if every open covering of E in X can be reduced to a subcovering with only finitely or countably many elements. More precisely, this means that if $\{U_\alpha\}_{\alpha \in A}$ is an open covering of E

in X, then there is a subset A_1 of A such that A_1 has only finitely or countably many elements and

$$E \subseteq \bigcup_{\alpha \in A_1} U_\alpha. \tag{4.4.10}$$

If there is a base $\mathcal{B}$ for the topology of X with only finitely or countably many elements, then Lindelöf's theorem implies that every subset of X has the Lindelöf property.

Suppose that X has the Lindelöf property, and let $\epsilon > 0$ be given. Of course, X is covered by the collection of all open balls in X of radius ϵ. Because X has the Lindelöf property, there is a subset $A(\epsilon)$ of X with only finitely or countably many elements such that (4.3.6) holds. This implies that X is separable, as mentioned in the previous section.

4.5 The limit point property

Let $(X, d(x,y))$ be a metric space. A subset E of X is said to have the *limit point property* if for every infinite subset A of E there is an $x \in E$ such that

$$x \text{ is a limit point of } A \text{ in } X. \tag{4.5.1}$$

If E is compact, then it is well known that

$$E \text{ has the limit point property.} \tag{4.5.2}$$

Otherwise, there is an infinite set $A \subseteq E$ which has no limit point in E. This means that for every $x \in E$ there is a positive real number $r(x)$ such that $B(x, r(x))$ does not contain any element of A, except perhaps for x itself. Because E is compact, there are finitely many elements $x_1, \ldots, x_n$ of E such that

$$E \subseteq \bigcup_{j=1}^{n} B(x_j, r(x_j)). \tag{4.5.3}$$

In particular, this implies that

$$A \subseteq \bigcup_{j=1}^{n} (A \cap B(x_j, r(x_j))). \tag{4.5.4}$$

This means that A has at most n elements, contradicting the hypothesis that A have infinitely many elements.

If $E \subseteq X$ has the limit point property, then

$$E \text{ is totally bounded in } X. \tag{4.5.5}$$

Otherwise, there is an $r > 0$ and an infinite sequence $\{x_j\}_{j=1}^{\infty}$ of elements of X such that

$$d(x_j, x_l) \geq r \text{ when } j < l, \tag{4.5.6}$$

as in Section 4.2. Let A be the set of x_j's, $j \geq 1$, which is an infinite subset of E, because the x_j's are distinct. If $x \in X$, then it is easy to see that

$$A \cap B(x, r/2) \text{ has at most two elements,} \tag{4.5.7}$$

using the triangle inequality. This implies that x is not a limit point of A, and hence that E does not have the limit point property.

Suppose that $E \subseteq X$ has the limit point property, and let $U_1, U_2, U_3, \ldots$ be an infinite sequence of open subsets of X such that

$$E \subseteq \bigcup_{j=1}^{\infty} U_j. \tag{4.5.8}$$

We would like to show that there is a positive integer n such that

$$E \subseteq \bigcup_{j=1}^{n} U_j. \tag{4.5.9}$$

Otherwise, for each $n \geq 1$, we can choose $x_n \in E$ such that $x_n \not\in \bigcup_{j=1}^{n} U_j$. Let A be the set of points x_n, $n \geq 1$, that have been chosen in this way. Let us check that A has infinitely many elements.

If $y \in E$, then $y \in U_l$ for some $l \geq 1$, by (4.5.8). This means that $x_n \neq y$ when $n \geq l$, because $x_n \not\in U_l$, by construction. Thus $x_n = y$ for at most finitely many $n \geq 1$, which implies that A has infinitely many elements.

Hence there is an $x \in E$ such that x is a limit point of A in X, because E has the limit point property. We also have that $x \in U_j$ for some $j \geq 1$, by (4.5.8). Because U_j is an open set in X, there is an $r > 0$ such that

$$B(x, r) \subseteq U_j. \tag{4.5.10}$$

There are infinitely many elements of A in $B(x, r)$, because x is a limit point of A. This means that there are infinitely many elements of A in U_j, by (4.5.10). Thus $x_n \in U_j$ for infinitely many $n \geq 1$. This contradicts the fact that $x_n \not\in U_j$ when $n \geq j$, by construction.

Suppose that X has the limit point property, and let us show that

$$X \text{ is compact.} \tag{4.5.11}$$

Remember that X is totally bounded in this case, as mentioned earlier. This implies that X is separable, as in the previous section, and hence that there is a base for the topology of X with only finitely or countably many elements. It follows that X has the Lindelöf property, as in the previous section again. Combining this with the argument in the preceding paragraphs, we get that X is compact, as desired.

Now let E be any subset of X with the limit point property, and let us verify that

$$E \text{ is compact.} \tag{4.5.12}$$

Remember that E may also be considered as a metric space, with respect to the restriction of $d(x,y)$ to $x, y \in E$. It is easy to see that E has the limit point property as a subset of itself when E has the limit point property as a subset of X. This implies that E is compact as a subset of itself, as in the previous paragraph. It is well known that this implies that E is compact as a subset of X, as desired.

4.6 Sequential compactness

Let $(X, d(x,y))$ be a metric space again. A subset E of X is said to be *sequentially compact* if for every sequence $\{x_j\}_{j=1}^\infty$ of elements of E there is a subsequence $\{x_{j_l}\}_{l=1}^\infty$ of $\{x_j\}_{j=1}^\infty$ and an element x of E such that

$$\{x_{j_l}\}_{l=1}^\infty \text{ converges to } x \text{ in } X. \tag{4.6.1}$$

If E has the limit point property, then it is well known that

$$E \text{ is sequentially compact.} \tag{4.6.2}$$

More precisely, let $\{x_j\}_{j=1}^\infty$ be a sequence of elements of E, and let A be the set of x_j's, $j \geq 1$. If A has only finitely many elements, then there is a subsequence $\{x_{j_l}\}_{l=1}^\infty$ of $\{x_j\}_{j=1}^\infty$ and an element x of E such that $x_{j_l} = x$ for every $l \geq 1$. In particular, this implies that $\{x_{j_l}\}_{l=1}^\infty$ converges to x in X. Otherwise, if A has infinitely many elements, then there is an $x \in E$ such that x is a limit point of A in X, because E has the limit point property. In this case, one can show that there is a subsequence $\{x_{j_l}\}_{l=1}^\infty$ of $\{x_j\}_{j=1}^\infty$ that converges to x in X.

Conversely, if $E \subseteq X$ is sequentially compact, then E has the limit point property. To see this, let A be an infinite subset of E, and let $\{x_j\}_{j=1}^\infty$ be a sequence of distinct elements of A. By hypothesis, there is a subsequence $\{x_{j_l}\}_{l=1}^\infty$ of $\{x_j\}_{j=1}^\infty$ that converges to an element x of E. It is easy to see that x is a limit point of A in X, as desired.

Let us say that $E \subseteq X$ has the *Cauchy subsequence property* if every sequence $\{x_j\}_{j=1}^\infty$ of elements of E has a subsequence $\{x_{j_l}\}_{l=1}^\infty$ that is a Cauchy sequence in X. Sequential compactness automatically implies the Cauchy subsequence property, because convergent sequences are Cauchy sequences.

Let us say that a sequence $\{x_j\}_{j=1}^\infty$ of elements of X is an *ϵ-sequence* for some $\epsilon > 0$ if

$$d(x_j, x_l) < \epsilon \tag{4.6.3}$$

for every $j, l \geq 1$. Let us say that a subset E of X has the *small subsequence property* if for every sequence $\{x_j\}_{j=1}^\infty$ of elements of E and every $\epsilon > 0$ there is a subsequence $\{x_{j_l}\}_{l=1}^\infty$ of $\{x_j\}_{j=1}^\infty$ such that

$$\{x_{j_l}\}_{l=1}^\infty \text{ is an } \epsilon\text{-sequence.} \tag{4.6.4}$$

If E has the Cauchy subsequence property, then it is easy to see that E has the small subsequence property. This uses the fact that every Cauchy sequence in X is an ϵ-sequence for any $\epsilon > 0$ after skipping finitely many terms in the sequence.

If $E \subseteq X$ is totally bounded, then E has the small subsequence property. To see this, let a sequence $\{x_j\}_{j=1}^{\infty}$ of elements of E and an $\epsilon > 0$ be given. Because E is totally bounded, E can be covered by finitely many open balls in X of radius $\epsilon/2$. This implies that there is a single open ball of radius $\epsilon/2$ that contains x_j for infinitely many $j \geq 1$, so that there is a subsequence $\{x_{j_l}\}_{l=1}^{\infty}$ of $\{x_j\}_{j=1}^{\infty}$ such that all of the x_{j_l}'s are contained in that open ball. It follows that $\{x_{j_l}\}_{l=1}^{\infty}$ is an ϵ-sequence, as desired, by the triangle inequality.

Conversely, suppose that $E \subseteq X$ has the small subsequence property, and let us check that E is totally bounded. Otherwise, if E is not totally bounded, then there is an $r > 0$ and a sequence $\{x_j\}_{j=1}^{\infty}$ of elements of E such that $d(x_j, x_l) \geq r$ when $j < l$, as in Section 4.2. In this case, $\{x_j\}_{j=1}^{\infty}$ has no subsequence that is an r-subsequence, so that E does not have the small subsequence property.

If $E \subseteq X$ has the small subsequence property, then E has the Cauchy subsequence property. To see this, let a sequence $\{x_j\}_{j=1}^{\infty}$ of elements of E be given. By hypothesis, there is a subsequence of $\{x_j\}_{j=1}^{\infty}$ that is a $(1/2)$-sequence. We can repeat the process, to get a subsequence of the first subsequence that is a $(1/4)$-sequence. Continuing in this way, we get an infinite sequence of subsequences of $\{x_j\}_{j=1}^{\infty}$, such that the $(n+1)$th subsequence of $\{x_j\}_{j=1}^{\infty}$ is also a subsequence of the nth subsequence of $\{x_j\}_{j=1}^{\infty}$ for every $n \geq 1$, and the nth subsequence is a 2^{-n}-sequence for every $n \geq 1$. One can check that the sequence obtained by taking the rth term of the rth subsequence for each $r \geq 1$ is a subsequence of $\{x_j\}_{j=1}^{\infty}$ as well. This subsequence is a Cauchy sequence too, because the terms with $r \geq n$ may be considered as a subsequence of the nth subsequence for each $n \geq 1$.

4.7 A criterion for compactness

Let $(X, d(x, y))$ be a complete metric space. If

$$E \subseteq X \text{ is closed and totally bounded,} \tag{4.7.1}$$

then it is well known that E is compact. One way to show this is to use the fact that E has the Cauchy subsequence property, because E is totally bounded, as in the previous section. In this case, Cauchy sequences converge in X, because X is complete. We also have that the limit of a convergent sequence of elements of E is contained in E, because E is a closed set. It follows that E is sequentially compact under these conditions. This implies that E has the limit point property, as in the previous section, and hence that E is compact, as in Section 4.5.

Before describing another proof, let us reformulate the hypothesis that E be totally bounded. If r is a positive real number, then E is contained in the union of finitely many subsets of X with diameter less than or equal to r, as in Section 4.2. We may as well suppose that these are subsets of E, since we can replace them with their intersections with E. We can also take these subsets to be closed sets in X, by replacing them with their closures, which have the same diameters as before. The closures of these subsets of E are contained in E, because E is a closed set, by hypothesis.

Let $\{U_\alpha\}_{\alpha \in A}$ be an open covering of E in X for which there is no finite subcovering, so that $E \neq \emptyset$ in particular. As in the previous paragraph, E can be expressed as the union of finitely many closed sets, each of which has diameter less than or equal to $1/2$. At least one of these finitely many sets cannot be covered by finitely many U_α's. This implies that there is a subset E_1 of E such that E_1 is a closed set in X, the diameter of E_1 is less than or equal to $1/2$, and E_1 cannot be covered by finitely many U_α's.

Continuing in this way, we can get an infinite sequence $E_1, E_2, E_3, \ldots$ of subsets of E such that E_j is a closed set in X for every $j \geq 1$,

$$E_{j+1} \subseteq E_j \tag{4.7.2}$$

for every $j \geq 1$,

$$\operatorname{diam} E_j \leq 2^{-j} \tag{4.7.3}$$

for every $j \geq 1$, and E_j cannot be covered by finitely many U_α's for any $j \geq 1$. More precisely, suppose that E_j has been chosen for some $j \geq 1$, and let us choose E_{j+1}. By hypothesis, E_j is a closed set contained in E, and hence E_j is totally bounded. Thus E_j can be expressed as the union of finitely many closed sets, each of which has diameter less than or equal to 2^{-j-1}. At least one of these subsets cannot be covered by finitely many U_α's, because E_j cannot be covered by finitely many U_α's. Thus we take E_{j+1} to be one of these subsets of E_j, in such a way that E_{j+1} cannot be covered by finitely many U_α's. Clearly E_{j+1} satisfies (4.7.2), and the analogues of the other conditions for $j+1$.

Note that $\operatorname{diam} E_j \to 0$ as $j \to \infty$, by (4.7.3). This implies that there is an $x \in X$ that is contained in E_j for every $j \geq 1$, because X is complete, as before. In particular, $x \in E$, so that there is an $\alpha_0 \in A$ such that $x \in U_{\alpha_0}$. It follows that there is a positive real number r such that

$$B(x, r) \subseteq U_{\alpha_0}, \tag{4.7.4}$$

because U_{α_0} is an open set. If j is large enough so that $2^{-j} < r$, then E_j is contained in $B(x, r)$, because of (4.7.3) and the fact that $x \in E_j$. This means that $E_j \subseteq U_{\alpha_0}$, by (4.7.4). This contradicts the condition that E_j cannot be covered by finitely many U_α's, as desired.

Let $\{x_j\}_{j=1}^\infty$ be a Cauchy sequence of elements of a metric space X, which is not necessarily complete. If there is a subsequence $\{x_{j_l}\}_{l=1}^\infty$ of $\{x_j\}_{j=1}^\infty$ that converges to an element x of X, then it is well known and not too difficult to check that

$$\{x_j\}_{j=1}^\infty \text{ converges to } x \tag{4.7.5}$$

in X. In particular, if $\{x_j\}_{j=1}^\infty$ is a sequence of elements of a sequentially compact set $E \subseteq X$, then $\{x_j\}_{j=1}^\infty$ converges to an element of E. If X is sequentially compact, then it follows that X is complete. Remember that compact subsets of metric spaces have the limit point property, and hence are sequentially compact.

We can also use the second proof of this criterion for compactness as another way to show that sequential compactness implies compactness. More precisely, if a metric space X is sequentially compact, then X is complete and totally bounded, and hence compact, by this criterion. Now let E be a subset of an arbitrary metric space X, and remember that E may be considered as a metric space as well, by restricting the metric to elements of E. If E is sequentially compact in X, then it is easy to see that E is sequentially compact as a subset of itself too. This implies that E is compact as a subset of itself, as before, and hence that E is compact as a subset of X.

4.8 The Baire category theorem

Let $(X, d(x, y))$ be a metric space. One can check that $E \subseteq X$ is dense in X if and only if for every open set $V \subseteq X$ with $V \neq \emptyset$, we have that

$$E \cap V \neq \emptyset. \tag{4.8.1}$$

If E is dense in X and U is a dense open set in X, then one can verify that

$$E \cap U \text{ is dense in } X. \tag{4.8.2}$$

This implies that the intersection of two dense open sets in X is a dense open set in X as well. It follows that the intersection of finitely many dense open subsets of X is a dense open subset of X too.

Let $U_1, U_2, U_3, \ldots$ be an infinite sequence of dense open subsets of X. If X is complete, then the *Baire category theorem* states that

$$\bigcap_{j=1}^{\infty} U_j \text{ is dense in } X. \tag{4.8.3}$$

To see this, let $x \in X$ and $r > 0$ be given, and let us show that

$$\overline{B}(x, r) \cap \Big(\bigcap_{j=1}^{\infty} U_j\Big) \neq \emptyset. \tag{4.8.4}$$

Because U_1 is dense in X, there is a $y_1 \in U_1$ such that $d(x, y_1) < r$. Let us choose $r_1 > 0$ so that $r_1 \leq 1$,

$$\overline{B}(y_1, r_1) \subseteq U_1, \tag{4.8.5}$$

and

$$d(x, y_1) + r_1 \leq r. \tag{4.8.6}$$

This uses the hypothesis that U_1 be an open set to get (4.8.5). Note that

$$B(y_1, r_1) \subseteq B(x, r), \quad \overline{B}(y_1, r_1) \subseteq \overline{B}(x, r), \tag{4.8.7}$$

by (4.8.6).

Repeating the process, we get an infinite sequence $\{y_j\}_{j=1}^{\infty}$ of elements of X and an infinite sequence $\{r_j\}_{j=1}^{\infty}$ of positive real numbers with the following properties. First,

$$r_j \le 1/j \tag{4.8.8}$$

for each $j \ge 1$. Second,

$$\overline{B}(y_j, r_j) \subseteq U_j \tag{4.8.9}$$

for every $j \ge 1$. Third, (4.8.6) holds when $j = 1$, and otherwise

$$d(y_{j-1}, y_j) + r_j \le r_{j-1} \tag{4.8.10}$$

when $j \ge 2$. This implies that

$$B(y_j, r_j) \subseteq B(y_{j-1}, r_{j-1}), \quad \overline{B}(y_j, r_j) \subseteq \overline{B}(y_{j-1}, r_{j-1}) \tag{4.8.11}$$

when $j \ge 2$.

More precisely, we can do this when $j = 1$, as before. Suppose that $y_j \in X$ and $r_j > 0$ have been chosen with these properties for some $j \ge 1$, and let us see how we can choose y_{j+1} and r_{j+1}. Because U_{j+1} is dense in X, there is a $y_{j+1} \in U_{j+1}$ such that

$$d(y_j, y_{j+1}) < r_j. \tag{4.8.12}$$

We can choose $r_{j+1} > 0$ so that $r_{j+1} \le 1/(j+1)$,

$$\overline{B}(y_{j+1}, r_{j+1}) \subseteq U_{j+1}, \tag{4.8.13}$$

and

$$d(y_j, y_{j+1}) + r_{j+1} \le r_j. \tag{4.8.14}$$

This uses the hypothesis that U_{j+1} be an open set to get (4.8.13), as before.

If $1 \le j \le l$, then

$$B(y_l, r_l) \subseteq B(y_j, r_j), \quad \overline{B}(y_l, r_l) \subseteq \overline{B}(y_j, r_j), \tag{4.8.15}$$

by (4.8.11). In particular,

$$B(y_l, r_l) \subseteq B(x, r), \quad \overline{B}(y_l, r_l) \subseteq \overline{B}(x, r) \tag{4.8.16}$$

for every $l \ge 1$, by (4.8.7). Using (4.8.8) and (4.8.15), we get that

$$\{y_l\}_{l=1}^{\infty} \text{ is a Cauchy sequence in } X. \tag{4.8.17}$$

Hence $\{y_l\}_{l=1}^{\infty}$ converges to an element y of X, because X is complete.

Observe that $y_l \in \overline{B}(x, r)$ for every $l \ge 1$, by (4.8.16), so that

$$y \in \overline{B}(x, r). \tag{4.8.18}$$

Similarly, for each $j \ge 1$, we have that $y_l \in \overline{B}(y_j, r_j)$ when $l \ge 1$, by (4.8.15). This implies that

$$y \in \overline{B}(y_j, r_j) \tag{4.8.19}$$

for every $j \ge 1$. It follows that

$$y \in U_j \tag{4.8.20}$$

for every $j \ge 1$, because of (4.8.9). Thus

$$y \in \overline{B}(x, r) \cap \Big(\bigcap_{j=1}^{\infty} U_j\Big), \tag{4.8.21}$$

as desired.

Somse related results will be mentioned in the next section.

4.9 The interior of a set

Let $(X, d(\cdot, \cdot))$ be a metric space. The *interior* E° of a subset E of X is the set of $x \in E$ for which there is an $r > 0$ such that $B(x, r) \subseteq E$. It is easy to see that

$$X \setminus E^\circ = \overline{X \setminus E}. \tag{4.9.1}$$

In particular, $E^\circ = \emptyset$ if and only if $X \setminus E$ is dense in X.

Let $E_1, E_2, E_3, \ldots$ be an infinite sequence of closed subsets of X such that $E_j^\circ = \emptyset$ for every $j \ge 1$. If X is complete, then the Baire category theorem implies that

$$\bigcup_{j=1}^{\infty} E_j \text{ has empty interior} \tag{4.9.2}$$

too. More precisely, $X \setminus E_j$ is a dense open set in X for each j, so that

$$\bigcap_{j=1}^{\infty} (X \setminus E_j) \text{ is dense in } X \tag{4.9.3}$$

when X is complete. This means that $X \setminus \Big(\bigcup_{j=1}^{\infty} E_j\Big)$ is dense in X, as desired.

A subset A of X is said to be *nowhere dense* in X if

$$\text{the closure } \overline{A} \text{ of } A \text{ in } X \text{ has empty interior.} \tag{4.9.4}$$

A subset B of X is said to be of *first category* in X, or *meager*, if

$$\begin{array}{c} B \text{ can be expressed as the union} \\ \text{of a sequence of nowhere dense sets.} \end{array} \tag{4.9.5}$$

Otherwise, B is said to be of *second category* in X, or *nonmeager*. If $B \subseteq X$ is of first category and X is complete, then it follows that

$$B^\circ = \emptyset. \tag{4.9.6}$$

If $B_1, B_2, B_3, \ldots$ is an infinite sequences of subsets of X of first category, then it is not too difficult to show that

$$\bigcup_{l=1}^{\infty} B_l \tag{4.9.7}$$

is of first category in X as well.

Chapter 5

Equicontinuity and sequences of functions

Let $\{f_j\}_{j=1}^{\infty}$ be a sequence of mappings from a set X into a metric space Y. In some situations, we might like to find a subsequence $\{f_{j_l}\}_{l=1}^{\infty}$ of $\{f_j\}_{j=1}^{\infty}$ that converges to a mapping f from X into Y, at least in some sense. If X is a metric space too, then we may be interested in additional continuity conditions on the f_j's, and on f. In particular, X may be a subset of $\mathbf{R}^n$ for some positive integer n. These and related matters will be discussed in this chapter.

5.1 Pointwise convergent subsequences

Let E be a nonempty set with only finitely or countably many elements, and let Y be a metric space. Also let $\{f_j\}_{j=1}^{\infty}$ be a sequence of mappings from E into Y. Suppose that for each $x \in E$ there is a sequentially compact set $K(x) \subseteq Y$ such that

$$f_j(x) \in K(x) \tag{5.1.1}$$

for every $j \geq 1$. We would like to show that there is a subsequence $\{f_{j_l}\}_{l=1}^{\infty}$ of $\{f_j\}_{j=1}^{\infty}$ that converges pointwise to a mapping f from E into Y, with

$$f(x) \in K(x) \tag{5.1.2}$$

for every $x \in E$. Remember that compact subsets of Y have the limit point property, and thus are sequentially compact.

Of course, (5.1.1) implies that

$$\{f_j(x)\}_{j=1}^{\infty} \text{ is a bounded sequence in } Y. \tag{5.1.3}$$

If Y is the complex plane, or $\mathbf{R}^n$ for some positive integer n, with the standard Euclidean metric, and if $\{f_j(x)\}_{j=1}^{\infty}$ is a bounded sequence in Y, then there is a sequentially compact set $K(x) \subseteq Y$ such that (5.1.1) holds for each j.

Suppose first that E has only finitely many elements $x_1, \ldots, x_n$. Using sequential compactness of $K(x_1)$, we can get a subsequence $\{f_{j_l}\}_{l=1}^\infty$ of $\{f_j\}_{j=1}^\infty$ such that $\{f_{j_l}(x_1)\}_{l=1}^\infty$ converges to an element $f(x_1)$ of $K(x_1)$. If $n \geq 2$, then we can use sequential compactness of $K(x_2)$ to get a subsequence $\{f_{j_{l_r}}\}_{r=1}^\infty$ of $\{f_{j_l}\}_{l=1}^\infty$ such that $\{f_{j_{l_r}}(x_2)\}_{r=1}^\infty$ converges to an element $f(x_2)$ of $K(x_2)$. Note that $\{f_{j_{l_r}}(x_1)\}_{r=1}^\infty$ converges to $f(x_1)$ in Y, because a subsequence of a convergent sequence converges to the same limit. We can repeat the process until we get a subsequence of $\{f_j\}_{j=1}^\infty$ that converges pointwise on E in this case.

Suppose now that E is countably infinite, and let $x_1, x_2, x_3, \ldots$ be an enumeration of the elements of E. As before, we can use sequential compactness of $K(x_1)$ to get a subsequence of $\{f_j\}_{j=1}^\infty$ that converges pointwise at x_1. Repeating the process, we get a sequence of subsequences of $\{f_j\}_{j=1}^\infty$ with the following properties. First, for each $r \in \mathbf{Z}_+$,

(5.1.4) the rth subsequence converges pointwise at x_r to an element of $K(x_r)$.

Second, if $r \geq 2$, then

(5.1.5) the rth subsequence is a subsequence of the $(r-1)$th subsequence.

The second property ensures that the rth subsequence is a subsequence of all the previous subsequences, and of the initial sequence $\{f_j\}_{j=1}^\infty$. Combining this with the first property, we get that the rth subsequence converges pointwise at each of the previous points, with the same limits as the previous subsequences.

Let f be the mapping from E into Y such that for each $r \geq 1$, $f(x_r)$ is the limit of the rth subsequence of $\{f_j\}_{j=1}^\infty$ at x_r. Thus

$$f(x_r) \in K(x_r) \tag{5.1.6}$$

for every $r \geq 1$, by construction. We would like to show that there is a subsequence of $\{f_j\}_{j=1}^\infty$ that converges to f pointwise on E.

Consider the sequence of mappings $\{g_n\}_{n=1}^\infty$ from E into Y obtained by taking g_n to be the nth term of the nth subsequence mentioned before for each $n \in \mathbf{Z}_+$. One can check that

$$\{g_n\}_{n=1}^\infty \text{ is a subsequence of } \{f_j\}_{j=1}^\infty. \tag{5.1.7}$$

Similarly, for each $r \in \mathbf{Z}_+$,

(5.1.8) $\{g_n\}_{n=r}^\infty$ is a subsequence of the rth subsequence mentioned earlier.

This implies that

$$\{g_n(x_r)\}_{n=r}^\infty \text{ converges to } f(x_r) \tag{5.1.9}$$

in Y. It follows that

$$\{g_n(x_r)\}_{n=1}^\infty \text{ converges to } f(x_r) \tag{5.1.10}$$

in Y for each $r \in \mathbf{Z}_+$, as desired.

5.2 Equicontinuity

Let (X, d_X) and (Y, d_Y) be metric spaces. A collection $\mathcal{E}$ of mappings from X into Y is said to be *equicontinuous* at a point $x \in X$ if for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$d_Y(f(x), f(w)) < \epsilon \tag{5.2.1}$$

for every $f \in \mathcal{E}$ and $w \in X$ with $d_X(x, w) < \delta$. In particular, this implies that every $f \in \mathcal{E}$ is continuous at x.

Similarly, $\mathcal{E}$ is said to be *uniformly equicontinuous* on X if for every $\epsilon > 0$ there is a $\delta > 0$ such that (5.2.1) holds for every $x, w \in X$ with $d_X(x, w) < \delta$. This implies that $\mathcal{E}$ is equicontinuous at every point in X, and that every element of $\mathcal{E}$ is uniformly continuous on X.

If X is compact, then it is well known that every continuous mapping from X into Y is uniformly continuous. In this case, if $\mathcal{E}$ is equicontinuous at every point in X, then one can show that

$$\mathcal{E} \text{ is uniformly equicontinuous on } X, \tag{5.2.2}$$

using essentially the same argument.

Suppose for the moment that $\mathcal{E}$ has only finitely many elements. If $x \in X$ and every $f \in \mathcal{E}$ is continuous at x, then it is easy to see that $\mathcal{E}$ is equicontinuous at x. Similarly, if every element of $\mathcal{E}$ is uniformly continuous on X, then $\mathcal{E}$ is uniformly equicontinuous on X.

Let C be a nonnegative real number, and suppose that

$$\text{every element of } \mathcal{E} \text{ is Lipschitz with constant } C. \tag{5.2.3}$$

It is easy to see that $\mathcal{E}$ is uniformly equicontinuous on X in this case. There is an anlogous statement for Lipschitz or Hölder continuity conditions of any order $\alpha > 0$, as in Section A.2.

Let $\{f_j\}_{j=1}^{\infty}$ be a sequence of mappings from X into Y, and let x be an element of X. We say that $\{f_j\}_{j=1}^{\infty}$ is equicontinuous at x if the collection of f_j's, $j \in \mathbf{Z}_+$, is equicontinuous at x. Suppose that $\{f_j\}_{j=1}^{\infty}$ converges pointwise to a mapping f from X into Y, and that $\{f_j\}_{j=1}^{\infty}$ is equicontinuous at x. We would like to check that

$$f \text{ is continuous at } x \tag{5.2.4}$$

as well under these conditions. Let $\epsilon > 0$ be given, so that there is a $\delta = \delta(x, \epsilon) > 0$ such that

$$d_Y(f_j(x), f_j(w)) < \epsilon \tag{5.2.5}$$

for every $j \in \mathbf{Z}_+$ and $w \in X$ with $d_X(x, w) < \delta$. If $w \in X$ satisfies $d_X(x, w) < \delta$, then one can check that

$$d_Y(f(x), f(w)) \le \epsilon, \tag{5.2.6}$$

using (5.2.5) and the fact that $\{f_j(x)\}_{j=1}^{\infty}$ and $\{f_j(w)\}_{j=1}^{\infty}$ converge to $f(x)$ and $f(w)$ in Y, respectively, by hypothesis. More precisely,

$$\begin{aligned} d_Y(f(x), f(w)) &\le d_Y(f(x), f_j(x)) + d_Y(f_j(x), f_j(w)) + d_Y(f_j(w), f(w)) \\ &< d_Y(f_j(x), f(x)) + \epsilon + d_Y(f_j(w), f(w)) \end{aligned} \tag{5.2.7}$$

for every $j \in \mathbf{Z}_+$ and $w \in X$ with $d_X(x, w) < \delta$, and the right side is arbitrarily close to ϵ when j is sufficiently large. This implies that f is continuous at x, as desired.

Similarly, a sequence $\{f_j\}_{j=1}^{\infty}$ of mappings from X into Y is said to be uniformly equicontinuous on X if the collection of f_j's, $j \in \mathbf{Z}_+$, is uniformly equicontinuous on X. If $\{f_j\}_{j=1}^{\infty}$ is uniformly equicontinuous on X, and if $\{f_j\}_{j=1}^{\infty}$ converges pointwise to a mapping f from X into Y, then one can verify that

$$f \text{ is uniformly continuous on } X. \tag{5.2.8}$$

This is basically the same as the argument in the preceding paragraph, except that $\delta = \delta(\epsilon)$ does not depend on $x \in X$.

Let C be a nonnegative real number again, and let $\{f_j\}_{j=1}^{\infty}$ be a sequence of mappings from X into Y such that f_j is Lipschitz with constant C for each j. If $\{f_j\}_{j=1}^{\infty}$ converges pointwise to a mapping f from X into Y, then one can check that

$$f \text{ is Lipschitz with constant } C \tag{5.2.9}$$

too. There is an analogous statement for Lipschitz or Hölder continuity conditions of any order $\alpha > 0$, as before.

5.3 Uniformly Cauchy sequences

Let X be a set, and let (Y, d_Y) be a metric space. Let us say that a sequence $\{f_j\}_{j=1}^{\infty}$ of mappings from X into Y is *uniformly Cauchy* on X if for every $\epsilon > 0$ there is a positive integer $L(\epsilon)$ such that

$$d_Y(f_j(x), f_l(x)) < \epsilon \tag{5.3.1}$$

for every $x \in X$ and $j, l \geq L(\epsilon)$. If $X \neq \emptyset$ and the f_j's are bounded mappings from X into Y, then it is easy to see that this is equivalent to the condition that $\{f_j\}_{j=1}^{\infty}$ be a Cauchy sequence with respect to the supremum metric on the space of bounded mappings from X into Y. This is analogous to the equivalence between uniform convergence and convergence with respect to the supremum metric for bounded mappings from X into Y, as in Section 1.11.

If $\{f_j\}_{j=1}^{\infty}$ is any sequence of mappings from X into Y that converges uniformly to a mapping f from X into Y, then one can check that

$$\{f_j\}_{j=1}^{\infty} \text{ is uniformly Cauchy on } X. \tag{5.3.2}$$

This is analogous to the fact that convergent sequences in a metric space are Cauchy sequences.

Let $\{f_j\}_{j=1}^{\infty}$ be a uniformly Cauchy sequence of mappings from X into Y. In particular, this implies that

$$\{f_j(x)\}_{j=1}^{\infty} \text{ is a Cauchy sequence in } Y \tag{5.3.3}$$

for every $x \in X$. If Y is complete, then it follows that $\{f_j\}_{j=1}^\infty$ converges pointwise to a mapping f from X into Y. One can verify that

$$\{f_j\}_{j=1}^\infty \text{ converges uniformly to } f \text{ on } X \tag{5.3.4}$$

in this situation. This is analogous to the fact that the space $\mathcal{B}(X, Y)$ of bounded mappings from X into Y is complete with respect to the supremum metric when $X \neq \emptyset$ and Y is complete, as in Section 1.12.

Suppose now that (X, d_X) is a metric space too, and let $\{f_j\}_{j=1}^\infty$ be a uniformly Cauchy sequence of mappings from X into Y. Let $x \in X$ be given, and suppose that f_j is continuous at x for each $j \in \mathbf{Z}_+$. Under these conditions,

$$\{f_j\}_{j=1}^\infty \text{ is equicontinuous at } x. \tag{5.3.5}$$

To see this, let $\epsilon > 0$ be given, and let $L(\epsilon/3) \in \mathbf{Z}_+$ be as before, so that

$$d_Y(f_j(w), f_l(w)) < \epsilon/3 \tag{5.3.6}$$

for every $w \in X$ and $j, l \geq L(\epsilon/3)$. It follows that

$$\begin{aligned} d_Y(f_j(x), f_j(w)) &\leq d_Y(f_j(x), f_l(x)) + d_Y(f_l(x), f_l(w)) + d_Y(f_l(w), f_j(w)) \\ &< d_Y(f_l(x), f_l(w)) + 2\,\epsilon/3 \end{aligned} \tag{5.3.7}$$

for every $w \in X$ and $j, l \geq L(\epsilon/3)$.

If $l \geq L(\epsilon/3)$, then we can use the continuity of f_l at x to get a $\delta_l(x, \epsilon/3) > 0$ such that

$$d_Y(f_l(x), f_l(w)) < \epsilon/3 \tag{5.3.8}$$

for every $w \in X$ with

$$d_X(x, w) < \delta_l(x, \epsilon/3). \tag{5.3.9}$$

Combining this with (5.3.7), we get that

$$d_Y(f_j(x), f_j(w)) < \epsilon \tag{5.3.10}$$

for every $j \geq L(\epsilon/3)$ and $w \in X$ such that (5.3.9) holds. We may as well take $l = L(\epsilon/3)$ here.

We would like to find a $\delta(x, \epsilon) > 0$ such that (5.3.10) holds for every $j \in \mathbf{Z}_+$ and $w \in X$ with

$$d_X(x, w) < \delta(x, \epsilon). \tag{5.3.11}$$

This can be obtained from the previous statement for $j \geq L(\epsilon/3)$ and the continuity of f_j at x when $j < L(\epsilon/3)$.

Similarly, if f_j is uniformly continuous on X for each $j \geq 1$, then

$$\{f_j\}_{j=1}^\infty \text{ is uniformly equicontinuous on } X. \tag{5.3.12}$$

5.4 Equicontinuity and uniform convergence

Let (X, d_X) and (Y, d_Y) be metric spaces again, and let $\{f_j\}_{j=1}^\infty$ be a sequence of mappings from X into Y. Suppose that

$$\{f_j\}_{j=1}^\infty \text{ converges pointwise to a mapping } f \text{ from } X \text{ into } Y, \tag{5.4.1}$$

and that

$$\{f_j\}_{j=1}^\infty \text{ is equicontinuous at every } x \in X. \tag{5.4.2}$$

Let $\epsilon > 0$ be given, so that for each $x \in X$ there is a $\delta(x, \epsilon) > 0$ such that

$$d_Y(f_j(x), f_j(w)) < \epsilon/3 \tag{5.4.3}$$

for every $j \in \mathbf{Z}_+$ and $w \in X$ with $d_X(x, w) < \delta(x, \epsilon)$. This implies that

$$d_Y(f(x), f(w)) \le \epsilon/3 \tag{5.4.4}$$

for every $x, w \in X$ with $d_X(x, w) < \delta(x, \epsilon)$, as in Section 5.2. It follows that

$$\begin{aligned} d_Y(f_j(w), f(w)) & \\ \le \ & d_Y(f_j(w), f_j(x)) + d_Y(f_j(x), f(x)) + d_Y(f(x), f(w)) \\ < \ & 2\,\epsilon/3 + d_Y(f_j(x), f(x)) \end{aligned} \tag{5.4.5}$$

for every $j \in \mathbf{Z}_+$ and $x, w \in X$ with $d_X(x, w) < \delta(x, \epsilon)$, using the triangle inequality twice in the first step, and (5.4.3), (5.4.4) in the second step.

If $x \in X$, then there is a positive integer $L(x, \epsilon)$ such that

$$d_Y(f_j(x), f(x)) < \epsilon/3 \tag{5.4.6}$$

for every $j \ge L(x, \epsilon)$, because of (5.4.1). Combining this with (5.4.5), we obtain that

$$d_Y(f_j(w), f(w)) < \epsilon \tag{5.4.7}$$

for every $x \in X$, $j \ge L(x, \epsilon)$, and $w \in X$ with $d_X(x, w) < \delta(x, \epsilon)$.

Let K be a compact subset of X. The collection of open balls

$$B_X(x, \delta(x, \epsilon)) \tag{5.4.8}$$

in X centered at elements x of K is an open covering of K in X. Because K is compact, there are finitely many elements $x_1, \ldots, x_n$ of K such that

$$K \subseteq \bigcup_{l=1}^n B_X(x_l, \delta(x_l, \epsilon)). \tag{5.4.9}$$

Put

$$L_K(\epsilon) = \max_{1 \le l \le n} L(x_l, \epsilon). \tag{5.4.10}$$

It follows that (5.4.7) holds for every $w \in K$ and $j \ge L_K(\epsilon)$, so that

$$\{f_j\}_{j=1}^\infty \text{ converges to } f \text{ uniformly on } K \tag{5.4.11}$$

under these conditions.

If $\{f_j\}_{j=1}^\infty$ is uniformly equicontinuous on X, then we can take $\delta(x, \epsilon) = \delta(\epsilon)$ to be independent of $x \in X$. In this case, one can check that

(5.4.12) $\{f_j\}_{j=1}^\infty$ converges to f uniformly on totally bounded subsets of X,

using an argument like the one in the preceding paragraph.

5.5 Equicontinuity and Cauchy sequences

Let (X, d_X), (Y, d_Y) be metric spaces, and let $\{f_j\}_{j=1}^\infty$ be a sequence of mappings from X into Y. Also let E be a subset of X, and suppose that for every $w \in E$,

(5.5.1) $$\{f_j(w)\}_{j=1}^\infty \text{ is a Cauchy sequence in } Y.$$

If $x \in X$ is a limit point of E, and if

(5.5.2) $$\{f_j\}_{j=1}^\infty \text{ is equicontinuous at } x,$$

then

(5.5.3) $$\{f_j(x)\}_{j=1}^\infty \text{ is a Cauchy sequence in } Y.$$

To see this, let $\epsilon > 0$ be given. Because of (5.5.2), there is a $\delta > 0$ such that

(5.5.4) $$d_Y(f_j(x), f_j(w)) < \epsilon/3$$

for every $j \geq 1$ and $w \in X$ with $d_X(x, w) < \delta$. Thus

(5.5.5) $$\begin{aligned} d_Y(f_j(x), f_l(x)) & \\ &\leq d_Y(f_j(x), f_j(w)) + d_Y(f_j(w), f_l(w)) + d_Y(f_l(w), f_l(x)) \\ &< 2\,\epsilon/3 + d_Y(f_j(w), f_l(w)) \end{aligned}$$

for every $j, l \geq 1$ and $w \in X$ with $d_X(x, w) < \delta$. This uses the triangle inequality twice in the first step, and (5.5.4) twice in the second step, applied to j and l.

If x is a limit point of E, then there is a $w \in E$ such that $d_X(x, w) < \delta$. In this case, (5.5.1) implies that there is a positive integer L such that

(5.5.6) $$d_Y(f_j(w), f_l(w)) < \epsilon/3$$

for every $j, l \geq L$. Combining this with (5.5.5), we get that

(5.5.7) $$d_Y(f_j(x), f_l(x)) < \epsilon$$

for every $j, l \geq L$, as desired. If Y is complete as a metric space, then it follows that $\{f_j(x)\}_{j=1}^\infty$ converges in Y.

5.6 Equicontinuity and subsequences

Let (X, d_X) and (Y, d_Y) be metric spaces again, and let $\{f_j\}_{j=1}^{\infty}$ be a sequence of mappings from X into Y. Suppose that for every $x \in X$ there is a sequentially compact set $K(x) \subseteq Y$ such that

$$f_j(x) \in K(x) \tag{5.6.1}$$

for every $j \geq 1$. Suppose also that

$$X \text{ is separable} \tag{5.6.2}$$

as a metric space, and let E be a dense set in X with only finitely or countably many elements. Thus

$$\begin{array}{c}\text{there is a subsequence } \{f_{j_l}\}_{l=1}^{\infty} \text{ of } \{f_j\}_{j=1}^{\infty} \\ \text{that converges pointwise on } E,\end{array} \tag{5.6.3}$$

as in Section 5.1.

Of course, if

$$\{f_j\}_{j=1}^{\infty} \text{ is equicontinuous at every } x \in X, \tag{5.6.4}$$

then

$$\{f_{j_l}\}_{l=1}^{\infty} \text{ is equicontinuous at every } x \in X \tag{5.6.5}$$

too. In this case,

$$\{f_{j_l}(x)\}_{l=1}^{\infty} \text{ is a Cauchy sequence in } Y \tag{5.6.6}$$

for every $x \in X$, as in the previous section. It follows that

$$\{f_{j_l}(x)\}_{l=1}^{\infty} \text{ converges to an element of } K(x) \tag{5.6.7}$$

for each $x \in X$, because a Cauchy sequence of elements of a sequentially compact set converges to an element of that set, as mentioned near the end of Section 4.7.

Let $f(x)$ be the limit of $\{f_{j_l}(x)\}_{l=1}^{\infty}$ for every $x \in X$, so that f is a mapping from X into Y, and

$$\{f_{j_l}\}_{l=1}^{\infty} \text{ converges to } f \text{ pointwise on } X. \tag{5.6.8}$$

Note that

$$f \text{ is continuous on } X, \tag{5.6.9}$$

because of (5.6.5), as in Section 5.2. We also get that

$$\{f_{j_l}\}_{l=1}^{\infty} \text{ converges to } f \text{ uniformly on compact subsets of } X, \tag{5.6.10}$$

as in Section 5.4.

Similarly, if

$$\{f_j\}_{j=1}^\infty \text{ is uniformly equicontinuous on } X, \tag{5.6.11}$$

then

$$\{f_{j_l}\}_{l=1}^\infty \text{ is uniformly equicontinuous on } X, \tag{5.6.12}$$

$$f \text{ is uniformly continuous on } X, \tag{5.6.13}$$

and

$$\{f_{j_l}\}_{l=1}^\infty \text{ converges to } f \text{ uniformly on totally bounded subsets of } X. \tag{5.6.14}$$

Of course, this is all a bit simpler when X is compact. Note that X is automatically separable in this case, as in Section 4.3.

Let us now take $X = [0, 1]$ and $Y = \mathbf{R}$, using the standard Euclidean metric on $\mathbf{R}$ and its restriction to $[0, 1]$. Put $f_j(x) = x^j$ on $[0, 1]$ for every $j \geq 1$, and remember that $\{f_j\}_{j=1}^\infty$ converges pointwise on $[0, 1]$, as in Section 1.8. One can check that $\{f_j\}_{j=1}^\infty$ is equicontinuous at $x \in [0, 1]$ when $x < 1$, and not when $x = 1$. More precisely, if $0 \leq r < 1$, then $\{f_j\}_{j=1}^\infty$ is uniformly equicontinuous on $[0, r]$, and $\{f_j\}_{j=1}^\infty$ converges to 0 uniformly on $[0, r]$. However, there is no subsequence of $\{f_j\}_{j=1}^\infty$ that converges uniformly on $[0, 1]$.

One can use uniformly convergent subsequences to get existence of solutions to ordinary differential equations under suitable conditions, as in Exercises 25, 26 on p170f of [155], for instance. Note that uniqueness of solutions involves additional conditions on the differential equation, as in Exercises 27, 28 on p119 of [155].

The question of finding more elementary approaches to existence theorems like this was raised in [95]. Some responses to this question can be found in [55, 186, 187].

5.7 Pointwise and uniform boundedness

Let X be a set, let (Y, d_Y) be a metric space, and let $\mathcal{E}$ be a collection of mappings from X into Y. If $x \in X$, then put

$$\mathcal{E}(x) = \{f(x) : f \in \mathcal{E}\}, \tag{5.7.1}$$

which is a subset of Y. Let us say that $\mathcal{E}$ is *pointwise bounded* on a subset A of X if for every $x \in A$,

$$\mathcal{E}(x) \text{ is bounded in } Y. \tag{5.7.2}$$

Similarly, put

$$\mathcal{E}(A) = \bigcup_{x \in A} \mathcal{E}(x) = \bigcup_{f \in \mathcal{E}} f(A) = \{f(x) : x \in A,\ f \in \mathcal{E}\}, \tag{5.7.3}$$

which is a subset of Y as well. We say that $\mathcal{E}$ is *uniformly bounded* on A when

$$\mathcal{E}(A) \text{ is bounded in } Y. \tag{5.7.4}$$

If $\mathcal{E}$ is uniformly bounded on A, then $\mathcal{E}$ is pointwise bounded on A, and the restriction of every $f \in \mathcal{E}$ to A is bounded as a mapping from A into Y. Suppose that $X, Y \neq \emptyset$, and let $\mathcal{E}$ be a collection of bounded mappings from X into Y. One can check that

$$\mathcal{E} \text{ is uniformly bounded on } X \tag{5.7.5}$$

if and only if $\mathcal{E}$ is bounded as a subset of the space $\mathcal{B}(X, Y)$ of bounded mappings from X into Y, with respect to the supremum metric.

Now let (X, d_X) be a metric space too, and let $\mathcal{E}$ be a collection of mappings from X into Y that is equicontinuous at a point $x \in X$. Let $\epsilon > 0$ be given, so that there is a $\delta(x, \epsilon) > 0$ such that

$$d_Y(f(x), f(w)) < \epsilon \tag{5.7.6}$$

for every $f \in \mathcal{E}$ and $w \in X$ with $d_X(x, w) < \delta(x, \epsilon)$. If there is a $w_0 \in X$ such that $d_X(x, w_0) < \delta(x, \epsilon)$ and

$$\mathcal{E}(w_0) \text{ is bounded in } Y, \tag{5.7.7}$$

then it is easy to see that (5.7.2) holds. Similarly, if (5.7.2) holds, then

$$\mathcal{E} \text{ is uniformly bounded on the open ball } B_X(x, \delta(x, \epsilon)). \tag{5.7.8}$$

Suppose for the moment that

$$\mathcal{E} \text{ is equicontinuous at every } x \in X. \tag{5.7.9}$$

One can check that

$$\{u \in X : \mathcal{E}(u) \text{ is bounded in } Y\} \text{ is a closed set in } X, \tag{5.7.10}$$

using the remarks in the preceding paragraph. If $K \subseteq X$ is compact, and

$$\mathcal{E} \text{ is pointwise bounded on } K, \tag{5.7.11}$$

then one can verify that

$$\mathcal{E} \text{ is uniformly bounded on } K. \tag{5.7.12}$$

If

$$\mathcal{E} \text{ is uniformly equicontinuous on } X, \tag{5.7.13}$$

then we can take $\delta(x, \epsilon) = \delta(\epsilon)$ to be independent of $x \in X$ in the preceding paragraph. In this case, if

$$A \subseteq X \text{ is totally bounded} \tag{5.7.14}$$

and

$$\mathcal{E} \text{ is pointwise bounded on } A, \tag{5.7.15}$$

then

$$\mathcal{E} \text{ is uniformly bounded on } A. \tag{5.7.16}$$

Suppose that (5.7.9) holds again. One can check that the set of $u \in X$ such that $\mathcal{E}(u)$ is totally bounded in Y is a closed set in X. This uses the equicontinuity condition for all $\epsilon > 0$, while a single $\epsilon > 0$ would suffice for the remarks about boundedness in the previous paragraph. If $K \subseteq X$ is compact, and $\mathcal{E}(x)$ is totally bounded in Y for every $x \in K$, then $\mathcal{E}(K)$ is totally bounded in Y. If $\mathcal{E}$ is uniformly equicontinuous on X, $A \subseteq X$ is totally bounded, and $\mathcal{E}(x)$ is totally bounded in Y for every $x \in A$, then $\mathcal{E}(A)$ is totally bounded in Y.

5.8 Total boundedness in $\mathcal{B}(X,Y)$

Let X be a nonempty set, and let (Y, d_Y) be a nonempty metric space. Remember that $\mathcal{B}(X,Y)$ is the space of bounded mappings from X into Y, and that $\theta(f,g)$ denotes the supremum metric on $\mathcal{B}(X,Y)$, as in Section 1.11. Suppose that

$$\mathcal{E} \subseteq \mathcal{B}(X,Y) \text{ is totally bounded with respect to } \theta(\cdot,\cdot). \tag{5.8.1}$$

Let $x \in X$ be given, and remember that $\mathcal{E}(x)$ is the subset of Y defined in (5.7.1). It is easy to see that

$$\mathcal{E}(x) \text{ is totally bounded in } Y. \tag{5.8.2}$$

Suppose that (X, d_X) is a metric space too, and that

$$\text{every } f \in \mathcal{E} \text{ is continuous at } x. \tag{5.8.3}$$

We would like to check that

$$\mathcal{E} \text{ is equicontinuous at } x. \tag{5.8.4}$$

Let $\epsilon > 0$ be given. Because $\mathcal{E}$ is totally bounded with respect to the supremum metric, there are finitely many elements $f_1, \ldots, f_n$ of $\mathcal{E}$ such that for every $f \in \mathcal{E}$ there is a positive integer $j \le n$ such that

$$\theta(f, f_j) \le \epsilon/3. \tag{5.8.5}$$

By hypothesis, f_j is continuous at x for each $j = 1, \ldots, n$, so that there is a $\delta_j(x) > 0$ such that

$$d_Y(f_j(x), f_j(w)) < \epsilon/3 \tag{5.8.6}$$

for every $w \in X$ with $d_X(x,w) < \delta_j(x)$.

Put

$$\delta(x) = \min_{1 \le j \le n} \delta_j(x) > 0. \tag{5.8.7}$$

Let $f \in \mathcal{E}$ be given, and let $j \leq n$ be as in (5.8.5). Thus

$$\begin{aligned} (5.8.8) \quad & d_Y(f(x), f(w)) \\ & \leq \; d_Y(f(x), f_j(x)) + d_Y(f_j(x), f_j(w)) + d_Y(f_j(w), f(w)) \\ & \leq \; d_Y(f_j(x), f_j(w)) + 2\,\epsilon/3 \end{aligned}$$

for every $w \in X$, using the triangle inequality in the first step. It follows that

$$(5.8.9) \qquad d_Y(f(x), f(w)) < \epsilon/3 + 2\,\epsilon/3 = \epsilon$$

when $d_X(x, w) < \delta(x) \leq \delta_j(x)$, using (5.8.6) in the first step. This shows that (5.8.4) holds, as desired.

Similarly, if

(5.8.10) every $f \in \mathcal{E}$ is uniformly continuous on X,

then

(5.8.11) $\mathcal{E}$ is uniformly equicontinuous on X.

This is essentially the same as before, but with $\delta_j(x)$ and hence $\delta(x)$ independent of x.

Let A be a subset of X, and remember that $\mathcal{E}(A)$ is the subset of Y defined in (5.7.3). Suppose that

(5.8.12) for each $f \in \mathcal{E}$, $f(A)$ is totally bounded in Y.

One can check that

(5.8.13) $\mathcal{E}(A)$ is totally bounded in Y

as well under these conditions. Of course, (5.8.1) holds automatically when $\mathcal{E}$ has only finitely amy elements. However, (5.8.12) does not holds automatically in this case.

5.9 Equicontinuity and total boundedness

Let (X, d_X) and (Y, d_Y) be nonempty metric spaces again. If f is a uniformly continuous mapping from X into Y and A is a totally bounded subset of X, then it is not difficult to show that

(5.9.1) $f(A)$ is totally bounded in Y.

Suppose for the rest of the section that

(5.9.2) X is totally bounded with respect to d_X.

This implies that uniformly continuous mappings from X into Y are bounded, because totally bounded subsets of Y are bounded. Let $\theta(\cdot, \cdot)$ be the supremum metric on the space $\mathcal{B}(X, Y)$ of bounded mappings from X into Y, as usual.

Suppose that

$$\mathcal{E} \text{ is a uniformly equicontinuous collection of mappings from } X \text{ into } Y. \tag{5.9.3}$$

In particular, every $f \in \mathcal{E}$ is uniformly continuous as a mapping from X into Y, and hence bounded, as in the preceding paragraph. Let $\epsilon > 0$ be given, so that there is a $\delta > 0$ such that

$$d_Y(f(x), f(w)) < \epsilon/3 \tag{5.9.4}$$

holds for every $f \in \mathcal{E}$ and $x, w \in X$ with $d_X(x, w) < \delta$. Because X is totally bounded, there are finitely many elements $x_1, \ldots, x_n$ of X such that

$$X = \bigcup_{j=1}^{n} B_X(x_j, \delta), \tag{5.9.5}$$

where $B_X(x, r)$ is the open ball in X centered at $x \in X$ with radius $r > 0$, as usual.

Let $f, g \in \mathcal{E}$ and $x \in X$ be given, so that $d_X(x_j, x) < \delta$ for some $j \le n$, by (5.9.5). Under these conditions, we have that

$$d_Y(f(x_j), f(x)),\ d_Y(g(x_j), g(x)) < \epsilon/3, \tag{5.9.6}$$

as in (5.9.4). This implies that

$$\begin{aligned} d_Y(f(x), g(x)) & \\ \le\ & d_Y(f(x), f(x_j)) + d_Y(f(x_j), g(x_j)) + d_Y(g(x_j), g(x)) \\ <\ & d_Y(f(x_j), g(x_j)) + 2\,\epsilon/3, \end{aligned} \tag{5.9.7}$$

using the triangle inequality in the first step. Thus

$$d_Y(f(x), g(x)) < \max_{1 \le j \le n} d_Y(f(x_j), g(x_j)) + 2\,\epsilon/3 \tag{5.9.8}$$

for every $x \in X$. It follows that

$$\theta(f, g) \le \max_{1 \le j \le n} d_Y(f(x_j), g(x_j)) + 2\,\epsilon/3 \tag{5.9.9}$$

for every $f, g \in \mathcal{E}$.

Suppose that

$$\mathcal{E}(x) = \{f(x) : f \in \mathcal{E}\} \tag{5.9.10}$$

is totally bounded in Y for each $x \in X$, in addition to the other conditions mentioned in the previous paragraphs. Using this and (5.9.9), one can check that

$$\mathcal{E} \text{ is totally bounded with respect to the supremum metric.} \tag{5.9.11}$$

5.10 Equiconvergence of limits

Let (X, d_X) and (Y, d_Y) be metric spaces, let A be a subset of X, and suppose that $x \in X$ is a limit point of A in X. Also let $\mathcal{E}$ be a collection of mappings from E into Y, and suppose that for each $f \in \mathcal{E}$, the limit

$$\lim_{\substack{w \in A \\ w \to x}} f(w) = q_f \tag{5.10.1}$$

exists in Y. Let us say that we have *equiconvergence of the limit* for $f \in \mathcal{E}$ if for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$d_Y(f(w), q_f) < \epsilon \tag{5.10.2}$$

for every $f \in \mathcal{E}$ and $w \in A$ with $d_X(x, w) < \delta$ and $w \neq x$.

More precisely, this condition includes the existence of the limit as in (5.10.1) for each $f \in \mathcal{E}$. If $\mathcal{E}$ has only finitely many elements, and the limit as in (5.10.1) exists for each $f \in \mathcal{E}$, then one can check that we have equiconvergence of the limit for $f \in \mathcal{E}$.

Suppose for the moment that $A = X$, so that x is a limit point of X. In this case, it is well known and easy to see that a mapping f from X into Y is continuous at x if and only if

$$\lim_{w \to x} f(w) = f(x), \tag{5.10.3}$$

where the existence of the limit is part of this condition. One can verify that we have equiconvergence of the limit for $f \in \mathcal{E}$, with $q_f = f(x)$ for every $f \in \mathcal{E}$, exactly when $\mathcal{E}$ is equicontinuous at x.

Let A be any subset of X again, with $x \in X$ a limit point of A, and let $\{f_j\}_{j=1}^\infty$ be a sequence of mappings from A into Y. Suppose that for each positive integer j, the limit

$$\lim_{\substack{w \in A \\ w \to x}} f_j(w) = q_j \tag{5.10.4}$$

exists in Y. We can define *equiconvergence of the limit* for $\{f_j\}_{j=1}^\infty$ in the same way as before, by considering the collection of f_j's, $j \in \mathbf{Z}$.

Suppose for the moment that

$$\text{for each } w \in A,\ \{f_j(w)\}_{j=1}^\infty \text{ is a Cauchy sequence in } Y. \tag{5.10.5}$$

If we have equiconvergence of the limit in (5.10.4) for $\{f_j\}_{j=1}^\infty$, then one can show that

$$\{q_j\}_{j=1}^\infty \text{ is a Cauchy sequence in } Y. \tag{5.10.6}$$

This is very similar to the argument in Section 5.5. If Y is complete as a metric space, then it follows that

$$\{q_j\}_{j=1}^\infty \text{ converges to an element } q \text{ of } Y. \tag{5.10.7}$$

Suppose now that $\{f_j\}_{j=1}^\infty$ converges pointwise to a mapping f from A into Y. If we have equiconvergence of the limit in (5.10.4) for $\{f_j\}_{j=1}^\infty$, and if (5.10.7) holds, then one can show that

$$\lim_{\substack{w\in A\\ w\to x}} f(w) = q. \tag{5.10.8}$$

This is analogous to the argument for equicontinuous sequences of functions that converge pointwise, as in Section 5.2.

Let $\{f_j\}_{j=1}^\infty$ be a uniformly Cauchy sequence of mappings from A into Y, as in Section 5.3. Suppose that the limit as in (5.10.4) exists for each $j \in \mathbf{Z}_+$. Under these conditions, one can check that (5.10.6) holds. Using this, one can verify that we have equiconvergence of the limit as in (5.10.4) for $\{f_j\}_{j=1}^\infty$. This is analogous to the argument for equicontinuity in Section 5.3.

Suppose that $\{f_j\}_{j=1}^\infty$ converges uniformly to a mapping f from A into Y, so that $\{f_j\}_{j=1}^\infty$ is uniformly Cauchy on A in particular. Suppose also that the limit as in (5.10.4) exists for every j, and that (5.10.7) holds. Using the remarks in the previous two paragraphs, we get that (5.10.8) holds too.

5.11 Equiconvergence and differentiability

Let a, b be real numbers with $a < b$, and let f be a real-valued function on $[a,b]$. As usual, the derivative of f at $x \in [a,b]$ is defined by

$$f'(x) = \lim_{w\to x} \frac{f(w) - f(x)}{w - x}, \tag{5.11.1}$$

when the limit exists. Of course, this is really a one-sided limit and derivative when $x = a$ or b.

Let $\mathcal{E}$ be a collection of real-valued functions on $[a,b]$, each of which is differentiable at $x \in [a,b]$. In this case, we may be interested in the equiconvergence of the limit of difference quotients in (5.11.1) for $f \in \mathcal{E}$, as in the previous section.

Suppose for the moment that each $f \in \mathcal{E}$ is differentiable at every point in $[a,b]$. If the collection

$$\mathcal{E}' = \{f' : f \in \mathcal{E}\} \tag{5.11.2}$$

of derivatives of elements of $\mathcal{E}$ is equicontinuous at x, then one can check that we have equiconvergence of the limit as in (5.11.1) for $f \in \mathcal{E}$, using the mean value theorem.

Let $\{f_j\}_{j=1}^\infty$ be a sequence of real-valued functions on $[a,b]$, and suppose that f_j is differentiable at a point $x \in [a,b]$ for each j. As before, we may be interested in the equiconvergence of the limit

$$f_j'(x) = \lim_{w\to x} \frac{f_j(w) - f_j(x)}{w - x} \tag{5.11.3}$$

for $\{f_j\}_{j=1}^\infty$.

Suppose for the moment that f_j is differentiable at every point in $[a, b]$ for each j. If $\{f_j\}_{j=1}^\infty$ is equicontinuous at x, then we have equiconvergence of the limit as in (5.11.3) for $\{f_j\}_{j=1}^\infty$, as before.

Suppose now that $\{f_j\}_{j=1}^\infty$ converges pointwise to a real-valued function f on $[a, b]$. If $x, w \in [a, b]$ and $x \neq w$, then it follows that

$$\lim_{j \to \infty} \frac{f_j(w) - f_j(x)}{w - x} = \frac{f(w) - f(x)}{w - x}. \tag{5.11.4}$$

If f_j is differentiable at x for each j, and we have equiconvergence of the limit as in (5.11.3) for $\{f_j\}_{j=1}^\infty$, then $\{f_j'(x)\}_{j=1}^\infty$ is a Cauchy sequence in $\mathbf{R}$, as in the previous section. Of course, this means that $\{f_j'(x)\}_{j=1}^\infty$ converges in $\mathbf{R}$, because $\mathbf{R}$ is complete with respect to the standard Euclidean metric. Using this, we get that f is differentiable at x, with

$$f'(x) = \lim_{j \to \infty} f_j'(x), \tag{5.11.5}$$

as in the previous section.

Suppose that f_j is differentiable at a point $x \in [a, b]$ for each j again. Suppose also that the sequence of difference quotients

$$\frac{f_j(w) - f_j(x)}{w - x} \tag{5.11.6}$$

is uniformly Cauchy as a sequence of real-valued functions of w on $[a, b] \setminus \{x\}$. This implies that $\{f_j'(x)\}_{j=1}^\infty$ is a Cauchy sequence, as in the previous section. This means that $\{f_j'(x)\}_{j=1}^\infty$ converges in $\mathbf{R}$, because $\mathbf{R}$ is complete with respect to the standard Euclidean metric. In this case, we get equiconvergence of the limit as in (5.11.3) for $\{f_j\}_{j=1}^\infty$ too, as in the previous section.

Note that

$$\frac{f_j(w) - f_j(x)}{w - x} - \frac{f_l(w) - f_l(x)}{w - x} = \frac{(f_j(w) - f_l(w)) - (f_j(x) - f_l(x))}{w - x} \tag{5.11.7}$$

for every $w \in [a, b] \setminus \{x\}$ and $j, l \geq 1$. The uniform Cauchy condition for (5.11.6) on $[a, b] \setminus \{x\}$ means that if j, l are sufficiently large, then (5.11.7) is as small as we like, uniformly over $w \in [a, b] \setminus \{x\}$. Suppose that f_j is differentiable at every point in $[a, b]$ for each j, and that $\{f_j'\}_{j=1}^\infty$ is uniformly Cauchy as a sequence of real-valued functions on $[a, b]$. In this case, one can use the mean value theorem to get that (5.11.7) is as small as we like when j, l are sufficiently large, uniformly over $x, w \in [a, b]$ with $x \neq w$. This is related to Theorem 7.17 on p152 of [155].

Chapter 6

More on sums and norms

6.1 Weierstrass' criterion

Let X be a nonempty set, and let $a_1(x), a_2(x), a_3(x), \ldots$ be an infinite sequence of complex-valued functions on X. Also let $A_1, A_2, A_3, \ldots$ be an infinite sequence of nonnegative real numbers such that

$$|a_j(x)| \le A_j \tag{6.1.1}$$

for every $x \in X$ and $j \ge 1$. Suppose that

$$\sum_{j=1}^{\infty} A_j \tag{6.1.2}$$

converges, which implies that

$$\sum_{j=1}^{\infty} a_j(x) \tag{6.1.3}$$

converges absolutely for every $x \in X$. Under these conditions, the sequence of partial sums

$$\sum_{j=1}^{n} a_j(x) \tag{6.1.4}$$

converges to (6.1.3) uniformly on X. This is a well-known criterion of Weierstrass for uniform convergence.

To see this, observe that

$$\Big|\sum_{j=1}^{\infty} a_j(x) - \sum_{j=1}^{n} a_j(x)\Big| = \Big|\sum_{j=n+1}^{\infty} a_j(x)\Big| \le \sum_{j=n+1}^{\infty} |a_j(x)| \le \sum_{j=n+1}^{\infty} A_j \tag{6.1.5}$$

for every $x \in X$ and $n \ge 1$. The convergence of (6.1.2) implies that the right side of (6.1.5) tends to 0 as $n \to \infty$. It follows that (6.1.4) converges to (6.1.3) uniformly on X, because the right side of (6.1.5) does not depend on x.

Suppose that $(X, d(\cdot,\cdot))$ is a metric space, and that $a_j(x)$ is a continuous complex-valued function on X for each $j \geq 1$, with respect to the standard Euclidean metric on the complex plane $\mathbf{C}$. This implies that the partial sums (6.1.4) are continuous on X as well. If (6.1.2) converges, then it follows that (6.1.3) is continuous on X too.

Now let

$$\sum_{j=0}^{\infty} a_j\, z^j \tag{6.1.6}$$

be a power series with complex coefficients. Suppose that

$$\sum_{j=0}^{\infty} |a_j|\, r^j \tag{6.1.7}$$

converges for some nonnegative real number r, which implies that (6.1.6) converges absolutely for every $z \in \mathbf{C}$ with $|z| \leq r$. Using Weierstrass' criterion, we get that the partial sums

$$\sum_{j=0}^{n} a_j\, z^j \tag{6.1.8}$$

converge to (6.1.6) uniformly on the closed disk

$$\{z \in \mathbf{C} : |z| \leq r\}. \tag{6.1.9}$$

Of course, (6.1.8) is continuous as a mapping from $\mathbf{C}$ into itself for each $n \geq 0$, using the standard metric on $\mathbf{C}$. In particular, the restriction of (6.1.8) to (6.1.9) is continuous with respect to the restriction of the standard metric on $\mathbf{C}$ to (6.1.9), so that (6.1.6) is continuous on (6.1.9) as well.

Suppose that $0 < \rho \leq \infty$ has the property that (6.1.7) converges when $0 \leq r < \rho$. This implies that (6.1.6) converges absolutely for every $z \in \mathbf{C}$ with $|z| < \rho$. Under these conditions, (6.1.6) defines a continuous complex-valued function on

$$\{z \in \mathbf{C} : |z| < \rho\}, \tag{6.1.10}$$

with respect to the restriction of the standard metric on $\mathbf{C}$ to (6.1.10). To see this, let $z_0 \in \mathbf{C}$ with $|z_0| < \rho$ be given, and let us check that (6.1.6) is continuous at z_0. Let r be a positive real number such that $|z_0| < r < \rho$. The remarks in the preceding paragraph imply that (6.1.6) is continuous on (6.1.9). One can use this to verify that (6.1.6) is continuous at z_0 as a complex-valued function on (6.1.10), because $|z_0| < r$.

6.2 Radius of convergence

Let

$$\sum_{j=0}^{\infty} a_j\, z^j \tag{6.2.1}$$

be a power series with complex coefficients, and let A be the set of nonnegative real numbers r such that

$$\sum_{j=0}^{\infty} |a_j|\, r^j \tag{6.2.2}$$

converges. Of course, $0 \in A$. If $r \in A$, then

$$[0, r] \subseteq A, \tag{6.2.3}$$

by the comparison test. If t is a positive real number such that $t \not\in A$, then it follows that

$$[t, +\infty) \cap A = \emptyset. \tag{6.2.4}$$

The *radius of convergence* of (6.2.1) can be defined as a nonnegative extended real number by

$$R = \sup A. \tag{6.2.5}$$

If $A = [0, \infty)$, then $R = +\infty$. Otherwise, $R < +\infty$, and A is either $[0, R)$ or $[0, R]$. If $z \in \mathbf{C}$ and $|z| < R$, then (6.2.1) converges absolutely. In fact, (6.2.1) defines a continuous function on

$$\{z \in \mathbf{C} : |z| < R\}, \tag{6.2.6}$$

as in the previous section.

Let t be a positive real number such that $\{|a_j|\, t^j\}_{j=0}^{\infty}$ is a bounded sequence of nonnegative real numbers. This means that there is a nonnegative real number C such that

$$|a_j|\, t^j \le C \tag{6.2.7}$$

for every $j \ge 0$. If r is any nonnegative real number, then we get that

$$|a_j|\, r^j \le C\, (r/t)^j \tag{6.2.8}$$

for every $j \ge 0$. If $r < t$, then it follows that (6.2.2) converges, by comparison with the convergent geometric series $\sum_{j=0}^{\infty} (r/t)^j$. This implies that

$$[0, t) \subseteq A, \tag{6.2.9}$$

so that $t \le R$.

If (6.2.1) converges for some $z \in \mathbf{C}$, then

$$\lim_{j \to \infty} a_j\, z^j = 0. \tag{6.2.10}$$

This implies that $\{a_j\, z^j\}_{j=0}^{\infty}$ is a bounded sequence of complex numbers, which is the same as saying that $\{|a_j|\, |z|^j\}_{j=0}^{\infty}$ is a bounded sequence of nonnegative real numbers. It follows that $|z| \le R$, as in the preceding paragraph.

Suppose that (6.2.2) converges for some positive real number r. If r_0 is a nonnegative real number strictly less that r, then

$$\lim_{j \to \infty} j\, (r_0/r)^j = 0, \tag{6.2.11}$$

because $r_0/r < 1$. In particular, $\{j\,(r_0/r)^j\}_{j=0}^\infty$ is a bounded sequence of nonnegative real numbers, so that there is a nonnegative real number C_0 such that

$$j\,(r_0/r)^j \le C_0 \tag{6.2.12}$$

for every $j \ge 0$. Thus

$$j\,|a_j|\,r_0^j \le C_0\,|a_j|\,r^j \tag{6.2.13}$$

for every $j \ge 0$, which implies that

$$\sum_{j=0}^\infty j\,|a_j|\,r_0^j \tag{6.2.14}$$

converges, by the comparison test. It follows that (6.2.14) converges when r_0 is strictly less than the radius of convergence R.

6.3 Termwise differentiation

Let a, b be real numbers with $a < b$, and let $\{f_j\}_{j=1}^\infty$ be a sequence of continuous real or complex-valued functions on $[a,b]$ that converges uniformly to a real or complex-valued function f on $[a,b]$, as appropriate. Thus f is also continuous on $[a,b]$, as in Section 1.8. Of course, we are implicitly using the standard Euclidean metrics on $\mathbf{R}$ and $\mathbf{C}$ here, and the restriction of the standard Euclidean metric on $\mathbf{R}$ to $[a,b]$. If $x \in [a,b]$, then put

$$F_j(x) = \int_a^x f_j(t)\,dt \tag{6.3.1}$$

for every $j \in \mathbf{Z}_+$, and

$$F(x) = \int_a^x f(t)\,dt, \tag{6.3.2}$$

using standard Riemann integrals on the right sides of (6.3.1) and (6.3.2). Observe that

$$\begin{aligned} |F_j(x) - F(x)| &= \left|\int_a^x (f_j(t) - f(t))\,dt\right| \\ &\le \int_a^x |f_j(t) - f(t)|\,dt \le \int_a^b |f_j(t) - f(t)|\,dt \end{aligned} \tag{6.3.3}$$

for every $j \ge 1$ and $x \in [a,b]$. This implies that $\{F_j\}_{j=1}^\infty$ converges uniformly to F on $[a,b]$, because the right side of (6.3.3) tends to 0 as $j \to \infty$, and does not depend on x. We also have that $F_j'(x) = f_j(x)$ for every $j \ge 1$ and $x \in [a,b]$, and that $F'(x) = f(x)$ for every $x \in [a,b]$, using the appropriate one-sided derivative when $x = a$ or b.

Now let $\{g_j\}_{j=1}^\infty$ be a sequence of continuously-differentiable real or complex-valued functions on $[a,b]$. Thus, for each $j \in \mathbf{Z}_+$, the derivative $g_j'(x)$ of g_j exists at every $x \in [a, b]$, using the appropriate one-sided derivative when $x = a$ or b, and g_j' is continuous on $[a, b]$. It follows that

$$g_j(x) = g_j(a) + \int_a^x g_j'(t)\,dt \tag{6.3.4}$$

for every $j \geq 1$ and $x \in [a, b]$, by the fundamental theorem of calculus. Suppose that $\{g_j(a)\}_{j=1}^\infty$ converges to a real or complex number $g(a)$, as appropriate, and that $\{g_j'\}_{j=1}^\infty$ converges uniformly to a real or complex-valued function f on $[a, b]$, as appropriate. Note that f is continuous on $[a, b]$, as in Section 1.8. Let g be the real or complex-valued function defined on $[a, b]$ by

$$g(x) = g(a) + \int_a^x f(t)\,dt \tag{6.3.5}$$

for each $x \in [a, b]$. Under these conditions, $\{g_j\}_{j=1}^\infty$ converges uniformly to g on $[a, b]$, as in the preceding paragraph. Of course, $g' = f$ on $[a, b]$.

Let $\sum_{j=0}^\infty a_j\, x^j$ be a power series with real or complex coefficients, and suppose that

$$\sum_{j=0}^\infty j\, |a_j|\, r^j \tag{6.3.6}$$

converges for some positive real number r. In particular, this implies that

$$\sum_{j=0}^\infty |a_j|\, r^j \tag{6.3.7}$$

converges, and we put

$$f(x) = \sum_{j=0}^\infty a_j\, x^j \tag{6.3.8}$$

for every $x \in \mathbf{R}$ with $|x| \leq r$. Similarly, put

$$\phi(x) = \sum_{j=1}^\infty j\, a_j\, x^{j-1} \tag{6.3.9}$$

for every $x \in \mathbf{R}$ with $|x| \leq r$, where the series on the right converges absolutely because of the convergence of (6.3.6). Under these conditions, the partial sums of the series on the right sides of (6.3.8) and (6.3.9) converge uniformly on $[-r, r]$, as in Section 6.1. By construction, the partial sums of the right side of (6.3.9) are the same as the first derivatives of the partial sums of the right side of (6.3.8). Using the remarks in the previous paragraph, we get that f is differentiable on $[-r, r]$, with

$$f'(x) = \phi(x) \tag{6.3.10}$$

for every $x \in [-r, r]$. This uses the appropriate one-sided derivatives when $x = \pm r$, as usual.

Suppose that $0 < \rho \le +\infty$ has the property that (6.3.7) converges when $0 \le r < \rho$. This implies that (6.3.6) converges when $0 \le r < \rho$ too. More precisely, if $\sum_{j=0}^\infty |a_j|\, t^j$ converges for some $t > r$, then (6.3.6) converges, as in the previous section. In this situation, the series on the right sides of (6.3.8) and (6.3.9) converge absolutely for every $x \in \mathbf{R}$ with $|x| < \rho$, so that $f(x)$ and $\phi(x)$ may be defined on $(-\rho, \rho)$ as before. Using the remarks in the preceding paragraph, we get that f is differentiable on $(-\rho, \rho)$, with derivative given by (6.3.10).

6.4 Cauchy products

Let $\sum_{j=0}^\infty a_j$ and $\sum_{l=0}^\infty b_l$ be infinite seris of complex numbers, and put

$$c_n = \sum_{j=0}^n a_j\, b_{n-j} \tag{6.4.1}$$

for each nonnegative integer n. The infinite series $\sum_{n=0}^\infty c_n$ is the *Cauchy product* of the series $\sum_{j=0}^\infty a_j$ and $\sum_{l=0}^\infty b_l$. It is easy to see that

$$\sum_{n=0}^\infty c_n = \Big(\sum_{j=0}^\infty a_j\Big)\Big(\sum_{l=0}^\infty b_l\Big) \tag{6.4.2}$$

formally. In particular, if $a_j = 0$ for all but finitely many $j \ge 0$, and $b_l = 0$ for all but finitely many $l \ge 0$, then one can check that $c_n = 0$ for all but finitely many $n \ge 0$, and that (6.4.2) holds.

Suppose for the moment that the a_j's and b_l's are nonnegative real numbers, so that the c_n's are nonnegative real numbers too. Observe that

$$\sum_{n=0}^N c_n \le \Big(\sum_{j=0}^N a_j\Big)\Big(\sum_{l=0}^N b_l\Big) \tag{6.4.3}$$

for every nonnegative integer N. If $\sum_{j=0}^\infty a_j$ and $\sum_{l=0}^\infty b_l$ converge, then

$$\sum_{n=0}^N c_n \le \Big(\sum_{j=0}^\infty a_j\Big)\Big(\sum_{l=0}^\infty b_l\Big) \tag{6.4.4}$$

for every $N \ge 0$. This implies that $\sum_{n=0}^\infty c_n$ converges, with

$$\sum_{n=0}^\infty c_n \le \Big(\sum_{j=0}^\infty a_j\Big)\Big(\sum_{l=0}^\infty b_l\Big). \tag{6.4.5}$$

We also have that

$$\Big(\sum_{j=0}^N a_j\Big)\Big(\sum_{l=0}^N b_l\Big) \le \sum_{n=0}^{2N} c_n \le \sum_{n=0}^\infty c_n \tag{6.4.6}$$

for every $N \geq 0$. If $\sum_{j=0}^\infty a_j$ and $\sum_{l=0}^\infty b_l$ converge, then we get that

$$\Big(\sum_{j=0}^\infty a_j\Big)\Big(\sum_{l=0}^\infty b_l\Big) \leq \sum_{n=0}^\infty c_n. \tag{6.4.7}$$

Of course, (6.4.2) follows from (6.4.5) and (6.4.7) in this situation.

Suppose now that $\sum_{j=0}^\infty a_j$ and $\sum_{l=0}^\infty b_l$ are absolutely convergent series of complex numbers. Clearly

$$|c_n| \leq \sum_{j=0}^n |a_j|\,|b_{n-j}| \tag{6.4.8}$$

for each $n \geq 0$, by the triangle inequality. The right side of (6.4.8) is the same as the nth term of the Cauchy product of $\sum_{j=0}^\infty |a_j|$ and $\sum_{l=0}^\infty |b_l|$. These two series converge, by hypothesis, and so their Cauchy product converges as well, as in the previous paragraph. This implies that $\sum_{n=0}^\infty c_n$ converges absolutely, with

$$\sum_{n=0}^\infty |c_n| \leq \sum_{n=0}^\infty\Big(\sum_{j=0}^n |a_j|\,|b_l|\Big) = \Big(\sum_{j=0}^\infty |a_j|\Big)\Big(\sum_{l=0}^\infty |b_l|\Big). \tag{6.4.9}$$

If the a_j's and b_l's are real numbers, then $\sum_{j=0}^\infty a_j$ and $\sum_{l=0}^\infty b_l$ can be expressed as differences of convergent series of nonnegative real numbers. In this case, (6.4.2) can be obtained from the corresponding statement for nonnegative real numbers, as in the previous paragraph. If the a_j's and b_l's are complex numbers, then one can get (6.4.2) using the analogous statements for the real and imaginary parts of the a_j's and b_l's.

Alternatively, consider

$$\sum_{j,l\geq 0} a_j\, b_l, \tag{6.4.10}$$

where more precisely the sum is taken over all ordered pairs (j,l) of nonnegative integers j, l. This sum can be identified formally with both sides of (6.4.2). The left side of (6.4.2) corresponds to summing first over (j,l) such that $j+l=n$, and then summing over $n \geq 0$. The right side of (6.4.2) can be obtained by summing over j and l separately. If $a_j = 0$ for all but finitely many $j \geq 0$, and $b_l = 0$ for all but finitely many $l \geq 0$, then $a_j\, b_l = 0$ for all but finitely many (j,l), and all of these sums can be reduced to finite sums. If the a_j's and b_l's are nonnegative real numbers, then (6.4.10) can be defined as a nonnegative extended real number, as in Section 11.2. This sum can be expressed in terms of iterated sums, as in Section 11.15. In particular, one can check that (6.4.10) is finite when $\sum_{j=0}^\infty a_j$ and $\sum_{l=0}^\infty b_l$ converge, in which case (6.4.10) is the same as both sides of (6.4.2). Similarly, if $\sum_{j=0}^\infty a_j$ and $\sum_{l=0}^\infty b_l$ are absolutely convergent series of complex numbers, then

$$\sum_{j,l\geq 0} |a_j|\,|b_l| = \Big(\sum_{j=0}^\infty |a_j|\Big)\Big(\sum_{l=0}^\infty |b_l|\Big). \tag{6.4.11}$$

Thus $a_j\, b_l$ is a summable complex-valued function of (j, l), so that (6.4.10) can be defined as in Section 11.8. In this situation, (6.4.10) can be expressed in terms of iterated sums, as in Section 11.16. This can be used to get that $\sum_{n=0}^{\infty} c_n$ converges absolutely, and that both sides of (6.4.2) are equal to (6.4.10).

Suppose now that $\sum_{j=0}^{\infty} a_j\, z^j$ and $\sum_{l=0}^{\infty} b_l\, z^l$ are power series with complex coefficients. One can check that their Cauchy product is the power series

$$\sum_{n=0}^{\infty} c_n\, z^n, \tag{6.4.12}$$

where c_n is as in (6.4.1) for each $n \geq 0$.

6.5 Rearrangements

Let $\sum_{j=1}^{\infty} a_j$ be an infinite series of real or complex numbers, and let π be a one-to-one mapping from the set $\mathbf{Z}_+$ onto itself. Under these conditions, the infinite series

$$\sum_{l=1}^{\infty} a_{\pi(l)} \tag{6.5.1}$$

is called a *rearrangement* of $\sum_{j=1}^{\infty} a_j$. If $a_j = 0$ for all but finitely many positive integers j, then $a_{\pi(l)} = 0$ for all but finitely many l too, and it is easy to see that

$$\sum_{l=1}^{\infty} a_{\pi(l)} = \sum_{j=1}^{\infty} a_j. \tag{6.5.2}$$

Suppose for the moment that a_j is a nonnegative real number for each $j \geq 1$. If $n \in \mathbf{Z}_+$, then

$$\sum_{l=1}^{n} a_{\pi(l)} \leq \sum_{j=1}^{N} a_j \tag{6.5.3}$$

for every $N \geq \max_{1 \leq l \leq n} \pi(l)$. Similarly,

$$\sum_{j=1}^{n} a_j \leq \sum_{l=1}^{N} a_{\pi(l)} \tag{6.5.4}$$

for every $N \geq \max_{1 \leq j \leq n} \pi^{-1}(j)$. This implies that $\sum_{j=1}^{\infty} a_j$ converges if and only if $\sum_{l=1}^{\infty} a_{\pi(l)}$ converges, in which case (6.5.2) holds.

If $\sum_{j=1}^{\infty} a_j$ is an infinite series of real or complex numbers, then it follows that $\sum_{j=1}^{\infty} a_j$ converges absolutely if and only if $\sum_{l=1}^{\infty} a_{\pi(l)}$ converges absolutely. One can check that (6.5.2) holds in this situation as well. More precisely, if the a_j's are real numbers, then one can reduce to the previous case by expressing $\sum_{j=1}^{\infty} a_j$ as a difference of convergent series of nonnegative real numbers. If the a_j's are complex numbers, then one can consider their real and imaginary parts.

6.6 Comparing norms on $\mathbf{R}^n$, $\mathbf{C}^n$

Let n be a positive integer, and let

$$e_j = (e_{j,1}, \ldots, e_{j,n}) \tag{6.6.1}$$

be the jth standard basis vector in $\mathbf{R}^n$ for each $j = 1, \ldots, n$. Thus

$$\begin{aligned} e_{j,l} &= 1 \quad \text{when } j = l \\ &= 0 \quad \text{when } j \neq l. \end{aligned} \tag{6.6.2}$$

If $v = (v_1, \ldots, v_n) \in \mathbf{R}^n$, then v can be expressed as

$$v = \sum_{j=1}^{n} v_j \, e_j, \tag{6.6.3}$$

where the right side is a linear combination of elements of $\mathbf{R}^n$. The e_j's may also be considered as standard basis vectors in $\mathbf{C}^n$. If $v = (v_1, \ldots, v_n) \in \mathbf{C}^n$, then v can be expressed as in (6.6.3) again, where now the right side is a linear combination of elements of $\mathbf{C}^n$.

Let N be a norm on $\mathbf{R}^n$ or $\mathbf{C}^n$, as in Sections 1.3 and 1.4. If $v = (v_1, \ldots, v_n)$ is an element of $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, then

$$N(v) = N\Big(\sum_{j=1}^{n} v_j \, e_j\Big), \tag{6.6.4}$$

by (6.6.3). This implies that

$$N(v) \le \sum_{j=1}^{n} N(v_j \, e_j) = \sum_{j=1}^{n} |v_j| \, N(e_j), \tag{6.6.5}$$

using the triangle inequality for N in the first step, and the homogeneity condition for N in the second step. Note that this argument also works for seminorms instead of norms.

It follows that

$$N(v) \le \Big(\max_{1 \le l \le n} N(e_l)\Big) \sum_{j=1}^{n} |v_j| = \Big(\max_{1 \le l \le n} N(e_l)\Big) \, \|v\|_1, \tag{6.6.6}$$

where $\|v\|_1$ is as in (1.3.6) or (1.4.2), as appropriate. Remember that

$$d_N(v, w) = N(v - w) \tag{6.6.7}$$

is the metric on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, associated to N, as in (1.3.9) and (1.4.4). Similarly, we let

$$d_1(v, w) = \|v - w\|_1 \tag{6.6.8}$$

be the metric on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, associated to $\|\cdot\|_1$, as in (1.3.11) and (1.4.6). Thus

$$d_N(v,w) \le \Big(\max_{1\le l\le n} N(e_l)\Big)\, d_1(v,w) \tag{6.6.9}$$

for every v, w in $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. In particular, the identity mapping on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, is Lipschitz with respect to (6.6.8) on the domain and (6.6.7) on the range.

Observe that

$$N(v) \le \Big(\sum_{l=1}^n N(e_j)^2\Big)^{1/2} \Big(\sum_{j=1}^n |v_j|^2\Big)^{1/2} \tag{6.6.10}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, using the Cauchy–Schwarz inequality (1.3.13) on the right side of (6.6.5). Equivalently, this means that

$$N(v) \le \Big(\sum_{l=1}^n N(e_j)^2\Big)^{1/2} \|v\|_2, \tag{6.6.11}$$

where $\|v\|_2$ is the standard Euclidean norm of v, as in (1.3.5) or (1.4.1), as appropriate. Remember that

$$d_2(v,w) = \|v-w\|_2 \tag{6.6.12}$$

is the standard Euclidean metric on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, as in (1.3.10) and (1.4.5). It follows that

$$d_N(v,w) \le \Big(\sum_{l=1}^n N(e_l)^2\Big)^{1/2} d_2(v,w) \tag{6.6.13}$$

for every v, w in $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. This implies that the identity mapping on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, is Lipschitz with respect to (6.6.12) on the domain and (6.6.7) on the range.

Using (6.6.5), we also get that

$$N(v) \le \Big(\sum_{j=1}^n N(e_j)\Big) \max_{1\le l\le n} |v_l| = \Big(\sum_{j=1}^n N(e_j)\Big) \|v\|_\infty \tag{6.6.14}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Here $\|v\|_\infty$ is as in (1.3.7) or (1.4.3), as appropriate. Let

$$d_\infty(v,w) = \|v-w\|_\infty \tag{6.6.15}$$

be the metric on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, associated to $\|\cdot\|_\infty$ as in (1.3.12) and (1.4.7). Note that

$$d_N(v,w) \le \Big(\sum_{j=1}^n N(e_j)\Big)\, d_\infty(v,w) \tag{6.6.16}$$

for every v, w in $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Hence the identity mapping on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, is Lipschitz with respect to (6.6.15) on the domain and (6.6.7) on the range.

6.7 Another comparison

Let N be a norm on $\mathbf{R}^n$ or $\mathbf{C}^n$ for some $n \in \mathbf{Z}_+$ again. Observe that

$$N(v) \le N(w) + N(v - w) \tag{6.7.1}$$

and

$$N(w) \le N(v) + N(v - w) \tag{6.7.2}$$

for every v, w in $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. This implies that

$$|N(v) - N(w)| = \max(N(v) - N(w), N(w) - N(v)) \le N(v - w), \tag{6.7.3}$$

for every v, w in $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. More precisely, the left side of (6.7.3) is the usual absolute value of $N(v) - N(w)$, as a real number.

Remember that

$$N(v) \le C\,\|v\|_2 \tag{6.7.4}$$

for some nonnegative real number C and every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, as in (6.6.11). Combining this with (6.7.3), we get that

$$|N(v) - N(w)| \le C\,\|v - w\|_2 = C\,d_2(v, w) \tag{6.7.5}$$

for every v, w in $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Thus N is Lipschitz as a real-valued function on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, using the corresponding Euclidean metrics on the domain and range. In particular, N is continuous. This argument works for seminorms instead of norms as well.

The extreme value theorem implies that there is a $u \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, such that $\|u\|_2 = 1$ and

$$N(u) \le N(w) \tag{6.7.6}$$

for every $w \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, with $\|w\|_2 = 1$. This uses the fact that the unit sphere in $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, is compact with respect to the Euclidean metric. Put

$$c = N(u), \tag{6.7.7}$$

and note that $c > 0$, because N is a norm. Let us check that

$$c\,\|v\|_2 \le N(v) \tag{6.7.8}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$. Of course, (6.7.8) is trivial when $v = 0$, and so we may suppose that $v \ne 0$. In this case, $\|v\|_2 > 0$, and

$$w = \frac{v}{\|v\|_2} \tag{6.7.9}$$

satisfies $\|w\|_2 = 1$. Thus

$$N(v)/\|v\|_2 = N(v/\|v\|_2) = N(w) \ge N(u) = c, \tag{6.7.10}$$

using (6.7.6) in the third step. This implies (6.7.8), as desired.

Put $C' = 1/c$, so that (6.7.8) is the same as saying that

$$\|v\|_2 \le C' \, N(v) \tag{6.7.11}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. It follows that

$$d_2(v, w) = \|v - w\|_2 \le C' \, N(v - w) = C' \, d_N(v, w) \tag{6.7.12}$$

for every v, w in $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. In particular, this implies that the identity mapping on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, is Lipschitz with respect to the metric $d_N(v, w)$ associated to N on the domain, and the standard Euclidean metric on the range. Of course, there are analogous arguments using $\| \cdot \|_1$ or $\| \cdot \|_\infty$ here, instead of the Euclidean norm $\| \cdot \|_2$.

6.8 Inner products on $\mathbf{R}^n$, $\mathbf{C}^n$

Let n be a positive integer, and let $x, y \in \mathbf{R}^n$ be given. Thus

$$\langle x, y \rangle = \langle x, y \rangle_{\mathbf{R}^n} = \sum_{j=1}^n x_j \, y_j \tag{6.8.1}$$

is defined as a real number. This is the standard *inner product* on $\mathbf{R}^n$. Note that

$$\langle x, x \rangle_{\mathbf{R}^n} = \sum_{j=1}^n x_j^2 = \|x\|_2^2, \tag{6.8.2}$$

where $\|x\|_2$ is the standard Euclidean norm of x, as in (1.3.5). We also have that

$$|\langle x, y \rangle_{\mathbf{R}^n}| = \left| \sum_{j=1}^n x_j \, y_j \right| \le \sum_{j=1}^n |x_j| \, |y_j| \le \|x\|_2 \, \|y\|_2, \tag{6.8.3}$$

using the Cauchy–Schwarz inequality (1.3.13) in the third step.

Let $z = x + i\, y$ be a complex number, where x and y are real numbers. Remember that the *complex conjugate* of z is the complex number defined by

$$\overline{z} = x - i\, y. \tag{6.8.4}$$

If w is another complex number, then

$$\overline{(w + z)} = \overline{w} + \overline{z} \tag{6.8.5}$$

and

$$\overline{(w\, z)} = \overline{w}\, \overline{z}. \tag{6.8.6}$$

It is easy to see that

$$z\, \overline{z} = x^2 + y^2 = |z|^2, \tag{6.8.7}$$

where $|z|$ is the usual absolute value of z. The well-known fact that

$$|w\,z| = |z|\,|w| \tag{6.8.8}$$

follows from (6.8.6) and (6.8.7).

Clearly

$$\overline{(\overline{z})} = z \tag{6.8.9}$$

and

$$|\overline{z}| = |z|. \tag{6.8.10}$$

Let $u, v \in \mathbf{R}$ be the real and imaginary parts of w, so that $w = u + i\,v$. The real part of $w\,\overline{z}$ is given by

$$\mathrm{Re}(w\,\overline{z}) = u\,x + v\,y. \tag{6.8.11}$$

As usual, w and z correspond to (u, v) and (x, y), respectively, as elements of $\mathbf{R}^2$. The right side of (6.8.11) is the same as the inner product of (u, v) and (x, y) as elements of $\mathbf{R}^2$.

If $w, z \in \mathbf{C}^n$, then

$$\langle w, z\rangle = \langle w, z\rangle_{\mathbf{C}^n} = \sum_{j=1}^{n} w_j\,\overline{z_j} \tag{6.8.12}$$

is defined as a complex number. This is the standard inner product on $\mathbf{C}^n$. Observe that

$$\overline{\langle w, z\rangle_{\mathbf{C}^n}} = \overline{\Big(\sum_{j=1}^{n} w_j\,\overline{z_j}\Big)} = \sum_{j=1}^{n} z_j\,\overline{w_j} = \langle z, w\rangle_{\mathbf{C}^n}. \tag{6.8.13}$$

As before,

$$\langle z, z\rangle_{\mathbf{C}^n} = \sum_{j=1}^{n} z_j\,\overline{z_j} = \sum_{j=1}^{n} |z_j|^2 = \|z\|_2^2, \tag{6.8.14}$$

where $\|z\|_2$ is the standard Euclidean norm of z, as in (1.4.1). Moreover,

$$|\langle w, z\rangle_{\mathbf{C}^n}| = \Big|\sum_{j=1}^{n} w_j\,\overline{z_j}\Big| \le \sum_{j=1}^{n} |w_j|\,|z_j| \le \|w\|_2\,\|z\|_2, \tag{6.8.15}$$

using the Cauchy–Schwarz inequality (1.3.13) in the third step.

6.9 Sums and inner products

Let X be a nonempty set, and let f, g be real-valued functions on X with finite support in X. Put

$$\langle f, g\rangle = \langle f, g\rangle_{c_{00}(X,\mathbf{R})} = \sum_{x \in X} f(x)\,g(x), \tag{6.9.1}$$

where the sum on the right reduces to a finite sum of real numbers, as in Section 1.6. This may be considered as the standard inner product on $c_{00}(X, \mathbf{R})$. Of course,

$$\langle f, f\rangle_{c_{00}(X,\mathbf{R})} = \sum_{x\in X} f(x)^2 = \|f\|_2^2, \tag{6.9.2}$$

where $\|f\|_2$ is as in Section 1.6. Observe that

$$|\langle f, g\rangle_{c_{00}(X,\mathbf{R})}| = \Big|\sum_{x\in X} f(x)\, g(x)\Big| \le \sum_{x\in X} |f(x)|\, |g(x)| \le \|f\|_2\, \|g\|_2, \tag{6.9.3}$$

by the Cauchy–Schwarz inequality.

Similarly, if f, g are complex-valued functions on X with finite support, then

$$\langle f, g\rangle = \langle f, g\rangle_{c_{00}(X,\mathbf{C})} = \sum_{x\in X} f(x)\, \overline{g(x)} \tag{6.9.4}$$

reduces to a finite sum of complex numbers. This may be considered as the standard inner product on $c_{00}(X, \mathbf{C})$. As before,

$$\langle f, f\rangle_{c_{00}(X,\mathbf{C})} = \sum_{x\in X} |f(x)|^2 = \|f\|_2^2 \tag{6.9.5}$$

for every $f \in c_{00}(X, \mathbf{C})$. Using the Cauchy–Schwarz inequality again, we get that

$$|\langle f, g\rangle_{c_{00}(X,\mathbf{C})}| = \Big|\sum_{x\in X} f(x)\, \overline{g(x)}\Big| \le \sum_{x\in X} |f(x)|\, |g(x)| \le \|f\|_2\, \|g\|_2 \tag{6.9.6}$$

for every $f, g \in c_{00}(X, \mathbf{C})$. We also have that

$$\overline{\langle f, g\rangle_{c_{00}(X,\mathbf{C})}} = \overline{\Big(\sum_{x\in X} f(x)\, \overline{g(x)}\Big)} = \sum_{x\in X} g(x)\, \overline{f(x)} = \langle g, f\rangle_{c_{00}(X,\mathbf{C})} \tag{6.9.7}$$

for every $f, g \in c_{00}(X, \mathbf{C})$.

Suppose now that $f, g \in \ell^2(\mathbf{Z}_+, \mathbf{R})$, and remember that $f\, g \in \ell^1(\mathbf{Z}_+, \mathbf{R})$, as in Section 2.3. In this case, we put

$$\langle f, g\rangle = \langle f, g\rangle_{\ell^2(\mathbf{Z}_+,\mathbf{R})} = \sum_{j=1}^{\infty} f(j)\, g(j), \tag{6.9.8}$$

where the right side converge absolutely. This may be considered as the standard inner product on $\ell^2(\mathbf{Z}_+, \mathbf{R})$. Note that

$$\langle f, f\rangle_{\ell^2(\mathbf{Z}_+,\mathbf{R})} = \sum_{j=1}^{\infty} f(j)^2 = \|f\|_2^2, \tag{6.9.9}$$

where $\|f\|_2$ is as in Section 2.3. The version of the Cauchy–Schwarz inequality mentioned in Section 2.3 implies that

$$(6.9.10)\quad |\langle f, g\rangle_{\ell^2(\mathbf{Z}_+,\mathbf{R})}| = \Big|\sum_{j=1}^{\infty} f(j)\, g(j)\Big| \le \sum_{j=1}^{\infty} |f(j)|\,|g(j)| \le \|f\|_2\,\|g\|_2.$$

Let $f, g \in \ell^2(\mathbf{Z}_+, \mathbf{C})$ be given, so that $|f|\,|g| \in \ell^1(\mathbf{Z}_+, \mathbf{C})$, and hence $f\,\overline{g}$ is an element of $\ell^1(\mathbf{Z}_+, \mathbf{C})$. Put

$$(6.9.11)\qquad \langle f, g\rangle = \langle f, g\rangle_{\ell^2(\mathbf{Z}_+,\mathbf{C})} = \sum_{j=1}^{\infty} f(j)\,\overline{g(j)},$$

which may be considered as the standard inner product on $\ell^2(\mathbf{Z}_+, \mathbf{C})$. As usual,

$$(6.9.12)\qquad \langle f, f\rangle_{\ell^2(\mathbf{Z}_+,\mathbf{C})} = \sum_{j=1}^{\infty} |f(j)|^2 = \|f\|_2^2.$$

We also have that

$$(6.9.13)\quad |\langle f, g\rangle_{\ell^2(\mathbf{Z}_+,\mathbf{C})}| = \Big|\sum_{j=1}^{\infty} f(j)\,\overline{g(j)}\Big| \le \sum_{j=1}^{\infty} |f(j)|\,|g(j)| \le \|f\|_2\,\|g\|_2,$$

by the version of the Cauchy–Schwarz inequality in Section 2.3. Moreover,

$$(6.9.14)\quad \overline{\langle f, g\rangle_{\ell^2(\mathbf{Z}_+,\mathbf{C})}} = \overline{\Big(\sum_{j=1}^{\infty} f(j)\,\overline{g(j)}\Big)} = \sum_{j=1}^{\infty} g(j)\,\overline{f(j)} = \langle g, f\rangle_{\ell^2(\mathbf{Z}_+,\mathbf{C})}.$$

6.10 Integral inner products

Let a, b be real numbers with $a < b$, and let α be a monotonically increasing real-valued function on $[a, b]$. If f, g are continuous real-valued functions on $[a, b]$, then

$$(6.10.1)\qquad \langle f, g\rangle = \langle f, g\rangle_\alpha = \int_a^b f(x)\, g(x)\, d\alpha(x)$$

is defined as a real number, using a Riemann–Stieltjes integral on the right side. Clearly

$$(6.10.2)\qquad \langle f, f\rangle_\alpha = \int_a^b f(x)^2\, d\alpha(x) = \|f\|_{2,\alpha}^2,$$

where $\|f\|_{2,\alpha}$ is as in (3.5.1). Observe that

$$(6.10.3)\qquad \begin{aligned} |\langle f, g\rangle_\alpha| &= \Big|\int_a^b f(x)\, g(x)\, d\alpha(x)\Big| \\ &\le \int_a^b |f(x)|\,|g(x)|\, d\alpha(x) \le \|f\|_{2,\alpha}\,\|g\|_{2,\alpha}, \end{aligned}$$

by the integral version (3.5.8) of the Cauchy–Schwarz inequality for Riemann–Stieltjes integrals. Of course, (6.10.1) is symmetric in f and g, and linear in each of f and g, because of linearity of the integral. If α is strictly increasing on $[a, b]$, then (6.10.2) is positive when $f \not\equiv 0$ on $[a, b]$, as before. In this case, (6.10.1) defines an inner product on the space $C([a, b], \mathbf{R})$ of continuous real-valued functions on $[a, b]$.

If f and g are continuous complex-valued functions on $[a, b]$, then

$$\langle f, g \rangle = \langle f, g \rangle_\alpha = \int_a^b f(x)\,\overline{g(x)}\,d\alpha(x) \tag{6.10.4}$$

is defined as a complex number, using a Riemann–Stieltjes integral on the right side again. In this situation, we also have that

$$\langle f, f \rangle_\alpha = \int_a^b |f(x)|^2\,d\alpha(x) = \|f\|_{2,\alpha}^2, \tag{6.10.5}$$

where $\|f\|_{2,\alpha}$ is as in (3.5.1). As before,

$$\begin{aligned} |\langle f, g \rangle_\alpha| &= \left|\int_a^b f(x)\,\overline{g(x)}\,d\alpha(x)\right| \\ &\le \int_a^b |f(x)|\,|g(x)|\,d\alpha(x) \le \|f\|_{2,\alpha}\,\|g\|_{2,\alpha}, \end{aligned} \tag{6.10.6}$$

using (3.5.8) in the third step. Note that

$$\overline{\langle f, g \rangle_\alpha} = \overline{\Big(\int_a^b f(x)\,\overline{g(x)}\,d\alpha(x)\Big)} = \int_a^b g(x)\,\overline{f(x)}\,d\alpha(x) = \langle g, f \rangle_\alpha. \tag{6.10.7}$$

It is easy to see that (6.10.4) is linear in f, because of linearity of the integral. Similarly, (6.10.4) is conjugate-linear in g, which is to say that it is additive in g, while multiplying g by a complex number t corresponds to multiplying (6.10.4) by $\bar{t}$. If α is strictly increasing on $[a, b]$, then (6.10.5) is positive when $f \not\equiv 0$ on $[a, b]$, and (6.10.4) defines an inner product on the space $C([a, b], \mathbf{C})$ of continuous complex-valued functions on $[a, b]$. If $\alpha(x) = x$ for every $x \in [a, b]$, then (6.10.1) and (6.10.4) are called the standard integral inner products on $C([a, b], \mathbf{R})$ and $C([a, b], \mathbf{C})$, respectively.

Now let f, g be continuous real-valued functions on the real line with compact support. In this case,

$$\langle f, g \rangle = \langle f, g \rangle_{C_{com}(\mathbf{R}, \mathbf{R})} = \int_{-\infty}^{\infty} f(x)\,g(x)\,dx \tag{6.10.8}$$

is defined as a real number, where the integral on the right reduces to a Riemann integral over a bounded interval, as in Section 3.8. This is the standard integral inner product on the space $C_{com}(\mathbf{R}, \mathbf{R})$ of continuous real-valued functions on $\mathbf{R}$ with compact support. As before,

$$\langle f, f \rangle_{C_{com}(\mathbf{R}, \mathbf{R})} = \int_{-\infty}^{\infty} f(x)^2\,dx = \|f\|_2^2, \tag{6.10.9}$$

where $\|f\|_2$ is as defined in (3.8.15). We also have that

$$(6.10.10) \qquad |\langle f, g\rangle_{C_{com}(\mathbf{R},\mathbf{R})}| = \biggl|\int_{-\infty}^{\infty} f(x)\, g(x)\, dx\biggr| \le \int_{\infty}^{\infty} |f(x)|\, |g(x)|\, dx \le \|f\|_2\, \|g\|_2,$$

where the third step reduces to the integral version (3.2.4) of the Cauchy–Schwarz inequality on bounded intervals. Clearly (6.10.8) is symmetric in f and g. The inner product (6.10.8) is linear in f and g, because of the linearity of the integral.

Similarly, if f and g are continuous complex-valued functions on $\mathbf{R}$ with compact support, then

$$(6.10.11) \qquad \langle f, g\rangle = \langle f, g\rangle_{C_{com}(\mathbf{R},\mathbf{C})} = \int_{-\infty}^{\infty} f(x)\, \overline{g(x)}\, dx$$

is defined as a complex number. This is the standard integral inner product on the space $C_{com}(\mathbf{R}, \mathbf{C})$ of continuous complex-valued functions on $\mathbf{R}$ with compact support. As usual,

$$(6.10.12) \qquad \langle f, f\rangle_{C_{com}(\mathbf{R},\mathbf{C})} = \int_{-\infty}^{\infty} |f(x)|^2\, dx = \|f\|_2^2,$$

where $\|f\|_2$ is as in (3.8.15) again. In addition,

$$(6.10.13) \qquad |\langle f, g\rangle_{C_{com}(\mathbf{R},\mathbf{C})}| = \biggl|\int_{-\infty}^{\infty} f(x)\, \overline{g(x)}\, dx\biggr| \le \int_{-\infty}^{\infty} |f(x)|\, |g(x)|\, dx \le \|f\|_2\, \|g\|_2,$$

using (3.2.4) in the third step. Observe that

$$(6.10.14) \qquad \overline{\langle f, g\rangle_{C_{com}(\mathbf{R},\mathbf{C})}} = \overline{\Bigl(\int_{-\infty}^{\infty} f(x)\, \overline{g(x)}\, dx\Bigr)} = \int_{-\infty}^{\infty} g(x)\, \overline{f(x)}\, dx = \langle g, f\rangle_{C_{com}(\mathbf{R},\mathbf{C})}.$$

As before, (6.10.11) is linear in f, because of the linearity of the integral. The inner product (6.10.11) is also conjugate-linear in g.

6.11 Some remarks about n-dimensional volume

Let n be a positive integer. If E is a reasonably nice subset of $\mathbf{R}^n$, then the n-dimensional volume

$$(6.11.1) \qquad \mathrm{Vol}_n(E)$$

of E may be defined in the usual way. In fact, if n-dimensional volume on $\mathbf{R}^n$ is interpreted as n-dimensional Lebesgue measure, then this is defined for Lebesgue measurable subsets of $\mathbf{R}^n$. This includes all open and closed subsets of $\mathbf{R}^n$, as well as their countable unions or intersections. Note that $\mathrm{Vol}_n(E)$ may be $+\infty$ when E is not bounded.

If n-dimensional volume on $\mathbf{R}^n$ is interpreted as Lebesgue outer measure, then $\mathrm{Vol}_n(E)$ is defined for all subsets E of $\mathbf{R}^n$. However, this may not always behave as one might expect in terms of additivity of volumes of unions of disjoint subsets of $\mathbf{R}^n$. This is related to the Banach–Tarski paradox, as in [172, 176, 182].

Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be real numbers with $a_j \le b_j$ for each $j = 1, \ldots, n$. The set

$$\mathcal{C} = \{x \in \mathbf{R}^n : a_j \le x_j \le b_j \text{ for each } j = 1, \ldots, n\} \tag{6.11.2}$$

may be called a *cell* in $\mathbf{R}^n$, as on p31 of [155]. This is the same as the Cartesian product of the closed intervals $[a_j, b_j]$, $j = 1, \ldots, n$. In this case,

$$\mathrm{Vol}_n(\mathcal{C}) = \prod_{j=1}^{n} (b_j - a_j). \tag{6.11.3}$$

In particular, this is equal to 0 when $a_j = b_j$ for any j.

If $a \in \mathbf{R}^n$ and $E \subseteq \mathbf{R}^n$, then put

$$E + a = \{x + a : x \in E\}, \tag{6.11.4}$$

which is the translation of E by a in $\mathbf{R}^n$. It is well known that

$$\mathrm{Vol}_n(E + a) = \mathrm{Vol}_n(E), \tag{6.11.5}$$

which is to say that the n-dimensional volume on $\mathbf{R}^n$ is invariant under translations. It is easy to check directly that this is compatible with (6.11.3).

If $t \in \mathbf{R}$ and $E \subseteq \mathbf{R}^n$, then put

$$t\,E = \{t\,x : x \in E\}, \tag{6.11.6}$$

which corresponds to dilating E by t in $\mathbf{R}^n$. It is also well known that

$$\mathrm{Vol}_n(t\,E) = |t|^n \, \mathrm{Vol}_n(E). \tag{6.11.7}$$

One can check directly that this is compatible with (6.11.3) as well.

6.12 Volumes and Lipschitz mappings

Let n be a positive integer, and let N be a norm on $\mathbf{R}^n$, so that $d_N(x, y) = N(x - y)$ defines a metric on $\mathbf{R}^n$, as usual. Observe that

$$d_N(x + a, y + a) = N((x + a) - (y + a)) = N(x - y) = d_N(x, y) \tag{6.12.1}$$

for all $a, x, y \in \mathbf{R}^n$, so that $d_N(\cdot,\cdot)$ is invariant under translations on $\mathbf{R}^n$. Similarly,

$$d_N(t\,x, t\,y) = N(t\,x - t\,y) = N(t\,(x-y)) = |t|\,N(x-y) = |t|\,d_N(x,y) \tag{6.12.2}$$

for every $x, y \in \mathbf{R}^n$ and $t \in \mathbf{R}$.

Suppose that f is a Lipschitz mapping from $E \subseteq \mathbf{R}^n$ into $\mathbf{R}^n$ with constant $C \geq 0$ with respect to $d_N(\cdot,\cdot)$ and its restriction to E, so that

$$d_N(f(x), f(y)) \leq C\,d_N(x,y) \tag{6.12.3}$$

for every $x, y \in E$. It is well known that

$$\mathrm{Vol}_n(f(E)) \leq C^n\,\mathrm{Vol}_n(E). \tag{6.12.4}$$

One can look at this in terms of the standard way of changing variables in n-dimensional integrals, under suitable conditions. Although the latter is discussed in many textbooks, one may also be interested in the articles [25, 115, 116, 169], as well as [47, 184, 185], in connection with Lebesgue measure and integration. Of course, Lebesgue measure and integration are discussed in many textbooks too, and one may be interested in the articles [40, 58, 117] as well.

Alternatively, one can consider (6.12.4) in terms of n-dimensional Hausdorff measure on $\mathbf{R}^n$ with respect to $d_N(\cdot,\cdot)$. It is well known that this is equal to a constant multiple of Lebesgue outer measure on $\mathbf{R}^n$.

If α is any positive real number, then α-dimensional Hausdorff measure may be defined on any metric space. There is an analogue of (6.12.4) for Lipschitz mappings between arbitrary metric spaces, using Hausdorff measures of the same dimension on the domain and range.

If k is a positive integer less than or equal to n, then k-dimensional Hausdorff measure on $\mathbf{R}^n$ with respect to the standard Euclidean metric is related to the usual k-dimensional volume of reasonably nice k-dimensional submanifolds of $\mathbf{R}^n$.

6.13 Bounded vector-valued functions

Let X be a nonempty set, and let m be a positive integer. Consider the spaces

$$c(X, \mathbf{R}^m),\ c(X, \mathbf{C}^m) \tag{6.13.1}$$

of functions on X with values in $\mathbf{R}^m$, $\mathbf{C}^m$, respectively. These spaces with $m = 1$ were discussed in Section 3.10. As before, if f and g are functions on X with values in $\mathbf{R}^m$ or $\mathbf{C}^m$, then $f + g$ defines a function on X with values in $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. Similarly, if t is a real or complex number, as appropriate, then $t\,f$ is a function on X with values in $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. As usual, $c(X, \mathbf{R}^m)$ and $c(X, \mathbf{C}^m)$ are basic classes of examples of vector spaces over the real and complex numbers, respectively. We shall also be interested in linear subspaces of these spaces, which are subsets of the spaces that contain 0 and are closed under addition and scalar multiplcation, as before.

If f is a function on X with values in $\mathbf{R}^m$ or $\mathbf{C}^m$, then let $f_j(x)$ be the jth coordinate of $f(x)$ for each $j = 1, \ldots, m$ and $x \in X$. Thus f_j is a real or complex-valued function on X, as appropriate, for each j. Of course, any m-tuple of real or complex-valued functions on X determines a function on X with values in $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, in this way.

Let N be a norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, so that $d_N(v, w) = N(v - w)$ defines a metric on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. Using this, we get the corresponding space

$$\mathcal{B}(X, \mathbf{R}^m) = \mathcal{B}_N(X, \mathbf{R}^m) \text{ or } \mathcal{B}(X, \mathbf{C}^m) = \mathcal{B}_N(X, \mathbf{C}^m) \tag{6.13.2}$$

of mappings from X into $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, that are bounded with respect to $d_N(\cdot, \cdot)$, as in Section 1.11. This space does not depend on the particular norm N, because any norm on $\mathbf{R}^m$ or $\mathbf{C}^m$ can be compared with the standard Euclidean norm up to constant factors, as in Sections 6.6 and 6.7. The corresponding supremum metric does depend on N, but it can be compared with the supremum metric associated to the standard Euclidean norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, with the same constant factors.

We may also use the notation

$$\ell^\infty(X, \mathbf{R}^m) = \ell^\infty_N(X, \mathbf{R}^m) \text{ or } \ell^\infty(X, \mathbf{C}^m) = \ell^\infty_N(X, \mathbf{C}^m) \tag{6.13.3}$$

for (6.13.2), as appropriate. Note that a function f on X with values in $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, is bounded with respect to $d_N(\cdot, \cdot)$ if and only if

$$N(f(x)) \tag{6.13.4}$$

is bounded as a real-valued function on X. This happens if and only if f_j is bounded as a real or complex-valued function on X, as appropriate, for each j, because of the usual comparisons of N with the standard Euclidean norm. It is easy to see that (6.13.3) is a linear subspace of $c(X, \mathbf{R}^m)$ or $c(X, \mathbf{C}^m)$, as appropriate.

If f is an element of (6.13.3), as appropriate, then put

$$\|f\|_{\infty,N} = \sup\{N(f(x)) : x \in X\}. \tag{6.13.5}$$

This is a nonnegative real number, which is equal to 0 exactly when $f = 0$ on X. One can check that

$$\|t\, f\|_{\infty,N} = |t|\, \|f\|_{\infty,N} \tag{6.13.6}$$

for every $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate. If g is another element of (6.13.3), as appropriate, then one can verify that

$$\|f + g\|_{\infty,N} \le \|f\|_{\infty,N} + \|g\|_{\infty,N}. \tag{6.13.7}$$

This means that (6.13.5) defines a norm on (6.13.3), as appropriate. This is the *supremum norm* associated to N. This can be compared with the supremum norm associated to the standard Euclidean norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, with the same constant factors as before. Note that

$$\|f - g\|_{\infty,N} \tag{6.13.8}$$

is the same as the supremum metric on (6.13.3), as appropriate, corresponding to $d_N(\cdot,\cdot)$.

Suppose now that (X, d_X) is a nonempty metric space, and consider the space

$$C(X, \mathbf{R}^m) = C_N(X, \mathbf{R}^m) \text{ or } C(X, \mathbf{C}^m) = C_N(X, \mathbf{C}^m) \tag{6.13.9}$$

of continuous mappings from X into $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, with respect to $d_N(\cdot,\cdot)$ on the range. This is the same as the space of continuous mappings from X into $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, with respect to the standard Euclidean metric on the range, because of the usual comparisons between N and the standard Euclidean norm. Equivalently, one can check that a mapping f from X into $\mathbf{R}^m$ or $\mathbf{C}^m$ is continuous if and only if f_j is continuous as a real or complex-valued function on X, as appropriate, for each $j = 1, \ldots, m$. In particular, the space of continuous mappings from X into $\mathbf{R}^m$ or $\mathbf{C}^m$ is a linear subspace of $c(X, \mathbf{R}^m)$ or $c(X, \mathbf{C}^m)$, as appropriate.

Similarly, we get the space

$$C_b(X, \mathbf{R}^m) = C_{b,N}(X, \mathbf{R}^m) \text{ or } C_b(X, \mathbf{C}^m) = C_{b,N}(X, \mathbf{C}^m) \tag{6.13.10}$$

of bounded continuous mappings from X into $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, with respect to $d_N(\cdot,\cdot)$ on the range. This is the same as the analogous space defined using the standard Euclidean metric on the range, as before. This is a linear subspace of (6.13.3) and (6.13.9), as appropriate.

Chapter 7

Matrix norms and Lipschitz conditions

7.1 Real and complex matrices

Let m and n be positive integers, and let us consider $m \times n$ matrices with entries in the real or complex numbers. Such a matrix may be denoted as

$$[a_{j,l}] = [a_{j,l}]_{j,l=1}^{m,n}, \tag{7.1.1}$$

where $a_{j,l}$ is a real or complex number, as appropriate, for each $j = 1, \ldots, m$ and $l = 1, \ldots, n$. Let $M_{m,n}(\mathbf{R})$ and $M_{m,n}(\mathbf{C})$ be the spaces of $m \times n$ matrices with entries in $\mathbf{R}$ and $\mathbf{C}$, respectively.

If $[a_{j,l}]$ and $[b_{j,l}]$ are $m \times n$ matrices with real or complex entries, then their sum is defined as an $m \times n$ matrix by adding the corresponding entries, so that

$$[a_{j,l}] + [b_{j,l}] = [a_{j,l} + b_{j,l}]. \tag{7.1.2}$$

Similarly, if t is a real or complex number, as appropriate, then t times $[a_{j,l}]$ is defined as an $m \times n$ matrix by multiplying the entries of $[a_{j,l}]$ by t,

$$t\,[a_{j,l}] = [t\,a_{j,l}]. \tag{7.1.3}$$

This makes $M_{m,n}(\mathbf{R})$ and $M_{m,n}(\mathbf{C})$ into vector spaces over the real and complex numbers, respectively. Of course, one can also identify $M_{m,n}(\mathbf{R})$ and $M_{m,n}(\mathbf{C})$ with $\mathbf{R}^{m\,n}$ and $\mathbf{C}^{m\,n}$, respectively.

As usual, a nonnegative real-valued function N defined on $M_{m,n}(\mathbf{R})$ or $M_{m,n}(\mathbf{C})$ is said to be a *norm* if it satisfies the following three conditions. First,

$$N([a_{j,l}]) = 0 \tag{7.1.4}$$

if and only if $[a_{j,l}] = 0$ as a matrix, which means that $a_{j,l} = 0$ for every $j = 1, \ldots, m$ and $l = 1, \ldots, n$. Second,

$$N(t\,[a_{j,l}]) = |t|\,N([a_{j,l}]) \tag{7.1.5}$$

for every $[a_{j,l}] \in M_{m,n}(\mathbf{R})$ or $M_{m,n}(\mathbf{C})$ and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate. Third,

$$N([a_{j,l}] + [b_{j,l}]) \le N([a_{j,l}]) + N([b_{j,l}]) \tag{7.1.6}$$

for every $[a_{j,l}], [b_{j,l}] \in M_{m,n}(\mathbf{R})$ or $M_{m,n}(\mathbf{C})$, as appropriate. In this case,

$$d_N([a_{j,l}], [b_{j,l}]) = N([a_{j,l}] - [b_{j,l}]) = N([a_{j,l} - b_{j,l}]) \tag{7.1.7}$$

defines a metric on $M_{m,n}(\mathbf{R})$ or $M_{m,n}(\mathbf{C})$, as appropriate. If $M_{m,n}(\mathbf{R})$ and $M_{m,n}(\mathbf{C})$ are identified with $\mathbf{R}^{m\,n}$ and $\mathbf{C}^{m\,n}$, respectively, as before, then the definition of a norm on $M_{m,n}(\mathbf{R})$ or $M_{m,n}(\mathbf{C})$ corresponds exactly to the earlier definitions for $\mathbf{R}^{m\,n}$ and $\mathbf{C}^{m\,n}$ in Sections 1.3 and 1.4. Similarly, the metric associated to a norm on $M_{m,n}(\mathbf{R})$ or $M_{m,n}(\mathbf{C})$ corresponds to the analogous notions for $\mathbf{R}^{m\,n}$ and $\mathbf{C}^{m\,n}$.

Put

$$\|[a_{j,l}]\|_{HS} = \Big(\sum_{j=1}^{m} \sum_{l=1}^{n} |a_{j,l}|^2\Big)^{1/2} \tag{7.1.8}$$

for every $[a_{j,l}] \in M_{m,n}(\mathbf{R})$ or $M_{m,n}(\mathbf{C})$. This is known as the *Hilbert–Schmidt norm* on $M_{m,n}(\mathbf{R})$ and $M_{m,n}(\mathbf{C})$. More precisely, (7.1.8) corresponds to the standard Euclidean norm on $\mathbf{R}^{m\,n}$ and $\mathbf{C}^{m\,n}$, using the identifications mentioned earlier. In particular, the fact that (7.1.8) defines a norm on $M_{m,n}(\mathbf{R})$ and $M_{m,n}(\mathbf{C})$ follows from the analogous statements for the standard Euclidean norms on $\mathbf{R}^{m\,n}$ and $\mathbf{C}^{m\,n}$. More precisely, the triangle inequality for (7.1.8) follows from the triangle inequality for the Euclidean norm, and the first two requirements of a norm can be verified directly.

7.2 Matrices and linear mappings

Let m and n be positive integers again. As usual, a mapping A from $\mathbf{R}^n$ into $\mathbf{R}^m$ is said to be *linear* if

$$A(v + w) = A(v) + A(w) \tag{7.2.1}$$

for every $v, w \in \mathbf{R}^n$, and

$$A(t\,v) = t\,A(v) \tag{7.2.2}$$

for every $v \in \mathbf{R}^n$ and $t \in \mathbf{R}$. Similarly, a mapping A from $\mathbf{C}^n$ into $\mathbf{C}^m$ is said to be *(complex) linear* if (7.2.1) holds for every $v, w \in \mathbf{C}^n$, and (7.2.2) holds for every $v \in \mathbf{C}^n$ and $t \in \mathbf{C}$.

Let $[a_{j,l}]$ be an $m \times n$ matrix whose entries are real or complex numbers. If $v = (v_1, \ldots, v_n) \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, then put

$$(A(v))_j = \sum_{l=1}^{n} a_{j,l}\,v_l \tag{7.2.3}$$

for each $j = 1, \ldots, m$. This defines $A(v)$ as an element of $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, where the jth coordinate of $A(v)$ is given by (7.2.3) for each $j =$

$1, \ldots, m$. It is easy to see that A is linear as a mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate. Conversely, it is well known and not too difficult to show that every linear mapping A from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$ corresponds to a unique $m \times n$ matrix with real or complex entries, as appropriate, in this way.

Let $e_k = (e_{k,1}, \ldots, e_{k,n})$ be the kth standard basis vector in $\mathbf{R}^n$ or $\mathbf{C}^n$ for each $k = 1, \ldots, n$. Thus $e_{k,l} = 1$ when $k = l$, and $e_{k,l} = 0$ when $k \neq l$. If A corresponds to $[a_{j,l}] \in M_{m,n}(\mathbf{R})$ or $M_{m,n}(\mathbf{C})$ as in (7.2.3), then

$$(A(e_k))_j = a_{j,k} \tag{7.2.4}$$

for every $j = 1, \ldots, m$ and $k = 1, \ldots, n$. If A is any linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$, then one can use (7.2.4) to define an $m \times n$ matrix $[a_{j,l}]$ of real or complex numbers, as appropriate. This implies that (7.2.3) holds for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, because of linearity, and by expressing v as a linear combination of $e_1, \ldots, e_n$.

Let $[a_{j,l}] \in M_{m,n}(\mathbf{R})$ or $M_{m,n}(\mathbf{C})$ be given again, and let A be the corresponding linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as in (7.2.3). Using (7.2.4), we get that

$$\|A(e_k)\|_2 = \Big(\sum_{j=1}^{m} |a_{j,k}|^2\Big)^{1/2} \tag{7.2.5}$$

for every $k = 1, \ldots, n$. More precisely, the left side of (7.2.5) refers to the standard Euclidean norm of $A(e_k)$ in $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. It follows that the Hilbert–Schmidt norm of $[a_{j,l}]$ defined in the previous section can be given by

$$\|[a_{j,l}]\|_{HS} = \Big(\sum_{k=1}^{n} \|A(e_k)\|_2^2\Big)^{1/2}. \tag{7.2.6}$$

Let $v = (v_1, \ldots, v_n) \in \mathbf{R}^n$ or $\mathbf{C}^n$ be given, as appropriate. Remember that

$$v = \sum_{k=1}^{n} v_k\, e_k, \tag{7.2.7}$$

where the right side is a linear combination of $e_1, \ldots, e_n$ in $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Using the linearity of A, we get that

$$A(v) = \sum_{k=1}^{n} v_k\, A(e_k), \tag{7.2.8}$$

where the right side is a linear combination of $A(e_1), \ldots, A(e_k)$ in $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. It follows that

$$\|A(v)\|_2 \le \sum_{k=1}^{n} |v_k|\, \|A(e_k)\|_2, \tag{7.2.9}$$

where $\| \cdot \|_2$ refers to the standard Euclidean norm on $\mathbf{R}^m$ or $\mathbf{C}^m$ again, as appropriate. This implies that

$$\|A(v)\|_2 \le \Big(\sum_{k=1}^n |v_k|^2\Big)^{1/2} \|[a_{j,l}]\|_{HS}, \tag{7.2.10}$$

using the Cauchy–Schwarz inequality on the right side of (7.2.9).

7.3 Some related estimates

Let m and n be positive integers, and let A be a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$, or from $\mathbf{C}^n$ into $\mathbf{C}^m$. Also let N be a norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. If $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, then v can be expressed as a linear combination of the standard basis vectors $e_1, \ldots, e_n$ in $\mathbf{R}^n$ or $\mathbf{C}^n$ as in (7.2.7). Hence $A(v)$ can be expressed as in (7.2.8), so that

$$N(A(v)) \le \sum_{k=1}^n |v_k| \, N(A(e_k)). \tag{7.3.1}$$

This corresponds to (7.2.9) when N is the standard Euclidean norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate.

Let us now look at this in terms of various norms on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Let us start with the standard Euclidean norm

$$\|v\|_2 = \Big(\sum_{k=1}^n |v_k|^2\Big)^{1/2} \tag{7.3.2}$$

of v. As before, we can apply the Cauchy–Schwarz inequality to the right side of (7.3.1) to get that

$$N(A(v)) \le \Big(\sum_{k=1}^n N(A(e_k))^2\Big)^{1/2} \|v\|_2. \tag{7.3.3}$$

This is the same as (7.2.10) when N is the standard Euclidean norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate.

Consider the norm

$$\|v\|_1 = \sum_{k=1}^n |v_k| \tag{7.3.4}$$

of v discussed in Sections 1.3 and 1.4. It is easy to see that

$$N(A(v)) \le \Big(\max_{1 \le k \le n} N(A(e_k))\Big) \|v\|_1, \tag{7.3.5}$$

using (7.3.1). Similarly, we can consider the norm

$$\|v\|_\infty = \max_{1 \le k \le n} |v_k| \tag{7.3.6}$$

of v discussed in Sections 1.3 and 1.4 as well. Observe that

$$N(A(v)) \le \Big(\sum_{k=1}^{n} N(A(e_k))\Big) \, \|v\|_\infty, \tag{7.3.7}$$

by (7.3.1) again.

Let N_0 be any norm on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. One can check that there is a nonnegative real number C such that

$$N(A(v)) \le C \, N_0(v) \tag{7.3.8}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. This follows from the previous remarks when $N_0(v)$ is given by (7.3.2), (7.3.4), or (7.3.6). Otherwise, one can use the fact that (7.3.2) is bounded by a constant multiple of $N_0(v)$, as in Section 6.7.

If $v, w \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, then

$$N(A(v) - A(w)) = N(A(v - w)) \le C \, N_0(v - w). \tag{7.3.9}$$

This uses the linarity of A in the first step, and (7.3.8) in the second step. Thus

$$d_N(A(v), A(w)) \le C \, d_{N_0}(v, w), \tag{7.3.10}$$

where d_{N_0} is the metric associated to N_0 on $\mathbf{R}^n$ or $\mathbf{C}^n$, and d_N is the metric associated to N on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate.

7.4 Spaces of linear mappings

Let m and n be positive integers again. The space of linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$ may be denoted $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$. Similarly, the space of (complex) linear mappings from $\mathbf{C}^n$ into $\mathbf{C}^m$ may be denoted $\mathcal{L}(\mathbf{C}^n, \mathbf{C}^m)$.

Let A, B be linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$, and let t be a real or complex number, as appropriate. Thus $t\,A$ can be defined as a mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$, or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate, by putting

$$(t\,A)(v) = t\,A(v) \tag{7.4.1}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Similarly, $A + B$ can be defined as a mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$, or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate, by putting

$$(A + B)(v) = A(v) + B(v) \tag{7.4.2}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. It is easy to see that these are also linear as mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$, or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate.

As in Section 7.2, there are standard one-to-one correspondences between $M_{m,n}(\mathbf{R})$ and $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$, and between $M_{m,n}(\mathbf{C})$ and $\mathcal{L}(\mathbf{C}^n, \mathbf{C}^m)$. These correspondences are linear, in the sense that sums of matrices are associated to sums of linear mappings, and similarly for multiplication of matrices and linear mappings by real or complex numbers, as appropriate.

A nonnegative real-valued function N on $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ or $\mathcal{L}(\mathbf{C}^n, \mathbf{C}^m)$ is said to be a *norm* if it satisfies the usual three conditions, as follows. First, if A is a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate, then

$$N(A) = 0 \tag{7.4.3}$$

if and only if $A = 0$, which means that $A(v) = 0$ for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Second, if A is a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$ and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then

$$N(t\,A) = |t|\,N(A). \tag{7.4.4}$$

Third, if A and B are linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate, then

$$N(A + B) \le N(A) + N(B). \tag{7.4.5}$$

Under these conditions,

$$d_N(A, B) = N(A - B) \tag{7.4.6}$$

defines a metric on $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ or $\mathcal{L}(\mathbf{C}^n, \mathbf{C}^m)$, as appropriate. Observe that norms on $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ and $\mathcal{L}(\mathbf{C}^n, \mathbf{C}^m)$ correspond exactly to norms on $M_{m,n}(\mathbf{R})$ and $M_{m,n}(\mathbf{C})$, respectively, using the correspondence between matrices and linear mappings described in Section 7.2. The metrics associated to these norms correspond to each other in the same way.

Let A be a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$ again. Also let $e_1, \ldots, e_n$ be the standard basis vectors in $\mathbf{R}^n$ or $\mathbf{C}^n$, as in Section 7.2, and let $\|\cdot\|_2$ be the standard Euclidean norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. Put

$$\|A\|_{HS} = \Big(\sum_{k=1}^n \|A(e_k)\|_2^2\Big)^{1/2}, \tag{7.4.7}$$

which corresponds exactly to the Hilbert–Schmidt norm of the matrix associated to A, as in Section 7.2. This defines a norm on each of $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ and $\mathcal{L}(\mathbf{C}^n, \mathbf{C}^m)$, which may be called the *Hilbert–Schmidt norm* as well.

7.5 Operator norms

Let m and n be positive integers, and let A be a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$, or from $\mathbf{C}^n$ into $\mathbf{C}^m$. Also let N_0 be a norm on $\mathbf{R}^n$ or $\mathbf{C}^n$, and let N be a norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. The corresponding *operator norm*

$$\|A\|_{op} \tag{7.5.1}$$

of A is defined to be the supremum of

$$\frac{N(A(v))}{N_0(v)} \tag{7.5.2}$$

over all $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, with $v \neq 0$. Note that this ratio is bounded, by (7.3.8), so that the supremum is finite. More precisely, (7.3.8) is equivalent to saying that the nonnegative real number C is an upper bound for (7.5.2) when $v \neq 0$, because (7.3.8) holds automatically when $v = 0$. Thus (7.3.8) holds exactly when

$$\|A\|_{op} \le C. \tag{7.5.3}$$

In particular, (7.3.8) holds with $C = \|A\|_{op}$. Alternatively, $\|A\|_{op}$ is the infimum of the nonnegative real numbers C for which (7.3.8) holds.

Note that

$$\frac{N(A(v))}{N_0(v)} = N((1/N_0(v))\,A(v)) = N(A((1/N_0(v))\,v)) \tag{7.5.4}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, with $v \neq 0$. Similarly,

$$N_0((1/N_0(v))\,v) = N_0(v)/N_0(v) = 1 \tag{7.5.5}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, with $v \neq 0$. Using this, one can verify that $\|A\|_{op}$ is the same as the supremum of

$$N(A(v)) \tag{7.5.6}$$

over all $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, such that $N_0(v) = 1$. This is also the same as the supremum of (7.5.6) over all $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, with $N_0(v) \le 1$. Of course, (7.5.6) is equal to 0 when $v = 0$, and otherwise

$$N(A(v)) = N(A((1/N_0(v))\,v))\,N_0(v), \tag{7.5.7}$$

as in (7.5.4).

By construction, $\|A\|_{op}$ is a nonnegative real number, and $\|A\|_{op} = 0$ if and only if $A(v) = 0$ for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. If $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then $t\,A$ also defines a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate. It is easy to see that

$$\|t\,A\|_{op} = |t|\,\|A\|_{op}, \tag{7.5.8}$$

because

$$N(t\,A(v)) = |t|\,N(A(v)) \tag{7.5.9}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Let B be another linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate, so that $A + B$ defines another such linear mapping. Observe that

$$\begin{aligned} N((A+B)(v)) = N(A(v) + B(v)) &\le N(A(v)) + N(B(v)) \\ &\le \|A\|_{op}\,N_0(v) + \|B\|_{op}\,N_0(v) \\ &= (\|A\|_{op} + \|B\|_{op})\,N_0(v) \end{aligned} \tag{7.5.10}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. This implies that

$$\|A + B\|_{op} \le \|A\|_{op} + \|B\|_{op}. \tag{7.5.11}$$

Thus the operator norm satisfies the requirements of a norm on the space of linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate.

Suppose that N_0 is the standard Euclidean norm on $\mathbf{R}^n$ or $\mathbf{C}^n$, and that N is the standard Euclidean norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. In this case, we get that

$$\|A\|_{op} \le \|A\|_{HS}, \tag{7.5.12}$$

as in (7.2.10), where the right side is the Hilbert–Schmidt norm of A, as in the previous section. One can check that

$$\|A\|_{HS} \le n^{1/2}\, \|A\|_{op}, \tag{7.5.13}$$

directly from the definitions of these two norms.

7.6 Determinants and volume

Let n be a positive integer, and let $[a_{j,l}]$ be an $n \times n$ matrix of real or complex numbers. The *determinant*

$$\det[a_{j,l}] \tag{7.6.1}$$

can be defined as a real or complex number, as appropriate, in a standard way. If A is the linear mapping from $\mathbf{R}^n$ or $\mathbf{C}^n$ into itself, as appropriate, associated to $[a_{j,l}]$, then the *determinant* of A is defined by

$$\det A = \det[a_{j,l}]. \tag{7.6.2}$$

The determinant of the identity mapping I on $\mathbf{R}^n$ or $\mathbf{C}^n$ is equal to 1. Similarly, if t is a real or complex number, as appropriate, then

$$\det(t\, I) = t^n. \tag{7.6.3}$$

Let A be a linear mapping from $\mathbf{R}^n$ into itself, and let E be a subset of $\mathbf{R}^n$. It is well known that

$$\mathrm{Vol}_n(A(E)) = |\det A|\, \mathrm{Vol}_n(E), \tag{7.6.4}$$

where $\mathrm{Vol}_n(\cdot)$ is the usual n-dimensional volume of a subset of $\mathbf{R}^n$. This is discussed in many textbooks, and one may also be interested in the articles [31, 84].

More precisely, E should be sufficiently nice for the volume to be defined, depending on the definition of the volume being used. The right side of (7.6.4) should be interpreted as being equal to 0 when $\det A = 0$, even if $\mathrm{Vol}_n(E)$ may be $+\infty$. Note that $|\det A|$ is uniquely determined by (7.6.4) when $\mathrm{Vol}_n(E)$ is positive and finite.

Let N be a norm on $\mathbf{R}^n$, and let $\|A\|_{op}$ be the corresponding operator norm of A, as in the previous section. More precisely, this uses N as the norm on $\mathbf{R}^n$ as both the domain and the range of A. It is well known that

$$|\det A| \le \|A\|_{op}^n \tag{7.6.5}$$

under these conditions.

Suppose for the moment that N is the standard Euclidean norm on $\mathbf{R}^n$. If A can be diagonalized using an orthonormal basis for $\mathbf{R}^n$ with respect to the standard inner product, then (7.6.5) can be verified directly. Otherwise, one can reduce to that case, by considering the composition of A with its adjoint with respect to the standard inner product on $\mathbf{R}^n$.

If N is any norm on $\mathbf{R}^n$, then let $\overline{B}_N$ be the closed unit ball in $\mathbf{R}^n$ with respect to N, as in Section A.6. It is easy to see that

$$A(\overline{B}_N) \subseteq \|A\|_{op}\, \overline{B}_N, \tag{7.6.6}$$

by the definition of the operator norm. Note that $\|A\|_{op}\, \overline{B}_N$ is the same as the closed ball in $\mathbf{R}^n$ centered at 0 with radius $\|A\|_{op}$ with respect to the metric $d_N(v, w) = N(v - w)$ associated to N. Using (7.6.6), we get that

$$\mathrm{Vol}_n(A(\overline{B}_N)) \le \mathrm{Vol}_n(\|A\|_{op}\, \overline{B}_N) = \|A\|_{op}^n\, \mathrm{Vol}_n(\overline{B}_N). \tag{7.6.7}$$

One can use this to get (7.6.5) from (7.6.4), with $E = \overline{B}_N$.

Let A be a linear mapping from $\mathbf{R}^n$ or $\mathbf{C}^n$ into itself again. If $v \in \mathbf{R}^n$ or $\mathbf{C}^n$ satisfies

$$A(v) = \lambda\, v \tag{7.6.8}$$

for some $\lambda \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then v is said to be an *eigenvector* of A with *eigenvalue* λ. Let N be a norm on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, and let $\|A\|_{op}$ be the corresponding operator norm of A, using N on both the domain and the range. Observe that

$$|\lambda|\, N(v) = N(\lambda\, v) = N(A(v)) \le \|A\|_{op}\, N(v), \tag{7.6.9}$$

by (7.6.8) and the definition of the operator norm. This implies that

$$|\lambda| \le \|A\|_{op} \tag{7.6.10}$$

when $v \ne 0$.

In the complex case, it is well known that the determinant of A is equal to the product of its n eigenvalues, associated to nonzero eigenvectors, counted with their appropriate multiplicities. One can use this to get (7.6.5) from (7.6.10).

7.7 Lipschitz constants

Let (X, d_X) and (Y, d_Y) be (nonempty) metric spaces. Remember that a mapping f from X into Y is said to be *Lipschitz* if

$$d_Y(f(x), f(w)) \le C\, d_X(x, w) \tag{7.7.1}$$

for some nonnegative real number C and every $x, w \in X$. In this case, we may also say that f is Lipschitz with constant C, to make the role of C more explicit. Let $\mathrm{Lip}(X, Y)$ be the space of all Lipschitz mappings from X into Y.

Let f be a mapping from X into Y again, and note that (7.7.1) holds automatically when $x = w$. If x, w are distinct elements of X, then (7.7.1) is the same as saying that

$$\frac{d_Y(f(x), f(w))}{d_X(x, w)} \le C. \tag{7.7.2}$$

If f is Lipschitz, and X has at least two elements, then put

$$\mathrm{Lip}(f) = \sup\left\{\frac{d_Y(f(x), f(w))}{d_X(x, w)} : x, w \in X,\ x \ne w\right\}. \tag{7.7.3}$$

Otherwise, if X has only one element, then we take $\mathrm{Lip}(f) = 0$. Thus (7.7.1) holds for some $C \ge 0$ and every $x, w \in X$ if and only if f is Lipschitz and

$$\mathrm{Lip}(f) \le C. \tag{7.7.4}$$

In particular, if f is Lipschitz, then (7.7.1) holds with $C = \mathrm{Lip}(f)$. Equivalently, $\mathrm{Lip}(f)$ is the infimum of the nonnegative real numbers C such that (7.7.1) holds for every $x, w \in X$.

Let m and n be positive integers, and suppose for the moment that $X = \mathbf{R}^n$ and $Y = \mathbf{R}^m$, or that $X = \mathbf{C}^n$ and $Y = \mathbf{C}^m$. Also let N_0 be a norm on $\mathbf{R}^n$ or $\mathbf{C}^n$, and let N be a norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. Thus N_0 determines a metric d_{N_0} on $\mathbf{R}^n$ or $\mathbf{C}^n$, and N determines a metric d_N on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. Let A be a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$, or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate. Under these conditions, A is Lipschitz with respect to d_{N_0} and d_N, as in (7.3.10). More precisely,

$$\mathrm{Lip}(A) = \|A\|_{op}, \tag{7.7.5}$$

where $\mathrm{Lip}(A)$ is defined using the metrics d_{N_0} and d_N as in the preceding paragraph, and the operator norm $\|A\|_{op}$ of A is defined using N_0 and N as in Section 7.5. This is easy to verify, directly from the definitions.

Let (X, d_X) be any nonempty metric space again, and let m be a positive integer. Let us take $Y = \mathbf{R}^m$ or $\mathbf{C}^m$, equipped with a norm N, which determines a metric d_N in the usual way. In this situation, if f is a mapping from X into Y, then (7.7.1) is the same as saying that

$$N(f(x) - f(w)) = d_N(f(x), f(w)) \le C\, d_X(x, w). \tag{7.7.6}$$

Let t be a real or complex number, as appropriate, so that $t\,f$ also defines a mapping from X into Y, and

$$\begin{aligned} d_N(t\,f(x), t\,f(w)) &= N(t\,f(x) - t\,f(w)) = N(t\,(f(x) - f(y))) \\ &= |t|\, N(f(x) - f(w)) = |t|\, d_N(f(x), f(w)) \end{aligned} \tag{7.7.7}$$

for every $x, w \in X$. If f is Lipschitz, then it is easy to see that $t\,f$ is Lipschitz too, with

$$\mathrm{Lip}(t\,f) = |t|\ \mathrm{Lip}(f). \tag{7.7.8}$$

Similarly, let g be another mapping from X into Y, so that $f + g$ defines a mapping from X into Y as well. Observe that

$$\begin{aligned} d_N((f+g)(x),(f+g)(w)) &= N((f(x)+g(x)) - (f(w)+g(w))) \\ &= N((f(x)-f(w)) + (g(x)-g(w))) \\ &\le N(f(x)-f(w)) + N(g(x)-g(w)) \end{aligned} \tag{7.7.9}$$

for every $x, w \in X$. If f and g are both Lipschitz, then it follows that

$$\begin{aligned} d_N((f+g)(x),(f+g)(w)) &\le \mathrm{Lip}(f)\, d_X(x,w) + \mathrm{Lip}(g)\, d_X(x,w) \\ &= (\mathrm{Lip}(f) + \mathrm{Lip}(g))\, d_X(x,w) \end{aligned} \tag{7.7.10}$$

for every $x, w \in X$. This implies that $f + g$ is Lipschitz, with

$$\mathrm{Lip}(f+g) \le \mathrm{Lip}(f) + \mathrm{Lip}(g). \tag{7.7.11}$$

In particular, the space of Lipschitz mappings from X into $\mathbf{R}^m$ or $\mathbf{C}^m$ may be considered as a vector space over the real or complex numbers, as appropriate, with respect to pointwise addition and scalar multiplication. In the terminology of Section A.6, $\mathrm{Lip}(f)$ defines a seminorm on this vector space. More precisely, $\mathrm{Lip}(f) = 0$ if and only if f is a constant mapping on X.

7.8 Compositions and isometries

Let (X, d_X), (Y, d_Y), and (Z, d_Z) be (nonempty) metric spaces. If f is a mapping from X into Y, and g is a mapping from Y into Z, then the composition $g \circ f$ can be defined as a mapping from X into Z, as usual. Suppose that f and g are both Lipschitz, so that

$$\begin{aligned} d_Z((g\circ f)(x),(g\circ f)(w)) &= d_Z(g(f(x)), g(f(w))) \\ &\le \mathrm{Lip}(g)\, d_Y(f(x), f(w)) \\ &\le \mathrm{Lip}(f)\ \mathrm{Lip}(g)\, d_X(x,w) \end{aligned} \tag{7.8.1}$$

for every $x, w \in X$. This implies that $g \circ f$ is Lipschitz as well, with

$$\mathrm{Lip}(g \circ f) \le \mathrm{Lip}(f)\ \mathrm{Lip}(g). \tag{7.8.2}$$

More precisely, $\mathrm{Lip}(f)$, $\mathrm{Lip}(g)$, and $\mathrm{Lip}(g \circ f)$ are as defined in the previous section, using the appropriate metric spaces in the domains and ranges of these meppings.

A mapping f from X into Y is said to be an *isometry* if

$$d_Y(f(x), f(w)) = d_X(x, w) \tag{7.8.3}$$

for every $x, w \in X$. In particular, this implies that f is Lipschitz, with constant $C = 1$. If a mapping g from Y into Z is an isometry too, then it is easy to see that the composition $g \circ f$ of f and g is an isometric mapping from X into Z.

Note that an isometric mapping f from X into Y is automatically injective. If f maps X onto Y, then the corresponding inverse mapping f^{-1} is an isometry from Y onto X.

Let n, m, and k be positive integers. Suppose that either A is a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ and B is a linear mapping from $\mathbf{R}^m$ into $\mathbf{R}^k$, or that A is a linear mapping from $\mathbf{C}^n$ into $\mathbf{C}^m$ and B is a linear mapping from $\mathbf{C}^m$ into $\mathbf{C}^k$. Thus the composition $B \circ A$ is either defined as a mapping from $\mathbf{R}^n$ into $\mathbf{R}^k$, or as a mapping from $\mathbf{C}^n$ into $\mathbf{C}^k$. In both cases, $B \circ A$ is a linear mapping. Of course, the matrix associated to $B \circ A$ can be given in terms of the matrices associated to A and B using matrix multiplication.

Let N_1, N_2, and N_3 be norms on $\mathbf{R}^n$, $\mathbf{R}^m$, and $\mathbf{R}^k$, respectively, or on $\mathbf{C}^n$, $\mathbf{C}^m$, and $\mathbf{C}^k$, respectively, as appropriate. Using these norms, the operator norms $\|A\|_{op}$, $\|B\|_{op}$, and $\|B \circ A\|_{op}$ can be defined as in Section 7.5. If $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, then

$$\begin{aligned} N_3((B \circ A)(v)) = N_3(B(A(v))) &\le \|B\|_{op}\, N_2(A(v)) \\ &\le \|A\|_{op}\, \|B\|_{op}\, N_1(v). \end{aligned} \tag{7.8.4}$$

It follows that

$$\|B \circ A\|_{op} \le \|A\|_{op}\, \|B\|_{op}. \tag{7.8.5}$$

This could also be obtained from (7.8.2), using (7.7.5).

In this situation, A is said to be an *isometric linear mapping* with respect to N_1 and N_2 if

$$N_2(A(v)) = N_1(v) \tag{7.8.6}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. This implies that

$$\begin{aligned} d_{N_2}(A(v), A(w)) = N_2(A(v) - A(w)) &= N_2(A(v-w)) \\ &= N_1(v-w) = d_{N_1}(v, w) \end{aligned} \tag{7.8.7}$$

for every $v, w \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, where d_{N_1} and d_{N_2} are the metrics associated to N_1 and N_2, respectively. Conversely, (7.8.6) follows from (7.8.7) by taking $w = 0$. Thus A is isometric as a linear mapping with respect to N_1 and N_2 if and only if A is isometric with respect to the corresponding metrics d_{N_1} and d_{N_2}. If B is also an isometric linear mapping with respect to N_2 and N_3, then their composition $B \circ A$ is an isometric linear mapping with respect to N_1 and N_3.

7.9 Bilipschitz embeddings

Let (X, d_X) and (Y, d_Y) be (nonempty) metric spaces, and let f be a mapping from X into Y. Also let c be a positive real number, and consider the following condition: for every $x, w \in X$, we have that

$$c\, d_X(x, w) \le d_Y(f(x), f(w)). \tag{7.9.1}$$

In particular, this condition implies that f is injective. If f is a one-to-one mapping from X onto Y, then (7.9.1) is the same as saying that the corresponding inverse mapping f^{-1} is Lipschitz with constant $1/c$ as a mapping from Y into X. Otherwise, if f is injective but not necessarily surjective, then one can consider the inverse mapping f^{-1} as a mapping from the image $f(X)$ of X under f into X, and use the restriction of d_Y to $f(X)$.

A mapping f from X into Y is said to be *bilipschitz* if f is Lipschitz and (7.9.1) holds for some $c > 0$. To be more precise, one may say that f is bilipschitz with constant $C \geq 1$ if f is Lipschitz with constant C and (7.9.1) holds with $c = 1/C$. Using this terminology, an isometric mapping from X into Y is the same as a bilipschitz mapping with constant $C = 1$. A one-to-one mapping f from X onto Y is bilipschitz if and only if f is Lipschitz and the inverse mapping f^{-1} is Lipschitz as a mapping from Y into X. If f is not surjective, then one can consider the inverse mapping f^{-1} as a mapping from $f(X)$ into X, using the restriction of d_Y to $f(X)$, as before.

Let (Z, d_Z) be another metric space, and let g be a mapping from Y into Z. Suppose that there is a positive real number c' such that

$$c'\, d_Y(y, u) \leq d_Z(g(y), g(u)) \tag{7.9.2}$$

for every $u, y \in Y$. If $f : X \to Y$ satisfies (7.9.1) for some $c > 0$, then the composition $g \circ f$ satisfies an analogous condition as a mapping from X into Z. More precisely, for each $x, w \in X$, we have that

$$\begin{aligned} c\, c'\, d_X(x, w) \leq c'\, d_Y(f(x), f(w)) &\leq d_Z(g(f(x)), g(f(w))) \\ &= d_Z((g \circ f)(x), (g \circ f)(w)). \end{aligned} \tag{7.9.3}$$

In particular, if f and g are both bilipschitz, then $g \circ f$ is bilipschitz as a mapping from X into Z.

Let m and n be positive integers, and let A be a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$. Also let N_0 and N be norms on $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively, or on $\mathbf{C}^n$ and $\mathbf{C}^m$, respectively, as appropriate. Suppose that

$$c\, N_0(v) \leq N(A(v)) \tag{7.9.4}$$

for some positive real number c and every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. If $v, w \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, then we get that

$$\begin{aligned} c\, d_{N_0}(v, w) = c\, N_0(v - w) &\leq N(A(v - w)) \\ &= N(A(v) - A(w)) = d_N(A(v), A(w)), \end{aligned} \tag{7.9.5}$$

where d_{N_0} and d_N are the metrics associated to N_0 and N, respectively. Of course, (7.9.5) corresponds to (7.9.1) in this situation. Note that (7.9.5) implies (7.9.4), by taking $w = 0$. Observe too that (7.9.4) implies directly that $v = 0$ when $A(v) = 0$, which is to say that the kernel of A is trivial.

Let k be another positive integer, and let B be a linear mapping from $\mathbf{R}^m$ into $\mathbf{R}^k$ or from $\mathbf{C}^m$ into $\mathbf{C}^k$, as appropriate. Also let N_3 be a norm on $\mathbf{R}^k$ or

$\mathbf{C}^k$, as appropriate. Suppose that

$$c'\,N(y) \le N_3(B(y)) \tag{7.9.6}$$

for some positive real number c' and every $y \in \mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. If $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, then we get that

$$c\,c' N_0(v) \le c'\,N(A(v)) \le N_3(B(A(v))) = N_3((B \circ A)(v)). \tag{7.9.7}$$

This basically corresponds to (7.9.3) in this situation, as before.

7.10 Linear mappings and seminorms

Let m and n be positive integers, and let A be a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$. Also let N be a seminorm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, as in Section A.6. More precisely, the definition of a seminorm was mentioned previously in the real case, and the complex case is analogous. Under these conditions, one can check that $N(A(v))$ defines a seminorm on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Indeed, it is easy to see that $N(A(v))$ satisfies the homogeneity requirement for a seminorm on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, because of the analogous property for N and the linearity of A. Similarly, $N(A(v))$ satisfies the triangle inequality on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, because of the triangle inequality for N and the linearity of A. Suppose for the rest of the section that N is a norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, so that

$$N(A(v)) = 0 \tag{7.10.1}$$

only when $A(v) = 0$.

If the kernel of A is trivial, then (7.10.1) holds only when $v = 0$. This means that $N(A(v))$ defines a norm on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Let N_0 be another norm on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Under these conditions, one can get that (7.9.4) holds for some $c > 0$ using the remarks in Section 6.7.

Suppose now that $m = n$, so that A is a linear mapping from $\mathbf{R}^n$ into itself, or from $\mathbf{C}^n$ into itself. If the kernel of A is trivial, then it is well known that A maps $\mathbf{R}^n$ or $\mathbf{C}^n$ onto itself, as appropriate. This means that A has an inverse mapping A^{-1} on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Of course, the inverse mapping is also linear.

In this case, (7.9.4) can be reformulated as saying that

$$N_0(A^{-1}(u)) \le (1/c)\,N(u) \tag{7.10.2}$$

for every $u \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. We have seen previously that this type of condition holds, because A^{-1} is a linear mapping on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. More precisely, this condition holds with

$$1/c = \|A^{-1}\|_{op}, \tag{7.10.3}$$

where the operator norm of A^{-1} is defined using N on the domain of A^{-1}, and N_0 on the range of A^{-1}. Note that $\|A^{-1}\|_{op} > 0$, because $A^{-1} \neq 0$. Of course, if (7.10.2) holds for some $c > 0$, then we have that

$$\|A^{-1}\|_{op} \le 1/c. \tag{7.10.4}$$

7.11 Small perturbations

Let (X, d_X) be a nonempty metric space, let m be a positive integer, and let f be a mapping from X into $\mathbf{R}^m$ or $\mathbf{C}^m$. Also let N be a norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, which leads to a metric d_N in the usual way. Suppose that

$$c\, d_X(x,w) \le d_N(f(x), f(w)) = N(f(x) - f(w)) \tag{7.11.1}$$

for some positive real number c and all $x, w \in X$. Let g be another mapping from X into $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. We would like to show that g satisfies an analogous condition when g is sufficiently close to f, in a suitable sense. More precisely, we ask first that $f - g$ be Lipschitz with respect to d_N on the range. This means that

$$\begin{aligned} N((f(x) - g(x)) - (f(w) - g(w))) &= d_N(f(x) - g(x), f(w) - g(w)) \\ &\le \operatorname{Lip}(f-g)\, d_X(x,w) \end{aligned} \tag{7.11.2}$$

for every $x, w \in X$, where $\operatorname{Lip}(f-g)$ is as defined in Section 7.7. Of course,

$$\begin{aligned} N(f(x) - f(w)) &\le N((f(x) - g(x)) - (f(w) - g(w))) \\ &\quad + N(g(x) - g(w)) \end{aligned} \tag{7.11.3}$$

for every $x, w \in X$, by the triangle inequality. Combining this with (7.11.1) and (7.11.2), we get that

$$c\, d_X(x,w) \le N(g(x) - g(w)) + \operatorname{Lip}(f-g)\, d_X(x,w) \tag{7.11.4}$$

for every $x, w \in X$. It follows that

$$(c - \operatorname{Lip}(f-g))\, d_X(x,w) \le N(g(x) - g(w)) = d_N(g(x), g(w)) \tag{7.11.5}$$

for every $x, w \in X$. If $\operatorname{Lip}(f-g) < c$, then this is the same type of condition as before.

Let n be a positive integer, and let us now take $X = \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Let N_0 be a norm on $\mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, which leads to a metric d_{N_0}, as usual. Also let A be a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate. Suppose that

$$c\, N_0(v) \le N(A(v)) \tag{7.11.6}$$

for some $c > 0$ and all $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. Let B be another linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ or from $\mathbf{C}^n$ into $\mathbf{C}^m$, as appropriate. We would like to verify that B satisfies the same type of condition when B is sufficiently close to A. This could be obtained from the remarks in the preceding paragraph, but the analogous argument is a bit simpler in this case, as follows.

Remember that

$$N(A(v) - B(v)) \le \|A - B\|_{op} \, N_0(v) \tag{7.11.7}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, where the operator norm $\|A - B\|_{op}$ of $A - B$ is as defined in Section 7.5. Thus

$$\begin{aligned} N(A(v)) &\le N(A(v) - B(v)) + N(B(v)) \\ &\le N(B(v)) + \|A - B\|_{op} \, N_0(v) \end{aligned} \tag{7.11.8}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, using the triangle inequality in the first step. This implies that

$$c \, N_0(v) \le N(B(v)) + \|A - B\|_{op} \, N_0(v) \tag{7.11.9}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate, by (7.11.6). Hence

$$(c - \|A - B\|_{op}) \, N_0(v) \le N(B(v)) \tag{7.11.10}$$

for every $v \in \mathbf{R}^n$ or $\mathbf{C}^n$, as appropriate. This is the same type of condition as before when $\|A - B\|_{op} < c$.

In particular, if $\|A - B\|_{op} < c$, then (7.11.10) implies that the kernel of B is trivial. If $m = n$, then it follows that B is invertible.

7.12 The contraction mapping theorem

Let $(X, d(x, y))$ be a nonempty metric space, and let f be a mapping from X into itself. Suppose that f is Lipschitz with constant $c \ge 0$, so that

$$d(f(x), f(y)) \le c \, d(x, y) \tag{7.12.1}$$

for every $x, y \in X$. If $c < 1$, and if X is complete with respect to $d(\cdot, \cdot)$, then a famous theorem states that there is a unique point $x \in X$ such that $f(x) = x$. More precisely, uniqueness can be verified directly, without using completeness.

To get the existence of the fixed point, let x_0 be any element of X, and let $x_1, x_2, x_3, \ldots$ be the sequence of elements of X defined recursively by

$$x_j = f(x_{j-1}) \tag{7.12.2}$$

when $j \ge 1$. Observe that

$$d(x_j, x_{j+1}) = d(f(x_{j-1}), f(x_j)) \le c \, d(x_{j-1}, x_j) \tag{7.12.3}$$

for every $j \ge 1$. This implies that

$$d(x_j, x_{j+1}) \le c^j \, d(x_0, x_1) \tag{7.12.4}$$

for every $j \geq 1$. If $l < n$ are positive integers, then it follows that

$$\begin{aligned} d(x_l, x_n) \leq \sum_{j=l}^{n-1} d(x_j, x_{j+1}) &\leq \sum_{j=l}^{n-1} c^j \, d(x_0, x_1) \\ &= c^l \sum_{j=0}^{n-l-1} c^j \, d(x_0, x_1) \leq \frac{c^l}{1-c} \, d(x_0, x_1). \end{aligned} \tag{7.12.5}$$

This implies that $\{x_j\}_{j=0}^\infty$ is a Cauchy sequence in X, because $c^l \to 0$ as $l \to \infty$ when $c < 1$.

If X is complete, then it follows that $\{x_j\}_{j=0}^\infty$ converges to an element x of X. We also have that $\{f(x_j)\}_{j=0}^\infty$ converges to $f(x)$ in X, because f is continuous. By construction, $\{f(x_j)\}_{j=0}^\infty$ is the same as $\{x_{j+1}\}_{j=0}^\infty$, which converges to x. Hence $f(x) = x$, as desired.

Let n be a positive integer, and let $\overline{B}(0,1)$ be the closed unit ball in $\mathbf{R}^n$ with respect to the standard Euclidean metric. Also let f be a continuous mapping from $\overline{B}(0,1)$ into itself, with respect to the restriction of the standard Euclidean metric on $\mathbf{R}^n$ to $\overline{B}(0,1)$. Under these conditions, *Brouwer's fixed-point theorem* states that f has a fixed point, which is to say that there is an $x \in \overline{B}(0,1)$ such that $f(x) = x$. If $n = 1$, then this can be obtained from the intermediate value theorem. These and related matters are discussed in many textbooks, as well as the articles [22, 23, 26, 51, 110, 115, 130, 132, 135, 138, 150, 158, 164], for instance.

Let n be a positive integer again, and let N be a norm on $\mathbf{R}^n$. Note that $\mathbf{R}^n$ is complete with respect to the metric d_N associated to N. This can be obtained from the completeness of $\mathbf{R}^n$ with respect to the standard Euclidean metric, and the comparability of N with the standard Euclidean norm on $\mathbf{R}^n$, as in Section 6.7.

Let g be a Lipschitz mapping from $\mathbf{R}^n$ into itself, with respect to d_N, and with constant $c \geq 0$. This means that

$$N(g(x) - g(y)) \leq c \, N(x - y) \tag{7.12.6}$$

for every $x, y \in \mathbf{R}^n$ in this situation. Let $a \in \mathbf{R}^n$ be given, and put

$$g_a(x) = g(x) + a \tag{7.12.7}$$

for every $x \in \mathbf{R}^n$. Observe that

$$N(g_a(x) - g_a(y)) = N(g(x) - g(y)) \leq c \, N(x - y) \tag{7.12.8}$$

for every $x, y \in \mathbf{R}^n$. Thus g_a is also Lipschitz with constant c as a mapping from $\mathbf{R}^n$ into itself, with respect to d_N.

Suppose that $c < 1$, so that the contraction mapping theorem can be applied to g_a on $\mathbf{R}^n$. It follows that there is a unique point $x(a) \in \mathbf{R}^n$ such that

$$g(x(a)) + a = g_a(x(a)) = x(a). \tag{7.12.9}$$

Put

$$h(x) = x - g(x) \tag{7.12.10}$$

for every $x \in \mathbf{R}^n$, which defines a mapping from $\mathbf{R}^n$ into itself. The previous statement can be reformulated as saying that for every $a \in \mathbf{R}^n$ there is a unique $x(a) \in \mathbf{R}^n$ such that

$$h(x(a)) = a. \tag{7.12.11}$$

Of course, this is the same as saying that h is a one-to-one mapping from $\mathbf{R}^n$ onto itself. Note that h is Lipschitz with constant $1 + c$ on $\mathbf{R}^n$ with respect to d_N, as in Section 7.7. We also have that

$$(1 - c)\, N(x - w) \le N(h(x) - h(w)) \tag{7.12.12}$$

for every $x, y \in \mathbf{R}^n$, as in the previous section.

7.13 A localized condition

Let n be a positive integer again, and let N be a norm on $\mathbf{R}^n$. Also let

$$\overline{B}_N(r) = \{x \in \mathbf{R}^n : N(x) \le r\} \tag{7.13.1}$$

be the closed ball in $\mathbf{R}^n$ centered at 0 with radius $r \ge 0$ with respect to the metric d_N on $\mathbf{R}^n$ associated to N. Remember that this is a closed set in $\mathbf{R}^n$ with respect to d_N for each $r \ge 0$. It is easy to see that $\overline{B}_N(r)$ is complete as a metric space with respect to the restriction of d_N to $\overline{B}_N(r)$ for every $r \ge 0$, as in Section 1.7. More precisely, any Cauchy sequence of elements of $\overline{B}_N(r)$ with respect to d_N may be considered as a Cauchy sequence in $\mathbf{R}^n$ with respect to d_N as well. Such a sequence converges to an element of $\mathbf{R}^n$ with respect to d_N, because $\mathbf{R}^n$ is complete with respect to d_N, as in the previous section. The limit of this sequence is contained in $\overline{B}_N(r)$, because $\overline{B}_N(r)$ is a closed set in $\mathbf{R}^n$ with respect to d_N.

Let r be a positive real number, and let g be a Lipschitz mapping from $\overline{B}_N(r)$ into $\mathbf{R}^n$ with constant $c \ge 0$, with respect to d_N and its restriction to $\overline{B}_N(r)$. Thus

$$N(g(x) - g(y)) \le c\, N(x - y) \tag{7.13.2}$$

for every $x, y \in \overline{B}_N(r)$. Suppose also that

$$g(0) = 0. \tag{7.13.3}$$

This implies that

$$N(g(x)) \le c\, N(x) \le c\, r \tag{7.13.4}$$

for every $x \in \overline{B}_N(r)$, using (7.13.2) in the first step.

Suppose too that

$$c < 1, \tag{7.13.5}$$

and let $a \in \mathbf{R}^n$ be given, with

$$N(a) \le (1 - c)\, r. \tag{7.13.6}$$

Put

$$g_a(x) = g(x) + a \tag{7.13.7}$$

for each $x \in \overline{B}_N(r)$, as before. Observe that

$$N(g_a(x)) \le N(g(x)) + N(a) \le c\, r + (1 - c)\, r = r \tag{7.13.8}$$

for every $x \in \overline{B}_N(r)$, using (7.13.4) and (7.13.6) in the second step.

This shows that g_a maps $\overline{B}_N(r)$ into itself under these conditions. Note that g_a is Lipschitz with constant c with respect to d_N as well, as in (7.12.8). Hence we can apply the contraction mapping theorem to g_a on $\overline{B}_N(r)$, because of (7.13.5). This implies that there is a unique point $x(a) \in \overline{B}_N(r)$ such that

$$g(x(a)) + a = g_a(x(a)) = x(a). \tag{7.13.9}$$

Let h be the mapping from $\overline{B}_N(r)$ into $\mathbf{R}^n$ defined by

$$h(x) = x - g(x) \tag{7.13.10}$$

for every $x \in \overline{B}_N(r)$. Observe that

$$h(x(a)) = x(a) - g(x(a)) = a, \tag{7.13.11}$$

by (7.13.9). It follows that

$$h(\overline{B}_N(r)) \supseteq \overline{B}_N((1 - c)\, r) \tag{7.13.12}$$

in this situation. Note that h is Lipschitz with constant $1 + c$ on $\overline{B}_N(r)$ with respect to d_N, and (7.12.12) holds for all $x, w \in \overline{B}_N(r)$, as in Sections 7.7 and 7.11.

7.14 Open mappings

Let (X, d_X) and (Y, d_Y) be metric spaces. A mapping f from X into Y is said to be an *open mapping* if for every open set $U \subseteq X$,

$$f(U) \text{ is an open set in } Y. \tag{7.14.1}$$

One can check that this happens if and only if for every $x \in X$ and $r > 0$ there is a $t > 0$ such that

$$B_Y(f(x), t) \subseteq f(B_X(x, r)). \tag{7.14.2}$$

Let n be a positive integer, and let N be a norm on $\mathbf{R}^n$. If $v \in \mathbf{R}^n$ and $r > 0$, then let

$$B_N(v, r) = \{w \in \mathbf{R}^n : N(v - w) < r\} \tag{7.14.3}$$

and

$$\overline{B}_N(v,r) = \{w \in \mathbf{R}^n : N(v-w) \le r\} \tag{7.14.4}$$

be the open and closed balls in $\mathbf{R}^n$ centered at v with radius r with respect to the metric d_N associated to N.

Let $v_0 \in \mathbf{R}^n$ and $r_0 > 0$ be given, and let f be a mapping from $\overline{B}_N(v_0, r_0)$ into $\mathbf{R}^n$. Suppose that $f(x) - x$ is Lipschitz with constant $c \ge 0$ on $\overline{B}_N(v_0, r_0)$, with respect to d_N and its restriction to $\overline{B}_N(v_0, r_0)$. If $c < 1$, then

$$\overline{B}_N(f(v_0), (1-c)\, r_0) \subseteq f(\overline{B}_N(v_0, r_0)). \tag{7.14.5}$$

This follows from the remarks in the previous section when $v_0 = f(v_0) = 0$, and otherwise one can reduce to this case.

Similarly, let $v \in \mathbf{R}^n$ and $r > 0$ be given, and let f be a mapping from $B_N(v,r)$ into $\mathbf{R}^n$. Suppose that $f(x) - x$ is Lipschitz with constant c, $0 \le c < 1$, on $B_N(v,r)$, with respect to d_N and its restriction to $B_N(v,r)$. Let $w \in B_N(v,r)$ be given, and let r_1 be a positive real number such that

$$r_1 \le r - N(v-w). \tag{7.14.6}$$

Note that $r - N(v-w) > 0$, and that (7.14.6) implies that

$$B_N(w, r_1) \subseteq B_N(v,r). \tag{7.14.7}$$

If $0 < r_0 < r_1$, then $\overline{B}_N(w, r_0) \subseteq B_N(w, r_1)$, and

$$\overline{B}_N(f(w), (1-c)\, r_0) \subseteq f(\overline{B}_N(w, r_0)), \tag{7.14.8}$$

as in (7.14.5). This implies that

$$B_N(f(w), (1-c)\, r_1) \subseteq f(B_N(w, r_1)). \tag{7.14.9}$$

In particular, this means that f is an open mapping from $B_N(v,r)$ into $\mathbf{R}^n$, with respect to d_N and its restriction to $B_N(v,r)$.

Chapter 8

Some topics related to differentiability

8.1 An integral triangle inequality

Let m be a positive integer, and let N be a norm on $\mathbf{R}^m$ or $\mathbf{C}^m$. If $v_1, \ldots, v_l$ are finitely many elements of $\mathbf{R}^m$ or $\mathbf{C}^m$ and $t_1, \ldots, t_l$ are real or complex numbers, as appropriate, then it is easy to see that

$$N\Big(\sum_{k=1}^{l} t_k\, v_k\Big) \le \sum_{k=1}^{l} |t_k|\, N(v_k). \tag{8.1.1}$$

We would like to consider analogous statements for integrals instead of finite sums. Although this works for Riemann–Stieltjes integrals, it is sufficient to consider Riemann integrals for the result in the next section.

Let a and b be real numbers with $a < b$, and let f be a continuous function defined on the closed interval $[a, b]$ with values in $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. This implicitly uses the restriction of the standard Euclidean metric on the real line to $[a, b]$, and one can also use the standard Euclidean metric on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. Equivalently, f can be given as

$$f(t) = (f_1(t), \ldots, f_m(t)), \tag{8.1.2}$$

where $f_1, \ldots, f_m$ are continuous real or complex-valued functions on $[a, b]$, as appropriate. Of course, the continuity of a complex-valued function on $[a, b]$ is equivalent to the continuity of its real and imaginary parts, as real-valued functions on $[a, b]$.

Let α be a monotonically increasing real-valued function on $[a, b]$. As before, it suffices to consider the case where $\alpha(t) = t$ for every $t \in [a, b]$ for the result discussed in the next section, which corresponds to using ordinary Riemann integrals on $[a,b]$. To define the Riemann–Stieltjes integral

$$\int_a^b f(t)\, d\alpha(t) \tag{8.1.3}$$

as an element of $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, one can take the jth coordinate of (8.1.3) to be the Riemann–Stieltjes integral of f_j with respect to α on $[a,b]$ for each $j = 1, \ldots, m$. Similarly, the Riemann–Stieltjes integral of a continuous complex-valued function on $[a,b]$ can be reduced to the real case, by considering the real and imaginary parts of the function.

Remember that N is (uniformly) continuous as a real-valued function on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate, as in Section 6.7. Thus $N(f(t))$ defines a nonnegative real-valued continuous function on $[a,b]$. In particular, its Riemann–Stieltjes integral over $[a,b]$ with respect to α can be defined as a nonnegative real number in the usual way.

Under these conditions, one can check that

$$N\Big(\int_a^b f(t)\, d\alpha(t)\Big) \le \int_a^b N(f(t))\, d\alpha(t). \tag{8.1.4}$$

More precisely, these integrals can be approximated by Riemann–Stieltjes sums associated to sufficiently fine partitions of $[a,b]$. To get (8.1.4), one can use the analogous inequalities for Riemann–Stieltjes sums, which follow from (8.1.1). One can also consider Riemann–Stieltjes integrability conditions on $[a,b]$, instead of continuity.

8.2 A basic Lipschitz estimate

Let a and b be real numbers with $a < b$, and let m be a positive integer. Also let f be a function defined on $[a,b]$ with values in $\mathbf{R}^m$ or $\mathbf{C}^m$. The derivative $f'(t)$ of f at a point $t \in (a,b)$ can be defined in the usual way, as the limit of difference quotients, when the limit exists. Similarly, if $t = a$ or b, then one can consider the corresponding one-sided limit. The differentiability of f at any $t \in [a,b]$ is equivalent to the differentiability of the jth component f_j of f at t for each $j = 1, \ldots, m$, as a real or complex-valued function on $[a,b]$. Similarly, the differentiability of a complex-valued function on $[a,b]$ is equivalent to the differentiability of its real and imaginary parts. If f is differentiable at $t \in [a,b]$, then the jth coordinate of $f'(t)$ is equal to the derivative of f_j at t for each $j = 1, \ldots, m$, and their common value is denoted $f_j'(t)$.

Let us suppose from now on in this section that f is continuously differentiable on $[a,b]$. This means that the derivative $f'(t)$ exists for every $t \in [a,b]$, and that $f'(t)$ is continuous as a function on $[a,b]$ with values in $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. Here we use the restriction of the standard Euclidean metric on $\mathbf{R}$ to $[a,b]$, and the standard Euclidean metric on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. Equivalently, f_j should be continuously differentiable as a real or complex-valued function on $[a,b]$ for each $j = 1, \ldots, m$. Of course, $f'(t)$ is really a one-sided derivative when $t = a$ or b. If $x \in [a, b]$, then the fundamental theorem of calculus implies that

$$\int_a^x f'(t)\,dt = f(x) - f(a), \tag{8.2.1}$$

where the Riemann integral on the left side can be defined as in the previous section. Of course, this can be reduced to the case of real-valued functions in the usual way. Similarly, if $a \le w \le x \le b$, then

$$\int_w^x f'(t)\,dt = f(x) - f(w). \tag{8.2.2}$$

Let N be a norm on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate. If $a \le w \le x \le b$, then

$$N(f(x) - f(w)) = N\Big(\int_w^x f'(t)\,dt\Big) \le \int_w^x N(f'(t))\,dt. \tag{8.2.3}$$

This uses (8.2.2) in the first step, and (8.1.4) in the second step, applied to f' and $\alpha(t) = t$. Note that $N(f'(t))$ is continuous as a real-valued function on $[a, b]$, as in the previous section. In particular, $N(f'(t))$ is bounded on $[a, b]$, because $[a, b]$ is compact, and in fact the maximum of $N(f'(t))$ is attained on $[a, b]$. It follows from (8.2.3) that

$$N(f(x) - f(w)) \le |x - w| \sup_{a \le t \le b} N(f'(t)) \tag{8.2.4}$$

for every $x, w \in [a, b]$. Thus f is Lipschitz with respect to the restriction of the standard Euclidean metric on $\mathbf{R}$ to $[a, b]$ and the metric d_N associated to N on $\mathbf{R}^m$ or $\mathbf{C}^m$, as appropriate.

More precisely,

$$\mathrm{Lip}(f) = \sup_{a \le t \le b} N(f'(t)), \tag{8.2.5}$$

where $\mathrm{Lip}(f)$ is as defined in Section 7.7. The fact that $\mathrm{Lip}(f)$ is less than or equal to the right side of (8.2.5) follows from (8.2.4). To get the opposite inequality, it suffices to verify that $N(f'(t))$ is less than or equal to $\mathrm{Lip}(f)$ for every $t \in [a, b]$. The difference quotients used to define $f'(t)$ have norm less than or equal to $\mathrm{Lip}(f)$, by definition of $\mathrm{Lip}(f)$. This implies that $N(f'(t)) \le \mathrm{Lip}(f)$, by taking the limit of the difference quotient.

8.3 Some partial Lipschitz conditions

Let n be a positive integer, and let E be a subset of $\mathbf{R}^n$. Also let (Y, d_Y) be a metric space, and let f be a mapping from E into Y. Suppose that l is a positive integer less than or equal to n, and that C_l is a nonnegative real number. Let us say that f is *partially Lipschitz in the lth variable with constant C_l on E* if

$$d_Y(f(x), f(x')) \le C_l\,|x_l - x'_l| \tag{8.3.1}$$

for every $x, x' \in E$ such that $x_j = x'_j$ when $j \neq l$. This means that $f(x)$ is Lipschitz as a function of x_l with constant C_l, when the other coordinates of x are fixed. More precisely, this uses the restriction of the standard Euclidean metric on $\mathbf{R}$ to the set of real numbers that occur as lth coordinates of elements of E, when the other coordinates are fixed. This is the same as saying that if L is any line in $\mathbf{R}^n$ that is parallel to the lth coordinate axis, then the restriction of f to $E \cap L$ is Lipschitz with constant C_l.

Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be real numbers such that $a_k < b_k$ for each $k = 1, \ldots, n$. Remember that the corresponding set

$$\mathcal{C} = \{x \in \mathbf{R}^n : a_k \leq x_k \leq b_k \text{ for each } k = 1, \ldots, n\} \tag{8.3.2}$$

is called a cell in $\mathbf{R}^n$, as in [155]. This is the same as the Cartesian product of the closed intervals $[a_k, b_k]$, $k = 1, \ldots, n$, as mentioned in Section 6.11. Suppose now that f is a mapping from $\mathcal{C}$ into Y. Thus $f(x)$ may be considered as a function of x_l on $[a_l, b_l]$, when the other variables are fixed elements of $[a_k, b_k]$, $k \neq l$.

Let C_l be a nonnegative real number for each $l = 1, \ldots, n$, and suppose that f is partially Lipschitz in the lth variable with constant C_l for each $l = 1, \ldots, n$. Let $x, w \in \mathcal{C}$ be given, and observe that

$$d_Y(f(x), f(w)) \leq \sum_{l=1}^{n} C_l \, |x_l - w_l|. \tag{8.3.3}$$

To see this, one can go from x to w in n steps, only changing one coordinate in each step. It follows that

$$d_Y(f(x), f(w)) \leq \Big(\max_{1 \leq l \leq n} C_l\Big) \, \|x - w\|_1, \tag{8.3.4}$$

where $\| \cdot \|_1$ is defined on $\mathbf{R}^n$ as in Section 1.3. Similarly,

$$d_Y(f(x), f(w)) \leq \Big(\sum_{l=1}^{n} C_l^2\Big)^{1/2} \, \|x - w\|_2, \tag{8.3.5}$$

where $\| \cdot \|_2$ is the standard Euclidean norm on $\mathbf{R}^n$. This uses the Cauchy–Schwarz inequality, applied to the sum on the right side of (8.3.3). We also get that

$$d_Y(f(x), f(w)) \leq \Big(\sum_{l=1}^{n} C_l\Big) \, \|x - w\|_\infty, \tag{8.3.6}$$

where $\|\cdot\|_\infty$ is as defined in Section 1.3. In particular, f is Lipschitz with respect to the restrictions of the metrics on $\mathbf{R}^n$ associated to these norms to $\mathcal{C}$.

Let f be a mapping from a set $E \subseteq \mathbf{R}^n$ into Y again, and suppose that f is partially Lipschitz in the lth variable on E with constant $C_l \geq 0$ for every $l = 1, \ldots, n$. If $\mathcal{C}$ is a cell in $\mathbf{R}^n$ with

$$\mathcal{C} \subseteq E, \tag{8.3.7}$$

then the restriction of f to $\mathcal{C}$ has the same properties mentioned in the preceding paragraph. In particular, if $E = \mathbf{R}^n$, then these properties hold for every $x, w \in \mathbf{R}^n$.

8.4 Partial derivatives

Let m and n be positive integers, and let U be a nonempty open subset of $\mathbf{R}^n$, with respect to the standard Euclidean metric. Also let f be a mapping from U into $\mathbf{R}^m$, and let $x \in U$ be given. If l is a positive integer less than or equal to n, then the *lth partial derivative*

$$\partial_l f(x) = D_l f(x) = \frac{\partial f}{\partial x_l}(x) \tag{8.4.1}$$

of f at x can be defined, as usual, as the derivative of f in the lth variable at x_l, when it exists, and with the other variables being fixed. More precisely, if we consider f as a function of the lth variable, with the kth variable equal to x_k when $k \neq l$, then f is defined on an open subset of the real line that contains x_l. Note that $\partial_l f(x)$ exists if and only if the lth partial derivative $\partial_l f_j(x)$ of the jth component f_j of f at x exists for each $j = 1, \ldots, m$, in which case $\partial_l f_j(x)$ is the same as the jth component of $\partial_l f(x)$ for each $j = 1, \ldots, m$, as an element of $\mathbf{R}^m$.

Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be real numbers with $a_k < b_k$ for each $k = 1, \ldots, n$, and let $\mathcal{C}$ be the corresponding cell in $\mathbf{R}^n$, as in (8.3.2). Suppose now that f is a mapping from $\mathcal{C}$ into $\mathbf{R}^m$. Let $x \in \mathcal{C}$ and a positive integer $l \leq n$ be given. If $a_l < x_l < b_l$, then the lth partial derivative of f at x can be defined as the derivative of f as a function of the lth variable on $[a_l, b_l]$ at x_l, when it exists, and with the other variables being fixed. If $x_l = a_l$ or b_l, then one can use the corresponding one-sided derivative, as before.

Suppose that the lth partial derivative of f exists everywhere on $\mathcal{C}$, and that it is continuous as a mapping from $\mathcal{C}$ into $\mathbf{R}^m$, with respect to the restriction of the standard Euclidean metric on $\mathbf{R}^n$ to $\mathcal{C}$, and the standard Euclidean metric on $\mathbf{R}^m$. Let N be a norm on $\mathbf{R}^m$. Remember that N is continuous on $\mathbf{R}^m$, as in Section 6.7, so that $N(\partial_l f(x))$ is continuous as a real-valued function on $\mathcal{C}$. This implies that $N(\partial_l f(x))$ is bounded on $\mathcal{C}$, because $\mathcal{C}$ is compact, and that its maximum on $\mathcal{C}$ is attained.

Let d_N be the metric on $\mathbf{R}^m$ associated to N. Using the remarks in Section 8.2, we get that f is partially Lipschitz in the lth variable on $\mathcal{C}$, with respect to d_N on $\mathbf{R}^m$, with constant

$$C_l = \sup_{x \in \mathcal{C}} N(\partial_l f(x)). \tag{8.4.2}$$

This is the smallest value of C_l with this property, for the same reasons as before.

Let f be a mapping from an open subset U of $\mathbf{R}^n$ into $\mathbf{R}^m$ again. Let us say that f is *continuously differentiable* on U if for each $l = 1, \ldots, n$, the lth partial derivative $\partial_l f(x)$ of f exists at every $x \in U$, and defines a continuous mapping from U into $\mathbf{R}^m$. This uses the restriction of the standard Euclidean metric on $\mathbf{R}^n$ to U, and the standard Euclidean metric on $\mathbf{R}^m$, as usual. If $\mathcal{C}$ is a cell in $\mathbf{R}^n$ and $\mathcal{C} \subseteq U$, then the restriction of f to $\mathcal{C}$ satisfies the analogous continuous differentiability property on $\mathcal{C}$.

8.5 Using the mean value theorem

Let n be a positive integer, let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be real numbers with $a_k < b_k$ for each $k = 1, \ldots, n$, and let $\mathcal{C}$ be the corresponding cell in $\mathbf{R}^n$ again, as in (8.3.2). Also let f be a real-valued function on $\mathcal{C}$, and let $1 \le l \le n$ be given. Suppose that $f(x)$ is continuous as a real-valued function of x_l on $[a_l, b_l]$, when $x_k \in [a_k, b_k]$ is kept fixed for $k \ne l$.

Suppose in addition that $f(x)$ is differentiable as a function of x_l on (a_l, b_l) when $x_k \in [a_k, b_k]$ is kept fixed for $k \ne l$, so that $\partial_l f(x)$ is defined under these conditions. Let us suppose as well that $\partial_l f(x)$ is bounded, and put

(8.5.1) $$C_l = \sup\{|\partial_l f(x)| : x \in \mathcal{C},\ a_l < x_l < b_l\}.$$

Under these conditions, one can use the mean value theorem to get that f is partially Lipschitz in the lth variable on $\mathcal{C}$, with respect to the standard Euclidean metric on $\mathbf{R}$, with constant C_l. This is the smallest value of C_l with this property, as usual.

Now let U be a nonempty open subset of $\mathbf{R}^n$, with respect to the standard Euclidean metric, and let f be a real-valued function on U. Suppose that $\partial_l f(x)$ exists for every $x \in U$, which implies that f is continuous as a function of x_l, with the other variables kept fixed, everywhere on U. If

(8.5.2) $$\mathcal{C} \subseteq U$$

and $\partial_l f$ is bounded on $\mathcal{C}$, then we get that f is partially Lipschitz in the lth variable on $\mathcal{C}$, with constant C_l as in (8.5.1), as before.

If $\partial_l f$ is bounded on U, then this holds for every cell $\mathcal{C}$ as in (8.5.2). If $\partial_l f$ exists and is bounded on U for each $l = 1, \ldots, n$, then one can use this to get that f is continuous on U, as in Exercise 7 on p239 of [155].

8.6 Directional derivatives

Let m and n be positive integers, and let U be an open subset of $\mathbf{R}^n$, with respect to the standard Euclidean metric. Also let f be a mapping from U into $\mathbf{R}^m$, and let $x \in U$ and $v \in \mathbf{R}^n$ be given. It is easy to see that

(8.6.1) $$U(x, v) = \{t \in \mathbf{R} : x + t\,v \in U\}$$

is an open set in the real line, with respect to the standard Euclidean metric on $\mathbf{R}$. Of course, $0 \in U(x, v)$, because $x \in U$. Let us consider

(8.6.2) $$f_{x,v}(t) = f(x + t\,v)$$

as an $\mathbf{R}^m$-valued function of $t \in U(x, v)$. If this function is differentiable at 0, then put

$$D_v f(x) = f'_{x,v}(0). \tag{8.6.3}$$

This is the *directional derivative* of f at x in the direction v.

Let $r \in \mathbf{R}$ be given, so that $r\,v \in \mathbf{R}^n$ too. Observe that

$$U(x, r\,v) = \{t \in \mathbf{R} : r\,t \in U(x, v)\}, \tag{8.6.4}$$

and

$$f_{x,r\,v}(t) = f_{x,v}(r\,t) \tag{8.6.5}$$

on $U(x, r\,v)$. If $D_v f(x)$ exists, then it is easy to see that $D_{r\,v} f(x)$ exists, with

$$D_{r\,v} f(x) = r\,D_v f(x). \tag{8.6.6}$$

Let $e_1, \ldots, e_n$ be the usual standard basis vectors in $\mathbf{R}^n$, so that the lth coordinate of e_k is equal to 1 when $k = l$, and to 0 when $k \neq l$. The directional derivative

$$D_{e_k} f(x) \tag{8.6.7}$$

is the same as the kth partial derivative $\partial_k f(x)$, when it exists, for each $k = 1, \ldots, n$.

If $j \in \{1, \ldots, m\}$, then let $f_j(x) \in \mathbf{R}$ be the jth coordinate of $f(x)$, as an element of $\mathbf{R}^m$. Thus $f_j(x)$ defines a real-valued function on U for each $j = 1, \ldots, m$. Of course, the directional derivative $D_v f(x)$ exists if and only if the directional derivative $D_v f_j(x)$ exists for every $j = 1, \ldots, m$, in which case $D_v f_j(x)$ is the jth coordinate of $D_v f(x)$ for each j. Similarly, note that f is continuously differentiable on U if and only if for each $j = 1, \ldots, m$, f_j is continuously-differentiable as a real-valued function on U.

Suppose for the moment that $D_v f(x)$ exists for every $v \in \mathbf{R}^n$. In some situations, we may also have that

$$D_{v+w} f(x) = D_v f(x) + D_w f(x) \tag{8.6.8}$$

for every $v, w \in \mathbf{R}^n$. This means that

$$v \mapsto D_v f(x) \tag{8.6.9}$$

defines a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$, because of (8.6.6). In this case, we get that

$$D_v f(x) = \sum_{k=1}^{n} v_k\,D_{e_k} f(x) = \sum_{k=1}^{n} v_k\,\partial_k f(x) \tag{8.6.10}$$

for every $v \in \mathbf{R}^n$, because $v = \sum_{k=1}^{n} v_k\,e_k$.

8.7 Differentiable mappings

Let m and n be positive integers again, and let U be a nonempty open subset of $\mathbf{R}^n$, with respect to the standard Euclidean metric on $\mathbf{R}^n$. Also let f be a mapping from U into $\mathbf{R}^m$, and let $x \in U$ be given. We say that f is *differentiable* at x if there is a linear mapping A from $\mathbf{R}^n$ into $\mathbf{R}^m$ such that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - A(h)\|_{2,\mathbf{R}^m}}{\|h\|_{2,\mathbf{R}^n}} = 0. \tag{8.7.1}$$

Here $\|\cdot\|_{2,\mathbf{R}^m}$ and $\|\cdot\|_{2,\mathbf{R}^n}$ are the standard Euclidean norms on $\mathbf{R}^m$ and $\mathbf{R}^n$, respectively. It is easy to see that this implies that f is continuous at x.

Put

$$U_x = \{h \in \mathbf{R}^n : x + h \in U\}, \tag{8.7.2}$$

which is an open set in $\mathbf{R}^n$ that contains 0. Of course, $f(x+h)$ is defined for every $h \in U_x$, by construction. Put

$$a(h) = f(x+h) - f(x) - A(h) \tag{8.7.3}$$

for each $h \in U_x$, so that

$$f(x+h) = f(x) + A(h) + a(h) \tag{8.7.4}$$

for every $h \in U_x$. Using this, (8.7.1) says that

$$\lim_{h \to 0} \frac{\|a(h)\|_{2,\mathbf{R}^m}}{\|h\|_{2,\mathbf{R}^n}} = 0. \tag{8.7.5}$$

Similarly, if $h \in U_x$ and $h \neq 0$, then put

$$\alpha(h) = a(h)\, \|h\|_{2,\mathbf{R}^n}^{-1}. \tag{8.7.6}$$

Let us put $\alpha(0) = 0$, so that

$$f(x+h) = f(x) + A(h) + \alpha(h)\, \|h\|_{2,\mathbf{R}^n} \tag{8.7.7}$$

for every $h \in U_x$, as in (8.7.4). Clearly (8.7.5) is equivalent to

$$\lim_{h \to 0} \|\alpha(h)\|_{2,\mathbf{R}^m} = 0. \tag{8.7.8}$$

One can check directly that A is unique, when it exists. In this case, we put $f'(x) = A$, which may be called the *differential* of f at x.

If $n = 1$, then this reduces to the usual definition of the derivative of a function of one variable. More precisely, a linear mapping from $\mathbf{R}$ into $\mathbf{R}^m$ corresponds to multiplying a real number by a fixed element of $\mathbf{R}^m$. The differential of f at x is the linear mapping that corresponds to multiplying a real number by the usual derivative of f at x, as an element of $\mathbf{R}^m$.

Suppose that f is differentiable at x. If $v \in \mathbf{R}^n$, then one can verify that the directional derivative of f at x in the direction v exists, with

$$D_v f(x) = f'(x)(v). \tag{8.7.9}$$

In particular, if $k \in \{1, \ldots, n\}$, then the kth partial derivative of f at x exists, with

$$\frac{\partial f}{\partial x_k}(x) = f'(x)(e_k). \tag{8.7.10}$$

Here e_k is the kth standard basis vector in $\mathbf{R}^n$, as before.

Remember that linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$ are associated to $m \times n$ matrices of real numbers in a standard way, using the standard basis vectors in $\mathbf{R}^n$. The entries of the matrix associated to $f'(x)$ are given by the partial derivatives of the m components of f at x. If $m = 1$, then this is related to the *gradient* of f at x, as in (8.6.10).

If $t \in \mathbf{R}$, then $t\,f$ is a function defined on U with values in $\mathbf{R}^m$ too. It is easy to see $t\,f$ is differentiable at x as well, with

$$(t\,f)'(x) = t\,f'(x). \tag{8.7.11}$$

Similarly, let g be another mapping from U into $\mathbf{R}^m$ that is differentiable at x. One can check that $f + g$ is differentiable at x, with

$$(f + g)'(x) = f'(x) + g'(x). \tag{8.7.12}$$

One can verify that a mapping from U into $\mathbf{R}^m$ is differentiable at x if and only if its m components are differentiable at x as real-valued functions on U. Suppose now that f, g are real-valued functions on U that are differentiable at x. It is not too difficult to show that $f\,g$ is differentiable at x as well, with

$$(f\,g)'(x) = g(x)\,f'(x) + f(x)\,g'(x). \tag{8.7.13}$$

8.8 Pointwise Lipschitz conditions

Let (X, d_X) and (Y, d_Y) be metric spaces, and let f be a mapping from X into Y. Also let x be an element of X, let C be a nonnegative real number, and let r be a positive real number. Let us say that

(8.8.1) f is *pointwise Lipschitz at x with constant C up to the scale r*

if for every $w \in X$ with

$$d_X(x, w) < r, \tag{8.8.2}$$

we have that

$$d_Y(f(x), f(w)) \le C\,d_X(x, w). \tag{8.8.3}$$

Of course, this implies that

(8.8.4) f is continuous at x.

We may allow $r = +\infty$ here too, so that (8.8.2) holds for every $w \in X$.

Let m and n be positive integers, let U be a nonempty open subset of $\mathbf{R}^n$, and let f be a mapping from U into $\mathbf{R}^m$. Also let $x \in U$ be given, and suppose that

$$f \text{ is differentiable at } x. \tag{8.8.5}$$

If N and N_0 are norms on $\mathbf{R}^m$ and $\mathbf{R}^n$, respectively, then (8.7.1) is equivalent to

$$\lim_{h \to 0} \frac{N(f(x+h) - f(x) - f'(x)(h))}{N_0(h)} = 0. \tag{8.8.6}$$

This follows from the comparison between N, N_0 and the standard Euclidean norms on $\mathbf{R}^m$, $\mathbf{R}^n$, respectively, as in Section 6.7.

Let $\epsilon > 0$ be given, so that there is a $\delta > 0$ such that

$$\frac{N(f(x+h) - f(x) - f'(x)(h))}{N_0(h)} < \epsilon \tag{8.8.7}$$

for every $h \in \mathbf{R}^n$ such that $h \neq 0$,

$$N_0(h) < \delta, \tag{8.8.8}$$

and $x+h \in U$. This uses the comparison between N_0 and the standard Euclidean norm on $\mathbf{R}^n$ again, to express (8.8.8) in terms of $N_0(h)$. We may as well take δ small enough so that (8.8.8) implies that $x + h \in U$.

If $h \in \mathbf{R}^n$ satisfies (8.8.8), then it follows that

$$N(f(x+h) - f(x) - f'(x)(h)) \leq \epsilon \, N_0(h). \tag{8.8.9}$$

This implies that

$$\begin{aligned} N(f(x+h) - f(x)) &\leq N(f(x+h) - f(x) - f'(x)(h)) + N(f'(x)(h)) \\ &\leq N(f'(x)(h)) + \epsilon \, N_0(h) \end{aligned} \tag{8.8.10}$$

when (8.8.8) holds. Using this, we get that

$$\begin{aligned} N(f(x+h) - f(x)) &\leq \|f'(x)\|_{op} \, N_0(h) + \epsilon \, N_0(h) \\ &= (\|f'(x)\|_{op} + \epsilon) \, N_0(h) \end{aligned} \tag{8.8.11}$$

when (8.8.8) holds, where $\|f'(x)\|_{op}$ is the operator norm of $f'(x)$ corresponding to N_0 and N, as in Section 7.5. This is a pointwise Lipschitz condition for f at x, with respect to the metrics associated to N_0 and N.

Using (8.8.9) and the triangle inequality for N again, we get that

$$N(f'(x)(h)) \leq N(f(x+h) - f(x)) + \epsilon \, N_0(h) \tag{8.8.12}$$

when $h \in \mathbf{R}^n$ satisfies (8.8.8). Let r be a positive real number, let $C(r)$ be a nonnegative real number, and suppose that

(8.8.13) f is pointwise Lipschitz at x with constant $C(r)$ up to scale r, with respect to the metric on $\mathbf{R}^m$ associated to N, and the restriction to U of the metric on $\mathbf{R}^n$ associated to N_0. This means that

$$N(f(x+h) - f(x)) \le C(r)\, N_0(h) \tag{8.8.14}$$

for every $h \in \mathbf{R}^n$ such that

$$N_0(h) < r \tag{8.8.15}$$

and $x + h \in U$. As before, we may as well take r small enough so that (8.8.15) implies that $x + h \in U$, using the comparison between N_0 and the standard Euclidean norm on $\mathbf{R}^n$. Combining (8.8.12) and (8.8.14), we get that

$$N(f'(x)(h)) \le C(r)\, N_0(h) + \epsilon\, N_0(h) = (C(r) + \epsilon)\, N_0(h) \tag{8.8.16}$$

for every $h \in \mathbf{R}^n$ with

$$N_0(h) < \min(\delta, r). \tag{8.8.17}$$

It is easy to see that this implies that (8.8.16) holds for every $h \in \mathbf{R}^n$, because $f'(x)(h)$ is linear in h. More precisely, one can reduce to the case where (8.8.17) holds, by multiplying h by a sufficiently small positive real number. It follows that

$$\|f'(x)\|_{op} \le C(r) + \epsilon. \tag{8.8.18}$$

This implies that

$$\|f'(x)\|_{op} \le C(r), \tag{8.8.19}$$

because $\epsilon > 0$ is arbitrary.

8.9 The chain rule

Let m, n, and p be positive integers, let U be a nonempty open subset of $\mathbf{R}^n$, and let f be a mapping from U into $\mathbf{R}^m$. Also let V be an open set in $\mathbf{R}^m$, and suppose that

$$f(U) \subseteq V. \tag{8.9.1}$$

If g is a mapping from V into $\mathbf{R}^p$, then the composition $g \circ f$ of f and g is defined as a mapping from U into $\mathbf{R}^p$. Suppose that

$$f \text{ is differentiable at } x \in U, \tag{8.9.2}$$

and that

$$g \text{ is differentiable at } f(x) \in V. \tag{8.9.3}$$

Thus the differential $f'(x)$ of f at x is defined as a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$, and the differential $g'(f(x))$ of g at $f(x)$ is defined as a linear mapping from $\mathbf{R}^m$ into $\mathbf{R}^p$.

Under these conditions, it is not too difficult to show that

$$g \circ f \text{ is differentiable at } x, \tag{8.9.4}$$

with

$$(g \circ f)'(x) = g'(f(x)) \circ f'(x). \tag{8.9.5}$$

This is the analogue of the chain rule in this situation. One can check that this reduces to the usual version of the chain rule for real-valued functions of one real varaible when $m = n = p = 1$.

If g is a linear mapping from $\mathbf{R}^m$ into $\mathbf{R}^p$, then one can check the chain rule directly. Otherwise, one can use this to reduce to the case where

$$g'(f(x)) = 0, \tag{8.9.6}$$

and one can verify that

$$(g \circ f)'(x) = 0 \tag{8.9.7}$$

when f satisfies a pointwise Lipschitz condition at x.

In particular, (8.9.5) can be used to obtain the partial derivatives of $g \circ f$ at x in terms of the partial derivatives of f at x and the partial derivatives of g at $f(x)$. Similarly, let $v \in \mathbf{R}^n$ be given, so that

$$w = f'(x)(v) = D_v f(x) \tag{8.9.8}$$

is an element of $\mathbf{R}^m$. Using (8.9.5), we get that

$$\begin{aligned} D_v(g \circ f)(x) = (g \circ f)'(x)(v) &= g'(f(x))(f'(x)(v)) \\ &= g'(f(x))(w) = (D_w g)(f(x)). \end{aligned} \tag{8.9.9}$$

If $m = 1$, then this can be obtained from the usual chain rule for differentiable real-valued functions of one real variable.

Let l_1, l_2 be positive integers. Of course, constant mappings from $\mathbf{R}^{l_1}$ into $\mathbf{R}^{l_2}$ are differentiable at every point in $\mathbf{R}^{l_1}$, with differential equal to 0. If $f = A$ is a linear mapping from $\mathbf{R}^{l_1}$ into $\mathbf{R}^{l_2}$, then f is differentiable at every point $x \in \mathbf{R}^{l_1}$, with $f'(x) = A$.

It is well known that $t \mapsto t^2$ is differentiable as a real-valued function on the real line, with derivative equal to $2\,t$. Let f be a real-valued function on U, and suppose that f is differentiable at $x \in U$. Using the chain rule, we get that

$$f^2 \text{ is differentiable at } x \tag{8.9.10}$$

too, with

$$(f^2)'(x) = 2\, f(x)\, f'(x). \tag{8.9.11}$$

Let g be another real-valued function on U that is differentiable at x, so that g^2 is differentiable at x as well, with

$$(g^2)'(x) = 2\, g(x)\, g'(x). \tag{8.9.12}$$

Similarly, $f + g$ is differentiable at x, so that $(f + g)^2$ is differentiable at x, with

$$((f + g)^2)'(x) = 2\,(f(x) + g(x))\,(f'(x) + g'(x)). \tag{8.9.13}$$

Note that

$$f\,g = (1/2)\,((f+g)^2 - f^2 - g^2) \tag{8.9.14}$$

on U. It follows that $f\,g$ is differentiable at x, with $(f\,g)'(x)$ as in (8.7.13).

Similarly, it is well known that $t \mapsto 1/t$ is differentiable as a real-valued function on $\mathbf{R} \setminus \{0\}$, with derivative equal to $-1/t^2$. Let f be a real-valued function on U such that $f(w) \neq 0$ for every $w \in U$, and suppose that f is differentiable at $x \in U$. The chain rule implies that

$$1/f \text{ is differentiable at } x, \tag{8.9.15}$$

with

$$(1/f)'(x) = (-1/f(x)^2)\,f'(x). \tag{8.9.16}$$

In particular, one can use these remarks to get that polynomial functions on $\mathbf{R}^n$ are differentiable at every point. Similarly, rational functions are differentiable on open sets where the denominator is not zero.

8.10 Continuous differentiability

Let m and n be positive integers, let U be an open set in $\mathbf{R}^n$, and let f be a mapping from U into $\mathbf{R}^m$. If

$$f \text{ is differentiable at every point in } U, \tag{8.10.1}$$

then f is said to be *differentiable* on U. Remember that f is said to be continuously differentiable on U if for each $k = 1, \ldots, n$, the kth partial derivative $\partial_k f(x)$ exists at every $x \in U$, and is continuous as a mapping from U into $\mathbf{R}^m$, as in Section 8.4. It is well known that

(8.10.2) continuously-differentiable mappings on U are differentiable

in the sense just mentioned. More precisely, if $x \in U$, then $f'(x)$ is the linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ defined by

$$f'(x)(h) = \sum_{k=1}^{n} h_k\,\partial_k f(x) \tag{8.10.3}$$

for every $h \in \mathbf{R}^n$.

To see this, suppose that f is continuously differentiable on U, and let $x \in U$ be given. Suppose for the moment that

$$\partial_k f(x) = 0 \quad \text{for each } k = 1, \ldots, n, \tag{8.10.4}$$

so that we would like to show that f is differentiable at x, with $f'(x) = 0$. This is equivalent to saying that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x)\|_{2,\mathbf{R}^m}}{\|h\|_{2,\mathbf{R}^n}} = 0. \tag{8.10.5}$$

Because $\partial_k f$ is continuous on U, (8.10.4) implies that $\partial_k f$ is as small as we want near x. One can use this to get (8.10.5), as in Sections 8.3 and 8.4.

Alternatively, it is not difficult to reduce to the case where $m = 1$. This permits one to use the mean value theorem, as in Section 8.5. In order to estimate

$$f(x+h) - f(x), \tag{8.10.6}$$

one can go from x to $x + h$ in n steps, only changing one coordinate in each step, as in Section 8.3.

We can reduce to the case where (8.10.4) holds, as follows. If $w \in U$, then put

$$\widetilde{f}(w) = f(w) - \sum_{k=1}^{n} (w_k - x_k)\, \partial_k f(x). \tag{8.10.7}$$

This defines a mapping from U into $\mathbf{R}^m$, as a function of w. The partial derivatives of $\widetilde{f}$ are given by

$$\partial_l \widetilde{f}(w) = \partial_l f(w) - \partial_l f(x) \tag{8.10.8}$$

for every $w \in U$ and $l = 1, \dots, n$. In particular, $\widetilde{f}$ is continuously differentiable on U, with

$$\partial_l \widetilde{f}(x) = \partial_l f(x) - \partial_l f(x) = 0 \tag{8.10.9}$$

for every $l = 1, \dots, n$. This implies that

$$\lim_{h \to 0} \frac{\|\widetilde{f}(x+h) - \widetilde{f}(x)\|_{2,\mathbf{R}^n}}{\|h\|_{2,\mathbf{R}^m}} = 0, \tag{8.10.10}$$

as before. This is the same as saying that f is differentiable at x, with $f'(x)$ as in (8.10.3).

We can think of $f'(x)$ as a function of $x \in U$ with values in the space $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ of linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$. One can get nice metrics on $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ from norms in the usual way, such as the Hilbert–Schmidt norm or the operator norm associated to the standard Euclidean norms on $\mathbf{R}^n$ and $\mathbf{R}^m$. If f is continuously differentiable on U, then

$$f' \text{ is continuous as a mapping from } U \text{ into } \mathcal{L}(\mathbf{R}^n, \mathbf{R}^m), \tag{8.10.11}$$

with respect to such a metric. Conversely, if f is differentiable on U, and if f' is continuous on U in this sense, then

$$\text{the partial derivatives of } f \text{ are continuous on } U. \tag{8.10.12}$$

Thus continuous differentiability on U can be defined equivalently in this way.

8.11 Another basic Lipschitz estimate

Let m and n be positive integers, let U be a nonempty open subset of $\mathbf{R}^n$, and let f be a continuously-differentiable mapping from U into $\mathbf{R}^m$. Thus f is differentiable at every $x \in U$, as in the previous section, so that $f'(x)$ is defined as a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ for every $x \in U$. We also have that f' is continuous on U, as before.

Let N and N_0 be norms on $\mathbf{R}^m$ and $\mathbf{R}^n$, respectively, and let $\|\cdot\|_{op}$ be the corresponding operator norm for linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$, as in Section 7.5. Remember that N and N_0 can be compared with the standard Euclidean norms on $\mathbf{R}^m$ and $\mathbf{R}^n$, respectively, as in Section 6.7. This leads to an analogous comparison between $\|\cdot\|_{op}$ and the operator norm for linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$ with respect to the standard Euclidean norms on $\mathbf{R}^n$ and $\mathbf{R}^m$. In particular, one can use this to check that $\|\cdot\|_{op}$ is continuous on the space of linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$, with respect to the metric that corresponds to the Hilbert–Schmidt norm on $m \times n$ matrices of real numbers. This implies that

$$\|f'(x)\|_{op} \text{ is continuous as a real-valued function on } U, \tag{8.11.1}$$

because $f'(x)$ is continuous on U. If $x \in U$ and $v \in \mathbf{R}^n$, then the directional derivative $D_v f(x)$ of f at x in the direction of v is the same as $f'(x)$ applied to v, as in (8.7.9). Hence

$$N(D_v f(x)) = N(f'(x)(v)) \le \|f'(x)\|_{op}\, N_0(v), \tag{8.11.2}$$

by the definition of $\|\cdot\|_{op}$.

Let us suppose from now on in this section that U is also convex in $\mathbf{R}^n$. This means that for each $x, w \in U$ and $t \in \mathbf{R}$ with $0 \le t \le 1$ we have that

$$t\,x + (1-t)\,w \in U. \tag{8.11.3}$$

Let $x, w \in U$ be given, and consider

$$f(t\,x + (1-t)\,w) = f(w + t\,(x-w)) \tag{8.11.4}$$

as a function of $t \in [0,1]$ with values in $\mathbf{R}^m$. Because f is differentiable on U, we can differentiate (8.11.4) in t, to get that

$$\frac{d}{dt}(f(t\,x + (1-t)\,w)) = (D_{(x-w)} f)(t\,x + (1-t)\,w) \tag{8.11.5}$$

for every $t \in [0,1]$. It follows that

$$\int_0^1 (D_{(x-w)} f)(t\,x + (1-t)\,w)\,dt = f(x) - f(w), \tag{8.11.6}$$

by the fundamental theorem of calculus.

This implies that

$$N(f(x) - f(w)) \le \int_0^1 N((D_{(x-w)} f)(t\,x + (1-t)\,w))\,dt, \tag{8.11.7}$$

as in Section 8.1. Combining this with (8.11.2), we get that

$$N(f(x) - f(w)) \le N_0(x - w) \int_0^1 \|f'(t\,x + (1-t)\,w)\|_{op}\,dt. \tag{8.11.8}$$

Suppose that f' is bounded on U, and observe that

$$N(f(x) - f(w)) \le \Big(\sup_{u \in U} \|f'(u)\|_{op}\Big)\, N_0(x - w), \tag{8.11.9}$$

by (8.11.8). This means that f is Lipschitz as a mapping from U into $\mathbf{R}^m$, using the restriction to U of the metric on $\mathbf{R}^n$ associated to N_0, and the metric on $\mathbf{R}^m$ associated to N. More precisely, the corresponding Lipschitz constant is given by

$$\mathrm{Lip}(f) = \sup_{u \in U} \|f'(u)\|_{op}. \tag{8.11.10}$$

Indeed, $\mathrm{Lip}(f)$ is less than or equal to the right side of (8.11.10), because of (8.11.9). In order to get the opposite inequality, one can check directly that $\|f'(u)\|_{op}$ is less than or equal to $\mathrm{Lip}(f)$ for every $u \in U$, as in Section 8.8.

If $m = 1$, then we can take N to be the usual absolute value function on $\mathbf{R}$. In this case, one can use the mean value theorem to get that

$$|f(x) - f(w)| \le \Big(\sup_{u \in U} \|f'(u)\|_{op}\Big)\, N_0(x - w) \tag{8.11.11}$$

for every $x, w \in U$ when f is differentiable on U, and f' is bounded on U. The corresponding Lipschitz constant of f on U is given by (8.11.10), as before.

8.12 Some remarks about connectedness

Let (X, d_X) be a metric space. Remember that subsets A, B of X are said to be *separated* in X if

$$\overline{A} \cap B = A \cap \overline{B} = \emptyset, \tag{8.12.1}$$

where $\overline{A}$, $\overline{B}$ are the closures of A, B in X, respectively. A subset E of X is said to be *connected* if E cannot be expressed as the union of two nonempty separated subsets of X.

Let X_0 be a subset of X, and remember that X_0 may be considered as a metric space with respect to the restriction of $d_X(x, w)$ to $x, w \in X_0$. If $A \subseteq X_0$, then let $\overline{A}_X$ be the closure of A in X, and let $\overline{A}_{X_0}$ be the closure of A in X_0. One can check that

$$\overline{A}_{X_0} = \overline{A}_X \cap X_0 \tag{8.12.2}$$

for every $A \subseteq X_0$. More precisely, if x is any element of X_0, then one can verify that x is a limit point of A as a subset of X if and only if x is a limit point of A as a subset of X_0.

If A, B are subsets of X_0, then it follows that

(8.12.3) $$A,\ B \text{ are separated as subsets of } X_0$$

if and only if

(8.12.4) $$A,\ B \text{ are separated as subsets of } X.$$

If E is a subset of X_0, then the previous statement implies that

(8.12.5) $$E \text{ is connected as a subset of } X_0 \text{ if and only if } E \text{ is connected as a subset of } X.$$

Note that disjoint closed subsets of X are separated in X. One can check that

(8.12.6) $$\text{disjoint open subsets of } X \text{ are separated in } X$$

too. If X is not connected, then X can be expressed as the union of two nonempty separated sets A, B. In this case, one can verify that

(8.12.7) $$A \text{ and } B \text{ are each both open and closed in } X.$$

Suppose that $E \subseteq X$ is not connected, so that $E = A \cup B$ for some nonempty separated subsets A, B of X. If E is a closed set in X, then one can check that

(8.12.8) $$A \text{ and } B \text{ are closed sets in } X.$$

If E is an open set in X, then one can verify that

(8.12.9) $$A \text{ and } B \text{ are open sets in } X.$$

8.13 Locally constant mappings

Let (X, d_X) be a metric space, let Y be a set, and let f be a mapping from X into Y. We say that f is *locally constant* on X if for every $x \in X$ there is a positive real number $r(x)$ such that

(8.13.1) $$f(w) = f(x)$$

for every $w \in X$ with $d_X(x, w) < r(x)$. If Y is equipped with a metric d_Y, then

(8.13.2) any locally constant mapping from X into Y is continuous.

If Y is equipped with the discrete metric, as mentioned in Section 3.7, then it is easy to see that

(8.13.3) every continuous mapping from X into Y is locally constant.

Let f be a locally constant mapping from X into Y, and let E be a subset of X. If E is connected as a subset of X, then

$$f \text{ is constant on } E. \tag{8.13.4}$$

Equivalently, if f is not constant on E, then E is not connected. One way to see this is to use the fact that f is continuous with respect to the discrete metric on Y. It is well known that any continuous mapping from X into another metric space maps connected subsets of X to connected subsets of the range. If E is connected in X, then it follows that

$$f(E) \text{ is connected with respect to the discrete metric on } Y. \tag{8.13.5}$$

However, one can check that a subset of Y is connected with respect to the discrete metric if and only if it has at most one element.

If X is not connected, then X can be expressed as the union of two nonempty disjoint open sets, as in the previous section. One can use this to get a mapping f from X into any set Y with at least two elements such that f is locally constant on X, and not constant on X.

Let m and n be positive integers, let U be a nonempty open subset of $\mathbf{R}^n$, and let f be a mapping from U into $\mathbf{R}^m$. Suppose that f is differentiable on U, and that

$$f'(x) = 0 \tag{8.13.6}$$

for every $x \in U$. If U is convex, then it follows that f is constant on U, as in Section 8.11. Similarly,

$$\text{the restriction of } f \text{ to any convex open subset of } U \text{ is constant.} \tag{8.13.7}$$

One can check that open balls in $\mathbf{R}^n$ with respect to the metric associated to any norm on $\mathbf{R}^n$ are convex. This implies that

$$f \text{ is locally constant on } U, \tag{8.13.8}$$

with respect to the restriction to U of the standard Euclidean metric on $\mathbf{R}^n$. Of course, if f is any mapping from U into $\mathbf{R}^m$ that is locally constant with respect to the restriction to U of the standard Euclidean metric on $\mathbf{R}^n$, then f is differentiable on U, and satisfies (8.13.6) for every $x \in U$.

If U is also connected as a subset of $\mathbf{R}^n$, then (8.13.8) implies that

$$f \text{ is constant on } U, \tag{8.13.9}$$

as before. More precisely, this uses the fact that U is connected as a subset of itself, with respect to the restriction of the standard Euclidean metric on $\mathbf{R}^n$ to U, as in the previous section. This is related to Exercise 9 on p239 of [155]. If U is not connected as a subset of $\mathbf{R}^n$, then U can be expressed as the union of two nonempty disjoint open subsets of $\mathbf{R}^n$, as in the previous section. One can use this to get a locally constant mapping from U into $\mathbf{R}^m$ that is not constant on U, as before.

8.14 Some local Lipschitz conditions

Let (X, d_X) and (Y, d_Y) be metric spaces, and let f be a mapping from X into Y. If $x \in X$ and $r > 0$, then we let

$$B_X(x, r) \tag{8.14.1}$$

be the open ball in X centered at x with radius r with respect to d_X, as usual. Let us say that

$$f \text{ is } \textit{locally Lipschitz with constant } C \geq 0 \text{ at } x \tag{8.14.2}$$

if there is a positive real number r such that

$$\text{the restriction of } f \text{ to } B_X(x, r) \text{ is Lipschitz with constant } C, \tag{8.14.3}$$

with respect to the restriction of $d_X(\cdot, \cdot)$ to $B_X(x, r)$. Note that this implies that f is pointwise Lipschitz at x with constant C up to the scale r, as in Section 8.8.

We may also simply say that

$$f \text{ is } \textit{locally Lipschitz} \text{ at } x \tag{8.14.4}$$

if there is a nonnegative real number C such that f is locally Lipschitz at x with constant C. Similarly, we may say that

$$f \text{ is } \textit{locally Lipschitz} \text{ on } X \tag{8.14.5}$$

if f is locally Lipschitz at every $x \in X$. Of course, this implies that

$$f \text{ is continuous on } X. \tag{8.14.6}$$

Note that f is locally constant on X exactly when

$$f \text{ is locally Lipschitz with constant } C = 0 \text{ at every } x \in X. \tag{8.14.7}$$

Let W be an open subset of X, and suppose that the restriction of f to W is Lipschitz with constant $C \geq 0$, with respect to the restriction of $d_X(\cdot, \cdot)$ to W. This implies that for each $w \in W$, f is locally Lipschitz at w with constant C.

Let K be a compact subset of X, and suppose that

$$\text{for each } x \in K,\ f \text{ is locally Lipschitz at } x. \tag{8.14.8}$$

This means that for every $x \in K$, there is a positive real number $r(x)$ and a nonnegative real number $C(x)$ such that the restriction of f to $B_X(x, r(x))$ is Lipschitz with constant $C(x)$. Thus K can be covered by open balls of this type, and so there are finitely many elements $x_1, \ldots, x_l$ of K such that

$$K \subseteq \bigcup_{j=1}^{l} B_X(x_j, r(x_j)), \tag{8.14.9}$$

because K is compact. If we put

$$C = \max_{1 \le j \le l} C(x_j), \tag{8.14.10}$$

then it is easy to see that

$$f \text{ is locally Lipschitz with constant } C \tag{8.14.11}$$

at every $x \in K$. More precisely, this holds at every x in the right side of (8.14.9).

Let m and n be positive integers, let U be a nonempty open subset of $\mathbf{R}^n$, and let f be a continuously-differentiable mapping from U into $\mathbf{R}^m$. Also let N and N_0 be norms on $\mathbf{R}^m$ and $\mathbf{R}^n$, respectively, and let $\| \cdot \|_{op}$ be the corresponding operator norm for linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$. Suppose that E is a nonempty convex subset of $\mathbf{R}^n$ such that $E \subseteq U$, and that

$$\|f'(u)\|_{op} \text{ is bounded on } E. \tag{8.14.12}$$

If $x, w \in E$, then

$$N(f(x) - f(w)) \le \Big(\sup_{u \in E} \|f'(u)\|_{op}\Big) N_0(x - w), \tag{8.14.13}$$

as in Section 8.11. This means that

$$\text{the restriction of } f \text{ to } E \text{ is Lipschitz} \tag{8.14.14}$$

with respect to the metric on $\mathbf{R}^m$ associated to N, and the restriction to E of the metric on $\mathbf{R}^n$ associated to N_0.

If E is also closed and bounded in $\mathbf{R}^n$, then E is compact in $\mathbf{R}^n$. It is well known that this implies that E is compact as a subset of U, with respect to the restriction to U of the standard Euclidean metric on $\mathbf{R}^n$. Remember that $\|f'(u)\|_{op}$ is continuous as a real-valued function on U, as in Section 8.11. It follows that (8.14.12) holds under these conditions.

One can check that open and closed balls in $\mathbf{R}^n$ with respect to the metric associated to a norm are convex subsets of $\mathbf{R}^n$, as before. If $v \in U$ and ϵ is a positive real number, then

$$\|f'(u)\|_{op} < \|f'(v)\|_{op} + \epsilon \tag{8.14.15}$$

when $u \in U$ is sufficiently close to v, because $\|f'(u)\|_{op}$ is continuous at v. One can use this to get that f is locally Lipschitz at v with constant

$$C = \|f'(v)\|_{op} + \epsilon, \tag{8.14.16}$$

with respect to the metric on $\mathbf{R}^m$ associated to N, and the restriction to U of the metric on $\mathbf{R}^n$ associated to N_0. We also have that $\|f'(u)\|_{op}$ is bounded on any closed ball contained in U, as in the preceding paragraph. This implies that the restriction of f to such a ball is Lipschitz, as before.

Chapter 9

More on differentiable mappings

9.1 Continuously-differentiable mappings

Let m and n be positive integers, let U be an open subset of $\mathbf{R}^n$, and let f be a continuously-differentiable mapping from U into $\mathbf{R}^m$. Also let x be an element of U, and put

$$A = f'(x), \tag{9.1.1}$$

for convenience. Observe that $f - A$ is continuously differentiable as a mapping from U into $\mathbf{R}^m$ too. If $w \in U$, then

$$(f - A)'(w) = f'(w) - A = f'(w) - f'(x). \tag{9.1.2}$$

This tends to 0 as w approaches x, because f is continuously differentiable on U.

Let N_0 and N be norms on $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively. Thus $d_{N_0}(v, w) = N_0(v-w)$ and $d_N(y, z) = N(y-z)$ are metrics on $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively. The corresponding operator norm for linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$ is denoted $\| \cdot \|_{op}$, as usual. In particular,

$$\|f'(w) - A\|_{op} = \|f'(w) - f'(x)\|_{op} \to 0 \quad \text{as } w \to x, \tag{9.1.3}$$

because f is continuously-differentiable on U.

Let U_1 be a convex open subset of $\mathbf{R}^n$ such that $x \in U_1$ and $U_1 \subseteq U$. Suppose that f' is bounded on U_1, so that $f' - A$ is bounded on U_1 as well. Put

$$b_1 = \sup_{w \in U_1} \|f'(w) - A\|_{op}. \tag{9.1.4}$$

The restriction of $f - A$ to U_1 is Lipschitz with constant b_1, as in Sections 8.11 and 8.14. This uses the restriction of d_{N_0} to U_1, and d_N on $\mathbf{R}^m$.

Suppose that

$$N(A(v)) = N(f'(x)(v)) \geq c\, N_0(v) \tag{9.1.5}$$

for some $c > 0$ and all $v \in \mathbf{R}^n$. This implies that

$$d_N(A(u), A(v)) \geq c\, d_{N_0}(u, v) \tag{9.1.6}$$

for every $u, v \in \mathbf{R}^n$, as in (7.9.5). Remember that there is a positive real number c such that (9.1.5) holds for all $v \in \mathbf{R}^n$ exactly when A is one-to-one on $\mathbf{R}^n$, as in Sections 7.9 and 7.10.

Suppose that

$$b_1 < c. \tag{9.1.7}$$

Under these conditions, we have that

$$d_N(f(u), f(v)) \geq (c - b_1)\, d_{N_0}(u, v) \tag{9.1.8}$$

for every $u, v \in U_1$. More precisely, this corresponds to (7.11.5), with some changes in notation.

Similarly, if $w \in U_1$, then

$$N(f'(w)(v)) \geq (c - b_1)\, N_0(v) \tag{9.1.9}$$

for every $v \in \mathbf{R}^n$. This corresponds to (7.11.10).

Suppose for instance that U_1 is the open ball in $\mathbf{R}^n$ centered at x with radius $r > 0$ with respect to d_{N_0}. This is a convex open subset of $\mathbf{R}^n$, as before. If r is small enough, then $U_1 \subseteq U$, because U is an open set in $\mathbf{R}^n$ that contains x. We can make b_1 as small as we want by taking r small enough, because of (9.1.3). In particular, (9.1.7) holds when r is sufficiently small.

Some more properties of mappings like these will be discussed in Section 9.7.

9.2 Invertible linear mappings

Let n be a positive integer, and let

$$\mathcal{L}(\mathbf{R}^n) = \mathcal{L}(\mathbf{R}^n, \mathbf{R}^n) \tag{9.2.1}$$

be the space of linear mappings from $\mathbf{R}^n$ into itself. Also let N be a norm on $\mathbf{R}^n$, and let $\|A\|_{op}$ be the corresponding operator norm for $A \in \mathcal{L}(\mathbf{R}^n)$. More precisely, this uses N on $\mathbf{R}^n$ as both the domain and range of A. Remember that

$$d_{op}(A, B) = \|A - B\|_{op} \tag{9.2.2}$$

defines a metric on $\mathcal{L}(\mathbf{R}^n)$.

Let I be the identity mapping on $\mathbf{R}^n$, which sends every element of $\mathbf{R}^n$ to itself. It is easy to see that

$$\|I\|_{op} = 1. \tag{9.2.3}$$

If $A, B \in \mathcal{L}(\mathbf{R}^n)$, then their composition $B \circ A$ defines a linear mapping from $\mathbf{R}^n$ into itself as well, so that $B \circ A \in \mathcal{L}(\mathbf{R}^n)$. Remember that

$$\|B \circ A\|_{op} \le \|A\|_{op} \, \|B\|_{op}, \tag{9.2.4}$$

as in Section 7.8.

A linear mapping A from $\mathbf{R}^n$ into itself is said to be *invertible* if A is a one-to-one mapping from $\mathbf{R}^n$ onto itself. In this case, the inverse mapping A^{-1} is linear as well. The collection of invertible linear mappings on $\mathbf{R}^n$ may be denoted $GL(\mathbf{R}^n)$. It is well known that $A \in \mathcal{L}(\mathbf{R}^n)$ is one-to-one if and only if $A(\mathbf{R}^n) = \mathbf{R}^n$.

If $A \in GL(\mathbf{R}^n)$, then

$$N(A^{-1}(u)) \le \|A^{-1}\|_{op} \, N(u) \tag{9.2.5}$$

for every $u \in \mathbf{R}^n$. More precisely, $\|A^{-1}\|_{op}$ is the smallest nonnegative real number with this property, by definition of the operator norm. Equivalently, (9.2.5) says that

$$N(v) \le \|A^{-1}\|_{op} \, N(A(v)) \tag{9.2.6}$$

for every $v \in \mathbf{R}^n$. Note that $\|A^{-1}\|_{op} > 0$, because $A^{-1} \neq 0$. Thus (9.2.6) is the same as saying that

$$\|A^{-1}\|_{op}^{-1} \, N(v) \le N(A(v)) \tag{9.2.7}$$

for every $v \in \mathbf{R}^n$.

Suppose that $B \in \mathcal{L}(\mathbf{R}^n)$ satisfies

$$\|A - B\|_{op} < 1/\|A^{-1}\|_{op}. \tag{9.2.8}$$

Using (9.2.7), we get that

$$(\|A^{-1}\|_{op}^{-1} - \|A - B\|_{op}) \, N(v) \le N(B(v)) \tag{9.2.9}$$

for every $v \in \mathbf{R}^n$, as in Section 7.11. In particular, this implies that $B(v) = 0$ only when $v = 0$, so that B is injective. It follows that $B(\mathbf{R}^n) = \mathbf{R}^n$, and hence

$$B \in GL(\mathbf{R}^n), \tag{9.2.10}$$

as before. This shows that

$$GL(\mathbf{R}^n) \text{ is an open set in } \mathcal{L}(\mathbf{R}^n), \tag{9.2.11}$$

with respect to the metric (9.2.2) associated to the operator norm.

Note that

$$N(B^{-1}(u)) \le (\|A^{-1}\|_{op}^{-1} - \|A - B\|_{op})^{-1} \, N(u) \tag{9.2.12}$$

for every $u \in \mathbf{R}^n$, by (9.2.9). This means that

$$\begin{aligned} \|B^{-1}\|_{op} &\le (\|A^{-1}\|_{op}^{-1} - \|A - B\|_{op})^{-1} \\ &= \frac{\|A^{-1}\|_{op}}{1 - \|A^{-1}\|_{op} \, \|A - B\|_{op}} \end{aligned} \tag{9.2.13}$$

in this situation. If

$$\|A - B\|_{op} \le 1/(2\|A^{-1}\|_{op}), \tag{9.2.14}$$

then we get that

$$\|B^{-1}\|_{op} \le 2\,\|A^{-1}\|_{op}. \tag{9.2.15}$$

Of course, (9.2.14) implies (9.2.8).

We would like to look at continuity properties of

$$A \mapsto A^{-1} \tag{9.2.16}$$

on $GL(\mathbf{R}^n)$. If $A, B \in GL(\mathbf{R}^n)$, then it is easy to see that

$$A^{-1} - B^{-1} = A^{-1} \circ (B - A) \circ B^{-1}. \tag{9.2.17}$$

This implies that

$$\|A^{-1} - B^{-1}\|_{op} \le \|A^{-1}\|_{op}\,\|B^{-1}\|_{op}\,\|A - B\|_{op}. \tag{9.2.18}$$

If (9.2.14) holds, then we get that

$$\|A^{-1} - B^{-1}\|_{op} \le 2\,\|A^{-1}\|_{op}^2\,\|A - B\|_{op}, \tag{9.2.19}$$

by (9.2.15). It follows that $B \mapsto B^{-1}$ is continuous at A, with respect to the metric (9.2.2) associated to the operator norm, and its restriction to $GL(\mathbf{R}^n)$.

9.3 The inverse function theorem

Let n be a positive integer, and let N be a norm on $\mathbf{R}^n$, so that $d_N(v, w) = N(v - w)$ is a metric on $\mathbf{R}^n$. Using N, we get the corresponding operator norm $\|\cdot\|_{op}$ on the space $\mathcal{L}(\mathbf{R}^n)$ of linear mappings from $\mathbf{R}^n$ into itself, and its associated metric, as in (9.2.2). Of course, one can simply take N to be the standard Euclidean norm on $\mathbf{R}^n$.

Let W be an open subset of $\mathbf{R}^n$, and let f be a continuously-differentiable mapping from W into $\mathbf{R}^n$. Also let $w \in W$ be given, and suppose that

$$f'(w) \text{ is invertible} \tag{9.3.1}$$

as a linear mapping from $\mathbf{R}^n$ into itself. Under these conditions, the *inverse function theorem* states that there are open subsets U and V of $\mathbf{R}^n$ with the following properties, which we describe in two parts.

In the first part,

$$w \in U,\ U \subseteq W, \text{ and } f(w) \in V. \tag{9.3.2}$$

We are able to choose U and V so that

$$f \text{ is a one-to-one mapping from } U \text{ onto } V. \tag{9.3.3}$$

We can also choose U so that

(9.3.4) $f'(x)$ is invertible as a linear mapping on $\mathbf{R}^n$ for each $x \in U$.

In particular, we can define a mapping g from V onto U to be the inverse of the restriction of f to U, so that

$$g(f(x)) = x \tag{9.3.5}$$

for every $x \in U$. The second part of the inverse function theorem says that

(9.3.6) g is continuously-differentiable as a mapping from V into $\mathbf{R}^n$,

with

$$g'(y) = (f'(g(y)))^{-1} \tag{9.3.7}$$

for every $y \in V$. More precisely, the right side of (9.3.7) is the inverse of $f'(g(y))$, as a linear mapping on $\mathbf{R}^n$.

If $x \in U$ and g is differentiable at $f(x) \in V$, then the chain rule implies that

$$g'(f(x)) \circ f'(x) = I, \tag{9.3.8}$$

because of (9.3.5). This is the same as (9.3.7), with $y = f(x)$. The proof of the inverse function theorem will be discussed in the next section. Note that the proof could be simplified when $n = 1$, using the mean value theorem and the intermediate value theorem, which is related to Exercise 2 on p114 of [155].

9.4 Proving the inverse function theorem

Let us continue with the same notation and hypotheses as in the previous section. Remember that $f'(x)$ is continuous as a function of $x \in W$ with values in $\mathcal{L}(\mathbf{R}^n)$, as in Section 8.10. We have also seen that the collection $GL(\mathbf{R}^n)$ of invertible linear mapping from $\mathbf{R}^n$ into itself is an open set in $\mathcal{L}(\mathbf{R}^n)$, as in Section 9.2. It follows that

$$\{x \in W : f'(x) \in GL(\mathbf{R}^n)\} \tag{9.4.1}$$

is an open set in $\mathbf{R}^n$, because W is an open set, by hypothesis. Of course, (9.4.1) is the same as

$$(f')^{-1}(GL(\mathbf{R}^n)), \tag{9.4.2}$$

where f' is considered as a mapping from W into $\mathcal{L}(\mathbf{R}^n)$.

To prove the inverse function theorem, we can reduce to the case where

(9.4.3) $f'(w)$ is the identity mapping I on $\mathbf{R}^n$.

More precisely, this can be obtained by replacing f with

$$(f'(w))^{-1} \circ f, \tag{9.4.4}$$

as a continuously-differentiable mapping from W into $\mathbf{R}n$. It is easy to see that the differential of (9.4.4) at w is equal to I, by construction. If we can prove the inverse function theorem in this case, then the analogous conclusions for f can be obtained by composing (9.4.4) with $f'(w)$, as a linear mapping from $\mathbf{R}^n$ into itself.

Let $B_N(w, r)$ be the open ball in $\mathbf{R}^n$ centered at w with radius $r > 0$ with respect to $d_N(\cdot, \cdot)$, as before. Note that

$$B_N(w, r) \subseteq W \tag{9.4.5}$$

when r is sufficiently small, because W is an open set that contains w. We also have that $f'(x) - I$ is as small as we want when $x \in \mathbf{R}^n$ is close enough to w, because $f'(x)$ is continuous at w. In particular, we can choose $r > 0$ small enough so that (9.4.5) holds and

$$\|f'(x) - I\|_{op} \leq 1/2 \tag{9.4.6}$$

for every $x \in \mathbf{R}^n$ with $N(x - w) < r$. This implies that $f'(x)$ is invertible when $N(x - w) < r$, as in Section 9.2.

Using (9.4.6), we get that

(9.4.7) the restriction of $f(x) - x$ to $B_N(w, r)$
is Lipschitz with constant $c = 1/2$,

with respect to $d_N(\cdot, \cdot)$ and its restriction to $B_N(w, r)$, as in Section 8.11. This uses the fact that the differential of $f(x)-x$ is equal to $f'(x)-I$ for each $x \in W$. It follows that

(9.4.8) f is bilipschitz on $B_N(w, r)$,

as in Section 7.11, and in particular that f is one-to-one on $B_N(w, r)$. The restriction of f to $B_N(w, r)$ is an open mapping too, as in Section 7.14. The first part of the inverse function theorem now follows by taking

$$U = B_N(w, r) \tag{9.4.9}$$

and

$$V = f(B_N(w, r)). \tag{9.4.10}$$

Let g be the inverse of the restriction of f to $B_N(w, r)$, as before. This is a Lipschitz mapping from V into $\mathbf{R}^n$, with respect to $d_N(\cdot, \cdot)$ and its restriction to V, because f is bilipschitz on $B_N(w, r)$, as in the preceding paragraph. One can verify that

(9.4.11) g is differentiable on V,

with differential as in (9.3.7), using the differentiability of f. This corresponds to part of part (b) of Theorem 9.24 on p221 of [155], for instance. The continuity of g' on V follows from (9.3.7), because f' and g are continuous, and using the continuity of (9.2.16) on $GL(\mathbf{R}^n)$.

9.5 Some remarks about $\mathbf{R}^{n+m}$

Let n and m be positive integers, so that $\mathbf{R}^n$, $\mathbf{R}^m$, and $\mathbf{R}^{n+m}$ are the usual spaces of n-tuples, m-tuples, and $(n+m)$-tuples of real numbers, respectively. It will sometimes be convenient for us to identify $\mathbf{R}^{n+m}$ with the Cartesian product $\mathbf{R}^n \times \mathbf{R}^m$ of $\mathbf{R}^n$ and $\mathbf{R}^m$. Thus, if $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$ and $y = (y_1, \ldots, y_m) \in \mathbf{R}^m$, then we may identify

$$(x, y) \in \mathbf{R}^n \times \mathbf{R}^m \tag{9.5.1}$$

with

$$(x_1, \ldots, x_n, y_1, \ldots, y_m) \in \mathbf{R}^{n+m}. \tag{9.5.2}$$

In particular, $(x, 0)$ and $(0, y)$ can be identified with elements of $\mathbf{R}^{n+m}$.

If f is a function defined on a subset of $\mathbf{R}^{n+m}$, then we may use

$$f(x, y) \tag{9.5.3}$$

to denote the value of f at the point in $\mathbf{R}^{n+m}$ identified with (x, y), when that point is in the domain of f.

Let A be a linear mapping from $\mathbf{R}^{n+m}$ into $\mathbf{R}^n$. If $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$, then put

$$A_1(x) = A(x, 0) \tag{9.5.4}$$

and

$$A_2(y) = A(0, y). \tag{9.5.5}$$

This defines A_1 and A_2 as linear mappings from $\mathbf{R}^n$ and $\mathbf{R}^m$ into $\mathbf{R}^n$, respectively. Note that

$$A(x, y) = A_1(x) + A_2(y) \tag{9.5.6}$$

for every $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$. Conversely, if A_1 and A_2 are linear mappings from $\mathbf{R}^n$ and $\mathbf{R}^m$ into $\mathbf{R}^n$, respectively, then (9.5.6) defines a linear mapping from $\mathbf{R}^{n+m}$ into $\mathbf{R}^n$.

Let A be a linear mapping from $\mathbf{R}^{n+m}$ into $\mathbf{R}^n$, and let A_1 and A_2 be as in (9.5.4) and (9.5.5), respectively. Also let $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$ be given, and observe that

$$A(x, y) = 0 \tag{9.5.7}$$

if and only if

$$A_1(x) = -A_2(y), \tag{9.5.8}$$

by (9.5.6). If A_1 is a one-to-one mapping from $\mathbf{R}^n$ onto itself, then (9.5.8) is the same as saying that

$$x = -A_1^{-1}(A_2(y)). \tag{9.5.9}$$

In particular, for each $y \in \mathbf{R}^m$ there is a unique $x \in \mathbf{R}^n$ such that (9.5.7) holds in this situation.

Clearly

$$A_1(\mathbf{R}^n) \subseteq A(\mathbf{R}^{n+m}), \tag{9.5.10}$$

by the definition (9.5.4) of A_1. In particular, if $A_1(\mathbf{R}^n) = \mathbf{R}^n$, then

$$A(\mathbf{R}^{n+m}) = \mathbf{R}^n. \tag{9.5.11}$$

Consider the mapping $\widehat{A}$ from $\mathbf{R}^{n+m}$ into itself defined by

$$\widehat{A}(x, y) = (A(x, y), y) \tag{9.5.12}$$

for every $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$. This is a linear mapping from $\mathbf{R}^{n+m}$ into itself, which can also be expressed as

$$\widehat{A}(x, y) = (A_1(x) + A_2(y), y) \tag{9.5.13}$$

for every $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$, by (9.5.6). If A_1 is a one-to-one mapping from $\mathbf{R}^n$ onto itself, then one can check that

$$\widehat{A} \text{ is a one-to-one mapping from } \mathbf{R}^{n+m} \text{ onto itself.} \tag{9.5.14}$$

It is easy to see that the converse holds as well.

Let $u_1, \ldots, u_{n+m}$ be the standard basis vectors in $\mathbf{R}^{n+m}$, so that for each $l = 1, \ldots, m+n$, the lth coordinate of u_l is equal to 1, and the other coordinates of u_l are equal to 0. If A is a linear mapping from $\mathbf{R}^{n+m}$ onto $\mathbf{R}^n$, then

$$\mathbf{R}^n \text{ is spanned by } A(u_1), \ldots, A(u_{n+m}). \tag{9.5.15}$$

Under these conditions, it is well known that there is a subset K of $\{1, \ldots, n+m\}$ such that K has exactly n elements, and

$$A(u_k),\ k \in K, \text{ forms a basis for } \mathbf{R}^n. \tag{9.5.16}$$

If $K = \{1, \ldots, n\}$, then the linear mapping A_1 on $\mathbf{R}^n$ associated to A as in (9.5.4) is invertible. Otherwise, one could rearrange the coordinates on $\mathbf{R}^{n+m}$ to reduce to this case.

9.6 The implicit function theorem

Let n and m be positive integers again, and let O be an open subset of $\mathbf{R}^{n+m}$, with respect to the standard Euclidean metric. Also let f be a continuously-differentiable mapping from O into $\mathbf{R}^n$. Suppose that $a \in \mathbf{R}^n$ and $b \in \mathbf{R}^m$ have the properties that $(a, b) \in O$ and

$$f(a, b) = 0, \tag{9.6.1}$$

where (a, b) is identified with an element of $\mathbf{R}^{n+m}$, as in the previous section.

Put

$$A = f'(a, b), \tag{9.6.2}$$

which is a linear mapping from $\mathbf{R}^{n+m}$ into $\mathbf{R}^n$. This leads to a linear mappings A_1 and A_2 from $\mathbf{R}^n$ and $\mathbf{R}^m$ into $\mathbf{R}^n$, respectively, as in (9.5.4) and (9.5.5).

Suppose that

(9.6.3) A_1 is a one-to-one mapping from $\mathbf{R}^n$ onto itself.

Under these conditions, the *implicit function theorem* states that there are open sets $U \subseteq \mathbf{R}^{n+m}$ and $W \subseteq \mathbf{R}^m$ with the following properties, as in Theorem 9.28 on p224 of [155].

First,

$$(a, b) \in U,\ U \subseteq O, \text{ and } b \in W. \tag{9.6.4}$$

Second, for each $y \in W$ there is a unique $x \in \mathbf{R}^n$ such that

$$(x, y) \in U \quad \text{and} \quad f(x, y) = 0. \tag{9.6.5}$$

If $y \in W$, then let $g(y)$ be the element of $\mathbf{R}^n$ just mentioned, so that g is a mapping from W into $\mathbf{R}^n$, $g(b) = a$, and for every $y \in W$ we have that

$$(g(y), y) \in U \tag{9.6.6}$$

and

$$f(g(y), y) = 0. \tag{9.6.7}$$

We also have that g is continuously differentiable on W, with

$$g'(b) = -A_1^{-1} \circ A_2. \tag{9.6.8}$$

To prove the implicit function theorem, we put

$$F(x, y) = (f(x, y), y) \tag{9.6.9}$$

for every $(x, y) \in O$, where the right side is identified with an element of $\mathbf{R}^{n+m}$, as before. This defines F as a continuously-differentiable mapping from O into $\mathbf{R}^{n+m}$, with

$$F'(a, b) = \widehat{A}, \tag{9.6.10}$$

where $\widehat{A}$ is as in (9.5.12). In this situation, $\widehat{A}$ is a one-to-one linear mapping from $\mathbf{R}^{n+m}$ onto itself, so that the inverse function theorem can be applied to F at (a, b).

Thus there are open sets U, V in $\mathbf{R}^{n+m}$ such that $(a, b) \in U$, $U \subseteq O$,

$$F(a, b) = (f(a, b), b) = (0, b) \in V, \tag{9.6.11}$$

and F is a one-to-one mapping from U onto V. The inverse G of the restriction of F to U is a continuously-differentiable mapping from V onto U. We also have that G preserves the last m coordinates of points in its domain, because of the analogous property of F.

Observe that

$$W = \{y \in \mathbf{R}^m : (0, y) \in V\} \tag{9.6.12}$$

is an open subset of $\mathbf{R}^m$ that contains b. If $y \in W$, then we can take $g(y) \in \mathbf{R}^n$ so that

$$(g(y), y) = G(0, y). \tag{9.6.13}$$

See [155] or other texts for more details.

9.7 Local embeddings

Let n and m be positive integers, and let us continue to identify $\mathbf{R}^{n+m}$ with $\mathbf{R}^n \times \mathbf{R}^m$, as in Section 9.5. Suppose that A is a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^{n+m}$. If $v \in \mathbf{R}^n$, then $A(v) \in \mathbf{R}^{n+m}$ can be expressed as

$$A(v) = (A_1(v), A_2(v)), \tag{9.7.1}$$

where $A_1(v) \in \mathbf{R}^n$ and $A_2(v) \in \mathbf{R}^m$, using the identification just mentioned. More precisely, this defines A_1 and A_2 as linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively. Conversely if A_1 and A_2 are linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively, then (9.7.1) defines a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^{n+m}$.

Suppose now that

$$A_1 \text{ is injective on } \mathbf{R}^n, \tag{9.7.2}$$

which implies that

$$A_1 \text{ is invertible on } \mathbf{R}^n, \tag{9.7.3}$$

as before. Of course, (9.7.2) also implies that

$$A \text{ is injective on } \mathbf{R}^n. \tag{9.7.4}$$

It can be shown that (9.7.4) implies a condition like (9.7.2), after rearranging the coordinates on $\mathbf{R}^{n+m}$ is a suitable way. We shall discuss this further at the end of the section.

Note that

$$A_2 \circ A_1^{-1} \tag{9.7.5}$$

defines a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$. Put

$$B(v) = (v, A_2(A_1^{-1}(v))) \tag{9.7.6}$$

for every $v \in \mathbf{R}^n$, which defines a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^{n+m}$. Thus

$$A = B \circ A_1, \tag{9.7.7}$$

as a mapping from $\mathbf{R}^n$ into $\mathbf{R}^{n+m}$.

Let W be an open subset of $\mathbf{R}^n$, and let f be a mapping from W into $\mathbf{R}^{n+m}$. If $w \in W$, then $f(w) \in \mathbf{R}^{n+m}$ can be expressed as

$$f(w) = (f_1(w), f_2(w)), \tag{9.7.8}$$

where $f_1(w) \in \mathbf{R}^n$ and $f_2(w) \in \mathbf{R}^m$, as before. This defines f_1 and f_2 as mappings from W into $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively. Conversely, if f_1 and f_2 are mappings from W into $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively, then (9.7.8) defines f as a mapping from W into $\mathbf{R}^{n+m}$.

It is easy to see that f is differentiable at a point $w \in W$ if and only if

$$f_1 \text{ and } f_2 \text{ are differentiable at } w. \tag{9.7.9}$$

In this case,

$$f'(w)(v) = (f_1'(w)(v), f_2'(w)(v)) \tag{9.7.10}$$

for every $v \in \mathbf{R}^n$. Equivalently, this means that $A = f'(w)$ corresponds to $A_1 = f_1'(w)$ and $A_2 = f_2'(w)$ as in (9.7.1).

Similarly, one can check that f is continuously-differentiable as a mapping from W into $\mathbf{R}^{n+m}$ if and only if

$$f_1 \text{ and } f_2 \text{ are continuously-differentiable} \tag{9.7.11}$$

as mappings from W into $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively. In this situation, suppose that $w \in W$ has the property that

$$f_1'(w) \text{ is invertible} \tag{9.7.12}$$

as a linear mapping from $\mathbf{R}^n$ into itself. Thus the inverse function theorem implies that the restriction of f_1 to a suitable neighborhood U of w is a one-to-one mapping onto a neighborhood V of $f_1(w)$, and that the corresponding inverse mapping has some additional nice properties.

This can be used to obtain more information about the behavior of f near w. Of course, this is all much simpler when f_1 is the identity mapping on $\mathbf{R}^n$. Otherwise, one can try to reduce to this case, at least locally near w. If f_1 has a local inverse near w, as in the preceding paragraph, then f can be expressed near w as the composition of f_1 with a simpler mapping into $\mathbf{R}^{n+m}$.

Let $e_1, \ldots, e_n$ be the standard basis vectors in $\mathbf{R}^n$, and let $u_1, \ldots, u_{n+m}$ be the standard basis vectors for $\mathbf{R}^{n+m}$. If (9.7.4) holds, then

$$A(e_1), \ldots, A(e_n) \text{ are linearly independent in } \mathbf{R}^{n+m}. \tag{9.7.13}$$

In this case, it is well known that there is a set $L \subseteq \{1, \ldots, n+m\}$ with exactly m elements such that

$$A(e_1), \ldots, A(e_n) \text{ together with } u_k,\ k \in L, \text{ is a basis for } \mathbf{R}^{n+m}. \tag{9.7.14}$$

If $L = \{n+1, \ldots, n+m\}$, then it is easy to see that

$$A_1(\mathbf{R}^n) = \mathbf{R}^n, \tag{9.7.15}$$

so that (9.7.3) holds. Otherwise, one can get an analogous condition with the coordinates on $\mathbf{R}^{n+m}$ rearranged.

9.8 Ranks of linear mappings

In this section, we suppose that the reader has some familiarity with linear algebra on Euclidean spaces. Let m and n be arbitrary positive integers, and let A be a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$. It is easy to see that the image $A(\mathbf{R}^n)$ of $\mathbf{R}^n$ under A is a linear subspace of $\mathbf{R}^m$. The *rank* of A is defined to be

$$\text{the dimension of } A(\mathbf{R}^n). \tag{9.8.1}$$

Thus the rank of A is a nonnegative integer less than or equal to m.

Of course, the rank of A is equal to 0 exactly when $A = 0$ on $\mathbf{R}^n$. The rank of A is equal to m exactly when $A(\mathbf{R}^n) = \mathbf{R}^m$.

Suppose that A has rank $r \geq 1$, and let $w_1, \ldots, w_r$ be a basis for $A(\mathbf{R}^n)$. If $j \in \{1, \ldots, r\}$, then choose $v_j \in \mathbf{R}^n$ such that

$$A(v_j) = w_j. \tag{9.8.2}$$

Observe that

$$v_1, \ldots, v_r \text{ are linearly independent in } \mathbf{R}^n, \tag{9.8.3}$$

because $w_1, \ldots, w_r$ are linearly independent in $\mathbf{R}^m$. Let V be the linear span of $v_1, \ldots, v_r$ in $\mathbf{R}^n$, so that

$$V \text{ is a linear subspace of } \mathbf{R}^n \text{ of dimension } r. \tag{9.8.4}$$

In particular,

$$r \leq n \tag{9.8.5}$$

automatically. By construction, $A(V) = A(\mathbf{R}^n)$. One can check that

$$\text{the restriction of } A \text{ to } V \text{ is one-to-one} \tag{9.8.6}$$

in this situation.

Alternatively, let $u_1, \ldots, u_n$ be any basis for $\mathbf{R}^n$, such as the standard basis. Thus

$$A(\mathbf{R}^n) \text{ is spanned by } A(u_1), \ldots, A(u_n). \tag{9.8.7}$$

It is well known that a subset of the vectors $A(u_1), \ldots, A(u_n)$ forms a basis for $A(\mathbf{R}^n)$. Such a subset has exactly r elements, by hypothesis. The corresponding subset of the vectors $u_1, \ldots, u_n$ can be used as a basis for a linear subspace V of $\mathbf{R}^n$ as in the preceding paragraph.

Now let V_0 be a linear subspace of $\mathbf{R}^n$, and suppose that

$$\text{the restriction of } A \text{ to } V_0 \text{ is one-to-one.} \tag{9.8.8}$$

The dimension of V_0 is equal to the dimension of $A(V_0)$, which is less than or equal to the dimension of $A(\mathbf{R}^n)$. Equivalently, the dimension of V_0 is less than or equal to r. The remarks in the previous paragraphs show that it is always possible to choose V_0 so that its dimension is equal to r. This shows that

$$\begin{aligned} &\text{the rank of } A \text{ is equal to the maximum of the dimensions} \\ &\text{of the linear subspaces } V_0 \text{ of } \mathbf{R}^n \text{ on which } A \text{ is one-to-one.} \end{aligned} \tag{9.8.9}$$

If A is one-to-one on $\mathbf{R}^n$, then the rank of A is equal to n. Conversely, if the rank of A is equal to n, then A is one-to-one on $\mathbf{R}^n$.

Let V_0 be a linear subspace of $\mathbf{R}^n$ on which A is one-to-one again, and let B be another linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$. If B is sufficiently close to A, as linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$, then

$$\text{the restriction of } B \text{ to } V_0 \text{ is one-to-one} \tag{9.8.10}$$

as well. This can be obtained from the remarks in Section 7.11. It follows that

(9.8.11) the rank of B is greater than or equal to the rank of A

when B is sufficiently close to A. Note that the rank of B may be greater than the rank of A in this case.

Let f be a continuously-differentiable mapping from an open subset W of $\mathbf{R}^n$ into $\mathbf{R}^m$, and let w be an element of W. Suppose that if $x \in W$ is sufficiently close to w, then

(9.8.12) the rank of $f'(x)$ is equal to the rank of $f'(w)$,

as linear mappings from $\mathbf{R}^n$ into $\mathbf{R}^m$. The *rank theorem* describes the behavior of f near w in this situation, and we shall return to this in Section 9.12.

9.9 Some remarks about determinants

Let n be a positive integer, and let $[a_{j,l}]$ be an $n \times n$ matrix of real or complex numbers. The *determinant* $\det[a_{j,l}]$ of $[a_{j,l}]$ can be defined as a real or complex number, as appropriate, in a standard way, as mentioned in Section 7.6. More precisely, $\det[a_{j,l}]$ is a polynomial of degree n in the entries $a_{j,l}$. In particular,

(9.9.1) $\det[a_{j,l}]$ is continuous

as a real or complex-valued function on the spaces $M_{n,n}(\mathbf{R})$, $M_{n,n}(\mathbf{C})$ of $n \times n$ matrices with entries in $\mathbf{R}$, $\mathbf{C}$, respectively, with respect to the metric associated to the Hilbert–Schmidt norm.

Now let A be a linear mapping from $\mathbf{R}^n$ or $\mathbf{C}^n$ into itself. Remember that A corresponds to an $n \times n$ matrix of real or complex numbers, as appropriate, as in Section 7.2. The *determinant* $\det A$ of A is defined to be the determinant of the corresponding matrix, as in Section 7.6 again. This defines continuous real and complex-valued functions on the spaces $\mathcal{L}(\mathbf{R}^n)$, $\mathcal{L}(\mathbf{C}^n)$ of linear mappings from $\mathbf{R}^n$, $\mathbf{C}^n$ into themselves, respectively, with respect to the metrics associated to the corresponding Hilbert–Schmidt norms. One could also use the metrics associated to the operator norms corresponding to the standard Euclidean norms on $\mathbf{R}^n$, $\mathbf{C}^n$, or to any other norms on $\mathbf{R}^n$, $\mathbf{C}^n$.

It is well known that a linear mapping A from $\mathbf{R}^n$ or $\mathbf{C}^n$ into itself is invertible if and only if

(9.9.2) $\det A \neq 0.$

The fact that the sets of invertible elements of $\mathcal{L}(\mathbf{R}^n)$, $\mathcal{L}(\mathbf{C}^n)$ are open sets with respect to the metrics associated to the Hilbert–Schmidt or operator norms can be obtained from this, and the continuity of the determinant.

If A is an invertible linear mapping on $\mathbf{R}^n$ or $\mathbf{C}^n$, then it is well known that the matrix corresponding to the inverse of A can be obtained from the determinant of A and the determinants of $(n-1) \times (n-1)$ submatrices of the matrix corresponding to A, as in *Cramer's rule*. This can be used to get the continuity of

$$A \mapsto A^{-1} \tag{9.9.3}$$

on the spaces of invertible elements of $\mathcal{L}(\mathbf{R}^n)$ or $\mathcal{L}(\mathbf{C}^n)$, with respect to the restrictions of the metrics associated to the Hilbert–Schmidt or operator norms to these spaces.

9.10 Some remarks about linear subspaces

In this section, we suppose again that the reader has some familiarity with linear algebra on Euclidean spaces, although some basic notions will also be reviewed here. In particular, if n is a positive integer, then a *linear subspace* of $\mathbf{R}^n$ is a subset V of $\mathbf{R}^n$ such that

$$u + v \in V \tag{9.10.1}$$

for every $u, v \in V$, and

$$t\, v \in V \tag{9.10.2}$$

for every $v \in \mathbf{R}^n$ and $t \in \mathbf{R}$. If $V \neq \{0\}$, then it is well known that V can be spanned by finitely many of its elements. The smallest number of elements of V that span V is known as the *dimension* of V, and is denoted $\dim V$. This is a nonnegative integer less than or equal to n, which is interpreted as being equal to 0 when $V = \{0\}$.

If V, W be linear subspaces of $\mathbf{R}^n$, then their *sum* may be defined as the subset of $\mathbf{R}^n$ given by

$$V + W = \{v + w : v \in V,\, w \in W\}. \tag{9.10.3}$$

It is easy to see that this is a linear subspace of $\mathbf{R}^n$ too. Note that

$$\dim(V + W) \le \dim V + \dim W. \tag{9.10.4}$$

It is easy to see that the intersection of V and W is a linear subspace of $\mathbf{R}^n$ as well. If

$$V \cap W = \{0\}, \tag{9.10.5}$$

then V and W may be said to be *transverse* as linear subspaces of $\mathbf{R}^n$. This means that the elements of $V + W$ can be expressed in a unique way as $v + w$, with $v \in V$ and $w \in W$. In this case,

$$\dim(V + W) = \dim V + \dim W, \tag{9.10.6}$$

because one can combine bases for V and W to get a basis for $V + W$. Conversely, it is not too difficult to show that (9.10.6) implies (9.10.5).

Let m be another positive integer, and let A be a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$. The *kernel* or *null space* is defined by

$$\ker A = \{v \in \mathbf{R}^n : A(v) = 0\}. \tag{9.10.7}$$

It is easy to see that this is also a linear subspace of $\mathbf{R}^n$. One can check that

$$\ker A = \{0\} \tag{9.10.8}$$

if and only if A is one-to-one on $\mathbf{R}^n$. Of course, $A(\mathbf{R}^m)$ is a linear subspace of $\mathbf{R}^m$, as mentioned in Section 9.8.

Suppose that $A \neq 0$, so that the dimension r of $A(\mathbf{R}^n)$ is positive. One can choose $v_1, \ldots, v_r \in \mathbf{R}^n$ such that

$$A(v_1), \ldots, A(v_r) \text{ is a basis for } A(\mathbf{R}^n), \tag{9.10.9}$$

as in Section 9.8. In this case, $v_1, \ldots, v_r$ are linearly independent in $\mathbf{R}^n$, as before. If V is the linear span of $v_1, \ldots, v_r$ in $\mathbf{R}^n$, then one can check that

$$V + \ker A = \mathbf{R}^n \tag{9.10.10}$$

and

$$V \cap (\ker A) = \{0\}. \tag{9.10.11}$$

This implies the well-known fact that

$$\dim \ker A + \dim A(\mathbf{R}^n) = n, \tag{9.10.12}$$

because $\dim A(\mathbf{R}^n) = r = \dim V$.

9.11 Complementary linear subspaces

Let n be a positive integer, and let V, W be linear subspaces of $\mathbf{R}^n$. We say that V and W are *complementary* linear subspaces of $\mathbf{R}^n$ if

$$V + W = \mathbf{R}^n \tag{9.11.1}$$

and (9.10.5) holds. This means that every element of $\mathbf{R}^n$ can be expressed in a unique way as a sum of elements of V and W, as before. Note that

$$\dim V + \dim W = n \tag{9.11.2}$$

in this case. If V and W are any two linear subspaces of $\mathbf{R}^n$ that satisfy (9.11.2), then one can check that (9.10.5) is equivalent to (9.11.1).

If V is any linear subspace of $\mathbf{R}^n$, then one can find a linear subspace W of $\mathbf{R}^n$ such that

$$V \text{ and } W \text{ are complementary linear subspaces of } \mathbf{R}^n. \tag{9.11.3}$$

One way to do this is to start with a basis for V, and extend it to a basis for $\mathbf{R}^n$. In this case, one can check that the linear span of the additional basis elements is complementary to V in $\mathbf{R}^n$.

A linear mapping P from $\mathbf{R}^n$ into itself is said to be a *projection* if

$$P \circ P = P \tag{9.11.4}$$

on $\mathbf{R}^n$. In this case, one can check that

(9.11.5) $\ker P$ and $P(\mathbf{R}^n)$ are complementary linear subspaces of $\mathbf{R}^n$.

One can also verify that

$$I - P \text{ is a projection on } \mathbf{R}^n \tag{9.11.6}$$

as well. In fact,

$$\ker(I - P) = P(\mathbf{R}^n) \tag{9.11.7}$$

and

$$(I - P)(\mathbf{R}^n) = \ker P. \tag{9.11.8}$$

If V and W are comlementary linear subspaces of $\mathbf{R}^n$, then there is a unique projection P on $\mathbf{R}^n$ such that

$$\ker P = V \tag{9.11.9}$$

and

$$P(\mathbf{R}^n) = W. \tag{9.11.10}$$

More precisely,

$$P(v + w) = w \tag{9.11.11}$$

for every $v \in V$ and $w \in W$. This can be used to define P on $\mathbf{R}^n$, because every element of $\mathbf{R}^n$ can be expressed in a unique way as a sum of elements of V and W, as before.

9.12 The rank theorem

Let m, n, and r be positive integers with

$$r \le \min(m, n). \tag{9.12.1}$$

Also let W be a nonempty open subset of $\mathbf{R}^n$, and let F be a continuously-differentiable mapping from W into $\mathbf{R}^m$. Suppose that

$$\text{the rank of } F'(w) \text{ is equal to } r \text{ for every } w \in W. \tag{9.12.2}$$

Let $a \in W$ be given, and put

$$A = F'(a). \tag{9.12.3}$$

Also put

$$Y_1 = A(\mathbf{R}^n), \tag{9.12.4}$$

which is a linear subspace of $\mathbf{R}^m$ of dimension r. Let Y_2 be a linear subspace of $\mathbf{R}^m$ such that

(9.12.5) Y_1 and Y_2 are complementary linear subspaces of $\mathbf{R}^m$, as in the previous section. This leads to a projection P on $\mathbf{R}^m$ such that

(9.12.6) $$P(\mathbf{R}^m) = Y_1$$

and

(9.12.7) $$\ker P = Y_2,$$

as before.

Under these conditions, there are open subsets U, V of $\mathbf{R}^n$ with the following properties. First,

(9.12.8) $$a \in U \text{ and } U \subseteq W.$$

Second, there is a one-to-one mapping H from V onto U such that

(9.12.9) $$H \text{ is continuously differentiable on } V,$$

(9.12.10) $$H^{-1} \text{ is continuously differentiable on } U,$$

and

(9.12.11) $$P(F(H(x))) = A(x)$$

for every $x \in V$. This is part of the *rank theorem*, as on p229 of [155]. More precisely, (9.12.11) corresponds to (71) on p230 of [155].

Of course, (9.12.2) implies that

(9.12.12) $$\text{the rank of } A = F'(a) \text{ is equal to } r.$$

If $w \in W$ is sufficiently close to a, then (9.12.12) implies that

(9.12.13) $$\text{the rank of } F'(w) \text{ is greater than or equal to } r,$$

because $F'(w)$ is close to $F'(a)$, as in Section 9.8. Remember that the rank of any linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$ is less than or equal to $\min(m, n)$. If $r = \min(m, n)$, then (9.12.12) implies that

(9.12.14) $$\text{the rank of } F'(w) \text{ is equal to } r \quad \text{when } w \in W \text{ is sufficiently close to } a.$$

In this case, we can get (9.12.2) by replacing W with a sufficiently small open set in $\mathbf{R}^n$ that contains a.

If $r = m$, then (9.12.12) is the same as saying that

(9.12.15) $$Y_1 = A(\mathbf{R}^n) = \mathbf{R}^m.$$

This implies that $Y_2 = \{0\}$, so that P is the identity mapping on $\mathbf{R}^m$. In this case, (9.12.11) means that

(9.12.16) $$F(H(x)) = A(x)$$

for every $x \in V$.

If $r = m = n$, then $A = F'(a)$ is invertible on $\mathbf{R}^n$, and the rank theorem is basically the same as the inverse function theorem. If $r = m < n$, then the rank theorem is closely related to the implicit function theorem.

If $r = n$, then (9.12.12) is the same as saying that

(9.12.17) $$A = F'(a) \text{ is injective on } \mathbf{R}^n.$$

If $r = n < m$, then the rank theorem is related to the remarks in Section 9.7.

9.13 Proving this part

Let us continue with the same notation and hypotheses as in the previous section. Note that

(9.13.1) $$\dim Y_1 = r,$$

by (9.12.4) and (9.12.12). Let $y_1, \ldots, y_r$ be a basis for Y_1.

Choose $z_j \in \mathbf{R}^n$ for each $j = 1, \ldots, r$ so that

(9.13.2) $$A(z_j) = y_j.$$

Let T be the unique linear mapping from Y_1 into $\mathbf{R}^n$ such that

(9.13.3) $$T(y_j) = z_j$$

for each $j = 1, \ldots, r$. Thus $A(T(y_j)) = A(z_j) = y_j$ for each $j = 1, \ldots, r$, so that

(9.13.4) $$A(T(y)) = y$$

for every $y \in Y_1$.

If $w \in W$, then put

(9.13.5) $$G(w) = w + T(P(F(w) - A(w))),$$

as in (69) on p229 of [155]. This defines a continuously-differentiable mapping from W into $\mathbf{R}^n$. It is easy to see that

(9.13.6) $$G'(a) = I_{\mathbf{R}^n},$$

the identity mapping on $\mathbf{R}^n$, because $A = F'(a)$. Thus the inverse function theorem implies that there are open subsets U and V of $\mathbf{R}^n$ such that U satisfies (9.12.8), and the restriction of G to U is a one-to-one mapping from U onto V whose inverse is continuously differentiable too. Let H be the inverse of the restriction of G to U, which is a one-to-one mapping from V onto U that satisfies (9.12.9) and (9.12.10).

Note that

(9.13.7) $$H'(x) \text{ is invertible for every } x \in V,$$

as in Section 9.3. We may also suppose that

(9.13.8) $$V \text{ is convex},$$

by replacing it with a convex open subset that contains $G(a)$, if necessary, and adjusting U appropriately.

Let us check that

(9.13.9) $$A \circ T \circ P \circ A = A,$$

as a linear mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$. More precisely,

(9.13.10) $$P \circ A = A,$$

as a linear mapping from $\mathbf{R}^n$ into $Y_1 = A(\mathbf{R}^n)$, because P is a projection of $\mathbf{R}^m$ onto Y_1, as in the previous section. We also have that $A \circ T$ is the identity mapping on Y_1, as in (9.13.4). This implies (9.13.9).

If $w \in W$, then

$$A(G(w)) = A(w) + A(T(P(F(w) - A(w)))), \tag{9.13.11}$$

by (9.13.5). This implies that

$$A(G(w)) = A(T(P(F(w)))), \tag{9.13.12}$$

because of (9.13.9). It follows that

$$A(G(w)) = P(F(w)), \tag{9.13.13}$$

as in (70) on p230 of [155]. This uses (9.13.4), and the fact that P maps $\mathbf{R}^m$ into Y_1.

It is easy to see that (9.12.11) follows from (9.13.13), by taking $w = H(x)$, $x \in V$. Some more properties of $F \circ H$ will be discussed in the next section.

9.14 Some more properties of $F \circ H$

We continue with the same notation and hypotheses as in the previous two sections. Put

$$\Phi(x) = F(H(x)) \tag{9.14.1}$$

for each $x \in V$, so that Φ is a continuously-differentiable mapping from V into $\mathbf{R}^m$. Of course,

$$\Phi'(x) = F'(H(x)) \circ H'(x) \tag{9.14.2}$$

for every $x \in V$, by the chain rule. This implies that

$$\text{the rank of } \Phi'(x) \text{ is equal to the rank of } F'(H(x)) \tag{9.14.3}$$

for every $x \in V$, because of (9.13.7). It follows that

$$\text{the rank of } \Phi'(x) \text{ is equal to } r \tag{9.14.4}$$

for every $x \in V$, because of (9.12.2), as in (75) on p230 of [155].

Observe that

$$P \circ \Phi = A \tag{9.14.5}$$

on V, by (9.12.11). This implies that

$$P \circ \Phi'(x) = A \tag{9.14.6}$$

for every $x \in V$, by the chain rule.

Let $x \in V$ be given, and put

$$M = (\Phi'(x))(\mathbf{R}^n) \tag{9.14.7}$$

which is a linear subspace of $\mathbf{R}^m$. Note that

$$\dim M = r, \tag{9.14.8}$$

by (9.14.4). We also have that

$$P(M) = A(\mathbf{R}^n) = Y_1, \tag{9.14.9}$$

using (9.14.6) in the first step, and the definition (9.12.4) of Y_1 in the second step. It follows that

$$P \text{ is one-to-one on } M, \tag{9.14.10}$$

because M and Y_1 have the same dimension r.

Suppose that $h \in \mathbf{R}^n$ satisfies

$$A(h) = 0. \tag{9.14.11}$$

This means that

$$P((\Phi'(x))(h)) = 0, \tag{9.14.12}$$

because of (9.14.6). Using (9.14.10), we get that

$$(\Phi'(x))(h) = 0, \tag{9.14.13}$$

because $h \in M$.

Suppose that $y \in V$ satisfies

$$A(y) = A(x). \tag{9.14.14}$$

We would like to show that

$$\Phi(x) = \Phi(y), \tag{9.14.15}$$

which corresponds to (74) on p230 of [155]. If $t \in \mathbf{R}$ and $0 \le t \le 1$, then

$$(1-t)\,x + t\,y \in V, \tag{9.14.16}$$

because of (9.13.8). In this case,

$$\frac{d}{dt}\Phi((1-t)\,x + t\,y) = (\Phi'((1-t)\,x + t\,y))(y-x) = 0, \tag{9.14.17}$$

using (9.14.13) in the second step, and the fact that $A(y-x) = 0$, by (9.14.14). This implies (9.14.15).

Put

$$\psi(x) = \Phi(x) - A(x) = F(H(x)) - A(x) \tag{9.14.18}$$

for every $x \in V$, as in (72) on p230 of [155]. This defines a continuously-differentiable mapping from V into $\mathbf{R}^m$ with

$$P(\psi(x)) = 0 \tag{9.14.19}$$

for every $x \in V$, because of (9.12.11), or equivalently (9.14.5). This means that

$$\psi(V) \subseteq \ker P = Y_2, \tag{9.14.20}$$

using (9.12.7) in the second step. Note that

$$\psi(x) = \psi(y) \tag{9.14.21}$$

for all $x, y \in V$ that satisfy (9.14.14), because of (9.14.15), as in (74) on p230 of [155].

One may prefer to consider $\psi(x)$ as a function of $A(x) \in A(\mathbf{R}^n) = Y_1$, as in [155].

Chapter 10

Product spaces and related matters

10.1 Products of metric spaces

Let $(X_1, d_{X_1}), \ldots, (X_n, d_{X_n})$ be finitely many metric spaces. Also let

$$X = X_1 \times X_2 \times \cdots \times X_n = \prod_{j=1}^{n} X_j \tag{10.1.1}$$

be the *Cartesian product* of $X_1, \ldots, X_n$, as sets. This is the set of n-tuples $x = (x_1, \ldots, x_n)$ with $x_j \in X_j$ for each $j = 1, \ldots, n$.

We would like to define suitable metrics on X using $d_{X_1}, \ldots, d_{X_n}$, and indeed there are various ways to do this, as usual. If $x, y \in X$, then put

$$d_{X,1}(x, y) = \sum_{j=1}^{n} d_{X_j}(x_j, y_j). \tag{10.1.2}$$

It is easy to see that this defines a metric on X.

Similarly, put

$$d_{X,\infty}(x, y) = \max_{1 \le j \le n} d_{X_j}(x_j, y_j) \tag{10.1.3}$$

for every $x, y \in X$. One can check that this defines a metric on X as well. More precisely, the triangle inequality for $d_{X,\infty}$ follows from the triangle inequality for the norm $\| \cdot \|_\infty$ on $\mathbf{R}^n$, as in Section 1.3.

Put

$$d_{X,2}(x, y) = \Big(\sum_{j=1}^{n} d_{X_j}(x_j, y_j)^2 \Big)^{1/2} \tag{10.1.4}$$

for each $x, y \in X$, using of course the nonnegative square root on the right side. One can verify that this satisfies the triangle inequality, using the triangle inequality for the standard Euclidean norm on $\mathbf{R}^n$. It follows that $d_{X,2}$ defines another metric on X.

If $n = 1$, then $X = X_1$, and $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ are all the same as d_{X_1}.

Suppose for the moment that $X_j = \mathbf{R}$ for each $j = 1, \ldots, n$, equipped with the standard Euclidean metric. In this case, X is the same as $\mathbf{R}^n$. The metrics $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ on X are the same as the metrics associated to the norms $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ on $\mathbf{R}^n$, respectively, as in Section 1.3.

Let (X_j, d_{X_j}) be any metric space for $j = 1, \ldots, n$ again. Observe that

$$d_{X,\infty}(x, y) \le d_{X,1}(x, y),\ d_{X,2}(x, y) \tag{10.1.5}$$

for every $x, y \in X$. We also have that

$$d_{X,2}(x, y) \le d_{X,1}(x, y) \tag{10.1.6}$$

for every $x, y \in X$, because the standard Euclidean norm on $\mathbf{R}^n$ is less than or equal to the norm $\|\cdot\|_1$, as in Section 1.5.

Similarly,

$$d_{X,2}(x, y) \le n^{1/2}\, d_{X,\infty}(x, y) \tag{10.1.7}$$

and

$$d_{X,1}(x, y) \le n\, d_{X,\infty}(x, y) \tag{10.1.8}$$

for every $x, y \in X$. As before, one can use the Cauchy–Schwarz inequality to get that

$$d_{X,1}(x, y) \le n^{1/2}\, d_{X,2}(x, y) \tag{10.1.9}$$

for every $x, y \in X$.

One can use these inequalities to get that $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ have many of the same properties on X, in essentially the same way as for the analogous metrics on $\mathbf{R}^n$. In particular, these metrics determine the same collections of open sets, closed sets, compact sets, and so on, in X. They also determine the same limit points of subsets of X, and the same convergent sequences and Cauchy sequences. Some of these and related properties will be discussed further in the next sections.

These metrics have many of the same properties in terms of continuity conditions for mappings between X and other metric spaces too. More precisely, the identity mapping on X is Lipschitz as a mapping from X equipped with any of these three metrics into X equipped with any other of these three metrics. This implies that the identity mapping on X is uniformly continuous and thus continuous as a mapping from X equipped with any of these three metrics into X equipped with any other of these three metrics.

10.2 Open and closed sets

Let $(X_1, d_{X_1}), \ldots, (X_n, d_{X_n})$ be finitely many metric spaces again, put $X = \prod_{j=1}^n X_j$, and let $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ be the metrics defined on X as in the previous section. Let $x = (x_1, \ldots, x_n) \in X$ and $r > 0$ be given, and for each $j = 1, \ldots, n$, let

$$B_{X_j, d_{X_j}}(x_j, r),\ \overline{B}_{X_j, d_{X_j}}(x_j, r) \tag{10.2.1}$$

be the open and closed balls in X_j centered at x_j with radius r with respect to d_{X_j}, respectively. Similarly, let

$$B_{X, d_{X,1}}(x, r),\ B_{X, d_{X,2}}(x, r),\ B_{X, d_{X,\infty}}(x, r) \tag{10.2.2}$$

and

$$\overline{B}_{X, d_{X,1}}(x, r),\ \overline{B}_{X, d_{X,2}}(x, r),\ \overline{B}_{X, d_{X,\infty}}(x, r) \tag{10.2.3}$$

be the open and closed balls in X centered at x with radius r with respect to $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$, respectively.

It is easy to see that

$$B_{X, d_{X,\infty}}(x, r) = \prod_{j=1}^{n} B_{X_j, d_{X_j}}(x_j, r) \tag{10.2.4}$$

and

$$\overline{B}_{X, d_{X,\infty}}(x, r) = \prod_{j=1}^{n} \overline{B}_{X_j, d_{X_j}}(x_j, r), \tag{10.2.5}$$

by the definition (10.1.3) of $d_{X,\infty}$. Using (10.1.5) and (10.1.8), we get that

$$B_{X, d_{X,1}}(x, r) \subseteq B_{X, d_{X,\infty}}(x, r) \subseteq B_{X, d_{X,1}}(x, n\, r) \tag{10.2.6}$$

and

$$\overline{B}_{X, d_{X,1}}(x, r) \subseteq \overline{B}_{X, d_{X,\infty}}(x, r) \subseteq \overline{B}_{X, d_{X,1}}(x, n\, r). \tag{10.2.7}$$

Similarly,

$$B_{X, d_{X,2}}(x, r) \subseteq B_{X, d_{X,\infty}}(x, r) \subseteq B_{X, d_{X,2}}(x, n^{1/2}\, r) \tag{10.2.8}$$

and

$$\overline{B}_{X, d_{X,2}}(x, r) \subseteq \overline{B}_{X, d_{X,\infty}}(x, r) \subseteq \overline{B}_{X, d_{X,2}}(x, n^{1/2}\, r), \tag{10.2.9}$$

by (10.1.5) and (10.1.7). We also have that

$$B_{X, d_{X,1}}(x, r) \subseteq B_{X, d_{X,2}}(x, r) \subseteq B_{X, d_{X,1}}(x, n^{1/2}\, r) \tag{10.2.10}$$

and

$$\overline{B}_{X, d_{X,1}}(x, r) \subseteq \overline{B}_{X, d_{X,2}}(x, r) \subseteq \overline{B}_{X, d_{X,1}}(x, n^{1/2}\, r), \tag{10.2.11}$$

by (10.1.6) and (10.1.9).

Let U_j be an open subset of X_j with respect to $d_{X,j}$ for each $j = 1, \ldots, n$. One can check that

$$U = \prod_{j=1}^{n} U_j \tag{10.2.12}$$

is an open set in X with respect to $d_{X,\infty}$, using (10.2.4). It follows that U is an open set with respect to $d_{X,1}$ and $d_{X,2}$ as well, by (10.2.6) and (10.2.8).

Let $A_j \subseteq X_j$ be given for each $j = 1, \ldots, n$, and put

$$A = \prod_{j=1}^{n} A_j. \tag{10.2.13}$$

One can verify that the closure of A in X with respect to $d_{X,\infty}$ is equal to

$$\prod_{j=1}^{n} \overline{A_j}, \tag{10.2.14}$$

where $\overline{A_j}$ is the closure of A_j in X_j with respect to d_{X_j} for every $j = 1, \ldots, n$. This is the same as the closure of A in X with respect to $d_{X,1}$ and $d_{X,2}$. In particular, if

$$A_j \text{ is a closed set in } X_j \tag{10.2.15}$$

with respect to d_{X_j} for each $j = 1, \ldots, n$, then

$$A \text{ is a closed set in } X \tag{10.2.16}$$

with respect to $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$.

10.3 Sequences and bounded sets

Let $(X_1, d_{X_1}), \ldots, (X_n, d_{X_n})$ be finitely many nonempty metric spaces, put $X = \prod_{j=1}^{n} X_j$, and let $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ be the metrics defined on X as in Section 10.1. Also let $\{x(l)\}_{l=1}^{\infty}$ be a sequence of elements of X, so that

$$x(l) = (x_1(l), \ldots, x_n(l)) \tag{10.3.1}$$

for each $l \geq 1$. One can check that

$$\{x(l)\}_{l=1}^{\infty} \text{ converges to } x = (x_1, \ldots, x_n) \in X \tag{10.3.2}$$

with respect to $d_{X,\infty}$ if and only if

$$\{x_j(l)\}_{l=1}^{\infty} \text{ converges to } x_j \text{ in } X_j \tag{10.3.3}$$

with respect to d_{X_j} for each $j = 1, \ldots, n$. This is equivalent to the convergence of $\{x(l)\}_{l=1}^{\infty}$ to x in X with respect to $d_{X,1}$, and with respect to $d_{X,2}$, as well.

Similarly,

$$\{x(l)\}_{l=1}^{\infty} \text{ is a Cauchy sequence in } X \tag{10.3.4}$$

with respect to $d_{X,\infty}$ if and only if

$$\{x_j(l)\}_{l=1}^{\infty} \text{ is a Cauchy sequence in } X_j \tag{10.3.5}$$

with respect to d_{X_j} for each $j = 1, \ldots, n$. This is equivalent to $\{x(l)\}_{l=1}^{\infty}$ being a Cauchy sequence in X with respect to $d_{X,1}$, and with respect to $d_{X,2}$. Observe that the completeness of X with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ are equivalent. The completeness of X with respect to these metrics is equivalent to the completeness of X_j with respect to d_{X_j} for each $j = 1, \ldots, n$.

Note that the boundedness of any subset of X with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$ implies the boundedness of the set with respect to the other two metrics. Let E_j be a nonempty subset of X_j for each $j = 1, \ldots, n$, and put

$$E = \prod_{j=1}^{n} E_j. \tag{10.3.6}$$

If

$$E_j \text{ is bounded in } X_j \tag{10.3.7}$$

with respect to d_{X_j} for each $j = 1, \ldots, n$, then it is easy to see that

$$E \text{ is bounded in } X \tag{10.3.8}$$

with respect to $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$. Conversely, if E is bounded in X with respect to $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$, then E_j is bounded in X_j with respect to d_{X_j} for each $j = 1, \ldots, n$.

Suppose that E_j is bounded in X_j with respect to d_{X_j} for each $j = 1, \ldots, n$, and let

$$\operatorname{diam}_{X_j, d_{X_j}} E_j \tag{10.3.9}$$

be the diameter of E_j as a subset of X_j with respect to d_{X_j} for every $j = 1, \ldots, n$, as in Section 4.1. One can verify that the diameter of E as a subset of X with respect to $d_{X,\infty}$ is given by

$$\operatorname{diam}_{X, d_{X,\infty}} E = \max_{1 \le j \le n} \big(\operatorname{diam}_{X_j, d_{X_j}} E_j\big). \tag{10.3.10}$$

Similarly, the diameter of E with respect to $d_{X,1}$ is given by

$$\operatorname{diam}_{X, d_{X,1}} E = \sum_{j=1}^{n} \operatorname{diam}_{X_j, d_{X_j}} E_j. \tag{10.3.11}$$

The diameter of E with respect to $d_{X,2}$ is given by

$$\operatorname{diam}_{X, d_{X,2}} E = \Big(\sum_{j=1}^{n} \big(\operatorname{diam}_{X_j, d_{X_j}} E\big)^2\Big)^{1/2}. \tag{10.3.12}$$

The total boundedness of any subset of X with respect to any of $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ implies the total boundedness of that set with respect to the other two metrics. If

$$E_j \text{ is totally bounded in } X_j \tag{10.3.13}$$

with respect to d_{X_j} for each $j = 1, \ldots, n$, then one can check that

$$E \text{ is totally bounded in } X \tag{10.3.14}$$

with respect to $d_{X,\infty}$.

More precisely, let $r > 0$ be given, and suppose that

$$E_j \text{ can be covered by } L_j(r) \text{ balls of radius } r \tag{10.3.15}$$

in X_j for each $j = 1, \ldots, n$. One can verify that E can be covered by

$$L(r) = \prod_{j=1}^{n} L_j(r) \tag{10.3.16}$$

balls of radius r in X with respect to $d_{X,\infty}$. This uses the products of the balls in the covering of X_j for each j.

This implies that E is totally bounded with respect to $d_{X,1}$ and $d_{X,2}$, as before. Conversely, if E is totally bounded in X with respect to $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$, then E_j is totally bounded in X_j with respect to d_{X_j} for each $j = 1, \ldots, n$.

10.4 Products of compact sets

Let $(X_1, d_{X_1}), \ldots, (X_n, d_{X_n})$ be finitely many metric spaces, put $X = \prod_{j=1}^{n} X_j$, and let $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ be the metrics defined on X as in Section 10.1. If a subset of X is compact with respect to $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$, then it is compact with respect to the other two metrics. This follows from the fact that an open subset of X with respect to $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$ is an open set with respect to the other two metrics.

Similarly, if a subset of X is sequentially compact with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$, then one can check directly that it is sequentially compact with respect to the other two metrics. This uses the fact that convergence of sequences of elements of X with respect to $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ are the same, as in the previous section.

Let K_j be a compact subset of X_j with respect to d_{X_j} for each $j = 1, \ldots, n$, and put

$$K = \prod_{j=1}^{n} K_j. \tag{10.4.1}$$

It is well known that

$$K \text{ is compact in } X, \tag{10.4.2}$$

with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$. This may be considered as a particular case of a famous theorem of Tychonoff for arbitrary topological spaces, instead of metric spaces. Let us mention a couple of other ways to look at this.

If K_j is sequentially compact in X_j with respect to d_{X_j} for each $j = 1, \ldots, n$, then

$$K \text{ is sequentially compact in } X \tag{10.4.3}$$

with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$. This can be seen using an argument like one in Section 5.1, when the set E considered there is

$\{1, \ldots, n\}$. Remember that compactness and sequential compactness are equivalent in metric spaces, as in Sections 4.5 and 4.6.

Alternatively, if

$$K_j \text{ is compact in } X_j \tag{10.4.4}$$

with respect to d_{X_j} for each $j = 1, \ldots, n$, then

$$K_j \text{ is closed and totally bounded in } X_j \tag{10.4.5}$$

with respect to d_{X_j} for every $j = 1, \ldots, n$. This implies that

$$K \text{ is closed and totally bounded in } X \tag{10.4.6}$$

with respect to each of $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$, as in the previous two sections. If X is complete with respect to $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$, then it follows that K is compact, as in Section 4.7. If

$$X_j \text{ is complete with respect to } d_{X_j} \tag{10.4.7}$$

for each $j = 1, \ldots, n$, then X is complete with respect to each of $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$, as in the previous section.

Note that one can reduce to the case where $K_j = X_j$ for each $j = 1, \ldots, n$, by standard results. It is well known that compact metric spaces are complete, as in Section 4.7. Thus the argument mentioned in the preceding paragraph can be used when $K_j = X_j$ for each $j = 1, \ldots, n$.

If $K_j \neq \emptyset$ for each $j = 1, \ldots, n$, and K is compact in X with respect to $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$, then it is not too difficult to show that K_j is compact in X_j with respect to d_{X_j} for each $j = 1, \ldots, n$.

10.5 Some mappings on products

Let $(X_1, d_{X_1}), \ldots, (X_n, d_{X_n})$ be finitely many nonempty metric spaces, put $X = \prod_{j=1}^n X_j$, and let $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ be the metrics defined on X as in Section 10.1. If $x = (x_1, \ldots, x_n) \in X$ and $1 \leq l \leq n$, then put

$$p_l(x) = x_l. \tag{10.5.1}$$

This defines a mapping p_l from X onto X_l.

Clearly

$$\begin{aligned} d_{X_l}(p_l(x), p_l(w)) = d_{X_l}(x_l, w_l) &\leq d_{X,\infty}(x, w) \\ &\leq d_{X,1}(x, w),\ d_{X,2}(x, w) \end{aligned} \tag{10.5.2}$$

for every $x, w \in X$. This means that p_l is Lipschitz with constant 1 with respect to $d_{X,\infty}$ on X, and thus with respect to $d_{X,1}$ and $d_{X,2}$ on X. In particular, p_l is uniformly continuous on X with respect to $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$.

If $x \in X$ and $r > 0$, then

$$p_l\big(B_{X,d_{X,\infty}}(x,r)\big) = p_l\Big(\prod_{j=1}^n B_{X_j,d_{X_j}}(x_j,r)\Big) = B_{X_l,d_{X_l}}(x_l,r), \tag{10.5.3}$$

using (10.2.4) in the first step. It follows that

$$p_l\big(B_{X,d_{X,1}}(x,r)\big) \supseteq p_l\big(B_{X,d_{X,\infty}}(x,r/n)\big) = B_{X_l,d_{X_l}}(x_l,r/n), \tag{10.5.4}$$

using (10.2.6) in the first step. Similarly,

$$p_l\big(B_{X,d_{X,2}}(x,r)\big) \supseteq p_l\big(B_{X,d_{X,\infty}}(x,r/n^{1/2})\big) = B_{X_l,d_{X_l}}(x_l,r/n^{1/2}), \tag{10.5.5}$$

using (10.2.8) in the first step. This shows that p_l is an open mapping from X onto X_l, with respect to $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$ on X, as in Section 7.14.

Let (Y, d_Y) be another metric space, let E be a subset of X, and let f be a mapping from E into Y. Let us say that f is *partially Lipschitz in the lth variable with constant $C_l \geq 0$ on E* if

$$d_Y(f(x), f(x')) \leq C_l\, d_{X_l}(x_l, x'_l) \tag{10.5.6}$$

for every $x, x' \in E$ such that $x_j = x'_j$ when $j \neq l$. This is analogous to the condition discussed in Section 8.3 for functions defined on subsets of $\mathbf{R}^n$.

Suppose that $E = \prod_{j=1}^n E_j$ for some $E_j \subseteq X_j$, $1 \leq j \leq n$, and that f is partially Lipschitz in the lth variable with constant $C_l \geq 0$ on E for each $l = 1, \ldots, n$. If $x, w \in E$, then

$$d_Y(f(x), f(w)) \leq \sum_{l=1}^n C_l\, d_{X_l}(x_l, w_l), \tag{10.5.7}$$

as in Section 8.3. This implies that

$$d_Y(f(x), f(w)) \leq \Big(\max_{1 \leq l \leq n} C_l\Big)\, d_{X,1}(x, w), \tag{10.5.8}$$

as before. Similarly,

$$d_Y(f(x), f(w)) \leq \Big(\sum_{l=1}^n C_l^2\Big)^{1/2} d_{X,2}(x, w) \tag{10.5.9}$$

and

$$d_Y(f(x), f(w)) \leq \Big(\sum_{l=1}^n C_l\Big)\, d_{X,\infty}(x, w). \tag{10.5.10}$$

This shows that f is Lipschitz on E with respect to the restrictions of $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ to E under these conditions.

Now let f be any mapping from X into Y. If f is continuous at a point $x \in X$ with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$ on X, then it is easy to see that f is continuous at x with respect to the other two metrics. This is often called *joint continuity* of f at x.

It is often convenient to consider f as a function of n variables in $X_1, \ldots, X_n$. If $1 \le l \le n$, then we may consider f as a function of the lth variable on X_l with values in Y, with the jth variable equal to x_j when $j \ne l$. If f is continuous as a function of the lth variable on X_l at x_l in this way for each $l = 1, \ldots, n$, then f is said to be *separately continuous* at x. Note that joint continuity at x implies separate continuity.

10.6 Uniform continuity

Let (X, d_X) and (Y, d_Y) be metric spaces, let f be a mapping from X into Y, and let A be a subset of X. Let us say that f is *uniformly continuous along* A if for every $\epsilon > 0$ there is a $\delta > 0$ such that for every $x \in A$ and $w \in X$ with $d_X(x, w) < \delta$, we have that

$$d_Y(f(x), f(w)) < \epsilon. \tag{10.6.1}$$

This implies that the restriction of f to A is uniformly continuous, with respect to the restriction of d_X to A. This also implies that f is continuous at every element of A, as a mapping from X into Y.

Suppose for the moment that A is a compact subset of X. If f is continuous at every element of A, as a function on X, then one can show that f is uniformly continuous along A. This uses the same type of arguments as used to show that continuous mappings on compact metric spaces are uniformly continuous.

Let (X_1, d_{X_1}), (X_2, d_{X_2}) be metric spaces, and let us now take

$$X = X_1 \times X_2. \tag{10.6.2}$$

We can define the metrics $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ on X using d_{X_1} and d_{X_2} as in Section 10.1, with $n = 2$. If a mapping f from X into Y is uniformly continuous along a set $A \subseteq X$ with respect to any of these three metrics on X, then it is easy to see that f is uniformly continuous along A with respect to the other two metrics.

Let $x_1 \in X_1$ be given, and consider

$$A = \{x_1\} \times X_2. \tag{10.6.3}$$

Let f be a mapping from X into Y again, and for each $x_2 \in X_2$, let us consider $f(\cdot, x_2)$ as a mapping from X_1 into Y. Let

$$\mathcal{E} = \{f(\cdot, x_2) : x_2 \in X_2\} \tag{10.6.4}$$

be the collection of these mappings from X_1 into Y.

If f is uniformly continuous along A with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, or $d_{X,3}$ on X, then one can check that $\mathcal{E}$ is equicontinuous at x_1, as a collection of mappings from X_1 into Y. As before, uniform continuity along A

implies that the restriction of f to A is uniformly continuous, which in this case means that $f(x_1, \cdot)$ is uniformly continuous as a mapping from X_2 into Y.

Conversely, suppose that $f(x_1, \cdot)$ is uniformly continuous as a mapping from X_2 into Y, and that $\mathcal{E}$ is equicontinuous at x_1 as a collection of mappings from X_1 into Y. Under these conditions, one can verify that f is uniformly continuous along A with respect to $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ on X.

Observe that

$$x_2 \mapsto (x_1, x_2) \tag{10.6.5}$$

defines an isometry from X_2 into X, with respect to $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ on X. In particular, if X_2 is compact, then it follows that A is compact in X, with respect to $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$. In this case, if f is continuous at every point in A as a mapping from X into Y, then f is uniformly continuous along A, as before.

10.7 Continuity and integration

Let (X_1, d_{X_1}) be a metric space, and let a, b be real numbers, with $a < b$. Let us take

$$X_2 = [a, b], \tag{10.7.1}$$

and d_{X_2} to be the restriction of the standard Euclidean metric on $\mathbf{R}$ to X_2. As in the previous section, we take

$$X = X_1 \times X_2 = X_1 \times [a, b], \tag{10.7.2}$$

and $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ to be the metrics defined on X as in Section 10.1, with $n = 2$.

Let f be a continuous real-valued function on X, with respect to any of $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$, and thus with respect to each of these metrics. If $x_1 \in X_1$, then $f(x_1, \cdot)$ is a continuous real-valued function on X_2, and we put

$$F(x_1) = \int_a^b f(x_1, x_2)\, dx_2. \tag{10.7.3}$$

This defines a real-valued function on X_1. Note that

$$|F(x_1)| \le \int_a^b |f(x_1, x_2)|\, dx_2 \le (b - a) \sup_{a \le x_2 \le b} |f(x_1, x_2)|. \tag{10.7.4}$$

More precisely, the supremum on the right is finite, because $[a, b]$ is compact as a subset of the real line, and thus as a subset of itself, with respect to the standard Euclidean metric on $\mathbf{R}$ and its restriction to $[a, b]$.

If $x_1 \in X_1$, then

$$\{x_1\} \times [a, b] = \{x_1\} \times X_2 \tag{10.7.5}$$

is compact in X, with respect to $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$, as in the previous section. This implies that f is uniformly continuous along (10.7.5), as before. This means that

$$\mathcal{E} = \{f(\cdot, x_2) : a \le x_2 \le b\} \tag{10.7.6}$$

is equicontinuous at x_1 as a collection of real-valued functions on X_1, as in the previous section.

One can use this to check that F is continuous at x_1, as a real-valued function on X_1. Indeed, if $w_1 \in X_1$, then

$$|F(x_1) - F(w_1)| = \left| \int_a^b (f(x_1, x_2) - f(w_1, x_2))\, dx_2 \right| \le \int_a^b |f(x_1, x_2) - f(w_1, x_2)|\, dx_2 \le (b - a) \sup_{a \le x_2 \le b} |f(x_1, x_2) - f(w_1, x_2)|, \tag{10.7.7}$$

as before. The right side can be made arbitrarily small by taking w_1 sufficiently close to x_1 in X_1, because of the equicontinuity of $\mathcal{E}$ at x_1.

Alternatively, let $\{x_{1,j}\}_{j=1}^\infty$ be a sequence of elements of X_1 that converges to x_1. It is easy to see that $f(x_{1,j}, \cdot)$ converges to $f(x_1, \cdot)$ uniformly as a sequence of real-valued functions on $[a, b]$, because of the equicontinuity of $\mathcal{E}$ at x_1. This implies that

$$\lim_{j \to \infty} F(x_{1,j}) = F(x_1). \tag{10.7.8}$$

It follows that F is continuous at x_1.

If $\mathcal{E}$ is uniformly equicontinuous on X_1, then it is easy to see that F is uniformly continuous on X_1, using (10.7.7). In particular, this holds when X_1 is compact, because $\mathcal{E}$ is equicontinuous at every point in X_1, as in Section 5.2.

If X_1 is compact, then X is compact with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$, as in Section 10.4, because X_2 is compact. In this case, f is uniformly continuous with respect to any of these three metrics on X. Of course, uniform continuity of f on X with respect to any of these three metrics implies that $\mathcal{E}$ is uniformly equicontinuous on X_1.

10.8 Iterated integrals

Let n be a positive integer, and let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be real numbers, with

$$a_j < b_j \tag{10.8.1}$$

for each $j = 1, \ldots, n$. Let us consider

$$X_j = [a_j, b_j] \tag{10.8.2}$$

as a metric space for each $j = 1, \ldots, n$, with d_{X_j} equal to the restriction of the standard Euclidean metric on $\mathbf{R}$ to X_j.

Put

$$X^l = \prod_{j=1}^{l} X_j \tag{10.8.3}$$

for each $l = 1, \ldots, n$. We can define metrics $d_{X^l,1}$, $d_{X^l,2}$, and $d_{X^l,\infty}$ on X^l, as in Section 10.1. We shall refer to functions on X^l as being continuous if they are continuous with respect to any of these three metrics, and thus with respect to the other two metrics.

Of course, X^l is a subset of $\mathbf{R}^l$ for each $l = 1, \ldots, n$. Note that $d_{X^l,1}$, $d_{X^l,2}$, and $d_{X^l,\infty}$ are the same as the restrictions to X^l of the metrics on $\mathbf{R}^l$ associated to the norms $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$, respectively, as in Section 1.3.

Let f be a continuous real-valued function on X^n, and put $f_n = f$. Suppose that f_l has been defined as a continuous real-valued function on X^l for some $l = 1, \ldots, n$. If $l \geq 2$ and $x_j \in X_j$ for $j = 1, \ldots, l-1$, then put

$$f_{l-1}(x_1, \ldots, x_{l-1}) = \int_{a_l}^{b_l} f_l(x_1, \ldots, x_{l-1}, x_l)\, dx_l. \tag{10.8.4}$$

This defines f_{l-1} as a continuous real-valued function on X^{l-1}, as in the previous section. More precisely, this uses the obvious identification of X^l with $X^{l-1} \times X_l$.

Continuing in this way, we define f_l on X^l for each $l = 1, \ldots, n$. Similarly, put

$$f_0 = \int_{a_1}^{b_1} f_1(x_1)\, dx_1, \tag{10.8.5}$$

which is the analogue of (10.8.4) with $l = 1$, and which is simply a real number. This can be used to define the n-fold *iterated integral* of f over X^n, as on p246 of [155].

Alternatively, one can define the integral of f over X^n as an n-dimensional Riemann integral. Any reasonable approach to this will give the same answer, because f is uniformly continuous on X^n, since X^n is compact.

Of course, one could also define n-fold iterated integrals of f over X^n by integrating the variables in a different order, and one would like to verify that this leads to the same result. One way to do this is to show that the iterated integrals are all the same as the corresponding n-dimensional Riemann integral. Another approach is given in Theorem 10.2 on p246 of [155].

10.9 Partitions of intervals

Let a, b be real numbers, with $a < b$. Suppose that $\mathcal{P} = \{t_j\}_{j=0}^k$ is a *partition* of $[a, b]$, which is to say a finite sequence of real numbers such that

$$a = t_0 < t_1 < \cdots < t_{k-1} < t_k = b. \tag{10.9.1}$$

A real-valued function on $[a, b]$ is said to be *piecewise linear with breakpoints in* $\mathcal{P}$ if the function is linear on $[t_{j-1}, t_j]$ for each $j = 1, \ldots, k$. It is easy to see that such a function is continuous, with respect to the standard Euclidean metric on $\mathbf{R}$, and its restriction to $[a, b]$. Note that such a function is uniquely determined by its values at the points in $\mathcal{P}$. The values of such a function at the points in $\mathcal{P}$ can be arbitrary real numbers. This is because linear functions on **R** are uniquely determined by their values at any two distinct points, and those values can be arbitrary real numbers.

If f is any real-valued function on $[a, b]$, then we let $A_{\mathcal{P}}(f)$ be the unique piecewise-linear function on $[a, b]$ with breakpoints in $\mathcal{P}$ that is equal to f at the points in $\mathcal{P}$. If $x \in [a, b]$ and $t_{j-1} \le x \le t_j$ for some $j = 1, \ldots, k$, then

$$(A_{\mathcal{P}}(f))(x) = f(t_{j-1})\,(t_j - t_{j-1})^{-1}\,(t_j - x) + f(t_j)\,(t_j - t_{j-1})^{-1}\,(x - t_{j-1}). \tag{10.9.2}$$

In this case,

$$\begin{aligned} f(x) - (A_{\mathcal{P}}(f))(x) &= (f(x) - f(t_{j-1}))\,(t_j - t_{j-1})^{-1}\,(t_j - x) \\ &\quad + (f(x) - f(t_j))\,(t_j - t_{j-1})^{-1}\,(x - t_{j-1}). \end{aligned} \tag{10.9.3}$$

This implies that

$$\begin{aligned} |f(x) - (A_{\mathcal{P}}(f))(x)| &\le |f(x) - f(t_{j-1})|\,(t_j - t_{j-1})^{-1}\,(t_j - x) \\ &\quad + |f(x) - f(t_j)|\,(t_j - t_{j-1})^{-1}\,(x - t_{j-1}). \end{aligned} \tag{10.9.4}$$

It follows that

$$|f(x) - (A_{\mathcal{P}}(f))(x)| \le \max(|f(x) - f(t_{j-1})|, |f(x) - f(t_j)|), \tag{10.9.5}$$

because

$$\begin{aligned} & (t_j - t_{j-1})^{-1}\,(t_j - x) + (t_j - t_{j-1})^{-1}\,(x - t_{j-1}) \\ &\quad = (t_j - t_{j-1})^{-1}\,(t_j - t_{j-1}) = 1, \end{aligned} \tag{10.9.6}$$

where both terms on the first line are greater than or equal to 0.

Observe that

$$\begin{aligned} \int_a^b (A_{\mathcal{P}}(f))(x)\,dx &= \sum_{j=1}^{k} \int_{t_{j-1}}^{t_j} (A_{\mathcal{P}}(f))(x)\,dx \\ &= \sum_{j=1}^{k} (1/2)\,(f(t_{j-1}) + f(t_j))\,(t_j - t_{j-1}). \end{aligned} \tag{10.9.7}$$

Suppose that f is continuous on $[a, b]$, with respect to the standard Euclidean metric on **R** and its restriction to $[a, b]$. Under these conditions, we have that

$$\begin{aligned} \left| \int_a^b f(x)\,dx - \int_a^b (A_{\mathcal{P}}(f))(x)\,dx \right| &= \left| \int_a^b (f(x) - (A_{\mathcal{P}}(f))(x))\,dx \right| \\ &\le \int_a^b |f(x) - (A_{\mathcal{P}}(f))(x)|\,dx \\ &\le (b - a) \sup_{a \le x \le b} |f(x) - (A_{\mathcal{P}}(f))(x)|. \end{aligned} \tag{10.9.8}$$

Note that f is uniformly continuous on $[a, b]$, because $[a, b]$ is compact. Using this and (10.9.5), we get that f is uniformly approximated by $A_{\mathcal{P}}(f)$ on $[a, b]$ when the partition $\mathcal{P}$ of $[a, b]$ is sufficiently fine. More precisely, this means that

$$\sup_{a \le x \le b} |f(x) - (A_{\mathcal{P}}(f))(x)| \tag{10.9.9}$$

is as small as we want when

$$\max_{1 \le j \le k} (t_j - t_{j-1}) \tag{10.9.10}$$

is sufficiently small. It follows that the left side of (10.9.8) is as small as we want when (10.9.10) is sufficiently small. Similarly, one can check that the left side of (10.9.8) is arbitrarily small for suitable partitions $\mathcal{P}$ of $[a, b]$ when f is Riemann integrable on $[a, b]$.

10.10 Partitions and products

Let (X_1, d_{X_1}) be a metric space, and let a_2, b_2 be real numbers, with $a_2 < b_2$. Let us take

$$X_2 = [a_2, b_2], \tag{10.10.1}$$

and d_{X_2} to be the restriction of the standard Euclidean metric on $\mathbf{R}$ to X_2. We also take

$$X = X_1 \times X_2 = X_1 \times [a_2, b_2], \tag{10.10.2}$$

and $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ to be the metrics defined on X as in Section 10.1, with $n = 2$. We shall refer to functions on X as being continuous if they are continuous with respect to any of these three metrics, and thus with respect to the other two metrics, as before.

Let $\mathcal{P}_2 = \{t_{2,j}\}_{j=0}^k$ be a partition of $[a_2, b_2]$, so that

$$a_2 = t_{2,0} < \cdots < t_{2,k} = b_2. \tag{10.10.3}$$

Let f be a continuous real-valued function on X, and let us define $A_{2,\mathcal{P}_2}(f)$ as a real-valued function on X in essentially the same way as in the previous section, as a function of the second variable. Thus, if $x_1 \in X_1$, $x_2 \in [a_2, b_2]$, and $t_{2,j-1} \le x_2 \le t_{2,j}$ for some $j = 1, \ldots, k$, then we put

$$\begin{aligned} (A_{2,\mathcal{P}_2}(f))(x_1, x_2) &= f(x_1, t_{2,j-1})\,(t_{2,j} - t_{2,j-1})^{-1}\,(t_{2,j} - x_2) \\ &\quad + f(x_1, t_{2,j})\,(t_{2,j} - t_{2,j-1})^{-1}\,(x_2 - t_{2,j-1}). \end{aligned} \tag{10.10.4}$$

Under these conditions, we get that

$$\begin{aligned} &|f(x_1, x_2) - (A_{2,\mathcal{P}_2}(f))(x_1, x_2)| \\ &\le \max(|f(x_1, x_2) - f(x_1, t_{2,j-1})|, |f(x_1, x_2) - f(x_1, t_{2,j})|), \end{aligned} \tag{10.10.5}$$

as in (10.9.5).

As in Section 10.7,

$$\int_{a_2}^{b_2} f(x_1, x_2)\, dx_2 \tag{10.10.6}$$

defines a continuous real-valued function of $x_1 \in X_1$. The analogous statement for $A_{2,\mathcal{P}_2}(f)$ can be seen more directly. Namely,

$$\begin{aligned} \int_{a_2}^{b_2} (A_{2,\mathcal{P}_2}(f))(x_1, x_2)\, dx_2 \\ = \sum_{j=1}^{k} (1/2)\,(f(x_1, t_{2,j-1}) + f(x_1, t_{2,j}))\,(t_{2,j} - t_{2,j-1}) \end{aligned} \tag{10.10.7}$$

for every $x_1 \in X_1$, as in (10.9.7). We also have that

$$\begin{aligned} \Big| \int_{a_2}^{b_2} f(x_1, x_2)\, dx_2 - \int_{a_2}^{b_2} (A_{2,\mathcal{P}_2}(f))(x_1, x_2)\, dx_2 \Big| \\ \le \int_{a_2}^{b_2} |f(x_1, x_2) - (A_{2,\mathcal{P}_2}(f))(x_1, x_2)|\, dx_2 \\ \le (b_2 - a_2) \sup_{a_2 \le x_2 \le b_2} |f(x_1, x_2) - (A_{2,\mathcal{P}_2}(f))(x_1, x_2)| \end{aligned} \tag{10.10.8}$$

for every $x_1 \in X_1$, as in (10.9.8).

Suppose that

$$\mathcal{E}_2 = \{f(x_1, \cdot) : x_1 \in X_1\} \tag{10.10.9}$$

is uniformly equicontinuous as a collection of real-valued functions on $[a_2, b_2]$. This implies that

$$\sup_{a_2 \le x_2 \le b_2} |f(x_1, x_2) - (A_{2,\mathcal{P}_2}(f))(x_1, x_2)| \tag{10.10.10}$$

is as small as we want, uniformly over $x_1 \in X_1$, when

$$\max_{1 \le j \le k} (t_{2,j} - t_{2,j-1}) \tag{10.10.11}$$

is sufficiently small, because of (10.10.5). It follows that (10.10.6) can be uniformly approximated by (10.10.7), as real-valued functions of $x_1 \in X_1$, when (10.10.11) is sufficiently small, by (10.10.8).

If X_1 is compact, then $\mathcal{E}_2$ is equicontinuous at every point in $[a_2, b_2]$, as in Section 10.6. This implies that $\mathcal{E}_2$ is uniformly equicontinuous on $[a_2, b_2]$, as in Section 5.2, because $[a_2, b_2]$ is compact. Alternatively, if X_1 is compact, then X is compact with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$, as in Section 10.4, because $[a_2, b_2]$ is compact. This means that f is uniformly continuous on X, with respect to any of these three metrics. It is easy to see that this implies that $\mathcal{E}_2$ is uniformly equicontinuous on $[a_2, b_2]$.

10.11 Partitions and integrals

Let a_1, b_1 be real numbers, with $a_1 < b_1$. We would like to continue with the same notation and hypotheses as in the previous section, with

$$X_1 = [a_1, b_1], \tag{10.11.1}$$

and with d_{X_1} equal to the restriction of the standard Euclidean metric on $\mathbf{R}$ to X_1. Thus

$$X = X_1 \times X_2 = [a_1, b_1] \times [a_2, b_2], \tag{10.11.2}$$

and $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ are the same as the restrictions to X of the metrics on $\mathbf{R}^2$ associated to the norms $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$, respectively, as in Section 1.3. Note that f is uniformly continuous on X, because X is compact.

We can integrate $f(x_1, x_2)$ in x_1 to get a continuous real-valued function

$$\int_{a_1}^{b_1} f(x_1, x_2)\, dx_1 \tag{10.11.3}$$

of x_2 on $[a_2, b_2]$, as in Section 10.7. The analogous statement for $A_{2,\mathcal{P}_2}(f)$ can be seen more directly, as before. Indeed, if $x_2 \in [a_2, b_2]$ and $t_{2,j-1} \le x_2 \le t_{2,j}$ for some $j = 1, \ldots, k$, then

$$\begin{aligned} \int_{a_1}^{b_1} (A_{2,\mathcal{P}_2}(f))(x_1, x_2)\, dx_1 \qquad \qquad & \\ = \ (t_{2,j} - t_{2,j-1})^{-1} \, (t_{2,j} - x_2) \int_{a_1}^{b_1} f(x_1, t_{2,j-1})\, dx_1 & \\ + (t_{2,j} - t_{2,j-1})^{-1} \, (x_2 - t_{2,j-1}) \int_{a_1}^{b_1} f(x_1, t_{2,j})\, dx_1, & \end{aligned} \tag{10.11.4}$$

by (10.10.4). Of course, this is equal to (10.11.3) when $x_2 = t_{2,j-1}$ or $t_{2,j}$.

Observe that

$$\begin{aligned} & \biggl| \int_{a_1}^{b_1} f(x_1, x_2)\, dx_1 - \int_{a_1}^{b_1} (A_{2,\mathcal{P}_2}(f))(x_1, x_2)\, dx_1 \biggr| \\ & = \biggl| \int_{a_1}^{b_1} (f(x_1, x_2) - (A_{2,\mathcal{P}_2}(f))(x_1, x_2))\, dx_1 \biggr| \\ & \le \int_{a_1}^{b_1} |f(x_1, x_2) - (A_{2,\mathcal{P}_2}(f))(x_1, x_2)|\, dx_1 \\ & \le (b_1 - a_1) \sup_{a_1 \le x_1 \le b_1} |f(x_1, x_2) - (A_{2,\mathcal{P}_2}(f))(x_1, x_2)| \end{aligned} \tag{10.11.5}$$

for every $x_2 \in [a_2, b_2]$. We also have that

$$\sup_{a_1 \le x_1 \le b_1} |f(x_1, x_2) - (A_{2,\mathcal{P}_2}(f))(x_1, x_2)| \tag{10.11.6}$$

is as small as we want, uniformly over $x_2 \in [a_2, b_2]$, when (10.10.11) is sufficiently small. Equivalently, this means that

$$\sup_{a_1 \le x_1 \le b_1} \sup_{a_2 \le x_2 \le b_2} |f(x_1, x_2) - (A_{2,\mathcal{P}_2}(f))(x_1, x_2)| \tag{10.11.7}$$

is as small as we want when (10.10.11) is sufficiently small. This is the same as the analogous statement for (10.10.10), which follows from the uniform continuity of f on X, as before. Combining this with (10.11.5), we get that (10.11.3) can be uniformly approximated by (10.11.4), as real-valued functions of $x_2 \in [a_2, b_2]$, when (10.10.11) is sufficiently small.

One can check directly that

$$\int_{a_1}^{b_1} \Big(\int_{a_2}^{b_2} (A_{2,\mathcal{P}_2}(f))(x_1, x_2)\, dx_2 \Big)\, dx_1 = \int_{a_2}^{b_2} \Big(\int_{a_1}^{b_1} (A_{2,\mathcal{P}_2}(f))(x_1, x_2)\, dx_1 \Big)\, dx_2. \tag{10.11.8}$$

More precisely, one can verify that

$$\int_{a_1}^{b_1} \Big(\int_{t_{2,j-1}}^{t_{2,j}} (A_{2,\mathcal{P}_2}(f))(x_1, x_2)\, dx_2 \Big)\, dx_1 = \int_{t_{2,j-1}}^{t_{2,j}} \Big(\int_{a_1}^{b_1} (A_{2,\mathcal{P}_2}(f))(x_1, x_2)\, dx_1 \Big) dx_2 \tag{10.11.9}$$

for each $j = 1, \ldots, k$, using the definition (10.10.4) of $A_{2,\mathcal{P}_2}(f)$. It is easy to see that (10.11.8) follows from (10.11.9), by summing over j.

We can use (10.11.8) to get that

$$\int_{a_1}^{b_1} \Big(\int_{a_2}^{b_2} f(x_1, x_2)\, dx_2 \Big)\, dx_1 = \int_{a_2}^{b_2} \Big(\int_{a_1}^{b_1} f(x_1, x_2)\, dx_1 \Big)\, dx_2, \tag{10.11.10}$$

as follows. The left side of (10.11.10) can be approximated by the left side of (10.11.8) when (10.10.11) is sufficiently small, because (10.10.6) is uniformly approximated by (10.10.7), as before. Similarly, the right side of (10.11.10) can be approximated by the right side of (10.11.8) when (10.10.11) is sufficiently small, because (10.11.3) is uniformly approximated by (10.11.4). This implies (10.11.10), because we can take $\mathcal{P}_2$ to be a partition of $[a_2, b_2]$ for which (10.10.11) is arbitarily small.

10.12 A simpler approximation

Let us continue with the same notation and hypotheses as in the previous two sections. Put

$$a_{2,\mathcal{P}_2}(j) = \sup_{a_1 \le x_1 \le b_1} \sup_{t_{2,j-1} \le x_2 \le t_{2,j}} |f(x_1, x_2) - f(x_1, t_{2,j})| \tag{10.12.1}$$

for each $j = 1, \ldots, k$, and

$$a_{2,\mathcal{P}_2} = \max_{1 \le j \le k} a_{2,\mathcal{P}_2}(j). \tag{10.12.2}$$

It is easy to see that $a_{2,\mathcal{P}_2}$ is as small as we want when (10.10.11) is sufficiently small, because f is uniformly continuous on X.

Observe that

$$\begin{aligned} &\int_{a_1}^{b_1} \Big(\int_{t_{2,j-1}}^{t_{2,j}} f(x_1, x_2)\, dx_2 \Big)\, dx_1 - (t_{2,j} - t_{2,j-1}) \int_{a_1}^{b_1} f(x_1, t_{2,j})\, dx_1 \\ &\quad = \int_{a_1}^{b_1} \Big(\int_{t_{2,j-1}}^{t_{2,j}} (f(x_1, x_2) - f(x_1, t_{2,j}))\, dx_2 \Big)\, dx_1 \end{aligned} \tag{10.12.3}$$

for each $j = 1, \ldots, k$. This implies that

$$\begin{aligned} &\Big| \int_{a_1}^{b_1} \Big(\int_{t_{2,j-1}}^{t_{2,j}} f(x_1, x_2)\, dx_2 \Big)\, dx_1 - (t_{2,j} - t_{2,j-1}) \int_{a_1}^{b_1} f(x_1, t_{2,j})\, dx_1 \Big| \\ &\quad \le \int_{a_1}^{b_1} \Big(\int_{t_{2,j-1}}^{t_{2,j}} |f(x_1, x_2) - f(x_1, t_{2,j})|\, dx_2 \Big)\, dx_1 \\ &\quad \le (b_1 - a_1)\,(t_{2,j} - t_{2,j-1})\, a_{2,\mathcal{P}_2}(j). \end{aligned} \tag{10.12.4}$$

Similarly,

$$\begin{aligned} &\Big| \int_{t_{2,j-1}}^{t_{2,j}} \Big(\int_{a_1}^{b_1} f(x_1, x_2)\, dx_1 \Big)\, dx_2 - (t_{2,j} - t_{2,j-1}) \int_{a_1}^{b_1} f(x_1, t_{2,j})\, dx_1 \Big| \\ &\quad \le \int_{t_{2,j-1}}^{t_{2,j}} \Big(\int_{a_1}^{b_1} |f(x_1, x_2) - f(x_1, t_{2,j})|\, dx_1 \Big)\, dx_2 \\ &\quad \le (b_1 - a_1)\,(t_{2,j} - t_{2,j-1})\, a_{2,\mathcal{P}_2}(j) \end{aligned} \tag{10.12.5}$$

for each $j = 1, \ldots, k$.

Using (10.12.3) and (10.12.5), we get that

$$\begin{aligned} &\Big| \int_{a_1}^{b_1} \Big(\int_{t_{2,j-1}}^{t_{2,j}} f(x_1, x_2)\, dx_2 \Big)\, dx_1 - \int_{t_{2,j-1}}^{t_{2,j}} \Big(\int_{a_1}^{b_1} f(x_1, x_2)\, dx_1 \Big)\, dx_2 \Big| \\ &\quad \le 2\,(b_1 - a_1)\,(t_{2,j} - t_{2,j-1})\, a_{2,\mathcal{P}_2}(j). \end{aligned} \tag{10.12.6}$$

It follows that

$$\begin{aligned} &\Big| \int_{a_1}^{b_1} \Big(\int_{a_2}^{b_2} f(x_1, x_2)\, dx_2 \Big)\, dx_1 - \int_{a_2}^{b_2} \Big(\int_{a_1}^{b_1} f(x_1, x_2)\, dx_1 \Big)\, dx_2 \Big| \\ &\quad \le \sum_{j=1}^{k} 2\,(b_1 - a_1)\,(t_{2,j} - t_{2,j-1})\, a_{2,\mathcal{P}_2}(j) \\ &\quad \le 2\,(b_1 - a_1) \sum_{j=1}^{k} (t_{2,j} - t_{2,j-1})\, a_{2,\mathcal{P}_2} \\ &\quad = 2\,(b_1 - a_1)\,(b_2 - a_2)\, a_{2,\mathcal{P}_2}. \end{aligned} \tag{10.12.7}$$

The right side of (10.12.7) is as small as we want when (10.10.11) is sufficiently small, as before. This is another way to obtain (10.11.10), by taking $\mathcal{P}_2$ to be a partition of $[a_2, b_2]$ for which (10.10.11) is arbitrarily small again.

10.13 Another approximation

Let us continue with the same notation and hypotheses as in the previous three sections. Also let $\mathcal{P}_1 = \{t_{1,l}\}_{l=0}^m$ be a partition of $[a_1, b_1]$, so that

$$a_1 = t_{1,0} < \cdots < t_{1,m} = b_1. \tag{10.13.1}$$

Of course,

$$\begin{aligned} &\int_{a_1}^{b_1} \Big(\int_{a_2}^{b_2} f(x_1, x_2)\, dx_2 \Big)\, dx_1 \\ &\quad = \sum_{l=1}^{m} \sum_{j=1}^{k} \int_{t_{1,l-1}}^{t_{1,l}} \Big(\int_{t_{2,j-1}}^{t_{2,j}} f(x_1, x_2)\, dx_2 \Big)\, dx_1. \end{aligned} \tag{10.13.2}$$

Similarly,

$$\begin{aligned} &\int_{a_2}^{b_2} \Big(\int_{a_1}^{b_1} f(x_1, x_2)\, dx_1 \Big)\, dx_2 \\ &\quad = \sum_{l=1}^{m} \sum_{j=1}^{k} \int_{t_{2,j-1}}^{t_{2,j}} \Big(\int_{t_{1,l-1}}^{t_{1,l}} f(x_1, x_2)\, dx_1 \Big)\, dx_2. \end{aligned} \tag{10.13.3}$$

If $1 \le l \le m$ and $1 \le j \le k$, then we can approximate

$$\int_{t_{1,l-1}}^{t_{1,l}} \Big(\int_{t_{2,j-1}}^{t_{2,j}} f(x_1, x_2)\, dx_2 \Big) dx_1 \tag{10.13.4}$$

and

$$\int_{t_{2,j-1}}^{t_{2,j}} \Big(\int_{t_{1,l-1}}^{t_{1,l}} f(x_1, x_2)\, dx_1 \Big)\, dx_2 \tag{10.13.5}$$

by

$$(t_{1,l} - t_{1,l-1})\, (t_{2,j} - t_{2,j-1})\, f(t_{1,l}, t_{2,j}). \tag{10.13.6}$$

If $t_{1,l} - t_{1,l-1}$ and $t_{2,j} - t_{2,j-1}$ are sufficiently small, then the errors in these approximations will be small compared to

$$(t_{1,l} - t_{1,l-1})\, (t_{2,j} - t_{2,j-1}), \tag{10.13.7}$$

because of the uniform continuity of f on X. This means that the difference of (10.13.4) and (10.13.5) is small compared to (10.13.7) in this case.

Note that

$$\begin{aligned}(10.13.8)\qquad & \sum_{l=1}^{m}\sum_{j=1}^{k}(t_{1,l}-t_{1,l-1})\,(t_{2,j}-t_{2,j-1}) \\ &= \Big(\sum_{l=1}^{m}(t_{1,l}-t_{1,l-1})\Big)\Big(\sum_{j=1}^{k}(t_{2,j}-t_{2,j-1})\Big) \\ &= (b_1-a_1)\,(b_2-a_2).\end{aligned}$$

It follows that the difference of (10.13.2) and (10.13.3) is as small as we like when $\mathcal{P}_1$ and $\mathcal{P}_2$ are sufficiently fine partitions of $[a_1, b_1]$ and $[a_2, b_2]$, respectively. This implies that the iterated integrals are equal, as in (10.11.10), because they do not depend on $\mathcal{P}_1$ or $\mathcal{P}_2$. Basically, we are approximating the iterated integrals on the left sides of (10.13.2) and (10.13.3) by two-dimensional Riemann sums, and these two-dimensional Riemann sums are equal to each other.

10.14 Partitions of unity

Let (X, d_X) be a metric space. As in Section 3.7, the support $\operatorname{supp} f$ of a real or complex-valued function f on X is defined to be the closure in X of the set of $x \in X$ such that $f(x) \neq 0$.

It is often helpful to be able to find finite collections of continuous real-valued functions $\psi_1, \ldots, \psi_r$ on X with properties like the following. First,

$$(10.14.1)\qquad 0 \le \psi_j(x) \le 1$$

for every $j = 1, \ldots, r$. Second,

$$(10.14.2)\qquad \sum_{j=1}^{r}\psi_j(x) = 1$$

for all x in a particular subset of X, which may be X itself. If (10.14.2) holds on a proper subset of X, then one may ask that

$$(10.14.3)\qquad \sum_{j=1}^{r}\psi_j(x) \le 1$$

for every $x \in X$. Of course, the first inequality in (10.14.1) together with (10.14.3) implies the second inequality in (10.14.1).

In addition, one may want to have restrictions on the supports of the ψ_j's. One may ask that the supports of the ψ_j's be contained in open sets in a particular family, for instance. A collection of functions like this is called a *partition of unity* on the set where (10.14.2) holds.

Let a, b be real numbers with $a < b$, and let $\mathcal{P} = \{t_j\}_{j=0}^{k}$ be a partition of $[a, b]$. If $0 \le l \le r$, then there is a unique nonnegative real-valued function τ_l on

$[a, b]$ that is piecewise linear with breakpoints in $\mathcal{P}$ such that

$$\begin{aligned} \tau_l(t_j) &= 1 \quad \text{when } j = l \\ &= 0 \quad \text{when } j \neq l. \end{aligned} \tag{10.14.4}$$

It is easy to see that

$$\sum_{l=0}^{k} \tau_l(x) = 1 \tag{10.14.5}$$

for every $x \in [a, b]$, because the left side is a piecewise-linear function on $[a, b]$ with breakpoints in $\mathcal{P}$ that is equal to 1 at t_j for each $j = 0, \ldots, k$. The support of τ_l, as a real-valued function on $[a, b]$, is given by

$$\begin{aligned} \operatorname{supp} \tau_l &= [t_0, t_1] \quad && \text{when } l = 0 \\ &= [t_{l-1}, t_{l+1}] \quad && \text{when } 1 \le l \le k-1 \\ &= [t_{k-1}, t_k] \quad && \text{when } l = k. \end{aligned} \tag{10.14.6}$$

Let (X, d_X) be any metric space again, and let $\phi_1, \ldots, \phi_r$ be nonnegative continuous real-valued functions on X. There are a couple of common ways to get a partition of unity $\psi_1, \ldots, \psi_r$ from $\phi_1, \ldots, \phi_r$, under suitable conditions.

In the first approach, we put

$$\Phi(x) = \sum_{j=1}^{r} \phi_j(x) \tag{10.14.7}$$

for each $x \in X$, and suppose that

$$\Phi(x) > 0 \quad \text{for every } x \in X. \tag{10.14.8}$$

In this case,

$$\psi_j(x) = \phi_j(x)/\Phi(x) \tag{10.14.9}$$

defines a continuous real-valued function on X for each $j = 1, \ldots, r$, and (10.14.2) holds for every $x \in X$, by construction. We also have that

$$\operatorname{supp} \psi_j = \operatorname{supp} \phi_j \tag{10.14.10}$$

for every $j = 1, \ldots, r$.

In the second approach, we suppose that

$$0 \le \phi_j(x) \le 1 \tag{10.14.11}$$

for every $j = 1, \ldots, r$ and $x \in X$. Put $\psi_1 = \phi_1$, and

$$\psi_l = \Big(\prod_{j=1}^{l-1} (1 - \phi_j) \Big) \phi_l \tag{10.14.12}$$

for $l = 2, \ldots, r$, as on p251 of [155]. Clearly (10.14.1) holds for every $j = 1, \ldots, r$, and one can check that

$$\sum_{j=1}^{l} \psi_j = 1 - \prod_{j=1}^{l} (1 - \phi_j) \tag{10.14.13}$$

for every $l = 1, \ldots, r$, by induction. This implies that (10.14.3) holds for every $x \in X$, and that (10.14.2) holds for every $x \in X$ such that

$$\phi_j(x) = 1 \text{ for some } j = 1, \ldots, r. \tag{10.14.14}$$

Note that

$$\operatorname{supp} \psi_j \subseteq \operatorname{supp} \phi_j \tag{10.14.15}$$

for every $j = 1, \ldots, r$.

10.15 Another approximation argument

Let (X_1, d_{X_1}), (X_2, d_{X_2}) be metric spaces, and put $X = X_1 \times X_2$. Also let $d_{X,1}$, $d_{X,2}$ and $d_{X,\infty}$ be the corresponding metrics defined on X as in Section 10.1. As before, we shall refer to functions on X as being continuous if they are continuous with respect to any of these three metrics, and thus with respect to the other two metrics.

Let $\psi_{2,1}, \ldots, \psi_{2,r}$ be finitely many continuous nonnegative real-valued functions on X_2 that form a partition of unity on X_2, so that

$$\sum_{j=1}^{r} \psi_{2,j}(x_2) = 1 \tag{10.15.1}$$

for every $x_2 \in X_2$. If f is a real-valued function on X, then

$$f(x_1, x_2) = \sum_{j=1}^{r} f(x_1, x_2)\, \psi_{2,j}(x_2) \tag{10.15.2}$$

for every $x_1 \in X_1$ and $x_2 \in X_2$.

Suppose that for each $j = 1, \ldots, r$, $x_{2,j} \in X_2$ and

$$\psi_{2,j}(x_{2,j}) > 0. \tag{10.15.3}$$

Put

$$(A_2(f))(x_1, x_2) = \sum_{j=1}^{r} f(x_1, x_{2,j})\, \psi_{2,j}(x_2) \tag{10.15.4}$$

for every $x_1 \in X_1$ and $x_2 \in X_2$. Thus

$$f(x_1, x_2) - (A_2(f))(x_1, x_2) = \sum_{j=1}^{r} (f(x_1, x_2) - f(x_1, x_{2,j}))\, \psi_{2,j}(x_2) \tag{10.15.5}$$

for every $x_1 \in X_1$ and $x_2 \in X_2$, by (10.15.2). It follows that

$$|f(x_1, x_2) - (A_2(f))(x_1, x_2)| \le \sum_{j=1}^{r} |f(x_1, x_2) - f(x_1, x_{2,j})| \, \psi_{2,j}(x_2) \tag{10.15.6}$$

for every $x_1 \in X_1$ and $x_2 \in X_2$.

Put

$$a_{2,j}(x_1) = \sup\{|f(x_1, x_2) - f(x_1, x_{2,j})| : x_2 \in X_2, \, \psi_{2,j}(x_2) > 0\} \tag{10.15.7}$$

for each $j = 1, \ldots, r$. The right side is allowed to be $+\infty$ here, although we shall typically be concerned with situations where it is finite. Observe that

$$|f(x_1, x_2) - f(x_1, x_{2,j})| \, \psi_{2,j}(x_2) \le a_{2,j}(x_1) \, \psi_{2,j}(x_2) \tag{10.15.8}$$

for every $x_1 \in X_1$ and $x_2 \in X_2$, where the right side may be interpreted as being equal to 0 when $\psi_{2,j}(x_2) = 0$, even if (10.15.7) is $+\infty$. This implies that

$$|f(x_1, x_2) - (A_2(f))(x_1, x_2)| \le \sum_{j=1}^{r} a_{2,j}(x_1) \, \psi_{2,j}(x_2) \tag{10.15.9}$$

for every $x_1 \in X_1$ and $x_2 \in X_2$, by (10.15.6).

Put

$$a_2(x_1) = \max_{1 \le j \le r} a_{2,j}(x_1) \tag{10.15.10}$$

for each $x_1 \in X_1$. Using (10.15.9), we get that

$$|f(x_1, x_2) - (A_2(f))(x_1, x_2)| \le a_2(x_1) \tag{10.15.11}$$

for every $x_1 \in X_1$, because of (10.15.1).

We may frequently be interested in situations where X_2 is compact, or at least totally bounded, and

$$\{x_2 \in X_2 : \psi_{2,j}(x_2) > 0\} \tag{10.15.12}$$

is a small subset of X_2 for each $j = 1, \ldots, r$. This may mean that (10.15.12) is contained in a ball in X_2 centered at $x_{2,j}$ with small radius, or that (10.15.12) has small diameter with respect to d_{X_2}, which is nearly the same thing. If (10.15.12) is sufficiently small in this way, then we may be able to show that (10.15.7) is small, using suitable continuity properties of f.

Suppose that $\mathcal{E}_2 = \{f(x_1, \cdot) : x_1 \in X_1\}$ is uniformly equicontinuous as a collection of real-valued functions on X_2. In this case, (10.15.10) is as small as we like when (10.15.12) is contained in a ball centered at $x_{2,j}$ of sufficiently small radius for each $j = 1, \ldots, r$, or the diameter of (10.15.12) is sufficiently small for each $j = 1, \ldots, r$, which is almost the same thing, as before.

If X_1 is compact, then $\mathcal{E}_2$ is equicontinuous at every $x_2 \in X_2$, as in Section 10.6. If X_2 is compact, then the equicontinuity of $\mathcal{E}_2$ at every point in X_2 implies that $\mathcal{E}_2$ is uniformly equicontinuous on X_2, as in Section 5.2. If f is uniformly continuous on X with respect to $d_{X,1}$, $d_{X,2}$, or $d_{X,3}$, then it is easy to see that $\mathcal{E}_2$ is uniformly equicontinuous on X_2. If X_1 and X_2 are compact, then X is compact with respect to $d_{X,1}$, $d_{X,2}$, and $d_{X,3}$, as in Section 10.4. In this case, continuity of f on X implies uniform continuity, as usual.

10.16 Some continuous functions

If c is any real number, then one can check that

$$\max(t, c) \text{ and } \min(t, c) \tag{10.16.1}$$

are Lipschitz functions of $t \in \mathbf{R}$, with constant 1. Of course, this uses the standard Euclidean metric on the real line.

Let (X, d_X) be a metric space, and let A be a nonempty subset of X. If $x \in X$, then the distance from x to A with respect to d_X is defined by

$$\operatorname{dist}(x, A) = \inf\{d(x, a) : a \in A\}. \tag{10.16.2}$$

One can check that

$$\operatorname{dist}(x, A) = 0 \tag{10.16.3}$$

if and only if x is an element of the closure $\overline{A}$ of A in X. One can also verify that

$$\operatorname{dist}(x, \overline{A}) = \operatorname{dist}(x, A) \tag{10.16.4}$$

for every $x \in X$.

It is easy to see that

$$\operatorname{dist}(x, A) \leq d(x, w) + \operatorname{dist}(w, A) \tag{10.16.5}$$

for every $x, w \in X$. Using this, one can check that $\operatorname{dist}(x, A)$ is Lipschitz with constant 1, as a real-valued function on X.

Let A, B be nonempty disjoint closed subsets of X. Observe that

$$\operatorname{dist}(x, A) + \operatorname{dist}(x, B) > 0 \tag{10.16.6}$$

for every $x \in X$. It follows that

$$\frac{\operatorname{dist}(x, A)}{\operatorname{dist}(x, A) + \operatorname{dist}(x, B)} \tag{10.16.7}$$

defines a continuous real-valued function on X. This function is equal to 0 exactly on A, and it is equal to 1 exactly on B. It also takes values in $[0, 1]$ on X, by construction.

Let $r > 0$ be given, and note that

$$\min(\operatorname{dist}(x, A), r) \tag{10.16.8}$$

is Lipschitz with constant 1, as a real-valued function of $x \in X$. This implies that

$$(1/r) \min(\operatorname{dist}(x, A), r) \tag{10.16.9}$$

is Lipschitz with constant $1/r$ on X.

Suppose for the moment that

$$d(a, b) \geq r \tag{10.16.10}$$

for every $a \in A$ and $b \in B$. Equivalently, this means that

$$\text{dist}(b, A) \geq r \tag{10.16.11}$$

for every $b \in B$. This implies that (10.16.8) is equal to r when $x \in B$, so that (10.16.9) is equal to 1.

Conversely, suppose that there is a uniformly continuous real-valued function f on X such that

$$f(a) = 0 \text{ for every } a \in A, \text{ and } f(b) = 1 \text{ for every } b \in B. \tag{10.16.12}$$

Under these conditions, it is easy to see that there is an $r > 0$ such that (10.16.10) holds for every $a \in A$ and $b \in B$.

Let $a_0 \in X$ and $r_0 > 0$ be given, and note that

$$d(x, a_0) \tag{10.16.13}$$

is Lipschitz with constant 1 as a real-valued function of $x \in X$. This is the same as (10.16.2), with $A = \{a_0\}$. It follows that

$$r_0 - d(x, a_0) \tag{10.16.14}$$

is Lipschitz with constant 1 on X, so that

$$\max(r_0 - d(x, a_0), 0) \tag{10.16.15}$$

is Lipschitz with constant 1 on X as well. Of course, (10.16.15) is equal to 0 when

$$d(x, a_0) \geq r_0, \tag{10.16.16}$$

and otherwise (10.16.15) is strictly positive.

Let r_1 be a nonnegative real number with

$$r_1 < r_0, \tag{10.16.17}$$

and consider

$$\min(\max(r_0 - d(x, a_0), 0), r_0 - r_1). \tag{10.16.18}$$

This is another real-valued Lipschitz function on X with constant 1, so that

$$(1/(r_0 - r_1)) \min(\max(r_0 - d(x, a_0), 0), r_0 - r_1) \tag{10.16.19}$$

is Lipschitz with constant $1/(r_0 - r_1)$ on X. This function is equal to 0 when (10.16.16) holds, and it is equal to 1 when

$$d(x, a_0) \leq r_1. \tag{10.16.20}$$

If

$$r_1 < d(x, a_0) < r_0, \tag{10.16.21}$$

then (10.16.19) is in $(0, 1)$.

10.17 Products and reciprocals

Let (X, d_X) be a metric space, and let f, g be real-valued functions on X. If f and g are uniformly continuous on X, then it is easy to see that $f + g$ is uniformly continuous on X too. If f and g are Lipschitz functions on X, then $f + g$ is a Lipschitz function on X as well, as in Section 7.7.

Suppose for the moment that f and g are bounded on X. If f and g are also uniformly continuous on X, then one can check that their product $f\,g$ is uniformly continuous on X. Similarly, if f and g are Lipschitz functions on X, then one can verify that $f\,g$ is Lipschitz on X.

If t_1, t_2 are nonzero real numbers, then

$$1/t_1 - 1/t_2 = (t_2 - t_1)/(t_1\,t_2). \tag{10.17.1}$$

This implies that

$$|1/t_1 - 1/t_2| = |t_1 - t_2|/(|t_1|\,|t_2|). \tag{10.17.2}$$

If r is a positive real number, and $|t_1|, |t_2| \geq r$, then we get that

$$|1/t_1 - 1/t_2| \leq r^{-2}\,|t_1 - t_2|. \tag{10.17.3}$$

This means that $t \mapsto 1/t$ is Lipschitz with constant r^{-2} as a real-valued function on

$$\{t \in \mathbf{R} : |t| \geq r\}. \tag{10.17.4}$$

Of course, this uses the standard Euclidean metric on $\mathbf{R}$, and its restriction to (10.17.4).

Suppose now that

$$|f(x)| \geq r \tag{10.17.5}$$

for every $x \in X$. If f is uniformly continuous on X, then it is easy to see that

$$1/f \text{ is uniformly continuous on } X. \tag{10.17.6}$$

Similarly, if f is Lipschitz on X, then

$$1/f \text{ is Lipschitz on } X. \tag{10.17.7}$$

These statements can be verified directly, or by considering $1/f$ as the composition of f with $t \mapsto 1/t$ on (10.17.4). There are analogous statements for complex numbers, and complex-valued functions on X.

10.18 Graphs of mappings

Let X_1, X_2 be nonempty sets, and let $X = X_1 \times X_2$ be their Cartesian product. If f is a mapping from X_1 into X_2, then the *graph* of f is the subset of X given by

$$\{(x_1, f(x_1)) : x_1 \in X_1\}, \tag{10.18.1}$$

as usual. Put

$$F(x_1) = (x_1, f(x_1)) \tag{10.18.2}$$

for every $x_1 \in X_1$, which defines F as a mapping from X_1 into X. Note that (10.18.1) is the same as the image of X_1 under F.

Let p_1, p_2 be the usual coordinate projections from X onto X_1, X_2, respectively, as in Section 10.5. Thus $p_j(x) = x_j$ for each $x = (x_1, x_2) \in X$ and $j = 1, 2$. Clearly

$$f = p_2 \circ F \tag{10.18.3}$$

on X_1, and $p_1 \circ F$ is the identity mapping on X_1.

Suppose now that (X_1, d_{X_1}) and (X_2, d_{X_2}) are metric spaces, so that we can define the metrics $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ on X as in Section 10.1. Remember that these metrics determine the same collections of open sets, closed sets, and compact sets in X, and that convergence of sequences in X with respect to these metrics are equivalent, as before.

If

$$f \text{ is continuous as a mapping from } X_1 \text{ into } X_2, \tag{10.18.4}$$

then one can check that

$$F \text{ is continuous as a mapping from } X_1 \text{ into } X, \tag{10.18.5}$$

with respect to each of $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ on X. Conversely, if (10.18.5) holds, with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$ on X, then (10.18.4) holds. This can be seen using (10.18.3), and the fact that p_2 is a continuous mapping from X onto X_2 with respect to each of $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ on X, as in Section 10.5.

One can check that a subset E of a metric space M is a closed set if and only if for every sequence $\{w_j\}_{j=1}^\infty$ of elements of E that converges to an element w of M, we have that

$$w \in E. \tag{10.18.6}$$

More precisely, if $\{w_j\}_{j=1}^\infty$ is a sequence of elements of any set $E \subseteq M$ that converges to an element w of M, and if $w_j \neq w$ for each j, then w is a limit point of E. If $w \in M$ is a limit point of E, then one can find a sequence $\{w_j\}_{j=1}^\infty$ of elements of E that converges to w, with $w_j \neq w$ for each j.

It is well known that if (10.18.4) holds, then

$$\text{the graph of } f \text{ is a closed set in } X, \tag{10.18.7}$$

with respect to each of the metrics $d_{X,1}$, $d_{X,2}$, $d_{X,\infty}$. As in the preceding paragraph, (10.18.7) holds if and only if for every sequence of elements of the graph of f that converges to an element of X, the limit of the sequence is in the graph of f. In this case, this means that (10.18.7) holds if and only if for every sequence $\{x_{1,j}\}_{j=1}^\infty$ of elements of X_1 such that

$$\{(x_{1,j}, f(x_{1,j}))\}_{j=1}^\infty \tag{10.18.8}$$

converges to an element (x_1, x_2) of X, we have that

$$f(x_1) = x_2. \tag{10.18.9}$$

As in Section 10.3, the convergence of (10.18.8) to $(x_1, x_2) \in X$ with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, $d_{X,\infty}$ is equivalent to the convergence of $\{x_{1,j}\}_{j=1}^\infty$ to x_1 in X_1 and the convergence of $\{f(x_{1,j})\}_{j=1}^\infty$ to x_2 in X_2. If f is continuous at x_1, then the convergence of $\{x_{1,j}\}_{j=1}^\infty$ to x_1 in X_1 implies that $\{f(x_{1,j})\}_{j=1}^\infty$ converges to $f(x_1)$ in X_2. This implies that (10.18.9) holds when $\{f(x_{1,j})\}_{j=1}^\infty$ converges to x_2 in X_2.

If (10.18.4) holds, and if

$$X_1 \text{ is compact,} \tag{10.18.10}$$

then

$$\text{the graph of } f \text{ is a compact subset of } X, \tag{10.18.11}$$

with respect to each of $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$. Remember that the graph of f is the same as $F(X_1)$, and that F is continuous as a mapping from X_1 into X in this case. This implies that $F(X_1)$ is a compact subset of X when X_1 is compact.

Conversely, let f be any mapping from X_1 into X_2, and suppose that (10.18.11) holds, with respect to $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$. It is easy to see that (10.18.10) holds in this case, because p_1 maps the graph of f onto X_1. This also use the fact that p_1 is continuous as a mapping from X onto X_1, as in Section 10.5.

It is well known that (10.18.4) holds under these conditions as well. To see this, let $x_1 \in X_1$ be given, and suppose for the sake of a contradiction that f is not continuous at x_1. This means that there is an $\epsilon > 0$ such that for every $\delta > 0$ there is a point $w_1 \in X_1$ with

$$d_{X_1}(x_1, w_1) < \delta \tag{10.18.12}$$

and

$$d_{X_2}(f(x_1), f(w_1)) \ge \epsilon. \tag{10.18.13}$$

One can use this to get a sequence $\{x_{1,j}\}_{j=1}^\infty$ of elements of X_1 that converges to x_1, with

$$d_{X_2}(f(x_1), f(x_{1,j})) \ge \epsilon \tag{10.18.14}$$

for every j.

If (10.18.11) holds, then the graph of f is sequentially compact in X. Using this, we get that there is a subsequence $\{x_{1,j_l}\}_{l=1}^\infty$ of $\{x_{1,j}\}_{j=1}^\infty$ such that

$$\{(x_{1,j_l}, f(x_{1,j_l}))\}_{l=1}^\infty \tag{10.18.15}$$

converges to an element $(y_1, f(y_1))$ of the graph of f in X. This means that

$$\{x_{1,j_l}\}_{l=1}^\infty \text{ converges to } y_1 \text{ in } X_1, \tag{10.18.16}$$

and that

$$\{f(x_{1,j_l})\}_{l=1}^{\infty} \text{ converges to } f(y_1) \text{ in } X_2, \tag{10.18.17}$$

as in Section 10.3. Note that $\{x_{1,j_l}\}_{l=1}^{\infty}$ converges to x_1 in X_1, because $\{x_{1,j}\}_{j=1}^{\infty}$ converges to x_1, by construction. Thus

$$x_1 = y_1. \tag{10.18.18}$$

It follows that

$$\{f(x_{1,j_l})\}_{l=1}^{\infty} \text{ converges to } f(x_1) \text{ in } X_2. \tag{10.18.19}$$

This contradicts (10.18.14), as desired.

10.19 Semicontinuity

Let us continue with the same notation and hypotheses as in the previous section, except that now we take $X_2 = \mathbf{R}$, with the standard Euclidean metric. If f is a real-valued function on X_1, then

$$\{(x_1, x_2) \in X_1 \times \mathbf{R} : f(x_1) > x_2\} \tag{10.19.1}$$

and

$$\{(x_1, x_2) \in X_1 \times \mathbf{R} : f(x_1) \le x_2\} \tag{10.19.2}$$

are complementary subsets of $X = X_1 \times \mathbf{R}$. In particular, (10.19.1) is an open set in $X = X_1 \times \mathbf{R}$ if and only if (10.19.2) is a closed set in X, with respect to any, and thus each, of the metrics $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$.

Similarly,

$$\{(x_1, x_2) \in X_1 \times \mathbf{R} : f(x_1) \ge x_2\} \tag{10.19.3}$$

and

$$\{(x_1, x_2) \in X_1 \times \mathbf{R} : f(x_1) < x_2\} \tag{10.19.4}$$

are complementary subsets of X. It follows that (10.19.3) is a closed set in X if and only if (10.19.4) is an open set, as before. Note that the graph of f is the same as the intersection of (10.19.2) and (10.19.3).

If f is continuous on X, then one can show that (10.19.2) and (10.19.3) are closed sets in X, using the same type of argument as in the previous section. However, there are more precise statements, using notions of *semicontinuity*.

We say that f is *upper semicontinuous* at a point $x_1 \in X_1$ if for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$f(w_1) < f(x_1) + \epsilon \tag{10.19.5}$$

for every $w_1 \in X_1$ with $d_{X_1}(x_1, w_1) < \delta$. Similarly, we say that f is *lower semicontinuous* at x_1 if for every $\epsilon > 0$ there is a $\delta' > 0$ such that

$$f(w_1) > f(x_1) - \epsilon \tag{10.19.6}$$

for every $w_1 \in X_1$ with $d_{X_1}(x_1, w_1) < \delta'$.

Observe that f is continuous at x_1 if and only if f is both upper and lower semicontinuous at x_1. It is easy to see that f is upper semicontinuous at x_1 if and only if $-f$ is lower semicontinuous at x_1.

Let us say that f is *upper semicontinuous* on X_1 if f is upper semicontinuous at every $x_1 \in X_1$. Similarly, we say that f is *lower semicontinuous* on X_1 if f is lower semicontinuous at every $x_1 \in X_1$.

One can check that f is upper semicontinuous on X_1 if and only if for every real number b,

$$\{x_1 \in X_1 : f(x_1) < b\} \tag{10.19.7}$$

is an open subset of X_1. Similarly, f is lower semicontinuous on X_1 if and only if for every $a \in \mathbf{R}$,

$$\{x_1 \in X_1 : f(x_1) > a\} \tag{10.19.8}$$

is an open set in X_1.

One can also verify that f is upper semicontinuous on X_1 if and only if (10.19.4) is an open set in X. Similarly, f is lower semicontinuous on X_1 if and only if (10.19.1) is an open set in X.

Let K be a nonempty compact subset of X_1. If f is upper semi-continuous on X_1, then it is not too difficult to show that f attains its maximum on K. Similarly, if f is lower semicontinuous on X_1, then f attains its minimum on K. Of course, this is another version of the *extreme value theorem.*

Let $\{x_{1,j}\}_{j=1}^{\infty}$ be a sequence of elements of X_1 that converges to $x_1 \in X_1$. If f is upper semi-continuous at x_1, then one can check that

$$\limsup_{j \to \infty} f(x_j) \le f(x_1). \tag{10.19.9}$$

Similarly, if f is lower semicontinuous at x_1, then

$$\liminf_{j \to \infty} f(x_j) \ge f(x_1). \tag{10.19.10}$$

One can verify that these properties characterize upper and lower semicontinuity of f at x_1 too.

10.20 Homeomorphisms between metric spaces

Let (X, d_X) and (Y, d_Y) be metric spaces. A one-to-one mapping f from X onto Y is said to be a *homeomorphism* if f is continuous, and the corresponding inverse mapping f^{-1} from Y onto X is continuous. Of course, f^{-1} is a homeomorphism from Y onto X in this case.

Let (Z, d_Z) be another metric space. If f is a homeomorphism from X onto Y, and g is a homeomorphism from Y onto Z, then their composition $g \circ f$ is a homeomorphism from X onto Z.

Let f be a one-to-one continuous mapping from X onto Y. If X is compact, then it is well known that the inverse mapping f^{-1} is continuous, so that

$$f \text{ is a homeomorphism.} \tag{10.20.1}$$

One can show this using sequences, or by looking at closed sets.

To show this using sequences, let $\{y_j\}_{j=1}^\infty$ be a sequence of elements of Y that converges to a point $y \in Y$, and let us check that

$$\{f^{-1}(y_j)\}_{j=1}^\infty \text{ converges to } f^{-1}(y) \text{ in } X. \tag{10.20.2}$$

Put $x_j = f^{-1}(y_j)$ for each j, and $x = f^{-1}(y)$, for convenience. Suppose for the sake of a contradiction that $\{x_j\}_{j=1}^\infty$ does not converge to x in X. This implies that there is an $\epsilon > 0$ such that

$$d_X(x, x_j) \geq \epsilon \tag{10.20.3}$$

for infinitely many $j \geq 1$. Equivalently, this means that there is a subsequence $\{x_{j_l}\}_{l=1}^\infty$ of $\{x_j\}_{j=1}^\infty$ such that

$$d_X(x, x_{j_l}) \geq \epsilon \tag{10.20.4}$$

for every $l \geq 1$.

If X is compact, and thus sequentially compact, then there is a subsequence $\{x_{j_{l_n}}\}_{n=1}^\infty$ of $\{x_{j_l}\}_{l=1}^\infty$ that converges to an element w of X. It follows that

$$\{f(x_{j_{l_n}})\}_{n=1}^\infty \text{ converges to } f(w) \text{ in } Y, \tag{10.20.5}$$

because f is continuous at w, by hypothesis. However,

$$\{f(x_{j_{l_n}})\}_{n=1}^\infty = \{y_{j_{l_n}}\}_{n=1}^\infty \tag{10.20.6}$$

is a subsequence of $\{y_j\}_{j=1}^\infty$, which converges to $y = f(x)$. This implies that

$$f(w) = f(x), \tag{10.20.7}$$

so that $w = x$, because f is one-to-one. This means that $\{x_{j_{l_n}}\}_{n=1}^\infty$ converges to x in X, contradicting (10.20.4), as desired.

Alternatively, let E be a closed set in X. In order to check that f^{-1} is continuous as a mapping from Y into X, we would like to verify that

$$(f^{-1})^{-1}(E) = \{y \in Y : f^{-1}(y) \in E\} \tag{10.20.8}$$

is a closed set in Y. It is easy to see that

$$(f^{-1})^{-1}(E) = f(E). \tag{10.20.9}$$

Note that

$$E \text{ is compact in } X, \tag{10.20.10}$$

because X is compact, and E is a closed set. This implies that

$$f(E) \text{ is a compact set in } Y, \tag{10.20.11}$$

because f is continuous. It follows that

$$f(E) \text{ is a closed set in } Y. \tag{10.20.12}$$

This means that f^{-1} is continuous, by a standard characterization of continuous mappings between metric spaces.

One might notice that the first argument is very similar to one in Section 10.18. This will be discussed further in the next section.

10.21 Graphs and homeomorphisms

Let (X_1, d_{X_1}), (X_2, d_{X_2}) be nonempty metric spaces, and put $X = X_1 \times X_2$. As usual, we can define metrics $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ on X as in Section 10.1. Let f be a mapping from X_1 into X_2, and let F be the mapping from X_1 into X defined by putting $F(x_1) = (x_1, f(x_1))$ for every $x_1 \in X_1$, as in Section 10.18.

Let

$$Y = F(X_1) \tag{10.21.1}$$

be the graph of f in X, as before. We may consider Y as a metric space, using the restriction of $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$ to Y. Note that F is a one-to-one mapping from X_1 onto Y, by construction.

Let p_1, p_2 be the usual coordinate projections from X onto X_1, X_2, respectively, as in Section 10.5. The restriction of p_1 to Y is the same as the inverse mapping F^{-1} of F on Y. It follows that

$$F^{-1} \text{ is continuous on } Y, \tag{10.21.2}$$

with respect to the restriction of $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$ to Y, as in Section 10.5.

If f is continuous as a mapping from X_1 into X_2, then F is continuous as a mapping from X_1 into X with respect to each of $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$, as in Section 10.18. In this case, we get that

$$F \text{ is a homeomorphism from } X_1 \text{ onto } Y, \tag{10.21.3}$$

with respect to the restriction of any of $d_{X,1}$, $d_{X,2}$, and $d_{X,\infty}$ to Y.

If Y is compact with respect to any of $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$, then one can get that

$$F \text{ is continuous}, \tag{10.21.4}$$

because F^{-1} is continuous on Y, as in the previous section. Of course, this implies that

$$f = p_1 \circ F \text{ is continuous on } X_1, \tag{10.21.5}$$

because p_1 is continuous on X, as in Section 10.5. This is another way to look at how the compactness of Y implies the continuity of f, as in Section 10.18.

Let m be a positive integer, and let us now take

$$X_2 = \mathbf{R}^m, \tag{10.21.6}$$

equipped with the standard Euclidean metric, or the metric associated to a norm. Let f be a mapping from X_1 into $\mathbf{R}^m$, and let Φ be the mapping from $X = X_1 \times \mathbf{R}^m$ into itself defined by

$$\Phi(x) = (x_1, x_2 + f(x_1)) \tag{10.21.7}$$

for every $x = (x_1, x_2) \in X$. It is easy to see that Φ is a one-to-one mapping from X onto itself. More precisely, the inverse mapping is given by

$$\Phi^{-1}(x) = (x_1, x_2 - f(x_1)) \tag{10.21.8}$$

for every $x \in X$.

If f is continuous on X_1, then one can check that Φ and Φ^{-1} are continuous on X, with respect to any of the metrics $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$. This means that Φ is a homeomorphism from X onto itself, with respect to any of these three metrics. It is easy to see that the continuity of f is necessary for the continuity of Φ or Φ^{-1}.

If $x_1 \in X_1$ and $x = (x_1, 0)$, then $\Phi(x) = F(x_1)$. In particular,

$$\Phi(X_1 \times \{0\}) = F(X_1). \tag{10.21.9}$$

Equivalently,

$$F(X_1) = (\Phi^{-1})^{-1}(X_1 \times \{0\}). \tag{10.21.10}$$

It is easy to see that $X_1 \times \{0\}$ is a closed set in X, with respect to any of $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$. If f is continuous, then Φ^{-1} is continuous, and (10.21.10) gives another way to see that $F(X_1)$ is a closed set in X, as in Section 10.18.

Suppose now that $m = 1$, so that f is a real-valued function on X_1. Observe that

$$\Phi(X_1 \times (0, +\infty)) = \{(x_1, x_2) \in X_1 \times \mathbf{R} : x_2 > f(x_1)\} \tag{10.21.11}$$

and

$$\Phi(X_1 \times (-\infty, 0)) = \{(x_1, x_2) \in X_1 \times \mathbf{R} : x_2 < f(x_1)\}. \tag{10.21.12}$$

It is easy to see that $X_1 \times (0, +\infty)$ and $X \times (0, -\infty)$ are open sets in X, with respect to any of $d_{X,1}$, $d_{X,2}$, or $d_{X,\infty}$. If f is continuous, then one can use the continuity of Φ^{-1} to get another way to see that (10.21.11) and (10.21.12) are open sets in X, as in Section 10.19.

Chapter 11

Summable functions

Sums of real and complex-valued functions on arbitrary nonempty sets are considered in this chapter, extending sums of absolutely convergent series of real and complex numbers. The reader may choose to skip this chapter, at least initially.

11.1 Extended real numbers

As usual, the set of *extended real numbers* consists of the real numbers together with two additional elements, denoted $+\infty$ and $-\infty$. The standard ordering is extended to the set of extended real numbers by putting

$$-\infty < x < +\infty \tag{11.1.1}$$

for every $x \in \mathbf{R}$. Normally when we consider extended real numbers here, we shall only be concerned with nonnegative extended real numbers.

In some situations, addition and multiplication of extended real numbers can be defined in a natural way. In particular, we put

$$x + \infty = \infty + x = \infty + \infty = +\infty \tag{11.1.2}$$

for every $x \in \mathbf{R}$. Similarly, we put

$$x \cdot \infty = \infty \cdot x = \infty \cdot \infty = \infty \tag{11.1.3}$$

for every positive real number x. Although $0 \cdot \infty$ is not necessarily defined, it will normally correspond to 0 here.

The notions of upper and lower bounds, supremum, and infimum can be defined for sets of extended real numbers, in the same way as for sets of real numbers. In particular, if A is a nonempty set of real numbers that does not have a finite upper bound in $\mathbf{R}$, then the supremum of A can be defined as an extended real number to be $+\infty$. Of course, if $+\infty$ is an element of A, then the supremum of A is equal to $+\infty$ automatically. If $+\infty$ is the only element of A, then the infimum of A is equal to $+\infty$.

Let t be a positive real number, and let E be a nonempty set of extended real numbers. Put

$$t\,E = \{t\,x : x \in E\}, \tag{11.1.4}$$

which is another nonempty set of extended real numbers. One can check that

$$\sup(t\,E) = t\,(\sup E). \tag{11.1.5}$$

Let $\{x_j\}_{j=1}^\infty$ be a sequence of real numbers. If $\{x_j\}_{j=1}^\infty$ converges to a real number x with respect to the standard metric on $\mathbf{R}$, then we may express this by $x_j \to x$ as $j \to \infty$. As usual, we say that $\{x_j\}_{j=1}^\infty$ tends to $+\infty$ as $j \to \infty$, or $x_j \to +\infty$ as $j \to \infty$, if for every positive real number R there is a positive integer L such that

$$x_j > R \tag{11.1.6}$$

for each $j \geq L$.

Let $\{x_j\}_{j=1}^\infty$ be a monotonically increasing sequence of real numbers, so that

$$x_j \leq x_{j+1} \tag{11.1.7}$$

for every $j \geq 1$. Put

$$x = \sup\{x_j : j \in \mathbf{Z}_+\}, \tag{11.1.8}$$

which is a real number when the set of x_j's, $j \in \mathbf{Z}_+$, has a finite upper bound in $\mathbf{R}$, and otherwise is equal to $+\infty$. It is well known that

$$x_j \to x \quad \text{as } j \to \infty \tag{11.1.9}$$

under these conditions.

Now let $\{x_j\}_{j=1}^\infty$ be a sequence of real numbers such that $x_j \to +\infty$ as $j \to \infty$. If $\{y_j\}_{j=1}^\infty$ is a sequence of real numbers with a finite lower bound in $\mathbf{R}$, then it is easy to see that

$$x_j + y_j \to +\infty \quad \text{as } j \to \infty. \tag{11.1.10}$$

In particular, this holds when $y_j \to y$ as $j \to \infty$, where $y \in \mathbf{R}$ or $y = +\infty$.

11.2 Nonnegative sums

Let X be a nonempty set, and let f be a nonnegative real-valued function on X. If A is a nonempty finite subset of X, then the sum

$$\sum_{x \in A} f(x) \tag{11.2.1}$$

can be defined as a nonnegative real number in the usual way. Put

$$\sum_{x \in X} f(x) = \sup\left\{\sum_{x \in A} f(x) : A \text{ is a nonempty finite subset of } X\right\}, \tag{11.2.2}$$

where the supremum on the right side is defined as a nonnegative extended real number, as in the previous section. Of course, if X has only finitely many elements, then the supremum is attained with $A = X$. Similarly, if f has finite support in X, then the supremum is attained with any nonempty finite subset A of X such that A contains the support of f.

Suppose for the moment that X is the set $\mathbf{Z}_+$ of positive integers, and let f be a nonnegative real-valued function defined on $\mathbf{Z}_+$. In this case, one may put

$$\sum_{j=1}^{\infty} f(j) = \sup\left\{\sum_{j=1}^{n} f(j) : n \in \mathbf{Z}_+\right\}, \tag{11.2.3}$$

where the supremum on the right side is defined as a nonnegative extended real number again. Of course, the sequence of partial sums $\sum_{j=1}^{n} f(j)$ increases monotonically in this case. The sequence of partial sums tends to its supremum as $n \to \infty$, as in the previous section. If the sequence of partial sums has a finite upper bound in $\mathbf{R}$, then the sequence of partial sums converges in $\mathbf{R}$ in the usual sense. This means that the infinite series on the left side of (11.2.3) converges in the usual sense, with sum equal to the right side of (11.2.3). If the sequence of partial sums does not have a finite upper bound, so that the right side of (11.2.3) is $+\infty$, then one may interpret the sum on the left side of (11.2.3) as being $+\infty$ as well.

If f is any nonnegative real-valued function on $\mathbf{Z}_+$, then

$$\sum_{j=1}^{\infty} f(j) = \sum_{j \in \mathbf{Z}_+} f(j), \tag{11.2.4}$$

where these sums are as defined in (11.2.2) and (11.2.3). More precisely, one can check that

$$\sum_{j=1}^{\infty} f(j) \le \sum_{j \in \mathbf{Z}_+} f(j), \tag{11.2.5}$$

because each of the partial sums on the right side of (11.2.3) may be considered as one of the finite sums on the right side of (11.2.2). To get the other inequality, one can use the fact that every finite subset A of $\mathbf{Z}_+$ is contained in $\{1, \ldots, n\}$ for some $n \in \mathbf{Z}_+$.

It is sometimes convenient to consider a nonnegative extended real-valued function f on a nonempty set X. If A is a nonempty finite subset of X and $f(x) = +\infty$ for some $x \in A$, then the corresponding sum (11.2.1) is equal to $+\infty$. If $f(x) = +\infty$ for some $x \in X$, then the right side of (11.2.2) is equal to $+\infty$. Similarly, if $X = \mathbf{Z}_+$, then the partial sum $\sum_{j=1}^{n} f(j)$ is $+\infty$ when $f(j) = +\infty$ for some $j \le n$, so that the right side of (11.2.3) is $+\infty$ when $f(j) = +\infty$ for some j. Thus (11.2.4) also holds in this situation.

11.3 Compositions and subsets

Let X and Y be nonempty sets, and let ϕ be a one-to-one mapping from X onto Y. Also let f be a nonnegative extended real-valued function on Y, so that $f(\phi(x))$ defines a nonnegative extended real-valued function on Y. If A is a nonempty finite subset of X, then $\phi(A)$ is a nonempty finite subset of Y, and

$$\sum_{x \in A} f(\phi(x)) = \sum_{y \in \phi(A)} f(y). \tag{11.3.1}$$

If B is any nonempty finite subset of Y, then $A = \phi^{-1}(B)$ is a nonempty finite subset of X, and $B = \phi(A)$. It follows that

$$\sum_{x \in X} f(\phi(x)) = \sum_{y \in Y} f(y), \tag{11.3.2}$$

because both sums are defined by taking the supremum of the corresponding finite subsums in (11.3.1).

Let f be a nonnegative extended real-valued function on a nonempty set X. If E is a nonempty subset of X, then the sum

$$\sum_{x \in E} f(x) \tag{11.3.3}$$

can be defined as a nonnegative extended real number in the same way as before, as the supremum of the corresponding collection of finite subsums. This is the same as applying the definition in the previous section to the restriction of f to E. If E_1, E_2 are nonempty subsets of X and $E_1 \subseteq E_2$, then

$$\sum_{x \in E_1} f(x) \le \sum_{x \in E_2} f(x). \tag{11.3.4}$$

This uses the fact that every finite subsum of the sum on the left is also a finite subsum of the sum on the right. If we also have that $f(x) = 0$ for every $x \in E_2 \setminus E_1$, then it follows that

$$\sum_{x \in E_1} f(x) = \sum_{x \in E_2} f(x). \tag{11.3.5}$$

This is because finite subsums of the sum on the right side can be reduced to finite subsums of the sum on the left, except possibly for sums over nonempty finite subsets of $E_2 \setminus E_1$, which are equal to 0 in this case.

Let $\{x_j\}_{j=1}^\infty$ be a sequence of distinct elements of X, and let

$$E = \{x_j : j \in \mathbf{Z}_+\}. \tag{11.3.6}$$

If f is a nonnegative extended real-valued function on X again, then $f(x_j)$ may be considered as a nonnegative extended real-valued function on $\mathbf{Z}_+$. Thus

$$\sum_{j=1}^\infty f(x_j) = \sum_{j \in \mathbf{Z}_+} f(x_j), \tag{11.3.7}$$

as in (11.2.4). We also have that

$$\sum_{j \in \mathbf{Z}_+} f(x_j) = \sum_{x \in E} f(x), \tag{11.3.8}$$

as in (11.3.2). More precisely, this uses the fact that $j \mapsto x_j$ is a one-to-one mapping from $\mathbf{Z}_+$ onto E.

Let f, g be nonnegative extended real-valued functions on X, and suppose that

$$f(x) \le g(x) \tag{11.3.9}$$

for every $x \in X$. If A is a nonempty finite subset of X, then

$$\sum_{x \in A} f(x) \le \sum_{x \in A} g(x). \tag{11.3.10}$$

Using this, it is easy to see that

$$\sum_{x \in X} f(x) \le \sum_{x \in X} g(x). \tag{11.3.11}$$

11.4 Nonnegative summable functions

Let f be a nonnegative real-valued function on a nonempty set X. If

$$\sum_{x \in X} f(x) < \infty, \tag{11.4.1}$$

then f is said to be *summable* on X. Of course, if f has finite support in X, then f is summable on X. Suppose now that f is summable on X, and let us check that f vanishes at infinity on X, as in Section 2.5.

Let $\epsilon > 0$ be given, and put

$$E_\epsilon(f) = \{x \in X : f(x) \ge \epsilon\}. \tag{11.4.2}$$

If A is a nonempty finite subset of $E_\epsilon(f)$, then

$$\epsilon\,(\#A) \le \sum_{x \in A} f(x) \le \sum_{x \in X} f(x), \tag{11.4.3}$$

where $\#A$ denotes the number of elements of A. Thus

$$\#A \le (1/\epsilon) \sum_{x \in X} f(x). \tag{11.4.4}$$

It follows that $E_\epsilon(f)$ has only finitely many elements, with

$$\#E_\epsilon(f) \le (1/\epsilon) \sum_{x \in X} f(x). \tag{11.4.5}$$

Let us continue to suppose that f is summable on X, and let $\epsilon > 0$ be given again. Observe that there is a nonempty finite subset $A(\epsilon)$ of X such that

$$\sum_{x \in X} f(x) - \epsilon < \sum_{x \in A(\epsilon)} f(x), \tag{11.4.6}$$

by the definition (11.2.2) of the sum over X. If A is a finite subset of X that contains $A(\epsilon)$, then

$$\sum_{x \in A(\epsilon)} f(x) \le \sum_{x \in A} f(x) \le \sum_{x \in X} f(x) < \sum_{x \in A(\epsilon)} f(x) + \epsilon. \tag{11.4.7}$$

Let B be a nonempty finite subset of X that is disjoint from $A(\epsilon)$, so that

$$\sum_{x \in A(\epsilon)} f(x) + \sum_{x \in B} f(x) = \sum_{x \in A(\epsilon) \cup B} f(x) \le \sum_{x \in X} f(x). \tag{11.4.8}$$

This implies that

$$\sum_{x \in B} f(x) \le \sum_{x \in X} f(x) - \sum_{x \in A(\epsilon)} f(x). \tag{11.4.9}$$

If $A(\epsilon) \ne X$, then it follows that

$$\sum_{x \in X \setminus A(\epsilon)} f(x) \le \sum_{x \in X} f(x) - \sum_{x \in A(\epsilon)} f(x) < \epsilon. \tag{11.4.10}$$

If $A(\epsilon) = X$, then the sum on the left side of (11.4.10) may be interpreted as being equal to 0.

If f is summable on X and E is a nonempty subset of X, then the restriction of f to E is summable on E. More precisely, the sum of $f(x)$ over $x \in E$ is less than or equal to the sum of $f(x)$ over $x \in X$, as in (11.3.4). Similarly, let f and g be nonnegative real-valued functions on X such that $f(x) \le g(x)$ for every $x \in X$. If g is summable on X, then f is summable on X too, by (11.3.11).

11.5 Some linearity properties

Let X be a nonempty set, and let f be a nonnegative extended real-valued function on X. Also let t be a positive real number, so that $t\,f$ is a nonnegative extended real-valued function on X as well. If A is a nonempty finite subset of X, then

$$\sum_{x \in A} t\,f(x) = t \sum_{x \in A} f(x). \tag{11.5.1}$$

Using this, one can check that

$$\sum_{x \in X} t\,f(x) = t \sum_{x \in X} f(x). \tag{11.5.2}$$

More precisely, this can be obtained from (11.1.5) and the definition (11.2.2) of the sum over X.

Let g be another nonnegative extended real-valued function on X, so that $f+g$ is a nonnegative extended real-valued function on X too. If A is a nonempty finite subset of X, then

$$\sum_{x \in A} (f(x) + g(x)) = \sum_{x \in A} f(x) + \sum_{x \in A} g(x). \tag{11.5.3}$$

One can use this to show that

$$\sum_{x \in X} (f(x) + g(x)) = \sum_{x \in X} f(x) + \sum_{x \in X} g(x). \tag{11.5.4}$$

More precisely, if A is a nonempty finite subset of X, then

$$\sum_{x \in A} (f(x) + g(x)) \le \sum_{x \in X} f(x) + \sum_{x \in X} g(x), \tag{11.5.5}$$

by (11.5.3). This implies that

$$\sum_{x \in X} (f(x) + g(x)) \le \sum_{x \in X} f(x) + \sum_{x \in X} g(x). \tag{11.5.6}$$

We would like to verify that

$$\sum_{x \in X} f(x) + \sum_{x \in X} g(x) \le \sum_{x \in X} (f(x) + g(x)), \tag{11.5.7}$$

in order to get (11.5.4). This is trivial when the right side of (11.5.7) is $+\infty$, and so we may suppose that it is finite, so that $f + g$ is summable on X. It follows that f and g are summable on X, because $f, g \le f+g$, as in the previous section. In particular, $f(x)$ and $g(x)$ are finite for every $x \in X$. Let A and B be nonempty finite subsets of X, and observe that

$$\begin{aligned} \sum_{x \in A} f(x) + \sum_{x \in B} g(x) &\le \sum_{x \in A \cup B} f(x) + \sum_{x \in A \cup B} g(x) \\ &= \sum_{x \in A \cup B} (f(x) + g(x)) \le \sum_{x \in X} (f(x) + g(x)). \end{aligned} \tag{11.5.8}$$

Thus

$$\sum_{x \in A} f(x) \le \sum_{x \in X} (f(x) + g(x)) - \sum_{x \in B} g(x). \tag{11.5.9}$$

This implies that

$$\sum_{x \in X} f(x) \le \sum_{x \in X} (f(x) + g(x)) - \sum_{x \in B} g(x) \tag{11.5.10}$$

for every nonempty finite set $B \subseteq X$. Equivalently,

$$\sum_{x \in B} g(x) \le \sum_{x \in X} (f(x) + g(x)) - \sum_{x \in X} f(x) \tag{11.5.11}$$

for every nonempty finite set $B \subseteq X$, and hence

$$\sum_{x \in X} g(x) \le \sum_{x \in X} (f(x) + g(x)) - \sum_{x \in X} f(x). \tag{11.5.12}$$

This shows that (11.5.7) holds under these conditions, as desired.

Let f be a nonnegative extended real-valued function on X again, and let E_1, E_2 be disjoint nonempty subsets of X. It is easy to see that

$$\sum_{x \in E_1 \cup E_2} f(x) = \sum_{x \in E_1} f(x) + \sum_{x \in E_2} f(x), \tag{11.5.13}$$

using (11.5.4). More precisely, this can be obtained by expressing f on $E_1 \cup E_2$ as the sum of two functions, with supports contained in E_1 and E_2.

11.6 ℓ^1 Spaces

A real or complex–valued function f on a nonempty set X is said to be *summable* on X if $|f|$ is summable as a nonnegative real-valued function on X. Let $\ell^1(X, \mathbf{R})$ and $\ell^1(X, \mathbf{C})$ be the spaces of real and complex-valued summable functions on X, respectively. Of course, if a real or complex-valued function f on X has finite support, then f is summable on X, so that

$$c_{00}(X, \mathbf{R}) \subseteq \ell^1(X, \mathbf{R}), \quad c_{00}(X, \mathbf{C}) \subseteq \ell^1(X, \mathbf{C}). \tag{11.6.1}$$

If f is any real or complex-valued summable function on X, then $|f|$ vanishes at infinity on X, as in Section 11.4. This means that f vanishes at infinity on X too, so that

$$\ell^1(X, \mathbf{R}) \subseteq c_0(X, \mathbf{R}), \quad \ell^1(X, \mathbf{C}) \subseteq c_0(X, \mathbf{C}). \tag{11.6.2}$$

If f is a real or complex-valued summable function on X, then put

$$\|f\|_1 = \sum_{x \in X} |f(x)|. \tag{11.6.3}$$

This is a nonnegative real number, which is equal to 0 exactly when $f(x) = 0$ for every $x \in X$. If $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then $|t\, f| = |t|\, |f|$ is summable on X, by (11.5.2), and we have that

$$\|t\, f\|_1 = \sum_{x \in X} |t\, f(x)| = |t| \sum_{x \in X} |f(x)| = |t|\, \|f\|_1. \tag{11.6.4}$$

Similarly, if g is another real or complex-valued summable function on X, as appropriate, then it is easy to see that $f + g$ is summable on X as well, with

$$\|f+g\|_1 = \sum_{x \in X} |f(x) + g(x)| \le \sum_{x \in X} (|f(x)| + |g(x)|) = \|f\|_1 + \|g\|_1. \tag{11.6.5}$$

Thus $\ell^1(X, \mathbf{R})$ and $\ell^1(X, \mathbf{C})$ are linear subspaces of the real and complex vector spaces of all real and complex-valued functions on X, respectively, and (11.6.3) defines a norm on each of $\ell^1(X, \mathbf{R})$ and $\ell^1(X, \mathbf{C})$.

One can verify that

$$d_1(f, g) = \|f - g\|_1 \tag{11.6.6}$$

defines a metric on each of $\ell^1(X, \mathbf{R})$ and $\ell^1(X, \mathbf{C})$, using (11.6.4) and (11.6.5), as usual. Remember that $\|f\|_\infty$ denotes the supremum norm of a bounded real or complex-valued f on X, as in (1.13.4). If f is a real or complex-valued summable function on X, then

$$|f(x)| \le \|f\|_1 \tag{11.6.7}$$

for every $x \in X$. This implies that f is bounded on X, with

$$\|f\|_\infty \le \|f\|_1. \tag{11.6.8}$$

If f, g are real or complex-valued summable functions on X, then it follows that

$$d_\infty(f, g) \le d_1(f, g), \tag{11.6.9}$$

where $d_\infty(f, g)$ is the supremum metric, as in (1.13.8).

Let us check that $c_{00}(X, \mathbf{R})$, $c_{00}(X, \mathbf{C})$ are dense in $\ell^1(X, \mathbf{R})$, $\ell^1(X, \mathbf{C})$, respectively, with respect to the ℓ^1 metric (11.6.6). Of course, this is trivial when X has only finitely many elements, and so we suppose that X is an infinite set. Let f be a real or complex-valued summable function on X, and let $\epsilon > 0$ be given. Remember that there is a finite subset $A(\epsilon)$ of X such that

$$\sum_{x \in X \setminus A(\epsilon)} |f(x)| < \epsilon, \tag{11.6.10}$$

as in (11.4.10). Let f_ϵ be the real or complex-valued function, as appropriate, defined on X by

$$\begin{aligned} f_\epsilon(x) &= f(x) \quad \text{when } x \in A(\epsilon) \\ &= 0 \qquad \text{when } x \in X \setminus A(\epsilon). \end{aligned} \tag{11.6.11}$$

Thus f_ϵ has finite support in X. Note that $f - f_\epsilon$ is equal to 0 on $A(\epsilon)$, and to f on $X \setminus A(\epsilon)$. This implies that

$$\|f - f_\epsilon\|_1 = \sum_{x \in X \setminus A(\epsilon)} |f(x)| < \epsilon, \tag{11.6.12}$$

as desired.

11.7 Real-valued summable functions

Let f be a real-valued function on a nonempty set X, and put

$$f_+(x) = \max(f(x), 0) \quad \text{and} \quad f_-(x) = \max(-f(x), 0) \tag{11.7.1}$$

for each $x \in X$. These are nonnegative real-valued functions on X that satisfy

$$f_+(x) + f_-(x) = |f(x)| \tag{11.7.2}$$

and

$$f_+(x) - f_-(x) = f(x) \tag{11.7.3}$$

for every $x \in X$. If f is summable on X, so that $|f|$ is summable on X, then f_+ and f_- are summable as nonnegative real-valued functions on X. In this case, we put

$$\sum_{x \in X} f(x) = \sum_{x \in X} f_+(x) - \sum_{x \in X} f_-(x). \tag{11.7.4}$$

Note that

$$\left|\sum_{x \in X} f(x)\right| \le \sum_{x \in X} f_+(x) + \sum_{x \in X} f_-(x) = \sum_{x \in X} |f(x)|. \tag{11.7.5}$$

Let f_1, f_2 be nonnegative real-valued summable functions on X such that

$$f(x) = f_1(x) - f_2(x) \tag{11.7.6}$$

for every $x \in X$. This implies that

$$f_1(x) + f_-(x) = f_+(x) + f_2(x) \tag{11.7.7}$$

for every $x \in X$, because of (11.7.3). Hence

$$\sum_{x \in X} f_1(x) + \sum_{x \in X} f_-(x) = \sum_{x \in X} f_+(x) + \sum_{x \in X} f_2(x), \tag{11.7.8}$$

as in (11.5.7). It follows that

$$\sum_{x \in X} f_1(x) - \sum_{x \in X} f_2(x) = \sum_{x \in X} f_+(x) - \sum_{x \in X} f_-(x), \tag{11.7.9}$$

because the individual sums are all finite. This means that

$$\sum_{x \in X} f(x) = \sum_{x \in X} f_1(x) - \sum_{x \in X} f_2(x), \tag{11.7.10}$$

by (11.7.4).

If f is a real-valued summable function on X and $t \in \mathbf{R}$, then $t\,f$ is summable on X too, as in the previous section. One can check that

$$\sum_{x \in X} t\,f(x) = t \sum_{x \in X} f(x). \tag{11.7.11}$$

Of course, this is trivial when $t = 0$. If $t > 0$, then (11.7.11) can be derived from the analogous fact (11.5.2) for nonnegative real-valued functions on X, and the definition (11.7.4) of the sum for real-valued summable functions on X. If $t = -1$, then (11.7.11) can be obtained directly from (11.7.4).

If f, g are real-valued summable functions on X, then $f + g$ is summable on X as well, as in the previous section again. One can verify that

$$\sum_{x \in X} (f(x) + g(x)) = \sum_{x \in X} f(x) + \sum_{x \in X} g(x). \tag{11.7.12}$$

We have already seen in Section 11.5 that this holds when f and g are nonnegative. Otherwise, f and g can be expressed as differences of nonnegative real-valued summable functions on X, as in (11.7.3). This leads to an expression of $f + g$ as a difference of nonnegative real-valued summable functions on X. Each of the sums of f, g, and $f + g$ over X can be given as the difference of the corresponding sums of nonnegative summable functions, as in (11.7.4) and (11.7.10). Using this, one can reduce (11.7.12) to the analogous statement for nonnegative functions.

11.8 Complex-valued summable functions

Let f be a complex-valued summable function on a nonempty set X, and let $\operatorname{Re} f(x)$, $\operatorname{Im} f(x)$ be the real and imaginary parts of $f(x)$ for each $x \in X$, as usual. Note that

$$|\operatorname{Re} f(x)|, |\operatorname{Im} f(x)| \le |f(x)| \le |\operatorname{Re} f(x)| + |\operatorname{Im} f(x)| \tag{11.8.1}$$

for every $x \in X$. This implies that f is summable as a complex-valued function on X if and only if its real and imaginary parts are summable as real-valued functions on X. In this case, we put

$$\sum_{x \in X} f(x) = \sum_{x \in X} \operatorname{Re} f(x) + i \sum_{x \in X} \operatorname{Im} f(x), \tag{11.8.2}$$

where the sums on the right side are defined as in the preceding section.

If f, g are complex-valued summable functions on X, then $f+g$ is summable on X too, as in Section 11.6. It is easy to see that

$$\sum_{x \in X} (f(x) + g(x)) = \sum_{x \in X} f(x) + \sum_{x \in X} g(x). \tag{11.8.3}$$

Of course, the real and imaginary parts of $f + g$ are the same as the sums of the real and imaginary parts of f and g, respectively. Thus (11.8.3) follows from the definition (11.8.2) of the sum in the complex case and the analogous statement for the sum in the real case.

If f is a complex-valued summable function on X and $t \in \mathbf{C}$, then $t\,f$ is summable on X as well, as in Section 11.6 again. Let us check that

$$\sum_{x \in X} t\,f(x) = t \sum_{x \in X} f(x). \tag{11.8.4}$$

If $t \in \mathbf{R}$, then this can be obtained from the definition (11.8.2) of the sum in the complex case and the analogous property of the sum in the real case. Similarly, if t is imaginary, then (11.8.4) can be reduced to the analogous statement in the real case. If $t \in \mathbf{C}$, then (11.8.4) can be derived from the previous two cases, using (11.8.3).

Let f be a complex-valued summable function on X again, and observe that

$$\Big|\sum_{x \in X} f(x)\Big| \le \Big|\sum_{x \in X} \operatorname{Re} f(x)\Big| + \Big|\sum_{x \in X} \operatorname{Im} f(x)\Big|, \tag{11.8.5}$$

by the definition (11.8.2) of the sum. This implies that

$$\Big|\sum_{x \in X} f(x)\Big| \le \sum_{x \in X} |\operatorname{Re} f(x)| + \sum_{x \in X} |\operatorname{Im} f(x)|, \tag{11.8.6}$$

because of (11.7.5). It follows that

$$\Big|\sum_{x \in X} f(x)\Big| \le 2 \sum_{x \in X} |f(x)|, \tag{11.8.7}$$

using the first step in (11.8.1). Of course, we would rather have that

$$\Big|\sum_{x \in X} f(x)\Big| \le \sum_{x \in X} |f(x)|. \tag{11.8.8}$$

This corresponds to the ordinary triangle inequality for the standard absolute value function on $\mathbf{C}$ in the case of finite sums. In this situation, one can get (11.8.8) by approximating by finite sums. Some more details about this type of approximation will be given in the next section.

11.9 Some properties of the sum

Let f be a real or complex-valued function on a nonempty set X, and suppose for the moment that f has finite support in X. In this case, the sum $\sum_{x \in X} f(x)$ can be defined as a real or complex number by reducing to a finite sum. If f is a nonnegative real-valued function on X, then we have seen that the definition of the sum in Section 11.2 reduces to the same finite sum. If f is a real or complex-valued function on X, then we can apply this to $|f(x)|$, to get that f is summable on X. If f is a real-valued function on X, then the functions defined on X in (11.7.1) have finite support. This implies that the definition of the sum in Section 11.7 is the same as the finite sum, because of the analogous statement for nonnegative real-valued functions on X with finite support. Similarly, if f is a complex-valued function on X, then the real and imaginary parts of f have finite support in X. It follows that the definition of the sum in the preceding section is the same as the finite sum, because of the analogous statement for real-valued functions on X with finite support. In particular, (11.8.8) holds, because of the ordinary triangle inequality for finite sums.

Now let f, g be real or complex-valued summable functions on X. Observe that

$$\left|\sum_{x\in X} f(x) - \sum_{x\in X} g(x)\right| = \left|\sum_{x\in X}(f(x) - g(x))\right| \le 2 \sum_{x\in X} |f(x) - g(x)|. \tag{11.9.1}$$

This uses the linearity of the sum in the first step, and (11.8.7) in the second step. It follows that the mapping from a real or complex-valued summable function f on X to its sum $\sum_{x\in X} f(x)$ is uniformly continuous as a mapping from $\ell^1(X, \mathbf{R})$, $\ell^1(X, \mathbf{C})$ into $\mathbf{R}$, $\mathbf{C}$, respectively. Here we use the corresponding ℓ^1 metric on the domain, as in (11.6.6), and the standard Euclidean metric on the range.

One can use these properties of the sum to check that (11.8.8) holds for all complex-valued summable functions f on X. More precisely, if f has finite support in X, then we have already seen that (11.8.8) holds. Otherwise, if f is any complex-valued summable function on X, then f can be approximated by complex-valued functions on X with finite support with respect to the ℓ^1 metric, as in Section 11.6. To get that f satisfies (11.8.8) as well, one can use the continuity condition (11.9.1).

Once we have that (11.8.8) holds for every complex-valued summable function on X, we get that

$$\left|\sum_{x\in X} f(x) - \sum_{x\in X} g(x)\right| \le \sum_{x\in X} |f(x) - g(x)| \tag{11.9.2}$$

for all complex-valued summable functions f, g on X. This is basically the same as (11.9.1), except that we use (11.8.8) in the second step. Of course, if f and g are real-valued summable functions on X, then we could already get this using (11.7.5).

Let f be a real or complex-valued summable function on X again. If E is a nonempty subset of X, then the restriction of f to E is summable as a real or complex-valued function on E, as appropriate. This follows from the analogous statement for nonnegative real-valued summable functions, which was mentioned in Section 11.4. In particular, this means that $\sum_{x\in E} f(x)$ can be defined as a real or complex number, as appropriate, as in the previous two sections.

Note that

$$\sum_{x\in E} f(x) = \sum_{x\in X} f(x) \tag{11.9.3}$$

when $f(x) = 0$ for every $x \in X \setminus E$. This was mentioned in Section 11.3 when f is a nonnegative real-valued function on X. If f is a real-valued summable function on X with support contained in E, then the functions defined in (11.7.1) have support contained in E too. In this case, (11.9.3) can be reduced to the corresponding statement for nonnegative real-valued functions, because of the way that the sum was defined in Section 11.7. Similarly, if f is a complex-valued summable function on X with support contained in E, then (11.9.3) can be obtained from the corresponding statement for real-valued functions, applied to the real and imaginary parts of f.

If E_1 and E_2 are nonempty disjoint subsets of X, then

$$\sum_{x \in E_1 \cup E_2} f(x) = \sum_{x \in E_1} f(x) + \sum_{x \in E_2} f(x). \tag{11.9.4}$$

This was mentioned in Section 11.5 when f is nonnegative. If f is any real or complex-valued summable function on X, then (11.9.4) can be obtained from the linearity of the sum, by expressing f on $E_1 \cup E_2$ as a sum of functions supported in E_1 and E_2, as before. Alternatively, if f is a real-valued summable function on X, then one can reduce to the case of nonnegative real-valued summable functions on X, because of the way that the sum is defined in Section 11.7. Similarly, if f is a complex-valued summable function on X, then one can apply the previous statement to the real and imaginary parts of f.

11.10 Generalized convergence

Let f be a real or complex-valued function on a nonempty set X. The sum

$$\sum_{x \in X} f(x) \tag{11.10.1}$$

is said to *converge in the generalized sense* if there is a real or complex number λ, as appropriate, such that for every $\epsilon > 0$ there is a nonempty finite subset A_ϵ of X with the property that

$$\left| \sum_{x \in A} f(x) - \lambda \right| < \epsilon \tag{11.10.2}$$

for every nonempty finite subset A of X with $A_\epsilon \subseteq A$. In this case, the value of the sum (11.10.1) is defined to be λ. More precisely, one can check that λ is unique when it exists. Of course, if X has only finitely many elements, then one can take $A_\epsilon = X$ for every $\epsilon > 0$, to get that this condition holds trivially with λ equal to the finite sum (11.10.1).

Suppose that f is summable on X, and let us check that the sum (11.10.1) converges in the generalized sense, with the same value of the sum as defined in Sections 11.2, 11.7, and 11.8. This is trivial when X has only finitely many elements, and so we suppose now that X is an infinite set. If A is a nonempty finite subset of X, then

$$\sum_{x \in X} f(x) = \sum_{x \in A} f(x) + \sum_{x \in X \setminus A} f(x), \tag{11.10.3}$$

as in (11.9.4). Note that $X \setminus A \neq \emptyset$ in this situation, and that the restriction of f to $X \setminus A$ is summable, as in the previous section. It follows that

$$\left|\sum_{x \in A} f(x) - \sum_{x \in X} f(x)\right| = \left|\sum_{x \in X \setminus A} f(x)\right| \le \sum_{x \in X \setminus A} |f(x)|, \tag{11.10.4}$$

using (11.7.5) or (11.8.8) in the second step, as appropriate. Remember that for each $\epsilon > 0$ there is a nonempty finite subset $A(\epsilon)$ of X such that

$$\sum_{x \in X \setminus A(\epsilon)} |f(x)| < \epsilon, \tag{11.10.5}$$

as in (11.4.10). If $A(\epsilon) \subseteq A$, then $X \setminus A \subseteq X \setminus A(\epsilon)$, and hence

$$\sum_{x \in X \setminus A} |f(x)| \le \sum_{x \in X \setminus A(\epsilon)} |f(x)|, \tag{11.10.6}$$

as in (11.3.4). Thus

$$\left|\sum_{x \in A} f(x) - \sum_{x \in X} f(x)\right| \le \sum_{x \in X \setminus A} |f(x)| \le \sum_{x \in X \setminus A(\epsilon)} |f(x)| < \epsilon \tag{11.10.7}$$

when A is a nonempty finite subset of X with $A(\epsilon) \subseteq A$, as desired.

Alternatively, if f is a nonnegative real-valued summable function on X, then the convergence of the sum (11.10.1) in the generalized sense can be obtained from (11.4.7). If f is a real-valued summable function on X, then f can be expressed as a difference of nonnegative real-valued summable functions on X, as in (11.7.3). In this case, the convergence of the sum (11.10.1) in the generalized sense can be reduced to the previous statement. Similarly, if f is a complex-valued summable function on X, then the real and imaginary parts of f are summable on X too. This permits one to reduce the convergence of (11.10.1) in the generalized sense to the analogous statements for the real and imaginary parts of f.

Let f be a real or complex-valued function on X again, and suppose that the sum (11.10.1) converges in the generalized sense. Applying the earlier definition with $\epsilon = 1$, we get that there is a real or complex number λ, as appropriate, and a nonempty finite subset A_1 of X such that

$$\left|\sum_{x \in A} f(x) - \lambda\right| < 1 \tag{11.10.8}$$

for every nonempty finite subset A of X with $A_1 \subseteq A$. In particular,

$$\left|\sum_{x \in A} f(x)\right| < |\lambda| + 1 \tag{11.10.9}$$

for every nonempty finite subset A of X with $A_1 \subseteq A$. Let us check that

$$\Big|\sum_{x \in B} f(x)\Big| \le |\lambda| + 1 + \sum_{x \in A_1} |f(x)| \tag{11.10.10}$$

for every nonempty finite subset B of X. Of course, (11.10.10) follows from (11.10.9) when $A_1 \subseteq B$. Otherwise, we can apply (11.10.9) to $A = B \cup A_1$, to get that

$$\Big|\sum_{x \in B \cup A_1} f(x)\Big| < |\lambda| + 1. \tag{11.10.11}$$

This implies that

$$\begin{aligned} \Big|\sum_{x \in B} f(x)\Big| &= \Big|\sum_{x \in B \cup A_1} f(x) - \sum_{x \in A_1 \setminus B} f(x)\Big| \\ &\le \Big|\sum_{x \in B \cup A_1} f(x)\Big| + \Big|\sum_{x \in A_1 \setminus B} f(x)\Big| \\ &\le |\lambda| + 1 + \sum_{x \in A_1 \setminus B} |f(x)|. \end{aligned} \tag{11.10.12}$$

This shows that (11.10.10) also holds when $A_1 \setminus B \ne \emptyset$, as desired.

If f is a nonnegative real-valued function on X, then (11.10.10) implies that f is summable on X. If f is a real-valued function on X, then (11.10.10) implies that the positive and negative parts of f are summable on X. More precisely, one can get this by applying (11.10.10) to nonempty finite subsets B of X on which f has constant sign. It follows that f is summable on X too in this case. If f is a complex-valued function on X, then one can use (11.10.10) to get that the real and imaginary parts of f are summable on X, and hence that f is summable on X.

11.11 Compositions and sums

Let X and Y be nonempty sets, and let ϕ be a one-to-one mapping from X onto Y. Also let f be a real or complex-valued function on Y, so that $f(\phi(x))$ defines a real or complex-valued function on X, as appropriate. Note that

$$\sum_{x \in X} |f(\phi(x))| = \sum_{y \in Y} |f(y)|, \tag{11.11.1}$$

as in (11.3.2). Thus f is summable on Y if and only if $f(\phi(x))$ is summable on X. Let us check that

$$\sum_{x \in X} f(\phi(x)) = \sum_{y \in Y} f(y) \tag{11.11.2}$$

in this case. If f is a nonnegative real-valued function on Y, then (11.11.2) is the same as (11.11.1). If f is a real-valued function on Y, then one can get

(11.11.2) by expressing f as a difference of nonnegative real-valued summable functions on Y, as in Section 11.7. If f is a complex-valued summable function on Y, then (11.11.2) can be obtained by applying the previous statement to the real and imaginary parts of f.

Similarly, if f is a real or complex-valued function on Y, then one can check directly that $\sum_{x \in X} f(\phi(x))$ converges in the generalized sense if and only if $\sum_{y \in Y} f(y)$ converges in the generalized sense, with the same value of the sums. This uses the fact that $A \mapsto \phi(A)$ defines a one-to-one correspondence between nonempty finite subsets of X and nonempty finite subsets of Y.

Now let X be a countably infinite set, and let $\{x_j\}_{j=1}^\infty$ be a sequence of elements of X in which every element of X occurs exactly once. Thus $j \mapsto x_j$ is a one-to-one mapping from the set $\mathbf{Z}_+$ of positive integers onto X. Let f be a real or complex-valued function on X, so that $f(x_j)$ may be considered as a real or complex-valued function of $j \in \mathbf{Z}_+$, as appropriate. As before, f is summable on X if and only if $f(x_j)$ is summable on $\mathbf{Z}_+$, in which case

$$\sum_{j \in \mathbf{Z}_+} f(x_j) = \sum_{x \in X} f(x). \tag{11.11.3}$$

Alternatively, the sum on the left converges in the generalized sense if and only if the sum on the right converges in the generalized sense, with the same value of the sum, as in the preceding paragraph.

Remember that

$$\sum_{j=1}^\infty |f(x_j)| = \sum_{j \in \mathbf{Z}_+} |f(x_j)|, \tag{11.11.4}$$

as in (11.2.4). Thus $f(x_j)$ is summable on $\mathbf{Z}_+$ if and only if the infinite series on the left side of (11.11.4) converges in the usual sense, which means that $\sum_{j=1}^\infty f(x_j)$ converges absolutely. Under these conditions,

$$\sum_{j=1}^\infty f(x_j) = \sum_{j \in \mathbf{Z}_+} f(x_j). \tag{11.11.5}$$

More precisely, (11.11.5) is the same as (11.11.4) when f is real-valued and nonnegative. If f is real-valued, then (11.11.5) can be obtained from the previous statement by expressing f as a difference of nonnegative real-valued summable functions. If f is complex-valued, then one can get (11.11.5) by considering the real and imaginary parts separately. Alternatively, if the sum on the right side of (11.11.5) converges in the generalized sense, then it is easy to see that the sum on the left side of (11.11.5) converges, and with the same value of the sum.

11.12 Completeness of ℓ^1 spaces

Let X be a nonempty set, and let $\{f_j\}_{j=1}^\infty$ be a sequence of real or complex-valued functions on X that converge pointwise to a real or complex-valued function f on X, as appropriate. Suppose also that the f_j's are summable on X, with bounded ℓ^1 norms, so that there is a nonnegative real number C such that

$$\sum_{x \in X} |f_j(x)| \le C \tag{11.12.1}$$

for every $j \ge 1$. We would like to show that f is summable on X too under these conditions, with

$$\sum_{x \in X} |f(x)| \le C. \tag{11.12.2}$$

If A is any nonempty finite subset of X, then

$$\sum_{x \in A} |f(x)| = \lim_{j \to \infty} \sum_{x \in A} |f_j(x)| \le C, \tag{11.12.3}$$

using basic properties of limits in the first step, and (11.12.1) in the second step. This implies (11.12.2), as desired.

Let us now check that $\ell^1(X, \mathbf{R})$ and $\ell^1(X, \mathbf{C})$ are complete as metric spaces, with respect to the ℓ^1 metric (11.6.6). Let $\{f_j\}_{j=1}^\infty$ be a sequence of real or complex-valued summable functions on X that is a Cauchy sequence with respect to the ℓ^1 metric. This means that for each $\epsilon > 0$ there is an $L(\epsilon) \in \mathbf{Z}_+$ such that

$$\sum_{x \in X} |f_j(x) - f_l(x)| = \|f_j - f_l\|_1 < \epsilon \tag{11.12.4}$$

for every $j, l \ge L(\epsilon)$. In particular, if $x \in X$, then it follows that

$$|f_j(x) - f_l(x)| < \epsilon \tag{11.12.5}$$

for every $j, l \ge L(\epsilon)$. Thus $\{f_j(x)\}_{j=1}^\infty$ is a Cauchy sequence in $\mathbf{R}$ or $\mathbf{C}$, as appropriate, with respect to the standard Euclidean metric.

It is well known that $\mathbf{R}$ and $\mathbf{C}$ are complete as metric spaces with respect to the corresponding Euclidean metric. Thus $\{f_j(x)\}_{j=1}^\infty$ converges in $\mathbf{R}$ or $\mathbf{C}$, as appropriate, for each $x \in X$. Put

$$f(x) = \lim_{j \to \infty} f(x) \tag{11.12.6}$$

for every $x \in X$, which defines f as a real or complex-valued function on X, as appropriate. We would like to show that f is summable on X, and that $\{f_j\}_{j=1}^\infty$ converges to f with respect to the ℓ^1 metric. Of course, if X has only finitely many elements, then this follows easily from pointwise convergence.

Let $\epsilon > 0$ and $l \ge L(\epsilon)$ be given, and consider $\{f_j - f_l\}_{j=L(\epsilon)}^\infty$ as a sequence of summable functions on X that converges pointwise to $f - f_l$. Using (11.12.4) and the remarks at the beginning of the section, we get that $f - f_l$ is summable on X, with

$$\sum_{x \in X} |f(x) - f_l(x)| \le \epsilon. \tag{11.12.7}$$

In particular, f is summable on X, because f_l is summable on X. More precisely, we can simply take $\epsilon = 1$ and $l = L(1)$ in this step. It follows from (11.12.7) that $\{f_j\}_{j=1}^\infty$ converges to f with respect to the ℓ^1 metric, as desired.

11.13 Monotone convergence

Let X be a nonempty set, and let $\{f_j\}_{j=1}^\infty$ be a sequence of nonnegative real-valued functions on X. Suppose that the f_j's are monotonically increasing in j, so that

$$f_j(x) \le f_{j+1}(x) \tag{11.13.1}$$

for every $x \in X$ and $j \ge 1$. Put

$$f(x) = \sup_{j \ge 1} f_j(x) \tag{11.13.2}$$

for each $x \in X$, which defines f as a nonnegative extended real-valued function on X. Equivalently, this means that

$$f_j(x) \to f(x) \quad \text{as } j \to \infty \tag{11.13.3}$$

for every $x \in X$, as in Section 11.1. Note that

$$\sum_{x \in X} f_j(x) \le \sum_{x \in X} f_{j+1}(x) \tag{11.13.4}$$

for every $j \ge 1$, because of (11.13.1), as in (11.3.11).

Similarly,

$$\sum_{x \in X} f_j(x) \le \sum_{x \in X} f(x) \tag{11.13.5}$$

for every $j \ge 1$, because $f_j(x) \le f(x)$ for every $x \in X$ and $j \ge 1$, by construction. Thus

$$\sup_{j \ge 1} \Big(\sum_{x \in X} f_j(x) \Big) \le \sum_{x \in X} f(x). \tag{11.13.6}$$

We would like to show that

$$\sup_{j \ge 1} \Big(\sum_{x \in X} f_j(x) \Big) = \sum_{x \in X} f(x) \tag{11.13.7}$$

under these conditions. This is basically the same as saying that

$$\sum_{x \in X} f_j(x) \to \sum_{x \in X} f(x) \quad \text{as } j \to \infty, \tag{11.13.8}$$

because of (11.13.4). More precisely, if f_j is summable on X for each $j \ge 1$, then (11.13.7) is equivalent to (11.13.8), as in Section 11.1. Otherwise, if f_j is not summable on X for some $j \ge 1$, then f_j is not summable on X for all sufficiently large j, because of (11.13.1). In this case, it is easy to see that f is not summable on X, so that (11.13.7) holds, and that (11.13.8) holds in a suitable sense.

In order to show (11.13.7) or (11.13.8), let a be a real number such that

$$a < \sum_{x \in X} f(x). \tag{11.13.9}$$

By definition of the sum on the right, there is a nonempty finite subset A of X such that

$$a < \sum_{x \in A} f(x). \tag{11.13.10}$$

Observe that

$$\sum_{x \in A} f_j(x) \to \sum_{x \in A} f(x) \quad \text{as } j \to \infty, \tag{11.13.11}$$

because of (11.13.3). It follows that

$$a < \sum_{x \in A} f_j(x) \tag{11.13.12}$$

for all but finitely many $j \geq 1$. This implies that

$$a < \sum_{x \in X} f_j(x) \tag{11.13.13}$$

for the same j's, and hence all but finitely many $j \geq 1$. Using this, one can check that (11.13.7) or (11.13.8) holds, as desired. This is the *monotone convergence theorem* for sums. Of course, if X has only finitely many elements, then this is more elementary, as in (11.13.11).

11.14 Dominated convergence

Let X be a nonempty set, and let $\{f_j\}_{j=1}^\infty$ be a sequence of real or complex-valued functions on X that converges pointwise to a real or complex-valued function f on X, as appropriate. Suppose that there is a nonnegative real-valued summable function g on X such that

$$|f_j(x)| \leq g(x) \tag{11.14.1}$$

for every $x \in X$ and $j \geq 1$. This implies that

$$|f(x)| \leq g(x) \tag{11.14.2}$$

for every $x \in X$, because $\{f_j(x)\}_{j=1}^\infty$ converges to $f(x)$, by hypothesis. In particular, f_j is summable on X for each $j \geq 1$, because of (11.14.1), and f is summable on X too, because of (11.14.2). We would like to show that

$$\lim_{j \to \infty} \sum_{x \in X} |f_j(x) - f(x)| = 0 \tag{11.14.3}$$

under these conditions. Of course, if X has only finitely many elements, then this follows from standard results about sums of convergent sequences. Thus we may suppose that X has infinitely many elements.

Let $\epsilon > 0$ be given. Because g is summable on X, there is a nonempty finite subset A_0 of X such that

$$\sum_{x \in X \setminus A_0} g(x) < \epsilon/3, \tag{11.14.4}$$

as in (11.4.10). Note that

$$|f_j(x) - f(x)| \le |f_j(x)| + |f(x)| \le 2\,g(x) \tag{11.14.5}$$

for every $x \in X$ and $j \ge 1$, by (11.14.1) and (11.14.2). It follows that

$$\sum_{x \in X \setminus A_0} |f_j(x) - f(x)| \le \sum_{x \in X \setminus A_0} 2\,g(x) < 2\,\epsilon/3 \tag{11.14.6}$$

for every $j \ge 1$.

As before,

$$\lim_{j \to \infty} \sum_{x \in A_0} |f_j(x) - f(x)| = 0, \tag{11.14.7}$$

because A_0 has only finitely many elements. Hence there is an $L \in \mathbf{Z}_+$ such that

$$\sum_{x \in A_0} |f_j(x) - f(x)| < \epsilon/3 \tag{11.14.8}$$

for every $j \ge L$. Combining this with (11.14.6), we get that

$$\begin{aligned} \sum_{x \in X} |f_j(x) - f(x)| &= \sum_{x \in A_0} |f_j(x) - f(x)| + \sum_{x \in X \setminus A_0} |f_j(x) - f(x)| \\ &< \epsilon/3 + 2\,\epsilon/3 = \epsilon \end{aligned} \tag{11.14.9}$$

for every $j \ge L$, as desired.

Observe that

$$\begin{aligned} \left| \sum_{x \in X} f_j(x) - \sum_{x \in X} f(x) \right| &= \left| \sum_{x \in X} (f_j(x) - f(x)) \right| \\ &\le \sum_{x \in X} |f_j(x) - f(x)| \end{aligned} \tag{11.14.10}$$

for every $j \ge 1$. This uses (11.7.5) or (11.8.8), as appropriate, in the second step. Thus (11.14.3) implies that

$$\lim_{j \to \infty} \sum_{x \in X} f_j(x) = \sum_{x \in X} f(x). \tag{11.14.11}$$

The fact that this holds under these conditions is the *dominated convergence theorem* for sums. If X has only finitely many elements, then this is a standard result about sums of convergent sequences, as before.

11.15 Nonnegative sums of sums

Let X be a nonempty set, and let f be a nonnegative real-valued function on X. Also let I be a nonempty set, and let E_j be a nonempty subset of X for each $j \in I$. Suppose that

$$E_j \cap E_l = \emptyset \tag{11.15.1}$$

for every $j, l \in I$ with $j \neq l$, and put

$$E = \bigcup_{j \in I} E_j. \tag{11.15.2}$$

As usual,

$$\sum_{x \in E_j} f(x) \tag{11.15.3}$$

is defined as a nonnegative extended real number for each $j \in I$, so that (11.15.3) may be considered as a nonnegative extended real-valued function of j on I. Thus

$$\sum_{j \in I} \Big(\sum_{x \in E_j} f(x) \Big) \tag{11.15.4}$$

can also be defined as a nonnegative extended real number, as in Section 11.2. Of course,

$$\sum_{x \in E} f(x) \tag{11.15.5}$$

can be defined as a nonnegative extended real number as well. We would like to show that (11.15.4) is equal to (11.3.3). This is elementary when E is a finite set, which is the same as saying that I is a finite set, and that E_j is a finite set for each $j \in I$.

Let us first verify that (11.15.5) is less than or equal to (11.15.4). Let A be a nonempty finite subset of E. Put

$$I_A = \{j \in I : A \cap E_j \neq \emptyset\}, \tag{11.15.6}$$

which is a nonempty finite subset of I, and observe that

$$A = \bigcup_{j \in I_A} A \cap E_j. \tag{11.15.7}$$

Thus

$$\sum_{x \in A} f(x) = \sum_{j \in I_A} \Big(\sum_{x \in A \cap E_j} f(x) \Big), \tag{11.15.8}$$

as mentioned earlier. If $j \in I_A$, then

$$\sum_{x \in A \cap E_j} f(x) \le \sum_{x \in E_j} f(x), \tag{11.15.9}$$

because $A \cap E_j \subseteq E_j$. Combining this with (11.15.8), we get that

$$\sum_{x \in A} f(x) \le \sum_{j \in I_A} \Big(\sum_{x \in E_j} f(x) \Big). \tag{11.15.10}$$

It follows that

$$\sum_{x \in A} f(x) \le \sum_{j \in I} \Big(\sum_{x \in E_j} f(x) \Big), \tag{11.15.11}$$

because $I_A \subseteq I$. This implies that

$$\sum_{x \in E} f(x) \le \sum_{j \in I} \Big(\sum_{x \in E_j} f(x) \Big), \tag{11.15.12}$$

by the definition of the sum over E on the left.

Now let us check that (11.15.4) is less than or equal to (11.15.5). Let I_0 be a nonempty finite subset of I, and put

$$E(I_0) = \bigcup_{j \in I_0} E_j. \tag{11.15.13}$$

Observe that

$$\sum_{j \in I_0} \Big(\sum_{x \in E_j} f(x) \Big) = \sum_{x \in E(I_0)} f(x). \tag{11.15.14}$$

This follows from the analogous statement (11.5.13) for the union of two disjoint nonempty subsets of X. Hence

$$\sum_{j \in I_0} \Big(\sum_{x \in E_j} f(x) \Big) \le \sum_{x \in E} f(x), \tag{11.15.15}$$

because $E(I_0) \subseteq E$. It follows that

$$\sum_{j \in I} \Big(\sum_{x \in E} f(x) \Big) \le \sum_{x \in E} f(x), \tag{11.15.16}$$

by the definition of the sum over I on the left. Combining (11.15.12) and (11.15.16), we get that (11.15.4) is equal to (11.15.5), as desired.

11.16 Real and complex sums

Let X be a nonempty set again, and let f be a real or complex-valued summable function on X. As in the previous section, we let I be a nonempty set, and we let E_j be a nonempty subset of X for each $j \in I$. We suppose that the E_j's are pairwise disjoint, as in (11.15.1), and we let E be their union, as in (11.15.2). The sum (11.15.3) of $f(x)$ over $x \in E_j$ is now defined as a real or complex number for each $j \in I$, as appropriate. Similarly, the sum (11.15.5) of $f(x)$ over $x \in E$ is defined as a real or complex number, as appropriate.

If $j \in I$, then

$$\Big| \sum_{x \in E_j} f(x) \Big| \le \sum_{x \in E_j} |f(x)|, \tag{11.16.1}$$

as in (11.7.5) or (11.8.8), as appropriate. Thus

$$\sum_{j \in I} \Big| \sum_{x \in E_j} f(x) \Big| \le \sum_{j \in I} \Big(\sum_{x \in E_j} |f(x)| \Big), \tag{11.16.2}$$

where these sums over I are defined as nonnegative extended real numbers in the usual way. The remarks in the previous section imply that

$$\sum_{j\in I}\Big(\sum_{x\in E_j}|f(x)|\Big)=\sum_{x\in E}|f(x)|. \tag{11.16.3}$$

We also have that

$$\sum_{x\in E}|f(x)|\le\sum_{x\in X}|f(x)|<\infty, \tag{11.16.4}$$

because f is summable on X, by hypothesis. It follows that

$$\sum_{j\in I}\Big|\sum_{x\in E_j}f(x)\Big|<\infty. \tag{11.16.5}$$

This means that the sum (11.15.3) of $f(x)$ over $x\in E_j$ is summable as a real or complex-valued function of $j\in I$, as appropriate. Hence the sum (11.15.4) of this sum over $j\in I$ can be defined as a real or complex number, as appropriate. We would like to check that this sum is equal to the sum (11.15.5) of $f(x)$ over $x\in E$. If f is a nonnegative real-valued function on X, then this follows from the remarks in the previous section. If f is a real-valued summable function on X, then f can be expressed as the difference of two nonnegative real-valued summable functions on X, as in (11.7.3). In this case, the equality of (11.15.4) and (11.15.5) follows from the analogous statement for nonnegative real-valued summable functions on X. If f is a complex-valued summable function on X, then one can apply the previous statement to the real and imaginary parts of f.

Let Y and Z be nonempty sets, and suppose that $X=Y\times Z$ is their Cartesian product. If $f(y,z)$ is a nonnegative real-valued function on $Y\times Z$, then

$$\sum_{z\in Z}f(y,z) \tag{11.16.6}$$

defines a nonnegative extended real-valued function of $y\in Y$, and

$$\sum_{y\in Y}f(y,z) \tag{11.16.7}$$

defines a nonnegative extended real-valued function of $z\in Z$. Thus

$$\sum_{y\in Y}\Big(\sum_{z\in Z}f(y,z)\Big) \tag{11.16.8}$$

and

$$\sum_{z\in Z}\Big(\sum_{y\in Y}f(y,z)\Big) \tag{11.16.9}$$

are defined as nonnegative extended real numbers, as is

$$\sum_{(y,z)\in Y\times Z}f(y,z). \tag{11.16.10}$$

The remarks in the previous section imply that (11.16.8) is equal to (11.16.10). This corresponds to expressing $Y \times Z$ as the union of the pairwise-disjoint nonempty subsets of the form $\{y\} \times Z$, with $y \in Y$. Similarly, (11.16.9) is equal to (11.16.10), which corresponds to expressing $Y \times Z$ as the uion of the pairwise-disjoint nonempty subsets of the form $Y \times \{z\}$, with $z \in Z$. In particular, it follows that (11.16.8) is equal to (11.16.9).

Suppose now that $f(y, z)$ is a real or complex-valued summable function on $Y \times Z$. In this case, (11.16.6) defines a real or complex-valued function of $y \in Y$, and (11.16.7) defines a real or complex-valued function of $z \in Z$, as appropriate. These functions are summable on Y and Z, respectively, as in (11.16.5). Hence (11.16.8) and (11.16.9) are defined as real or complex numbers, as appropriate. The sum (11.16.10) is also defined as a real or complex number, as appropriate, and it is equal to (11.16.8) and (11.16.9), as before.

11.17 Square-summable functions

Let X be a nonempty set, and let f be a real or complex-valued function on X. If $|f(x)|^2$ is summable on X, then we say that f is *square-summable* on X. In this case, we put

$$\|f\|_2 = \Big(\sum_{x \in X} |f(x)|^2 \Big)^{1/2}, \tag{11.17.1}$$

using the nonnegative square root on the right side, as usual. If $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then $t\,f$ is square-summable on X too, because $|t\,f(x)|^2 = |t|^2\,|f(x)|^2$ is summable on X. We also have that

$$\|t\,f\|_2 = \Big(\sum_{x \in X} |t|^2\,|f(x)|^2 \Big)^{1/2} = |t|\,\|f\|_2. \tag{11.17.2}$$

Remember that

$$a\,b \le \frac{1}{2}\,(a^2 + b^2) \tag{11.17.3}$$

for all nonnegative real numbers a, b, as in (2.3.6). Let g be another real or complex-valued square-summable function on X, as appropriate. Observe that

$$|f(x)|\,|g(x)| \le \frac{1}{2}\,(|f(x)|^2 + |g(x)|^2) \tag{11.17.4}$$

for every $x \in X$, by (11.17.3). This implies that

$$\begin{aligned} \sum_{x \in X} |f(x)|\,|g(x)| &\le \sum_{x \in X} \frac{1}{2}\,(|f(x)|^2 + |g(x)|^2) \\ &= \frac{1}{2} \sum_{x \in X} |f(x)|^2 + \frac{1}{2} \sum_{x \in X} |g(x)|^2 = \frac{1}{2}\,\|f\|_2^2 + \frac{1}{2}\,\|g\|_2^2. \end{aligned} \tag{11.17.5}$$

In particular, $|f(x)|\,|g(x)|$ is summable on X when f and g are square-summable on X.

More precisely,

$$\sum_{x \in X} |f(x)|\,|g(x)| \le \|f\|_2\,\|g\|_2 \tag{11.17.6}$$

when f and g are square-summable on X. This is another version of the Cauchy–Schwarz inequality. If $\|f\|_2 = \|g\|_2 = 1$, then (11.17.6) follows from (11.17.5). If $\|f\|_2, \|g\|_2 > 0$, then one can reduce to the previous case, using (11.17.2). Otherwise, if $f(x) = 0$ for every $x \in X$, or $g(x) = 0$ for every $x \in X$, then (11.17.6) is trivial.

Note that

$$\begin{aligned} |f(x) + g(x)|^2 &\le (|f(x)| + |g(x)|)^2 \\ &= |f(x)|^2 + 2\,|f(x)|\,|g(x)| + |g(x)|^2 \end{aligned} \tag{11.17.7}$$

for every $x \in X$. Hence

$$\begin{aligned} &\sum_{x \in X} |f(x) + g(x)|^2 \\ &\qquad \le \sum_{x \in X} |f(x)|^2 + 2 \sum_{x \in X} |f(x)|\,|g(x)| + \sum_{x \in X} |g(x)|^2. \end{aligned} \tag{11.17.8}$$

If f and g are square-summable on X, then it follows that $f + g$ is square-summable on X, because $|f(x)|\,|g(x)|$ is summable on X, as before. Using (11.17.6), we get that

$$\|f + g\|_2^2 \le \|f\|_2^2 + 2\,\|f\|_2\,\|g\|_2 + \|g\|_2^2 = (\|f\|_2 + \|g\|_2)^2. \tag{11.17.9}$$

Equivalently, this means that

$$\|f + g\|_2 \le \|f\|_2 + \|g\|_2. \tag{11.17.10}$$

11.18 ℓ^2 Spaces

Let X be a nonempty set again, and let $\ell^2(X, \mathbf{R})$ and $\ell^2(X, \mathbf{C})$ be the spaces of real and complex-valued square-summable functions on X, respectively. These are linear subspaces of the real and complex vector spaces of all real and complex-valued functions on X, respectively, as in the previous section. Note that (11.17.1) defines a norm on each of $\ell^2(X, \mathbf{R})$ and $\ell^2(X, \mathbf{C})$, because of (11.17.2), (11.17.10), and the fact that $\|f\|_2 = 0$ if and only if $f(x) = 0$ for every $x \in X$. This implies that

$$d_2(f, g) = \|f - g\|_2 \tag{11.18.1}$$

defines a metric on each of $\ell^2(X, \mathbf{R})$ and $\ell^2(X, \mathbf{C})$, as usual.

If f is a real or complex-valued square-summable function on X, then

$$|f(x)| \le \|f\|_2 \tag{11.18.2}$$

for every $x \in X$, by the definition (11.17.1) of $\|f\|_2$. It follows that f is bounded on X, with

$$\|f\|_\infty \le \|f\|_2. \tag{11.18.3}$$

Here $\|f\|_\infty$ denotes the supremum norm of f, as in (1.13.4). If g is another real or complex-valued square-summable function on X, as appropriate, then we get that

$$d_\infty(f, g) \le d_2(f, g), \tag{11.18.4}$$

where $d_\infty(f, g)$ is the supremum metric, as in (1.13.8).

Let f be a real or complex-valued square-summable function on X again, so that $|f|^2$ is summable on X. This implies that $|f|^2$ vanishes at infinity on X, as in Section 11.4. Using this, it is easy to see that f vanishes at infinity on X as well. Thus

$$\ell^2(X, \mathbf{R}) \subseteq c_0(X, \mathbf{R}), \quad \ell^2(X, \mathbf{C}) \subseteq c_0(X, \mathbf{C}). \tag{11.18.5}$$

Let f be a real or complex-valued summable function on X, and remember that f is bounded on X, as in Section 11.6. Observe that

$$\sum_{x \in X} |f(x)|^2 \le \|f\|_\infty \sum_{x \in X} |f(x)| = \|f\|_\infty \, \|f\|_1 \le \|f\|_1^2, \tag{11.18.6}$$

where $\|f\|_1$ is the ℓ^1 norm of f, as in (11.6.3). This implies that f is square-summable on X, with

$$\|f\|_2 \le \|f\|_1. \tag{11.18.7}$$

Hence

$$\ell^1(X, \mathbf{R}) \subseteq \ell^2(X, \mathbf{R}), \quad \ell^1(X, \mathbf{C}) \subseteq \ell^2(X, \mathbf{C}). \tag{11.18.8}$$

If g is another real or complex-valued summable function on X, as appropriate, then we have that

$$d_2(f, g) \le d_1(f, g), \tag{11.18.9}$$

where $d_1(f, g)$ is the ℓ^1 metric, as in (11.6.6).

In particular, real or complex-valued functions on X with finite support are square-summable. Let us verify that $c_{00}(X, \mathbf{R})$, $c_{00}(X, \mathbf{C})$ are dense in $\ell^2(X, \mathbf{R})$, $\ell^2(X, \mathbf{C})$, respectively, with respect to the ℓ^2 metric (11.18.1). This is very similar to the analogous argument for ℓ^1 spaces, in Section 11.6. As before, there is nothing to do when X has only finitely many elements, and so we suppose that X is an infinite set. Let f be a real or complex-valued square-summable function on X, and let $\epsilon > 0$ be given. Because $|f|^2$ is summable on X, there is a finite subset $A(\epsilon)$ of X such that

$$\sum_{x \in X \setminus A(\epsilon)} |f(x)|^2 < \epsilon^2, \tag{11.18.10}$$

as in (11.4.10). Let f_ϵ be the real or complex-valued function on X, as appropriate, that is equal to f on $A(\epsilon)$ and to 0 on $X \setminus A(\epsilon)$. Thus f_ϵ has finite support in X, and $f - f_\epsilon$ is equal to 0 on $A(\epsilon)$, and to f on $X \setminus A(\epsilon)$. It follows that

$$\sum_{x \in X} |f(x) - f_\epsilon(x)|^2 = \sum_{x \in X \setminus A(\epsilon)} |f(x)|^2 < \epsilon^2, \tag{11.18.11}$$

so that

$$\|f - f_\epsilon\|_2 < \epsilon, \tag{11.18.12}$$

as desired.

11.19 Completeness of ℓ^2 spaces

Let X be a nonempty set, and let $\{f_j\}_{j=1}^\infty$ be a sequence of real or complex-valued functions on X that converges pointwise to a real or complex-valued function f on X, as appropriate. Suppose that the f_j's are square-summable on X, with bounded ℓ^2 norms, so that

$$\|f_j\|_2 \le C \tag{11.19.1}$$

for some nonnegative real number C and every $j \ge 1$. Equivalently, this means that

$$\sum_{x \in X} |f_j(x)|^2 \le C^2 \tag{11.19.2}$$

for every $j \ge 1$. Of course, $\{|f_j(x)|^2\}_{j=1}^\infty$ converges to $|f(x)|^2$ with respect to the standard Euclidean metric on $\mathbf{R}$ for every $x \in X$, by well-known results about convergent sequences of real or complex numbers. Under these conditions, f is a square-summable function on X, with

$$\sum_{x \in X} |f(x)|^2 \le C^2. \tag{11.19.3}$$

This follows from the remarks at the beginning of Section 11.12, applied to $|f_j|^2$. Of course, (11.19.3) is the same as saying that

$$\|f\|_2 \le C. \tag{11.19.4}$$

We would like to show that $\ell^2(X, \mathbf{R})$, $\ell^2(X, \mathbf{C})$ are complete with respect to the ℓ^2 metric (11.18.1). Let $\{f_j\}_{j=1}^\infty$ be a sequence of real or complex-valued square-summable functions on X that is a Cauchy sequence with respect to the ℓ^2 metric. Thus for each $\epsilon > 0$ there is an $L(\epsilon) \in \mathbf{Z}_+$ such that

$$\|f_j - f_l\|_2 < \epsilon \tag{11.19.5}$$

for every $j, l \ge L(\epsilon)$. If $x \in X$, then it follows that

$$|f_j(x) - f_l(x)| < \epsilon \tag{11.19.6}$$

for every $j, l \ge L(\epsilon)$, by (11.18.2). This implies that $\{f_j(x)\}_{j=1}^\infty$ is a Cauchy sequence in $\mathbf{R}$ or $\mathbf{C}$, as appropriate, with respect to the standard Euclidean metric.

Hence $\{f_j(x)\}_{j=1}^\infty$ converges in $\mathbf{R}$ or $\mathbf{C}$, as appropriate, for every $x \in X$, because $\mathbf{R}$ and $\mathbf{C}$ are complete as metric spaces. Let $f(x)$ be the limit of this sequence for each $x \in X$, so that f defines a real or complex-valued function on X, as appropriate. We want to verify that f is square-summable on X, and that $\{f_j\}_{j=1}^\infty$ converges to f with respect to the ℓ^2 metric. If X has only finitely many elements, then f is automatically square-summable on X, and convergence with respect to the ℓ^2 metric can be obtained from pointwise convergence.

Let $\epsilon > 0$ and $l \ge L(\epsilon)$ be given, and let us consider $\{f_j - f_l\}_{j=L(\epsilon)}^\infty$ as a sequence of square-summable functions on X that converges to $f - f_l$ pointwise on X. The remarks at the beginning of the section imply that $f - f_l$ is square-summable on X, with

$$\|f - f_l\|_2 \le \epsilon, \tag{11.19.7}$$

because of (11.19.5). In particular, this holds with $\epsilon = 1$ and $l = L(1)$. This implies that f is square-summable on X, because $f_{L(1)}$ is square-summable on X. Using (11.19.7) again, we get that $\{f_j\}_{j=1}^\infty$ converges to f with respect to the ℓ^2 metric, as desired.

11.20 Inner products on ℓ^2 spaces

Let X be a nonempty set, and let f, g be square-summable real-valued functions on X. Under these conditions, $|f(x)|\,|g(x)|$ is summable as a nonnegative real-valued function on X, as in Section 11.17. This means that $f(x)\,g(x)$ is summable as a real-valued function on X, so that

$$\langle f, g\rangle = \langle f, g\rangle_{\ell^2(X,\mathbf{R})} = \sum_{x \in X} f(x)\,g(x) \tag{11.20.1}$$

can be defined as a real number, as in Section 11.7. This is the standard inner product on $\ell^2(X, \mathbf{R})$. In particular,

$$\langle f, f\rangle_{\ell^2(X,\mathbf{R})} = \sum_{x \in X} f(x)^2 = \|f\|_2^2, \tag{11.20.2}$$

where $\|f\|_2$ is the ℓ^2 norm of f, as in (11.17.1). As before,

$$|\langle f, g\rangle_{\ell^2(X,\mathbf{R})}| = \Big|\sum_{x \in X} f(x)\,g(x)\Big| \le \sum_{x \in X} |f(x)|\,|g(x)| \le \|f\|_2\,\|g\|_2, \tag{11.20.3}$$

using the Cauchy–Schwarz inequality (11.17.6) in the third step. Of course, it is much easier to define this inner product on the space $c_{00}(X, \mathbf{R})$ of real-valued functions on X with finite support, by reducing the sum on the right side of (11.20.1) to a finite sum.

Note that

$$\langle f, g\rangle_{\ell^2(X,\mathbf{R})} = \sum_{x\in X} f(x)\, g(x) = \sum_{x\in X} g(x)\, f(x) = \langle g, f\rangle_{\ell^2(X,\mathbf{R})}. \tag{11.20.4}$$

If $t \in \mathbf{R}$, then $t\, f$ is square-summable on X too, and

$$\begin{aligned} \langle t\, f, g\rangle_{\ell^2(X,\mathbf{R})} &= \sum_{x\in X} t\, f(x)\, g(x) \\ &= t \sum_{x\in X} f(x)\, g(x) = t\, \langle f, g\rangle_{\ell^2(X,\mathbf{R})}. \end{aligned} \tag{11.20.5}$$

If f_1 and f_2 are square-summable real-valued functions on X, then we have seen in Section 11.17 that $f_1 + f_2$ is square-summable on X as well. In this case, we have that

$$\begin{aligned} \langle f_1 + f_2, g\rangle_{\ell^2(X,\mathbf{R})} &= \sum_{x\in X} (f_1(x) + f_2(x))\, g(x) \\ &= \sum_{x\in X} f_1(x)\, g(x) + \sum_{x\in X} f_2(x)\, g(x) \\ &= \langle f_1, g\rangle_{\ell^2(X,\mathbf{R})} + \langle f_2, g\rangle_{\ell^2(X,\mathbf{R})}, \end{aligned} \tag{11.20.6}$$

using the linearity of the sum in the second step. The analogous linearity properties of the inner product (11.20.1) in g can be obtained in the same way, or using (11.20.4).

Now let f, g be square-summable complex-valued functions on X. As before, $|f(x)|\, |g(x)|$ is summable as a nonnegative real-valued function on X, so that $f(x)\, \overline{g(x)}$ is summable as a complex-valued function on X. Hence

$$\langle f, g\rangle = \langle f, g\rangle_{\ell^2(X,\mathbf{C})} = \sum_{x\in X} f(x)\, \overline{g(x)} \tag{11.20.7}$$

can be defined as a complex number, as in Section 11.8, which is the standard inner product on $\ell^2(X, \mathbf{C})$. As usual,

$$\langle f, f\rangle_{\ell^2(X,\mathbf{C})} = \sum_{x\in X} |f(x)|^2 = \|f\|_2^2, \tag{11.20.8}$$

where $\|f\|_2$ is the ℓ^2 norm of f. We also have that

$$|\langle f, g\rangle_{\ell^2(X,\mathbf{C})}| = \left|\sum_{x\in X} f(x)\, \overline{g(x)}\right| \le \sum_{x\in X} |f(x)|\, |g(x)| \le \|f\|_2\, \|g\|_2, \tag{11.20.9}$$

using the Cauchy–Schwarz inequality (11.17.6) in the third step again.

In this situation, we have that

$$\overline{\langle f, g\rangle_{\ell^2(X,\mathbf{C})}} = \overline{\Big(\sum_{x\in X} f(x)\, \overline{g(x)}\Big)} = \sum_{x\in X} g(x)\, \overline{f(x)} = \langle g, f\rangle_{\ell^2(X,\mathbf{C})}. \tag{11.20.10}$$

If $t \in \mathbf{C}$, then $t\, f$ is square-summable on X, and it is easy to see that

$$\langle t\, f, g \rangle_{\ell^2(X, \mathbf{C})} = t\, \langle f, g \rangle_{\ell^2(X, \mathbf{C})}, \tag{11.20.11}$$

as before. Similarly, $t\, g$ is square-summable on X, and

$$\langle f, t\, g \rangle_{\ell^2(X, \mathbf{C})} = \bar{t}\, \langle f, g \rangle_{\ell^2(X, \mathbf{C})}, \tag{11.20.12}$$

by the same type of argument, or using (11.20.10). If f_1 and f_2 are square-summable complex-valued functions on X, then $f_1 + f_2$ are square-summable on X, and

$$\langle f_1 + f_2, g \rangle_{\ell^2(X, \mathbf{C})} = \langle f_1, g \rangle_{\ell^2(X, \mathbf{C})} + \langle f_2, g \rangle_{\ell^2(X, \mathbf{C})}, \tag{11.20.13}$$

as in the real case. The analogous additivity property of the inner product (11.20.7) in g can be obtained in the same way, or using (11.20.10).

Chapter 12

Some additional topics

12.1 Lebesgue measure and integration

Let a, b be real numbers with $a < b$, and remember that

$$d_1(f, g) = \int_a^b |f(x) - g(x)| \, dx \tag{12.1.1}$$

defines a metric on the space $C([a, b], \mathbf{R})$ of continuous real-valued functions on $[a, b]$, as in Section 3.1. It is not difficult to see that this space is not complete with respect to (12.1.1), as in Section A.5. In order to get completeness, one can use the *Lebesgue integral*. Although we shall not discuss this in detail here, let us mention a few other topics related to Lebesgue's theory of measure and integration.

Let $\{f_j\}_{j=1}^\infty$ be a sequence of continuous real-valued functions on $[a, b]$, and let f be another continuous real-valued function on $[a, b]$. If $\{f_j\}_{j=1}^\infty$ converges to f uniformly on $[a, b]$, then it is easy to see that $\{f_j\}_{j=1}^\infty$ converges to f with respect to (12.1.1). In particular, this implies that

$$\lim_{j \to \infty} \int_a^b f_j(x) \, dx = \int_a^b f(x) \, dx, \tag{12.1.2}$$

because

$$\left| \int_a^b f_j(x) \, dx - \int_a^b f(x) \, dx \right| \le \int_a^b |f_j(x) - f(x)| \, dx \tag{12.1.3}$$

for every $j \ge 1$. However, one can give examples to show that this does not work if we only ask that $\{f_j\}_{j=1}^\infty$ converge to f pointwise on $[a, b]$, as in Section A.4.

Suppose now that $\{f_j\}_{j=1}^\infty$ is uniformly bounded on $[a, b]$, in the sense that there is a nonnegative real number A such that

$$|f_j(x)| \le A \tag{12.1.4}$$

for every $x \in [a,b]$ and $j \geq 1$. If $\{f_j\}_{j=1}^\infty$ converges to f pointwise on $[a,b]$, then a classical theorem of Arzelà and Osgood implies that (12.1.2) holds, as in [37, 69, 122, 126, 180]. More precisely, this works when the f_j's and f are Riemann integrable on $[a,b]$, instead of continuous. One can also get that $\{f_j\}_{j=1}^\infty$ converges to f with respect to (12.1.1), by considering $\{|f_j - f|\}_{j=1}^\infty$ in the previous statement. Questions like these can be treated more extensively using the Lebesgue integral.

If f is a bounded real-valued function on $[a,b]$, then a famous theorem states that f is Riemann integrable on $[a,b]$ if and only if f is continuous at "almost every" point in $[a,b]$ with respect to Lebesgue measure. This means that the set of points in $[a,b]$ at which f is not continuous has Lebesgue measure equal to 0. This is discussed in many textbooks, as well as the articles [24, 147]. Some related results can be found in [121, 134], and also in in [177], for Riemann–Stieltjes integrability.

Now let f be a monotonically increasing real-valued function on $[a,b]$. It is well known that the one-sided limits of f exist at every point in (a,b), as well as the appropriate one-sided limits at the endpoints a, b. This can be used to show that f is continuous at all but finitely or countably many points in $[a,b]$. A famous theorem states that f is also differentiable at almost every point in $[a,b]$, with respect to Lebesgue measure. Of course, the derivative is nonnegative when it exists, because of monotonicity.

One can show that

$$\int_a^b f'(x)\,dx \leq f(b) - f(a), \tag{12.1.5}$$

where the integral on the left is defined as a Lebesgue integral. If f has a jump discontinuity at any point in $[a,b]$, then the inequality in (12.1.5) is strict. There are also examples where f is both continuous and monotonically increasing on $[a,b]$, and the inequality in (12.1.5) is still strict.

Suppose that f is a real-valued Lipschitz function on $[a,b]$, with respect to the standard Euclidean metric on $\mathbf{R}$ and its restriction to $[a,b]$, but not necessarily monotonic. Another famous theorem states that f is differentiable almost everywhere with respect to Lebesgue measure on $[a,b]$. It is easy to see that

$$|f'(x)| \leq \mathrm{Lip}(f), \tag{12.1.6}$$

when the derivative exists. This permits the integral on the left side of (12.1.5) to be defined as a Lebesgue integral, and in fact one has

$$\int_a^b f'(x)\,dx = f(b) - f(a) \tag{12.1.7}$$

in this situation. If $f$is a real-valued Lipschitz function on $\mathbf{R}^n$ for some $n \in \mathbf{Z}_+$, with respect to the standard Euclidean metrics on $\mathbf{R}^n$ and $\mathbf{R}$, then another famous theorem states that f is differentiable almost everywhere with respect to n-dimensional Lebesgue measure on $\mathbf{R}^n$.

12.2 Banach spaces

We shall now suppose that the reader has some familiarity with abstract vector spaces and related notions. Let V be a vector space over the real or complex numbers. Thus V is a set, on which operations of addition and scalar multiplication over $\mathbf{R}$ or $\mathbf{C}$ have been defined. These operations should satisfy a number of standard conditions, such as associativity and commutativity of addition, and compatibility of scalar multiplication with addition on V and addition and multiplication on $\mathbf{R}$ or $\mathbf{C}$, as appropriate. There should also be a distinguished additive element in V, denoted 0.

As usual, a nonnegative real-valued function N on V is said to be a *norm* on V if it satisfies the following three conditions. First, $N(v) = 0$ if and only if $v = 0$. Second,

$$N(t\, v) = |t|\, N(v) \tag{12.2.1}$$

for every $v \in V$ and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate. Third,

$$N(v + w) \le N(v) + N(w) \tag{12.2.2}$$

for every $v, w \in V$. In this case, it is easy to see that

$$d_N(v, w) = N(v - w) \tag{12.2.3}$$

defines a metric on V. If V is also complete as a metric space with respect to (12.2.3), then V is said to be a *Banach space* with respect to N. Otherwise, one can pass to a suitable completion of V, but we shall get into that now.

Suppose that $\{v_j\}_{j=1}^\infty$, $\{w_j\}_{j=1}^\infty$ are sequences of elements of V that converge to $v, w \in V$, respectively, with respect to (12.2.3). Under these conditions, one can check that $\{v_j + w_j\}_{j=1}^\infty$ converges to $v + w$ in V with respect to (12.2.3). Similarly, suppose that $\{t_j\}_{j=1}^\infty$ is also a sequence of real or complex numbers that converges to a real or complex number t, with respect to the standard metric on $\mathbf{R}$ or $\mathbf{C}$, as appropriate. In this case, one can verify that $\{t_j\, v_j\}_{j=1}^\infty$ converges to $t\, v$ in V with respect to (12.2.3). The proofs of these statements are analogous to those for the corresponding facts about sums and products of convergent sequences of real and complex numbers.

Let W be a *linear subspace* of V. This means that W is a subset of V that contains 0 and satisfies the following two properties. First, if $v, w \in W$, then

$$v + w \in W. \tag{12.2.4}$$

Second, if $v \in W$ and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then

$$t\, v \in W. \tag{12.2.5}$$

It follows that W is also a vector space over $\mathbf{R}$ or $\mathbf{C}$, as appropriate, with respect to the restriction of the vector space operations on V to W. If N is a norm on V, then it is easy to see that the restriction of N to W defines a norm on W. Of course, the metric on W associated to the restriction of N to W is the same as the restriction to W of the metric (12.2.3) associated to N on V. If V is a Banach space with respect to N and W is a closed set in V with respect to (12.2.3), then W is a Banach space with respect to the restriction of N to W. More precisely, W is complete with respect to the restriction of (12.2.3) to W in this situation.

12.3 Hilbert spaces

Let V be a vector space over the real numbers. A real-valued function $\langle v, w\rangle$ defined for $v, w \in V$ is said to be an *inner product* on V if it satisfies the following three conditions. First, for each $w \in V$, $\langle v, w\rangle$ should be linear in V, as a mapping from V into $\mathbf{R}$. This means that

$$\langle v + v', w\rangle = \langle v, w\rangle + \langle v', w\rangle \tag{12.3.1}$$

for every $v, v' \in V$, and that

$$\langle t\, v, w\rangle = t\, \langle v, w\rangle \tag{12.3.2}$$

for every $v \in V$ and $t \in \mathbf{R}$. Second, $\langle v, w\rangle$ should be symmetric in v and w, so that

$$\langle v, w\rangle = \langle w, v\rangle \tag{12.3.3}$$

for every $v, w \in V$. This implies that $\langle v, w\rangle$ is linear in w for every $v \in V$, because of the linearity in v. Third,

$$\langle v, v\rangle > 0 \tag{12.3.4}$$

for every $v \in V$ with $v \neq 0$. Of course, $\langle v, w\rangle = 0$ whenever $v = 0$ or $w = 0$, because of linearity in v and w.

Similarly, if V is a vector space over the complex numbers, then a complex-valued function $\langle v, w\rangle$ defined for $v, w \in V$ is said to be an *inner product* if it satisfies the following three conditions. First, for every $w \in V$, $\langle v, w\rangle$ should be (complex) linear in v, as a mapping from V into $\mathbf{C}$. This means that (12.3.1) should hold for every $v, v' \in V$, as before, and that (12.3.2) should hold for every $v \in V$ and $t \in \mathbf{C}$. Second, $\langle v, w\rangle$ should be Hermitian symmetric in v and w, which means that

$$\langle w, v\rangle = \overline{\langle v, w\rangle} \tag{12.3.5}$$

for every $v, w \in V$, where $\overline{a}$ is the complex conjugate of a complex number a. Combining this with the first condition, we get that $\langle v, w\rangle$ is conjugate-linear in w for each $v \in V$, so that

$$\langle v, w + w'\rangle = \langle v, w\rangle + \langle v, w'\rangle \tag{12.3.6}$$

for every $w, w' \in V$, and

$$\langle v, t\, w\rangle = \overline{t}\, \langle v, w\rangle \tag{12.3.7}$$

for every $w \in V$ and $t \in \mathbf{C}$. Observe that

$$\overline{\langle v, v \rangle} = \langle v, v \rangle \tag{12.3.8}$$

for every $v \in V$, which means that $\langle v, v \rangle \in \mathbf{R}$. The third condition is that (12.3.4) hold for every $v \in V$ with $v \neq 0$ again. As before, $\langle v, w \rangle = 0$ whenever $v = 0$ or $w = 0$.

Let $(V, \langle v, w \rangle)$ be a real or complex inner product space. Put

$$\|v\| = \langle v, v \rangle^{1/2} \tag{12.3.9}$$

for every $v \in V$, using the nonnegative square root on the right side. It is well known that

$$|\langle v, w \rangle| \leq \|v\| \, \|w\| \tag{12.3.10}$$

for every $v, w \in V$, which is another version of the Cauchy–Schwarz inequality. More precisely, this can be shown using the fact that

$$\langle v + t\, w, v + t\, w \rangle = \|v + t\, w\|^2 \geq 0 \tag{12.3.11}$$

for every $v, w \in V$ amd $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate. Note that

$$\|t\, v\| = |t| \, \|v\| \tag{12.3.12}$$

for every $v \in V$ and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate. One can also check that

$$\|v + w\| \leq \|v\| + \|w\| \tag{12.3.13}$$

for every $v, w \in V$, using (12.3.10). Thus (12.3.9) defines a norm on V as a vector space over $\mathbf{R}$ or $\mathbf{C}$, as appropriate.

If V is complete with respect to the metric associated to the norm (12.3.9), then V is said to be a *Hilbert space* with respect to the inner product $\langle v, w \rangle$. Otherwise, one can pass to a completion of V, as before.

12.4 Infinite series in Banach spaces

Let V be a vector space over the real or complex numbers, and let N be a norm on V. An infinite series

$$\sum_{j=1}^{\infty} v_j \tag{12.4.1}$$

with terms v_j in V for each $j \geq 1$ is said to *converge* in V with respect to N if the corresponding sequence of partial sums

$$\sum_{j=1}^{l} v_j \tag{12.4.2}$$

converges to an element of V with respect to the metric d_N associated to N. In this case, the value of the sum (12.4.1) is defined to be the limit of the sequence

(12.4.2). If (12.4.1) converges in V, and if $\sum_{j=1}^\infty w_j$ is another infinite series of elements of V that converges in V, then $\sum_{j=1}^\infty (v_j + w_j)$ converges in V too, with

$$\sum_{j=1}^\infty (v_j + w_j) = \sum_{j=1}^\infty v_j + \sum_{j=1}^\infty w_j. \tag{12.4.3}$$

This follows from the analogous statement for sums of convergent sequences in V, applied to the partial sums of these series. Similarly, if (12.4.1) converges in V, and $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate, then $\sum_{j=1}^\infty t\, v_j$ converges in V as well, with

$$\sum_{j=1}^\infty t\, v_j = t \sum_{j=1}^\infty v_j. \tag{12.4.4}$$

This uses the fact that t times a convergent sequence in V converges to t times the limit of the sequence.

The condition that the sequence of partial sums (12.4.2) of an infinite series (12.4.1) be a Cauchy sequence with respect to the metric d_N associated to N is equivalent to saying that for each $\epsilon > 0$ there is a positive integer L such that

$$N\Big(\sum_{j=k}^l v_j\Big) < \epsilon \tag{12.4.5}$$

for every $l \geq k \geq L$. In particular, this holds when (12.4.1) converges in V, because a convergent sequence in any metric space is a Cauchy sequence. Note that the Cauchy condition for the partial sums implies that

$$\lim_{l \to \infty} N(v_l) = 0, \tag{12.4.6}$$

by taking $k = l$ in (12.4.5). This is the same as saying that $\{v_j\}_{j=1}^\infty$ converges to 0 in V, with respect to d_N. If V is a Banach space with respect to N, then the Cauchy condition for the partial sums implies that the series converges in V.

An infinite series (12.4.1) with terms in V is said to be *absolutely convergent* with respect to N if

$$\sum_{j=1}^\infty N(v_j) \tag{12.4.7}$$

converges as an infinite series of nonnegative real numbers. Observe that

$$N\Big(\sum_{j=k}^l v_j\Big) \leq \sum_{j=k}^l N(v_j) \tag{12.4.8}$$

for every $l \geq k \geq 1$, by the triangle inequality for N. If (12.4.1) converges absolutely with respect to N, then the sequence of partial sums (12.4.2) is a Cauchy sequence with respect to d_N. More precisely, one verify (12.4.5) using

(12.4.8) and the analogous Cauchy condition for the partial sums of (12.4.7). If V is a Banach space with respect to N, then it follows that (12.4.1) converges in V.

Now let $(V, \langle v, w \rangle)$ be a real or complex inner product space, and let us use the corresponding norm $\| \cdot \|$, as in the previous section. A pair v, w of vectors in V are said to be *orthogonal* if

$$\langle v, w \rangle = 0. \tag{12.4.9}$$

In this case, it is easy to see that

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2. \tag{12.4.10}$$

Let (12.4.1) be an infinite series of pairwise-orthogonal vectors in V, so that v_j is orthogonal to v_k when $j \neq k$. This implies that

$$\left\| \sum_{j=k}^{l} v_j \right\|^2 = \sum_{j=k}^{l} \|v_j\|^2 \tag{12.4.11}$$

for every $l \geq k \geq 1$. If

$$\sum_{j=1}^{\infty} \|v_j\|^2 \tag{12.4.12}$$

converges as an infinite series of nonnegative real numbers, then the sequence (12.4.2) is a Cauchy sequence in V. This uses (12.4.11) to get the Cauchy condition (12.4.5) from the analogous Cauchy condition for the partial sums of (12.4.12). If V is a Hilbert space with respect to $\langle \cdot, \cdot \rangle$, then it follows that (12.4.1) converges in V.

12.5 Bounded linear mappings

Let V and W be vector spaces, where more precisely V and W should both be defined over the real numbers, or both defined over the complex numbers. Also let N_V and N_W be norms on V and W, respectively. Thus

$$d_V(v, v') = N_V(v - v') \quad \text{and} \quad d_W(w, w') = N_W(w, w') \tag{12.5.1}$$

define metrics on V and W, respectively. A linear mapping T from V into W is said to be *bounded* with respect to N_V and N_W if there is a nonnegative real number C such that

$$N_W(T(v)) \leq C\, N_V(v) \tag{12.5.2}$$

for every $v \in V$. In this case, we have that

$$\begin{aligned} d_W(T(u), T(v)) = N_W(T(u) - T(v)) &= N_W(T(u - v)) \\ &\leq C\, N_V(u - v) = C\, d_V(u, v) \end{aligned} \tag{12.5.3}$$

for every $u, v \in V$, so that T is Lipschitz with respect to the corresponding metrics on V and W.

Conversely, if a linear mapping T from V into W is Lipschitz with respect to the metrics $d_V(\cdot, \cdot)$ and $d_W(\cdot, \cdot)$, then it is easy to see that T is bounded with respect to N_V and N_W. More precisely, if a linear mapping T from V into W has the property that $N_W(T(v))$ is bounded on a ball in V of positive radius, then one can check that T is bounded as a linear mapping. This uses scalar multiplication to obtain (12.5.2) from the boundedness of $N_W(T(v))$ when $N_V(v)$ is less than a fixed radius. In particular, if T is continuous at 0 with respect to the metrics $d_V(\cdot, \cdot)$ and $d_W(\cdot, \cdot)$ on V and W, then there is a $\delta > 0$ such that $N_W(T(v)) < 1$ when $v \in V$ satisfies $N_V(v) < 1$. This implies that T is a bounded as a linear mapping, as before.

If V is a finite-dimensional vector space, then every linear mapping T from V into W is bounded. Of course, this is trivial when $V = \{0\}$. Otherwise, there is a positive integer n such that V is isomorphic to $\mathbf{R}^n$ or $\mathbf{C}^n$ as a vector space over the real or complex numbers, as appropriate. This permits one to reduce to the same type of arguments as in Section 7.3.

Suppose that $V \neq \{0\}$, and that T is a bounded linear mapping from V into W with respect to N_V and N_W. The corresponding *operator norm* is defined by

$$\|T\|_{op} = \sup\left\{\frac{N_W(T(v))}{N_V(v)} : v \in V,\, v \neq 0\right\}, \tag{12.5.4}$$

where the finiteness of the supremum follows from the boundedness of T. If $V = \{0\}$, then one can simply take $\|T\|_{op} = 0$. Equivalently, $\|T\|_{op}$ is the same as the infimum of the nonnegative real numbers C such that (12.5.2) holds. As in Section 7.7, $\|T\|_{op}$ is also the same as $\mathrm{Lip}(T)$, defined with respect to the metrics $d_V(\cdot, \cdot)$ and $d_W(\cdot, \cdot)$ associated to N_V and N_W.

If a is a real or complex number, as appropriate, then $a\,T$ also defines a linear mapping from V into W, where $(a\,T)(v) = a\,T(v)$ for every $v \in V$. It is easy to see that $a\,T$ is a bounded linear mapping when T is, with

$$\|a\,T\|_{op} = |a|\,\|T\|_{op}. \tag{12.5.5}$$

Note that $\|T\|_{op} = 0$ if and only if $T = 0$, which is to say that $T(v) = 0$ for every $v \in V$. If T_1 and T_2 are linear mappings from V into W, then $T_1 + T_2$ defines a linear mapping from V into W as well, with

$$(T_1 + T_2)(v) = T_1(v) + T_2(v). \tag{12.5.6}$$

If T_1, T_2 are bounded as linear mappings with respect to N_V and N_W, then one can check that $T_1 + T_2$ is bounded as well, with

$$\|T_1 + T_2\|_{op} \le \|T_1\|_{op} + \|T_2\|_{op}. \tag{12.5.7}$$

The space of linear mappings from V into W is a vector space over the real or complex numbers, as appropriate, with respect to pointwise addition and scalar multiplication of mappings. The space of bounded linear mappings from

V into W is a linear subspace of the space of all linear mappings from V into W, as in the preceding paragraph. The operator norm (12.5.4) defines a norm on the space of bounded linear mappings from V into W, as a vector space over the real or complex numbers, as appropriate.

Let Z be another vector space, which is defined over the real or complex numbers, depending on whether V and W are defined over the real or complex numbers. Also let N_Z be a norm on Z, let T_1 be a bounded linear mapping from V into W, and let T_2 be a bounded linear mapping from W into Z. Under these conditions, one can verify that the composition $T_2 \circ T_1$ is bounded as a linear mapping from V into Z, with

$$\|T_2 \circ T_1\|_{op} \le \|T_1\|_{op} \, \|T_2\|_{op}. \tag{12.5.8}$$

Here the various operator norms are defined with respect to the given norms on V, W, and Z, as appropriate.

Appendix A

Some more on mappings, metrics, and norms

A.1 A nice inequality

Let a be a positive real number, with $a \le 1$. If r and t are nonnegative real numbers, then it is well known that

$$(r + t)^a \le r^a + t^a. \tag{A.1.1}$$

To see this, observe first that

$$\max(r, t) \le (r^a + t^a)^{1/a}. \tag{A.1.2}$$

Using this, we get that

$$r + t \le \max(r, t)^{1-a}\,(r^a + t^a) \le (r^a + t^a)^{((1-a)/a)+1} = (r^a + t^a)^{1/a}. \tag{A.1.3}$$

This is equivalent to (A.1.1).

Let $(X, d(x, y))$ be a metric space. If $0 < a \le 1$, then it is easy to see that

$$d(x, y)^a \tag{A.1.4}$$

is a metric on X as well, using (A.1.1). This was mentioned in Section 1.1 when $a = 1/2$.

Let us use $B_d(x, r)$, $\overline{B}_d(x, r)$ to denote the open and closed balls in X centered at $x \in X$ with radius r with respect to $d(\cdot, \cdot)$, respectively, as in Section 1.9, and $B_{d^a}(x, r)$, $\overline{B}_{d^a}(x, r)$ for the open and closed balls in X centered at x with radius r with respect to $d(\cdot, \cdot)^a$, respectively. Observe that

$$B_{d^a}(x, r^a) = B_d(x, r) \tag{A.1.5}$$

for every $r > 0$, and that

$$\overline{B}_{d^a}(x, r^a) = \overline{B}_d(x, r) \tag{A.1.6}$$

for every $r \geq 0$.

It is easy to see that the identity mapping on X is uniformly continuous as a mapping from X equipped with $d(\cdot,\cdot)$ into X equipped with $d(\cdot,\cdot)^a$. Similarly, the identity mapping on X is uniformly continuous as a mapping from X equipped with $d(\cdot,\cdot)^a$ into X equipped with $d(\cdot,\cdot)$. These statements were also mentioned in Section 1.1 when $a = 1/2$.

If $a > 1$, then one can verify that

$$|x - y|^a \tag{A.1.7}$$

is not a metric on the real line. This corresponds to the first part of Exercise 11 at the end of Chapter 2 in [155] when $a = 2$.

A.2 Some more Lipschitz conditions

Let (X, d_X) and (Y, d_Y) be metric spaces, and let α be a positive real number. A mapping f from X into Y is said to be *Lipschitz of order* α if there is a nonnegative real number C such that

$$d_Y(f(x), f(w)) \leq C\, d_X(x, w)^\alpha \tag{A.2.1}$$

for every $x, w \in X$. This is the same as a Lipschitz mapping as in Section 1.1 when $\alpha = 1$. One may also say that f is *Hölder continuous of order* α in this case.

Lipschitz mappings of any order are uniformly continuous, as before. Of course, (A.2.1) holds with $C = 0$ if and only if f is constant on X.

Let a be a positive real number, and suppose that

$$d_X(x, w)^a \text{ is a metric on } X. \tag{A.2.2}$$

This holds automatically when $a \leq 1$, as in the previous section. Note that (A.2.1) is the same as saying that

$$d_Y(f(x), f(w)) \leq C\, (d_X(x, w)^a)^{\alpha/a} \tag{A.2.3}$$

for every $x, y \in X$. This means that

$$f \text{ is Lipschitz of order } \alpha \text{ with respect to } d_X \text{ on } X \tag{A.2.4}$$

if and only if

$$f \text{ is Lipschitz of order } \alpha/a \text{ with respect to } d_X(\cdot,\cdot)^a \text{ on } X, \tag{A.2.5}$$

and with the same constant C.

Similarly, let b be a positive real number, and suppose that

$$d_Y(\cdot,\cdot)^b \text{ is a metric on } Y. \tag{A.2.6}$$

Clearly (A.2.1) is the same as saying that

$$d_Y(f(x), f(w))^b \le C^b \, d_X(x, w)^{\alpha\, b} \tag{A.2.7}$$

for every $x, y \in X$. Thus

(A.2.8) f is Lipschitz of order α with constant C with respect to d_Y on Y

if and only if

(A.2.9) f is Lipschitz of order $\alpha\, b$ with constant C^b with respect to $d_Y(\cdot, \cdot)^b$ on Y.

Suppose that $X = Y = \mathbf{R}$, with the standard Euclidean metric. If f is Lipschitz of order $\alpha > 1$, then

(A.2.10) f is constant on $\mathbf{R}$.

More precisely, it is easy to see that the derivative of f is equal to 0 at every point in $\mathbf{R}$ under these conditions.

A.3 Another nice inequality

Let X be a nonempty set, and let f be a real or complex-valued function on X with finite support, as in Section 1.6. If p is a positive real number, then put

$$\|f\|_p = \Big(\sum_{x \in X} |f(x)|^p\Big)^{1/p}. \tag{A.3.1}$$

This is the same as in Section 1.6 when $p = 1$ or 2. If $X = \{1, \ldots, n\}$ for some positive integer n, then this corresponds to an analogous expression in Section 1.4. We also put

$$\|f\|_\infty = \max_{x \in X} |f(x)|, \tag{A.3.2}$$

as before.

Observe that

$$\|f\|_\infty \le \|f\|_p. \tag{A.3.3}$$

If $0 < p_1 \le p_2 < \infty$, then we would like to check that

$$\|f\|_{p_2} \le \|f\|_{p_1}. \tag{A.3.4}$$

Clearly

$$\|f\|_{p_2}^{p_2} = \sum_{x \in X} |f(x)|^{p_2} \le \|f\|_\infty^{p_2 - p_1} \sum_{x \in X} |f(x)|^{p_1} = \|f\|_\infty^{p_2 - p_1} \, \|f\|_{p_1}^{p_1}. \tag{A.3.5}$$

This implies that

$$\|f\|_{p_2}^{p_2} \le \|f\|_{p_1}^{p_2}, \tag{A.3.6}$$

because of (A.3.3). It follows that (A.3.4) holds, as desired.

It is easy to see that (A.1.1) follows from (A.3.4), with $p_1 = a$, $p_2 = 1$, and where X has two elements.

Observe that

$$\|f\|_p \le (\#\operatorname{supp} f)^{1/p}\,\|f\|_\infty, \tag{A.3.7}$$

where $\#\operatorname{supp} f$ is the number of elements in the support of f. One can use this and (A.3.3) to get that

$$\|f\|_p \to \|f\|_\infty \text{ as } p \to \infty. \tag{A.3.8}$$

Of course, $\|f\|_p = 0$ if and only if $f = 0$ on X, and

$$\|t\,f\|_p = |t|\,\|f\|_p \tag{A.3.9}$$

for every $t \in \mathbf{R}$ or $\mathbf{C}$, as appropriate. If g is another real or complex-valued function on X with finite support and $0 < p \le 1$, then

$$\begin{aligned} \|f+g\|_p^p &= \sum_{x\in X} |f(x)+g(x)|^p \le \sum_{x\in X} (|f(x)|+|g(x)|)^p \\ &\le \sum_{x\in X} (|f(x)|^p + |g(x)|^p) = \sum_{x\in X} |f(x)|^p + \sum_{x\in X} |g(x)|^p \\ &= \|f\|_p^p + \|g\|_p^p, \end{aligned} \tag{A.3.10}$$

using (A.1.1) in the third step, with $a = p$. This implies that

$$\|f-g\|_p^p \tag{A.3.11}$$

defines a metric on $c_{00}(X, \mathbf{R})$ and $c_{00}(X, \mathbf{C})$ when $0 < p \le 1$.

A.4 Some functions on $[0,1]$

Let j be a positive integer, and let f_j be the real-valued function defined on $[0,1]$ by

$$\begin{aligned} f_j(x) &= 2\,j\,x && \text{when } 0 \le x \le 1/(2\,j) \\ &= 2 - 2\,j\,x && \text{when } 1/(2\,j) \le x \le 1/j \\ &= 0 && \text{when } 1/j \le x \le 1. \end{aligned} \tag{A.4.1}$$

Thus

$$f_j(0) = 0,\ f_j(1/(2\,j)) = 1, \text{ and } f(1/j) = 0, \tag{A.4.2}$$

with the overlapping definitions of $f_j(x)$ agreeing at $x = 1/(2\,j), 1/j$. Equivalently, f_j is defined to be linear on $[0, 1/(2\,j)]$ and $[1/(2\,j), 1/j]$, with these values at the endpoints. Note that f_j is continuous on $[0,1]$. One can check that

$$\{f_j\}_{j=1}^\infty \text{ converges to 0 pointwise on } [0,1], \tag{A.4.3}$$

and not uniformly.

In fact,

$$\|f_j\|_\infty = 1 \tag{A.4.4}$$

for every j, where $\|\cdot\|_\infty$ is the supremum norm on $C([0,1],\mathbf{R})$, as before. Observe that

$$\|f_j\|_1 = \int_0^1 f_j(x)\,dx = 2\int_0^{1/(2\,j)} f_j(x)\,dx = 2\,j\,(1/(2\,j))^2 = 1/(2\,j) \tag{A.4.5}$$

for every j, where $\|\cdot\|_1$ is as in (1.15.4). More precisely,

$$f_j(x) = f_j((1/j) - x) \quad \text{when } 0 \le x \le 1/j, \tag{A.4.6}$$

which is to say that $f_j(x)$ is symmetric about $x = 1/(2\,j)$ on $[0,1/j]$. This implies that the integrals of $f_j(x)$ over $[0,1/(2\,j)]$ and $[1/(2\,j),1/j]$ are the same. Of course, (A.4.5) is the same as the area of the triangle determined by the graph of f_j on $[0,1/j]$.

Similarly,

$$\begin{aligned}\|f_j\|_2^2 = \int_0^1 f_j(x)^2\,dx &= 2\int_0^{1/(2\,j)} f_j(x)^2\,dx \\ &= (2/3)\,(2\,j)^2\,(1/(2\,j))^3 = 1/(3\,j)\end{aligned} \tag{A.4.7}$$

for every j, where $\|\cdot\|_2$ is as in (1.15.6). This uses (A.4.6) in the second step, to get that the integrals of $f_j(x)^2$ over $[0,1/(2\,j)]$ and $[1/(2\,j),1/j]$ are the same. It follows that

$$\|f_j\|_2 = 1/(3\,j)^{1/2} \tag{A.4.8}$$

for each j.

If $\alpha \in \mathbf{R}$, then one can verify that

$$\{j^{-\alpha}\,f_j\}_{j=1}^\infty \text{ converges to 0 pointwise on } [0,1]. \tag{A.4.9}$$

Note that

$$\|j^{-\alpha}\,f_j\|_\infty = j^{-\alpha} \tag{A.4.10}$$

for every j, by (A.4.4). This implies that $\{j^{-\alpha}\,f_j\}_{j=1}^\infty$ is bounded with respect to the supremum metric exactly when $\alpha \ge 0$, and that $\{f_j\}_{j=1}^\infty$ converges to 0 with respect to the supremum metric exactly when $\alpha > 0$.

Similarly,

$$\|j^{-\alpha}\,f_j\|_1 = (1/2)\,j^{-1-\alpha} \tag{A.4.11}$$

for every j, by (A.4.5). This means that $\{j^{-\alpha}\,f_j\}_{j=1}^\infty$ is bounded with respect to the metric d_1 on $C([0,1],\mathbf{R})$ associated to $\|\cdot\|_1$ as in (1.15.10) exactly when $\alpha \ge -1$, and that $\{j^{-\alpha}\,f_j\}_{j=1}^\infty$ converges to 0 with respect to d_1 exactly when $\alpha > -1$. We also have that

$$\|j^{-\alpha}\,f_j\|_2 = (1/\sqrt{3})\,j^{-(1/2)-\alpha} \tag{A.4.12}$$

for every j, by (A.4.8). It follows that $\{j^{-\alpha}\,f_j\}_{j=1}^\infty$ is bounded with respect to the metric d_2 on $C([0,1],\mathbf{R})$ associated to $\|\cdot\|_2$ as in (1.15.11) if and only if $\alpha \ge -1/2$, and that $\{j^{-\alpha}\,f_j\}_{j=1}^\infty$ converges to 0 with respect to d_2 if and only if $\alpha > -1/2$.

A.5 Some Cauchy sequences

Let j be a positive integer, and let f_j be the real-valued function defined on $[0,1]$ by

$$\begin{aligned} f_j(x) &= 0 && \text{when } 0 \le x \le 1/2 - 1/(2\,j) \\ &= 2\,j\,(x - (1/2 - 1/(2\,j))) && \text{when } 1/2 - 1/(2\,j) \le x \le 1/2 \\ &= 1 && \text{when } 1/2 \le x \le 1. \end{aligned} \tag{A.5.1}$$

The second case says that f_j is linear in that range, with the same values at the endpoints as in the other two cases. In particular, f_j is continuous on $[0,1]$ for each j. It is easy to see that $\{f_j\}_{j=1}^\infty$ converges pointwise on $[0,1]$ to the real-valued function f defined on $[0,1]$ by

$$\begin{aligned} f(x) &= 0 && \text{when } 0 \le x < 1/2 \\ &= 1 && \text{when } 1/2 \le x \le 1. \end{aligned} \tag{A.5.2}$$

Note that $\{f_j\}_{j=1}^\infty$ does not converge to f uniformly on $[0,1]$.

However, one can check that $\{f_j\}_{j=1}^\infty$ is a Cauchy sequence in $C([0,1],\mathbf{R})$ with respect to the metric d_1 defined in (1.15.10). Basically, $\{f_j\}_{j=1}^\infty$ converges to f with respect to a metric like d_1 on a larger space. Similarly, one can verify that $\{f_j\}_{j=1}^\infty$ is a Cauchy sequence in $C([0,1],\mathbf{R})$ with respect to the metric d_2 defined in (1.15.11).

Let j be a positive integer again, and let f_j be the real-valued function defined on $[0,1]$ by

$$f_j(x) = \min(1/\sqrt{x}, j), \tag{A.5.3}$$

where the right side is interpreted as being equal to j when $x = 0$. Note that f_j is continuous on $[0,1]$ for each j. One can check that $\{f_j\}_{j=1}^\infty$ is a Cauchy sequence in $C([0,1],\mathbf{R})$ with respect to d_1. However, $\{f_j\}_{j=1}^\infty$ is not bounded as a sequence in $C([0,1],\mathbf{R})$ with respect to d_2.

Now let f be a real or complex-valued function on $X = \mathbf{Z}_+$. If j is a positive integer, then let f_j be the real or complex-valued function, as appropriate, defined on $\mathbf{Z}_+$ by

$$\begin{aligned} f_j(l) &= f(l) && \text{when } l \le j \\ &= 0 && \text{when } l > j. \end{aligned} \tag{A.5.4}$$

Clearly $\{f_j\}_{j=1}^\infty$ converges to f pointwise on $\mathbf{Z}_+$. If

$$\lim_{l\to\infty} f(l) = 0, \tag{A.5.5}$$

then $\{f_j\}_{j=1}^\infty$ converges to f uniformly on $\mathbf{Z}_+$.

By construction, the support of f_j has only finitely many elements for each j. Suppose for the moment that

$$\sum_{l=1}^\infty |f(l)| \tag{A.5.6}$$

converges as an infinite series of nonnegative real numbers. In this case, one can check that $\{f_j\}_{j=1}^{\infty}$ is a Cauchy sequence in $c_{00}(\mathbf{Z}_+, \mathbf{R})$ or $c_{00}(\mathbf{Z}_+, \mathbf{C})$, as appropriate, with respect to the metric $d_1(\cdot, \cdot)$ defined in Section 1.6. Similarly, suppose instead that

$$\sum_{l=1}^{\infty} |f(l)|^2 \tag{A.5.7}$$

converges as an infinite series of nonnegative real numbers. One can verify that $\{f_j\}_{j=1}^{\infty}$ is a Cauchy sequence in $c_{00}(\mathbf{Z}_+, \mathbf{R})$ or $c_{00}(\mathbf{Z}_+, \mathbf{C})$, as appropriate, with respect to the metric $d_2(\cdot, \cdot)$ defined in Section 1.6 in this situation.

A.6 Norms and convexity

If the reader is familiar with the abstract notion of a vector space, then let V be a vector space over the real numbers. Otherwise, one can consider particular situations like $V = \mathbf{R}^n$ for some positive integer n, or a vector space of real-valued functions on a nonempty set. As usual, a subset E of V is said to be *convex* if for every $v, w \in E$ and $t \in \mathbf{R}$ with $0 \le t \le 1$ we have that

$$t\,v + (1 - t)\,w \in E. \tag{A.6.1}$$

Let N be a nonnegative real-valued function on V such that

$$N(t\,v) = |t|\,N(v) \tag{A.6.2}$$

for every $t \in \mathbf{R}$ and $v \in V$. If N also satisfies the triangle inequality

$$N(v + w) \le N(v) + N(w) \tag{A.6.3}$$

for every $v, w \in V$, then N is said to be a *seminorm* on V. Thus a norm on V is the same as a seminorm N such that $N(v) > 0$ for every $v \in V$ with $v \ne 0$. Put

$$\overline{B}_N = \{v \in V : N(v) \le 1\}, \tag{A.6.4}$$

which is the closed unit ball in V with respect to N. If N is a seminorm on V, then it is easy to see that $\overline{B}_N$ is a convex subset of V. Conversely, if N is a nonnegative real-valued function on V that satisfies (A.6.2), and if $\overline{B}_N$ is a convex set in V, then N satisfies (A.6.3), and hence is a seminorm on V. This is not too difficult to show, directly from the definitions.

Let X be a nonempty set, and let $c_{00}(X, \mathbf{R})$ be the space of real-valued functions on X with finite support, as in Section 1.6. If $f \in c_{00}(X, \mathbf{R})$ and p is a positive real number, then $\|f\|_p$ may be defined as in Section A.3. This satisfies the first two conditions in the definition of a norm on $c_{00}(X, \mathbf{R})$, as before. If $p \ge 1$, then one can show that the corresponding closed unit ball

$$\overline{B}_p = \{f \in c_{00}(X, \mathbf{R}) : \|f\|_p \le 1\} \tag{A.6.5}$$

is a convex set in $c_{00}(X, \mathbf{R})$. More precisely, (A.6.5) is the same as taking

$$\overline{B}_p = \Big\{f \in c_{00}(X, \mathbf{R}) : \|f\|_p^p = \sum_{x \in X} |f(x)|^p \le 1\Big\}. \tag{A.6.6}$$

One can show that this is a convex subset of $c_{00}(X, \mathbf{R})$ when $p \ge 1$, using the convexity of r^p for $r \ge 0$ when $p \ge 1$. This implies that $\|f\|_p$ defines a norm on $c_{00}(X, \mathbf{R})$ when $p \ge 1$, as in the preceding paragraph. If $0 < p < 1$, then (A.6.5) is not convex in $c_{00}(X, \mathbf{R})$ when X has at least two elements.

A vector space over the complex numbers may be considered as a vector space over the real numbers as well, to which the earlier remarks about convexity can be applied. In particular, there are analogues of the statements in the preceding paragraph for the space $c_{00}(X, \mathbf{C})$ of complex-valued functions on X with finite support.

Let n be a positive integer, and let N be a seminorm on $\mathbf{R}^n$. The closed unit ball $\overline{B}_N$ in $\mathbf{R}^n$ with respect to N is convex, as before, as is the open unit ball

$$B_N = \{v \in \mathbf{R}^n : N(v) < 1\}. \tag{A.6.7}$$

Note that B_N and $\overline{B}_N$ are also symmetric about the origin, which is to say that they are invariant under the mapping $v \mapsto -v$ on $\mathbf{R}^n$. One can show that N is continuous as a real-valued function on $\mathbf{R}^n$, and in fact it is Lipschitz with respect to the standard Euclidean metric on $\mathbf{R}^n$, as in Section 6.7. This implies that B_N is an open set, and $\overline{B}_N$ is a closed set, with respect to the standard Euclidean metric on $\mathbf{R}^n$. If N is a norm on $\mathbf{R}^n$, then the standard Euclidean norm on $\mathbf{R}^n$ is bounded by a constant multiple of N, as in Section 6.7. It follows that B_N and $\overline{B}_N$ are bounded sets with respect to the standard Euclidean metric on $\mathbf{R}^n$ in this case.

A.7 Path-connected sets

Let (X, d_X) be a metric space, and let a, b be real numbers with $a < b$. Suppose that

(A.7.1) p is a continuous mapping from the closed interval $[a, b]$ into X,

with respect to the restriction of the standard Euclidean metric on $\mathbf{R}$ to $[a, b]$. Under these conditions, one can check that

(A.7.2) $p([a, b])$ is a connected subset of X.

More precisely, it is well known that $[a, b]$ is a connected subset of the real line, with respect to the standard metric. This implies that $[a, b]$ is connected as a subset of itself, with respect to the restriction of the standard metric on $\mathbf{R}$ to $[a, b]$, as in Section 8.12. It follows that (A.7.2) holds, because of the well-known theorem that continuous mappings send connected sets to connected sets. Alternatively, one can extend p to a continuous mapping from $\mathbf{R}$ into X, by putting $p(t) = p(a)$ when $t < a$, and $p(t) = p(b)$ when $t > b$.

A subset E of X is said to be *path connected* if for every pair of points $x, w \in E$ there are real numbers a, b with $a < b$ and a continuous mapping p from $[a, b]$ into X such that $p(a) = x$, $p(b) = w$, and

$$p([a, b]) \subseteq E. \tag{A.7.3}$$

It is well known and not too difficult to show that

(A.7.4) path-connected sets are connected,

using (A.7.2).

If E is a convex set in $\mathbf{R}^n$ for some positive integer n, then

(A.7.5) E is path connected,

with respect to the standard Euclidean metric on $\mathbf{R}^n$. Note that connected subsets of $\mathbf{R}$ are convex.

If U is a connected open set in $\mathbf{R}^n$, then it is well known that

(A.7.6) U is path connected.

To see this, let $x \in U$ be given, and let U_x be the set of $w \in U$ for which there is a continuous path in U from x to w, as before. One can check that

(A.7.7) U_x is an open set in $\mathbf{R}^n$.

Similarly, one can verify that

(A.7.8) $U \setminus U_x$ is an open set in $\mathbf{R}^n$.

If U is connected, then it follows that

$$U_x = U. \tag{A.7.9}$$

It is well known that there are connected subsets of $\mathbf{R}^2$ that are not path connected.

If E is a connected set in any metric space X, then it is well known and not too difficult to show that

(A.7.10) $\overline{E}$ is connected in X

too. However, there are path-connected subsets of $\mathbf{R}^2$ whose closures are not path connected.

Let (Y, d_Y) be another metric space, and let f be a continuous mapping from X into Y. If E is a path-connected subset of X, then it is easy to see that

(A.7.11) $f(E)$ is path connected in Y.

Bibliography

[1] R. Agarwal, C. Flaut, and D. O'Regan, *An Introduction to Real Analysis*, CRC Press, 2018.

[2] T. Apostol, *Mathematical Analysis*, 2nd edition, Addison-Wesley, 1974.

[3] T. Archibald, *Connectivity and smoke-rings: Green's second identity in its first fifty years*, Mathematics Magazine **62** (1989), 219–232.

[4] L. Badger, *Notes: Generating the measures of n-balls*, American Mathematical Monthly **107** (2000), 256–258.

[5] J. Baker, *Isometries in normed spaces*, American Mathematical Monthly **78** (1971), 655–658.

[6] J. Baker, *Integration over spheres and the divergence theorem for balls*, American Mathematical Monthly **104** (1997), 36–47.

[7] J. Baker, *Integration of radial functions*, Mathematics Magazine **72** (1999), 392–395.

[8] R. Beals, *Advanced Mathematical Analysis: Periodic Functions and Distributions, Complex Analysis, Laplace Transform and Applications*, Springer-Verlag, 1973.

[9] R. Beals, *Analysis: An Introduction*, Cambridge University Press, 2004.

[10] R. Beals, *The scope of the power series method*, Mathematics Magazine **86** (2013), 56–62.

[11] P. Beesack, *On improper multiple integrals*, Mathematics Magazine **43** (1970), 113–123.

[12] M. Berger, *Convexity*, American Mathematical Monthly **97** (1990), 650–678.

[13] L. Bers, *Classroom notes: On avoiding the mean value theorem*, American Mathematical Monthly **74** (1967), 583.

[14] R. Boas, *A Primer of Real Functions*, 4th edition, revised and updated by H. Boas, Mathematical Association of America, 1996.

[15] S. Bochner, *Fourier series came first*, American Mathematicsl Monthly **86** (1979), 197–199.

[16] M. Botsko and R. Gosser, *The teaching of mathematics: On the differentiability of functions of several variables*, American Mathematical Monthly **92** (1985), 663–665.

[17] D. Bressoud, *A Radical Approach to Real Analysis*, 2nd edition, Mathematical Association of America, 2007.

[18] D. Bressoud, *A Radical Approach to Lebesgue's Theory of Integration*, MAA Textbooks, Cambridge University Press, 2008.

[19] D. Bressoud, *Historical reflections on teaching the fundamental theorem of integral calculus*, American Mathematical Monthly **118** (2011), 99–115.

[20] D. Bressoud, *Calculus Reordered: A History of the Big Ideas*, Princeton University Press, 2019.

[21] A. Browder, *Mathematical Analysis: An Introduction*, Springer-Verlag, 1996.

[22] A. Browder, *Topology in the complex plane*, American Mathematical Monthly **107** (2000), 393–401.

[23] A. Browder, *Complex numbers and the ham sandwich theorem*, American Mathematical Monthly **113** (2006), 935–937.

[24] A. Brown, *A proof of the Lebesgue condition for Riemann integrability*, American Mathematical Monthly **43** (1936), 396–398.

[25] A. Brown, *On transformation of multiple integrals*, American Mathematical Monthly **48** (1941), 29–33.

[26] A. Brown, *Mathematical notes: Extensions of the Brouwer fixed point theorem*, American Mathematical Monthly **69** (1962), 643.

[27] A. Bruckner, *Derivatives: why they elude classification*, Mathematics Magazine **49** (1976), 5–11.

[28] G. Bullock, *The teaching of mathematics: A geometric interpretation of the Riemann–Stieltjes integral*, American Mathematical Monthly **95** (1988), 448-455.

[29] R. Burckel and C. Goffman, *Rectifiable curves are of zero content*, Mathematics Magazine **44** (1971), 179–180.

[30] J. Burkill and H. Burkill, *A Second Course in Mathematical Analysis*, Cambridge University Press, 1970.

[31] T. Cameron, *The determinant from signed volume to the Laplace expansion*, American Mathematical Monthly **126** (2019), 437–447.

[32] D. Cannell, *Geoge Green: an enigmatic mathematician*, American Mathematical Monthly **106** (1999), 136–151.

[33] D. Cannell, *George Green, Mathematician & Physicist 1793–1841*, 2nd edition, Society for Industrial and Applied Mathematics (SIAM), 2001.

[34] L. Cohen, *Classroom notes: On being mean to the mean-value theorem*, American Mathematical Monthly **74** (1967), 581–582.

[35] R. Cohen, *Set isometries and their extensions in the max metric*, American Mathematical Monthly **97** (1990), 795–808.

[36] J. Conway, *A First Course in Analysis*, Cambridge University Press, 2018.

[37] N. de Silva, *A concise, elementary proof of Arzelà's bounded convergence theorem*, American Mathematical Monthly **117** (2010), 918–920.

[38] A. Devinatz, *Advanced Calculus*, Holt, Rinehart and Winston, 1968.

[39] T. Drammatico, *A new proof of the equality of mixed second partial derivatives*, American Mathematical Monthly **122** (2015), 602-603.

[40] R. Dressler and K. Stromberg, *The Tonelli integral*, American Mathematical Monthly **81** (1974), 67–68.

[41] W. Dunham, *Touring the calculus gallery*, American Mathematical Monthly **112** (2005), 1-19.

[42] W. Dunham, *The Calculus Gallery: Masterpieces from Newton to Lebesgue*, Princeton University Press, 2018.

[43] P. Duren, *Invitation to Classical Analysis*, American Mathematical Society, 2012.

[44] A. Fadell, *Classroom notes: A proof of the chain rule for derivatives in n-space*, American Mathematical Monthly **80**, 1134–1135.

[45] W. Felscher, *Bolzano, Cauchy, epsilon, delta*, American Mathematical Monthly **107** (2000), 844–862.

[46] L. Flatto, *Advanced Calculus*, Williams & Wilkins, 1976.

[47] T. Flett, *On transformations in $\mathbf{R}^n$ and a theorem of Sard*, American Mathematical Monthly **71** (1964), 623-629.

[48] G. Folland, *Fourier Analysis and its Applications*, Wadsworth & Brooks / Cole, 1992.

[49] G. Folland, *How to integrate a polynomial over a sphere*, American Mathematical Monthly **108** (2001), 446–448.

[50] G. Folland, *A Guide to Advanced Real Analysis*, Mathematical Association of America, 2009.

[51] M. Fort, Jr., *Some properties of continuous functions*, American Mathematical Monthly **59** (1952), 372–375.

[52] P. Franklin, *A Treatise on Advanced Calulus*, Dover, 1964.

[53] A. Friedman, *Advanced Calculus*, Holt, Rinehart and Winston, 1971.

[54] T. Gamelin and R. Greene, *Introduction to Topology*, 2nd edition, Dover, 1999.

[55] C. Gardner, *Another elementary proof of Peano's existence theorem*, American Mathematical Monthly **83** (1976), 556-560.

[56] H. Gaskill and P. Narayanaswami, *Foundations of Analysis: The Theory of Limits*, Harper & Row, 1989.

[57] B. Gelbaum and J. Olmsted, *Counterexamples in Analysis*, corrected reprint of the 2nd (1965) edition, Dover, 2003.

[58] C. Goffman, *Definition of the Lebesgue integral*, American Mathematical Monthly **60** (1953), 251–252.

[59] C. Goffman, *Real Functions*, Rinehart & Co., 1953.

[60] C. Goffman, *Classroom notes: Arc length*, American Mathematical Monthly **71** (1964), 303–304.

[61] C. Goffman, *Calculus of Several Variables*, Harper & Row, 1965.

[62] C. Goffman, *Introduction to Real Analysis*, Harper & Row, 1966.

[63] C. Goffman, *A note on integration*, Mathematics Magazine **44** (1971), 1–4.

[64] C. Goffman, *A bounded derivative which is not Riemann integrable*, American Mathematical Monthly **84** (1977), 205–206.

[65] R. Goldberg, *Methods of Real Analysis*, 2nd edition, Wiley, 1976.

[66] E. González-Velasco, *Connections in mathematical analysis: the case of Fourier series*, American Mathematical Monthly **99** (1992), 427–441.

[67] E. González-Velasco, *James Gregory's calculus in the* Geometriæ pars universalis, American Mathematical Monthly **114** (2007), 565–576.

[68] E. González-Velasco, *Journey Through Mathematics: Creative Episodes in its History*, Springer, 2011.

[69] R. Gordon, *A convergence theorem for the Riemann integral*, Mathematics Magazine **73** (2000), 141–147.

[70] R. Gordon, *A bounded derivative that is not Riemann integrable*, Mathematics Magazine **89** (2016), 364–370.

[71] F. Gouvêa, *Was Cantor surprised?*, American Mathematical Monthly **118** (2011), 198–209.

[72] J. Grabiner, *The Origins of Cauchy's Rigorous Calculus*, MIT Press, 1981.

[73] J. Grabiner, *Who gave you the epsilon? Cauchy and the origins of rigorous calculus*, American Mathematical Monthly **90** (1983), 185–194.

[74] J. Grabiner, *The changing concept of change: the derivative from Fermat to Weierstrass*, Mathematics Magazine **56** (1983), 195–206.

[75] J. Grabiner, *Was Newton's calculus a dead end? The continental influence of Maclauren's treatise of fluxions*, American Mathematical Monthly **104** (1997), 393–410.

[76] J. Grabiner, *A Historian Looks Back:* The Calculus as Algebra *and selected writings*, MAA Spectrum, Mathematical Association of America, 2010.

[77] I. Grattan-Guinness, *Why did George Green write his essay of 1828 on electricity and magnetism?*, American Mathematical Monthly **102** (1995), 387–396.

[78] J. Gray, *Miscellanea: Clear theory*, American Mathematical Monthly **92** (1985), 158–159.

[79] J. Gray, *The Real and the Complex: A History of Analysis in the 19th Century*, Springer, 2015.

[80] J. Gray, *Change and Variations — A History of Differential Equations to 1900*, Springer, 2021.

[81] J. Green, *The theory of functions of a real variable*, Mathematics Magazine **24** (1951), 209–217.

[82] J. Green and F. Valentine, *On the Arzelà–Ascoli theorem*, Mathematics Magazine **34** (1960/61), 199–202.

[83] R. Gunning, *An Introduction to Analysis*, Princeton University Press, 2018.

[84] J. Hannah, *A geometric approach to determinants*, American Mathematical Monthly **103** (1996), 401–409.

[85] E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, Springer-Verlag, 1975.

[86] T. Hildebrandt, *Definitions of Stieltjes integrals of the Riemann type*, American Mathematical Monthly **45** (1938), 265–278.

[87] K. Hoffman, *Analysis in Euclidean Space*, Prentice-Hall, 1975.

[88] F. Jones, *Lebesgue Integration on Euclidean Space*, Jones and Bartlett, 1993.

[89] H. Junghenn, *A Course in Real Analysis*, CRC Press, 2015.

[90] W. Kaczor and M. Nowak, *Problems in Mathematical Analysis I: Real Numbers, Sequences and Series*, translated and revised from the 1996 Polish original by the authors, American Mathematical Society, 2000.

[91] W. Kaczor and M. Nowak, *Problems in Mathematical Analysis II: Continuity and Differentiation*, translated from the 1998 Polish original, revised, and augments by the authors, American Mathematical Society, 2001.

[92] W. Kaczor and M. Nowak, *Problems in Mathematical Analysis III: Integration*, American Mathematical Society, 2003.

[93] I. Kaplansky, *Set Theory and Metric Spaces*, 2nd edition, AMS Chelsea, 1977.

[94] H. Kaptanoğlu, *In praise of* $x^\alpha \sin(\frac{1}{x})$, American Mathematical Monthly **108** (2001), 144–150.

[95] H. Kennedy, *Research problems: Is there an elementary proof of Peano's existence theorem for first order differential equations?*, American Mathematical Monthly **76** (1969), 1043–1045.

[96] H. Kennedy, *Peano: Life and Works of Giuseppe Peano*, Reidel, 1980.

[97] G. Klambauer, *Mathematical Analysis*, Dekker, 1975.

[98] A. Knapp, *Basic Real Analysis*, Birkhäuser, 2005.

[99] J. Koliha, *Mean, meaner, and meanest mean value theorem*, American Mathematical Monthly 116 (2009), 356–361.

[100] T. Körner, *Differentiable functions on the rationals*, Bulletin of the London Mathematical Society **23** (1991), 557–562.

[101] T. Körner, *Exercises for* Fourier Analysis ([102]). Cambridge University Press, 1993.

[102] T. Körner, *Butterfly hunting and the symmetry of mixed partial derivatives*, American Mathematical Monthly **105** (1998), 756–758.

[103] T. Körner, *A Companion to Analysis: A Second First and First Second Course in Analysis*, American Mathematical Society, 2004.

[104] T. Körner, *Fourier Analysis*, Cambridge University Press, 2022.

[105] S. Krantz, *A Guide to Real Variables*, Mathematical Association of America, 2009.

[106] S. Krantz, *Real Analysis and Foundations*, 4th edition, CRC Press, 2017.

[107] S. Krantz, *Elementary Introduction to the Lebesgue Integral*, CRC Press, 2018.

[108] S. Krantz and H. Parks, *The Implicit Function Theorem: History, Theory, and Applications*, Birkhäuser, 2002.

[109] K. Kreith, *Geometric interpretation of the Implicit Function Theorem*, Mathematics Magazine **36** (1963), 64–65.

[110] W. Kulpa, *The Poincaré–Miranda theorem*, American Mathematical Monthly **104** (1997), 545–550.

[111] M. Laczkovich and V. Sós, *Real Analysis — Foundations and Functions of One Variable*, 5th edition, translated from the third (2005) Hungarian edition by the authors, Springer, 2015.

[112] S. Lang, *Undergraduate Analysis*, 2nd edition, Springer-Verlag, 1997.

[113] R. Langer, *Fourier's Series: The Genesis and Evolution of a Theory*, The first Herbert Ellsworth Slaught Memorial Paper, American Mathematical Monthly **54** (1947), no. 7, part II.

[114] J. LaVita, *Classroom notes: A necessary and sufficient condition for Riemann integration*, American Mathematical Monthly **71** (1964), 193–196.

[115] P. Lax, *Change of variables in multiple integrals*, American Mathematical Monthly **106** (1999), 497–501.

[116] P. Lax, *Change of variables in multiple integrals II*, American Mathematical Monthly **108** (2001), 115–119.

[117] P. Lax, *Rethinking the Lebesgue integral*, American Mathematical Monthly **116** (2009), 863–881.

[118] T. Lehrer, *The Derivative Song*, American Mathematical Monthly **81** (1974), 490.

[119] T. Lehrer, *There's a Delta for Every Epsilon (Calypso)*, American Mathematical Monthly **81** (1974), 612.

[120] T. Lehrer, *The Professor's Song*, American Mathematical Monthly **81** (1974), 745.

[121] L. Levine, *On a necessary and sufficient condition for Riemann integrability*, American Mathematical Monthly **84** (1977), 205.

[122] J. Lewin, *The teaching of mathematics: A truly elementary approach to the bounded convergence theorem*, American Mathematical Monthly **93** (1986), 395–397.

[123] J. Lewin, *The teaching of mathematics: Some applications of the bounded convergence theorem for an introductory course in analysis*, American Mathematical Monthly **94** (1987), 988–993.

[124] C.-K. Li, *Norms, isometries, and isometry groups*, American Mathematical Monthly **107** (2000), 334–340.

[125] C.-K. Li and W. So, *Isometries of l_p-norm*, American Mathematical Monthly **101** (1994), 452–453.

[126] W. Luxemburg, *Arzelà's dominated convergence theorem for the Riemann integral*, American Mathematical Monthly **78** (1971), 970-979.

[127] S. Malik and S. Arora, *Mathematical Analysis*, 2nd edition, Wiley, 1992.

[128] S. Mandelbrojt, *Obituary: Emile Picard, 1856–1941*, American Mathematical Monthly **49** (1942), 277–278.

[129] S. Mandelbrojt, *The mathematical work of Jacques Hadamard*, American Mathematical Monthly **60** (1953), 599–604.

[130] J. Mawhin, *Simple proofs of the Hadamard and Poincaré–Miranda theorems using the Brouwer fixed point theorem*, American Mathematical Monthly **126** (2019), 260–263.

[131] P. McGrath, *Another proof of Clairaut's theorem*, American Mathematical Monthly **121** (2014), 165–166.

[132] J. McKnight, Jr., *Brown's method of extending fixed point theorems*, American Mathematical Monthly **72** (1965), 152–155.

[133] B. Mendelson, *Introduction to Topology*, 3rd edition, Dover, 1990.

[134] R. Metzler, *On Riemann integrability*, American Mathematical Monthly **78** (1971), 1129–1131.

[135] J. Milnor, *Analytic proofs of the "hairy ball theorem" and the Brouwer fixed-point theorem*, American Mathematical Monthly **85** (1978), 521-524.

[136] F. Morgan, *Real Analysis*, American Mathematical Society, 2005.

[137] F. Morgan, *Real Analysis and Applications: Including Fourier Series and the Calculus of Variations*, American Mathematical Society, 2005.

[138] C. Morrison and M. Stynes, *An intuitive proof of Brouwer's fixed point theorem in* $\mathbf{R^2}$, Mathematics Magazine **56** (1983), 38–41.

[139] R. Mortini, *A short proof of the chain rule for differentiable mappings in* $\mathbf{R}^n$, Mathematics Magazine **85** (2012), 136–141.

[140] S. Nadler, Jr., *A proof of Darboux's theorem*, American Mathematical Monthly **117** (2010), 174–175.

[141] A. Nijenhuis, *Strong derivatives and inverse mappings*, American Mathematical Monthly **81** (1974), 969–980; addendum, **83** (1975), 22.

[142] L. Olsen, *A new proof of Darboux's theorem*, American Mathematical Monthly **111** (2004), 713–715.

[143] S. Ponnusamy, *Foundations of Mathematical Analysis*, Birkhäuser / Springer, 2012.

[144] M. Protter, *Basic Elements of Real Analysis*, Springer-Verlag, 1998.

[145] M. Protter and C. Morrey, Jr., *A First Course in Real Analysis*, 2nd edition, Springer-Verlag, 1991.

[146] C. Pugh, *Real Mathematical Analysis*, 2nd edition, Springer, 2015.

[147] H. Rademacher, *On the condition of Riemann integrability*, American Mathematical Monthly **61** (1954), 1–8.

[148] M. Reed, *Fundamental Ideas in Analysis*, Wiley, 1998.

[149] H. Reiter and A. Reiter, *The space of closed subsets of a convergent sequence*, Mathematics Magazine **69** (1996), 217–221.

[150] C. Rogers, *A less strange version of Milnor's proof of Brouwer's fixed-point theorem*, American Mathematical Monthly **87** (1980), 525–527.

[151] M. Rosenlicht, *Introduction to Analysis*, Dover, 1986.

[152] C. Rosentrater, *Varieties of Integration*, Mathematical Association of America, 2015.

[153] K. Ross, *Another approach to Riemann–Stieltjes integrals*, American Mathematical Monthly **87** (1980), 660-662.

[154] K. Ross, *Elementary Analysis: The Theory of Calculus*, 2nd edition, in collaboration with J. López, Springer, 2013.

[155] W. Rudin, *Principles of Mathematical Analysis*, 3rd edition, McGraw-Hill, 1976.

[156] W. Rudin, *Real and Complex Analysis*, 3rd edition, McGraw-Hill, 1987.

[157] W. Rudin, *Functional Analysis*, 2nd edition, McGraw-Hill, 1991.

[158] D. Sanderson, *Advanced plane topology from an elementary standpoint*, Mathematics Magazine (1980), 81–89.

[159] U. Satyanarayana, *A note on Riemann-Stieltjes integrals*, American Mathematical Monthly **87** (1980), 477–478.

[160] D. Schattschneider, *A multiplicative metric*, Mathematics Magazine **49** (1976), 203–205.

[161] D. Schattschneider, *The taxicab group*, American Mathematical Monthly **91** (1984), 423–428.

[162] R. Shakarchi, *Problems and Solutions for* Undergraduate Analysis ([112]), Springer-Verlag, 1998.

[163] O. Shisha, *An approach to Darboux-Stieltjes integration*, American Mathematical Monthly **72** (1965), 890–892.

[164] H. Siegberg, *Some historical remarks concerning degree theory*, American Mathematical Monthly **88** (1981), 125–139.

[165] D. Smith and M. Vamanamurthy, *How small is a unit ball?*, Mathematics Magazine **62** (1989), 101–107.

[166] E. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press, 2003.

[167] E. Stein and R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*, Princeton University Press, 2005.

[168] P. Stein, *Classroom notes: A note on the volume of a simplex*, American Mathematical Monthly **73** (1966), 299-301.

[169] P. Stein, *A proof of a classical theorem in multiple integration*, American Mathematical Monthly **75** (1968), 160–163.

[170] G. Stoica, *Components used in the Stone–Weierstrass theorem*, American Mathematical Monthly **127** (2020), 658.

[171] R. Strichartz, *The Way of Analysis*, Revised edition, Jones and Bartlett, 2000.

[172] K. Stromberg, *The Banach–Tarski paradox*, American Mathematical Monthly **86** (1979), 151–161.

[173] K. Stromberg, *An Introduction to Classical Real Analysis*, AMS Chelsea, 1981.

[174] A. Taylor, *A note on an inequality for integrals*, American Mathematical Monthly **57** (1950), 93–96.

[175] A. Taylor and W. Mann, *Advanced Calculus*, 3rd edition, Wiley, 1983.

[176] A. Taylor and S. Wagon, *A paradox arising from the elimination of a paradox*, American Mathematical Monthly **126** (2019), 306–318.

[177] H. ter Horst, *Riemann–Stieltjes and Lebesgue–Stieltjes integrability*, American Mathematical Monthly **91** (1984), 551–559.

[178] A. Thompson, *Minkowski Geometry*, Cambridge University Press, 1996.

[179] B. Thomson, *Monotone convergence for the Riemann integral*, American Mathematical Monthly **117** (2010), 547–550.

[180] B. Thomson, *The bounded convergence theorem*, American Mathematical Monthly **127** (2020), 483–503.

[181] T. Tokieda, *A mean value theorem*, American Mathematical Monthly **106** (1999), 673–674.

[182] G. Tomkowicz and S. Wagon, *The Banach–Tarski Paradox*, 2nd edition, Cambridge University Press, 2016.

[183] J. Väisälä, *A proof of the Mazur–Ulam theorem*, American Mathematical Monthly **110** (2003), 633–635.

[184] D. Varberg, *On differentiable transformations in* $\mathbf{R}^n$, American Mathematical Monthly **73** (1966), no. 4, part II, 111-114.

[185] D. Varberg, *Change of variables in multiple integrals*, American Mathematical Monthly **78** (1971), 42–45.

[186] J. Walter, *On elementary proofs of Peano's existence theorems*, American Mathematical Monthly **80** (1973), 282-286.

[187] W. Walter, *There is an elementary proof of Peano's existence theorem*, American Mathematical Monthly **78** (1971), 170–173.

[188] W. Walter, *Ordinary Differential Equations*, translated from the sixth German (1996) edition by R. Thompson, Springer-Verlag, 1998.

[189] G. Weiss, *Complex methods in harmonic analysis*, American Mathematical Monthly **77** (1970), 465–474.

[190] D. Widder, *What is the Laplace transform?*, American Methematical Monthly **52** (1945), 410–425.

[191] D. Widder, *Advanced Calculus*, 3rd edition, Dover, 1989.

[192] B. Youse, *Introduction to Real Analysis*, Allyn and Bacon, 1972.

[193] H. Zatzkis, *Volume and surface of a sphere in* N*-dimensional Euclidean space*, Mathematics Magazine **30** (1957), 155–157.

Index