CHAPTER THREE

Metrics and Norms

In the beginning there were operations – hundreds of them – limits, derivatives, integrals, sums; all of the many operations on functions, sequences, sets, vectors, matrices, and whatever else you might have encountered in calculus. The hallmark of twentieth-century mathematics is that we now view these operations as functions defined on entire collections of "abstract" objects rather than as specific actions taken on individual objects, one at a time. Maurice Fréchet, in a short expository article from 1950, had this to say (the italics are his own):

> In modern times it has been recognized that it is possible to elaborate full mathematical theories dealing with elements of which the nature is not specified, that is, with abstract elements. A collection of these abstract elements will be called an *abstract set*. If to this set there is added some rule of association of these elements, or some relation between them, the set will be called an *abstract space*. A natural generalization of function consists in associating with any element x of an abstract set E a number $f(x)$. Functional analysis is the study of such "functionals" $f(x)$. More generally, *general analysis* is the theory of the transformations $y = F[x]$ of an element x of an abstract set E into an element y of another (or the same) abstract set F. It is obvious that the study of general analysis should be preceded by a discussion of abstract spaces.
>
> It is necessary to keep in mind that these notions are *not of a metaphysical nature*; that when we speak of an abstract element we mean that the nature of this element is indifferent, but *we do not mean at all that this element is unreal*. Our theory will apply to all elements; in particular, applications of it may be made to the natural sciences. Of course, due attention must be paid to any properties which depend essentially on the nature of any special category of elements under investigation.

Early examples of this type of abstraction appeared in 1906 in Fréchet's thesis, "Sur quelques points du calcul functionnel," in which he introduced a notion of distance defined on abstract sets of points. In particular, Fréchet considered the collection $C[0, 1]$, consisting of all continuous real-valued functions defined on the closed interval $[0, 1]$, where we measure the distance between two functions by taking the maximum vertical distance between their graphs; that is, $\text{dist}(f, g) = \max_{0 \le t \le 1} |f(t) - g(t)|$. (This distance function was actually well known in 1906, but Fréchet was the first to view it as a small part of a much bigger picture.) Given a notion of distance between elements of $C[0, 1]$, it makes sense to ask questions like: Is integration continuous? That is, are the numbers $\int_0^1 f(t)\,dt$ and $\int_0^1 g(t)\,dt$ "close" whenever f and g are "close"?

This new point of view proved to have immediate applications; in that same year Friedrich Riesz used Fréchet's ideas to give a new proof of a result of Erhardt Schmidt, stating that any orthonormal system in $C[0, 1]$ must be countable. In fact, Riesz extended this result to another collection of functions and in so doing introduced the L_p spaces. Riesz's techniques revolutionized the study of trigonometric series. To say that Fréchet's ideas caught on would be an understatement; the study of modern analysis would be lost without them. By 1928, Fréchet had compiled a monograph on his research on abstract spaces entitled *Les Espaces Abstraits*. (The word "space" has come to connote an abstract set of points that carries with it some additional structure.) Much of the terminology we will use, and certainly most of our examples of abstract spaces, can be found in Fréchet's monograph. By mathematical standards, 1928 is not so very long ago.

Metric Spaces

Given a set M, how might we define a distance function on M? What would we want a "reasonable" distance to do? Certainly we would want our distance to be (defined and) nonnegative for any pair of points in M. Let's start there: Let $d : M \times M \to [0, \infty)$ be a nonnegative, real-valued function defined on all pairs of elements from M. We would probably expect to have $d(x, x) = 0$ for any $x \in M$. And $d(x, y) = 0$ should mean that $x = y$. We would most likely want our distance to also satisfy $d(x, y) = d(y, x)$ for all pairs of points $x, y \in M$. Anything else? Well, in the hope of preserving at least a bit of the geometry granted by the familiar distances in $\mathbb{R}$ and $\mathbb{R}^n$, we might also require one last property. The distance function should satisfy **the triangle inequality**: For each triple of points x, y, z in M, we ask that $d(x, y) \le d(x, z) + d(z, y)$. The triangle inequality is the embodiment of that old saw, "The shortest distance between two points is a straight line." This timid little inequality will turn out to be immensely valuable.

A function d on $M \times M$ satisfying the following properties is called a **metric** on M.

(i) $0 \le d(x, y) < \infty$ for all pairs $x, y \in M$.
(ii) $d(x, y) = 0$ if and only if $x = y$.
(iii) $d(x, y) = d(y, x)$ for all pairs $x, y \in M$.
(iv) $d(x, y) \le d(x, z) + d(z, y)$ for all $x, y, z \in M$.

A function d on $M \times M$ that satisfies all of the above save item (ii) is sometimes called a *pseudometric*. Thus, a pseudometric will permit distinct points to be 0 distance apart.

The couple (M, d), consisting of a set M together with a metric d defined on M, is called a **metric space**. If a particular metric on M is understood, or if the argument at hand works equally well for any metric, we may forego this formality and simply refer to the set M as a metric space, with the tacit understanding that a metric d is available on demand.

Examples 3.1

(a) Every set M admits at least one metric. For example, check that the function defined by $d(x, y) = 1$ for any $x \neq y$ in M, and $d(x, x) = 0$ for all x in M, is a metric. This mundane, but always available, metric is called the **discrete metric** on M. It will prove to be much more interesting than first appearances suggest. A set supplied with its discrete metric will be called a **discrete space**.

(b) An important example for our purposes is the real line $\mathbb{R}$ together with its usual metric $d(a, b) = |a - b|$. Any time we refer to $\mathbb{R}$ without explicitly naming a metric, the absolute value metric is always understood to be the one that we have in mind.

(c) Any subset of a metric space is again a metric space in a very natural way. If d is a metric on M, and if A is a subset of M, then $d(x, y)$ is defined for any pair of points $x, y \in A$. Moreover, the restriction of d to $A \times A$ obviously still satisfies properties (i)–(iv). That is, the metric that is defined on M automatically defines a metric on A by restriction. We will even use the same letter d and simply refer to the metric space (A, d). Of particular interest in this regard is that $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R} \setminus \mathbb{Q}$ each come already supplied with a natural metric, namely, the restriction of the usual metric on $\mathbb{R}$. In each case, we will refer to this restriction as the usual metric.

EXERCISES

1. Show that

$$d(x, y) = \left| \frac{1}{x} - \frac{1}{y} \right|$$

defines a metric on $(0, \infty)$.

▷ **2.** If d is a metric on M, show that $|d(x, z) - d(y, z)| \le d(x, y)$ for any $x, y, z \in M$.

3. As it happens, some of our requirements for a metric are redundant. To see why this is so, let M be a set and suppose that $d : M \times M \to \mathbb{R}$ satisfies $d(x, y) = 0$ if and only if $x = y$, and $d(x, y) \le d(x, z) + d(y, z)$ for all $x, y, z \in M$. Prove that d is a metric; that is, show that $d(x, y) \ge 0$ and $d(x, y) = d(y, x)$ hold for all x, y.

4. Let M be a set and suppose that $d : M \times M \to [0, \infty)$ satisfies properties (i), (ii), and (iii) for a metric on M and the triangle inequality *reversed*: $d(x, y) \ge d(x, z) + d(z, y)$. Prove that M has at most one point.

▷ **5.** There are other, albeit less natural, choices for a metric on $\mathbb{R}$. For instance, check that $\rho(a, b) = \sqrt{|a - b|}$, $\sigma(a, b) = |a - b|/(1 + |a - b|)$, and $\tau(a, b) = \min\{|a - b|, 1\}$ each define metrics on $\mathbb{R}$. [Hint: To show that σ is a metric, you might first show that the function $F(t) = t/(1 + t)$ is increasing and satisfies $F(s + t) \le F(s) + F(t)$ for $s, t \ge 0$. A similar approach will also work for ρ and τ.]

▷ **6.** If d is any metric on M, show that $\rho(x, y) = \sqrt{d(x, y)}$, $\sigma(x, y) = d(x, y)/(1 + d(x, y))$, and $\tau(x, y) = \min\{d(x, y), 1\}$ are also metrics on M. [Hint: $\sigma(x, y) = F(d(x, y))$, where F is as in Exercise 5.]

7. Here is a generalization of Exercises 5 and 6. Let $f : [0, \infty) \to [0, \infty)$ be increasing and satisfy $f(0) = 0$, and $f(x) > 0$ for all $x > 0$. If f also satisfies

$f(x+y) \le f(x) + f(y)$ for all $x, y \ge 0$, then $f \circ d$ is a metric whenever d is a metric. Show that each of the following conditions is sufficient to ensure that $f(x+y) \le f(x) + f(y)$ for all $x, y \ge 0$:
(a) f has a second derivative satisfying $f'' \le 0$;
(b) f has a decreasing first derivative;
(c) $f(x)/x$ is decreasing for $x > 0$.
[Hint: First show that (a) $\Longrightarrow$ (b) $\Longrightarrow$ (c).]

8. If d_1 and d_2 are both metrics on the same set M, which of the following yield metrics on M: $d_1 + d_2$? $\max\{d_1, d_2\}$? $\min\{d_1, d_2\}$? If d is a metric, is d^2 a metric?

9. Recall that $2^{\mathbb{N}}$ denotes the set of all sequences (or "strings") of 0s and 1s. Show that $d(a, b) = \sum_{n=1}^{\infty} 2^{-n}|a_n - b_n|$, where $a = (a_n)$ and $b = (b_n)$ are sequences of 0s and 1s, defines a metric on $2^{\mathbb{N}}$.

10. The *Hilbert cube* H^∞ is the collection of all real sequences $x = (x_n)$ with $|x_n| \le 1$ for $n = 1, 2, \dots$.
(i) Show that $d(x, y) = \sum_{n=1}^{\infty} 2^{-n}|x_n - y_n|$ defines a metric on H^∞.
(ii) Given $x, y \in H^\infty$ and $k \in \mathbb{N}$, let $M_k = \max\{|x_1 - y_1|, \dots, |x_k - y_k|\}$. Show that $2^{-k}M_k \le d(x, y) \le M_k + 2^{-k}$.

11. Let $\mathbb{R}^\infty$ denote the collection of all real sequences $x = (x_n)$. Show that the expression

$$d(x, y) = \sum_{n=1}^{\infty} \frac{1}{n!} \frac{|x_n - y_n|}{1 + |x_n - y_n|}$$

defines a metric on $\mathbb{R}^\infty$.

12. Check that $d(f, g) = \max_{a \le t \le b} |f(t) - g(t)|$ defines a metric on $C[a, b]$, the collection of all continuous, real-valued functions defined on the closed interval $[a, b]$.

13. Fréchet's metric on $C[0, 1]$ is by no means the only choice (although we will see later that it is a good one). For example, show that $\rho(f, g) = \int_0^1 |f(t) - g(t)|\,dt$ and $\sigma(f, g) = \int_0^1 \min\{|f(t) - g(t)|, 1\}\,dt$ also define metrics on $C[0, 1]$.

▷ **14.** We say that a subset A of a metric space M is **bounded** if there is some $x_0 \in M$ and some constant $C < \infty$ such that $d(a, x_0) \le C$ for all $a \in A$. Show that a finite union of bounded sets is again bounded.

▷ **15.** We define the **diameter** of a nonempty subset A of M by $\operatorname{diam}(A) = \sup\{d(a, b) : a, b \in A\}$. Show that A is bounded if and only if $\operatorname{diam}(A)$ is finite.

Normed Vector Spaces

A large and important class of metric spaces are also vector spaces (over $\mathbb{R}$ or $\mathbb{C}$). Notice, for example, that $C[0, 1]$ is a vector space (and even a ring). An easy way to build a metric on a vector space is by way of a length function or norm. A **norm** on a vector space V is a function $\|\cdot\| : V \to [0, \infty)$ satisfying:

(i) $0 \le \|x\| < \infty$ for all $x \in V$;
(ii) $\|x\| = 0$ if and only if $x = 0$ (the zero vector in V);
(iii) $\|\alpha x\| = |\alpha|\,\|x\|$ for any scalar α and any $x \in V$; and
(iv) **the triangle inequality**: $\|x + y\| \le \|x\| + \|y\|$ for all $x, y \in V$.

A function $\|\cdot\| : V \to [0, \infty)$ satisfying all of the above properties except (ii) is called a *pseudonorm* on V; that is, a pseudonorm permits nonzero vectors to have 0 length.

The pair $(V, \|\cdot\|)$, consisting of a vector space V together with a norm on V, is called a **normed vector space** (or normed *linear* space). Just as with metric spaces, we may be a bit lax with this formality. Phrases such as "let V be a normed vector space" carry the tacit understanding that a norm is lurking about in the background.

It is easy to see that any norm induces a metric on V by setting $d(x, y) = \|x - y\|$. We will refer to this particular metric as the **usual** metric on $(V, \|\cdot\|)$. We may even be so bold as to refer to $(V, \|\cdot\|)$ as a metric space with the clear understanding that the usual metric induced by the norm is the one that we have in mind. Not all metrics on a vector space come from norms, however, so we cannot afford to be totally negligent (see Exercise 16).

Examples 3.2

(a) The absolute value function $|\cdot|$ clearly defines a norm on $\mathbb{R}$.

(b) Each of the following defines a norm on $\mathbb{R}^n$:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|, \qquad \|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2},$$

and $\|x\|_\infty = \max_{1\le i\le n} |x_i|$, where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. The first and last expressions are very easy to check while the second takes a bit more work. (Although this is probably familiar from calculus, we will supply a proof shortly.) The function $\|\cdot\|_2$ is often called *the Euclidean norm* and is generally accepted as the norm of choice on $\mathbb{R}^n$. As it happens, for any $1 \le p < \infty$, the expression $\|x\|_p = \left(\sum |x_i|^p\right)^{1/p}$ defines a norm on $\mathbb{R}^n$; see Theorem 3.8.

(c) Each of the following defines a norm on $C[a, b]$:

$$\|f\|_1 = \int_a^b |f(t)|\,dt, \qquad \|f\|_2 = \left(\int_a^b |f(t)|^2\,dt\right)^{1/2},$$

$$\text{and} \qquad \|f\|_\infty = \max_{a\le t\le b} |f(t)|.$$

Again, the second expression is hardest to check (and we will do so later; for now, see Exercise 25). The last expression is generally taken as "the" norm on $C[a, b]$.

(d) If $(V, \|\cdot\|)$ is a normed vector space, and if W is a *linear subspace* of V, then W is also normed by $\|\cdot\|$. That is, the restriction of $\|\cdot\|$ to W defines a norm on W.

(e) We might also consider the sequence space analogues of the "scale" of norms on $\mathbb{R}^n$ given in (b). For $1 \le p < \infty$, we define ℓ_p to be the collection of all real sequences $x = (x_n)$ for which $\sum_{n=1}^{\infty} |x_n|^p < \infty$, and we define ℓ_∞ to be the collection of all bounded real sequences. Each ℓ_p is a vector space under "coordinatewise" addition and scalar multiplication. Moreover, the expression $\|x\|_p = \left(\sum |x_n|^p\right)^{1/p}$ if $1 \le p < \infty$ or $\|x\|_\infty = \sup_n |x_n|$ if $p = \infty$ defines a norm on ℓ_p. The cases $p = 1$ and $p = \infty$ are easy to check (see Exercise 21), the case $p = 2$ is given as Theorem 3.4, while the case $1 < p < \infty$ is given as Theorem 3.8.

We can complete the details of several of our examples if we prove that ℓ_2 is a vector space and that $\|\cdot\|_2$ is a norm on ℓ_2. Now it is easy to see that if $\|x\|_2 = 0$, then $x_n = 0$ for all n and hence that $x = 0$ (the zero vector in ℓ_2). Also, given $x \in \ell_2$ and $\alpha \in \mathbb{R}$, it is easy to see that $\alpha x \in \ell_2$, where $\alpha x = (\alpha x_n)$, and that $\|\alpha x\|_2 = |\alpha|\|x\|_2$. What is not so clear is whether $x + y = (x_n + y_n)$ is in ℓ_2 whenever x and y are in ℓ_2. In other words, if x and y are square-summable, does it follow that $x + y$ is square-summable? A moment's reflection will convince you that to answer this question we will need to know something about the "dot product" $\sum x_n y_n$. This extra bit of information is supplied by the following lemma.

Lemma 3.3. (The Cauchy–Schwarz Inequality) $\sum_{i=1}^{\infty} |x_i y_i| \le \|x\|_2 \, \|y\|_2$ *for any* $x, y \in \ell_2$.

PROOF. To simplify our notation a bit, let's agree to write $\langle x, y\rangle = \sum x_i y_i$. We first consider the case where $x, y \in \mathbb{R}^n$ (that is, $x_i = 0 = y_i$ for all $i > n$). In this case, $\langle x, y\rangle$ is the usual "dot product" in $\mathbb{R}^n$. Also notice that we may suppose that $x, y \neq 0$. (There is nothing to show if either is 0.)

Now let $t \in \mathbb{R}$ and consider

$$0 \le \|x + ty\|_2^2 = \langle x + ty, x + ty\rangle = \|x\|_2^2 + 2t\langle x, y\rangle + t^2\|y\|_2^2.$$

Since this (nontrivial) quadratic in t is always nonnegative, it must have a nonpositive discriminant. (Why?) Thus, $\left(2\langle x, y\rangle\right)^2 - 4\|x\|_2^2\,\|y\|_2^2 \le 0$ or, after simplifying, $|\langle x, y\rangle| \le \|x\|_2\,\|y\|_2$. That is, $\left|\sum_{i=1}^{n} x_i y_i\right| \le \|x\|_2\,\|y\|_2$.

Now this isn't quite what we wanted, but it actually implies the stronger inequality in the statement of the lemma. Why? Because the inequality that we have shown must also hold for the vectors $(\,|x_i|\,)$ and $(\,|y_i|\,)$. That is,

$$\sum_{i=1}^{n} |x_i|\,|y_i| \;\le\; \|(\,|x_i|\,)\|_2\,\|(\,|y_i|\,)\|_2 \;=\; \|x\|_2\,\|y\|_2.$$

Finally, let $x, y \in \ell_2$. Then for each n we have

$$\sum_{i=1}^{n} |x_i y_i| \;\le\; \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} \left(\sum_{i=1}^{n} |y_i|^2\right)^{1/2} \;\le\; \|x\|_2\,\|y\|_2.$$

Thus, $\sum_{i=1}^{\infty} x_i y_i$ must be absolutely convergent and satisfy $\sum_{i=1}^{\infty} |x_i y_i| \le \|x\|_2\,\|y\|_2$. □

Now we are ready to prove the triangle inequality for the ℓ_2-norm.

Theorem 3.4. (Minkowski's Inequality) *If $x, y \in \ell_2$, then $x + y \in \ell_2$. Moreover, $\|x + y\|_2 \le \|x\|_2 + \|y\|_2$.*

PROOF. It follows from the Cauchy–Schwarz inequality that for each n we have

$$\begin{aligned}\sum_{i=1}^{n} |x_i + y_i|^2 &= \sum_{i=1}^{n} |x_i|^2 + 2\sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} |y_i|^2 \\ &\le \|x\|_2^2 + 2\|x\|_2\,\|y\|_2 + \|y\|_2^2 = (\|x\|_2 + \|y\|_2)^2 .\end{aligned}$$

Thus, since n is arbitrary, we have $x + y \in \ell_2$ and $\|x + y\|_2 \le \|x\|_2 + \|y\|_2$. □

We have now shown that ℓ_2 is a vector space and that $\|\cdot\|_2$ is a norm on ℓ_2. As you have no doubt already surmised, the proof is essentially identical to the one used to show that $\|\cdot\|_2$ is a norm on $\mathbb{R}^n$. In the next section a variation on this theme will be used to prove that ℓ_p is a vector space and that $\|\cdot\|_p$ is a norm.

EXERCISES

16. Let V be a vector space, and let d be a metric on V satisfying $d(x, y) = d(x - y, 0)$ and $d(\alpha x, \alpha y) = |\alpha|\, d(x, y)$ for every $x, y \in V$ and every scalar α. Show that $\|x\| = d(x, 0)$ defines a norm on V (that has d as its "usual" metric). Give an example of a metric on the vector space $\mathbb{R}$ that fails to be associated with a norm in this way.

17. Recall that for $x \in \mathbb{R}^n$ we have defined $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ and $\|x\|_\infty = \max_{1\le i\le n} |x_i|$. Check that each of these is indeed a norm on $\mathbb{R}^n$.

▷ **18.** Show that $\|x\|_\infty \le \|x\|_2 \le \|x\|_1$ for any $x \in \mathbb{R}^n$. Also check that $\|x\|_1 \le n\|x\|_\infty$ and $\|x\|_1 \le \sqrt{n}\,\|x\|_2$.

19. Show that we have $\sum_{i=1}^{n} x_i y_i = \|x\|_2\,\|y\|_2$ (equality in the Cauchy–Schwarz inequality) if and only if x and y are *proportional*, that is, if and only if either $x = \alpha y$ or $y = \alpha x$ for some $\alpha \ge 0$.

20. Show that $\|A\| = \max_{1\le i\le n} \left(\sum_{j=1}^{m} |a_{i,j}|^2\right)^{1/2}$ is a norm on the vector space $\mathbb{R}^{n\times m}$ of all $n \times m$ real matrices $A = [a_{i,j}]$.

21. Recall that we defined ℓ_1 to be the collection of all absolutely summable sequences under the norm $\|x\|_1 = \sum_{n=1}^{\infty} |x_n|$, and we defined ℓ_∞ to be the collection of all bounded sequences under the norm $\|x\|_\infty = \sup_{n\ge 1} |x_n|$. Fill in the details showing that each of these spaces is in fact a normed vector space.

22. Show that $\|x\|_\infty \le \|x\|_2$ for any $x \in \ell_2$, and that $\|x\|_2 \le \|x\|_1$ for any $x \in \ell_1$.

23. The subset of ℓ_∞ consisting of all sequences that converge to 0 is denoted by c_0. (Note that c_0 is actually a linear subspace of ℓ_∞; thus c_0 is also a normed vector space under $\|\cdot\|_\infty$.) Show that we have the following *proper* set inclusions: $\ell_1 \subset \ell_2 \subset c_0 \subset \ell_\infty$.

More Inequalities

We next supply the promised extension of Theorem 3.4 to the spaces ℓ_p, $1 < p < \infty$. Just as in the case of ℓ_2, notice that several facts are easy to check. For example, it is clear that $\|x\|_p = 0$ implies that $x = 0$, and it is easy to see that $\|\alpha x\|_p = |\alpha|\,\|x\|_p$ for any scalar α. Thus we lack only the triangle inequality. We begin with a few classical inequalities that are of interest in their own right. The first shows that ℓ_p is at least a vector space:

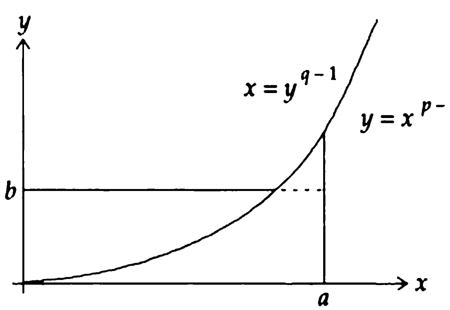
Lemma 3.5. *Let $1 < p < \infty$ and let $a, b \geq 0$. Then, $(a+b)^p \leq 2^p(a^p + b^p)$. Consequently, $x + y \in \ell_p$ whenever $x, y \in \ell_p$.*

PROOF. $(a+b)^p \leq (2\max\{a,b\})^p = 2^p \max\{a^p, b^p\} \leq 2^p(a^p + b^p)$. Thus, if $x, y \in \ell_p$, then $\sum_{n=1}^{\infty} |x_n + y_n|^p \leq 2^p \sum_{n=1}^{\infty} |x_n|^p + 2^p \sum_{n=1}^{\infty} |y_n|^p < \infty$. □

Lemma 3.6. (Young's Inequality) *Let $1 < p < \infty$ and let q be defined by $1/p + 1/q = 1$. Then, for any $a, b \geq 0$, we have $ab \leq a^p/p + b^q/q$, with equality occurring if and only if $a^{p-1} = b$.*

PROOF. Since the inequality trivially holds if either a or b is 0, we may certainly suppose that $a, b > 0$. Next notice that $q = p/(p-1)$ also satisfies $1 < q < \infty$ and $p - 1 = p/q = 1/(q-1)$. Thus, the functions $f(t) = t^{p-1}$ and $g(t) = t^{q-1}$ are *inverses* for $t \geq 0$.

The proof of the inequality follows from a comparison of areas (see Figure 3.1). The area of the rectangle with sides of lengths a and b is at most the sum of the areas under the graphs of the functions $y = x^{p-1}$ for $0 \leq x \leq a$ and $x = y^{q-1}$ for



$0 \leq y \leq b$. That is,

$$ab \leq \int_0^a x^{p-1}\,dx + \int_0^b y^{q-1}\,dy = \frac{a^p}{p} + \frac{b^q}{q}.$$

Clearly, equality can occur only if $a^{p-1} = b$. □

When $p = q = 2$, Young's inequality reduces to the *arithmetic–geometric mean inequality* (although it is usually stated in the form $\sqrt{ab} \leq (a+b)/2$). Young's inequality will supply the extension of the Cauchy–Schwarz inequality that we need.

Lemma 3.7. (Hölder's Inequality) *Let $1 < p < \infty$ and let q be defined by $1/p + 1/q = 1$. Given $x \in \ell_p$ and $y \in \ell_q$, we have $\sum_{i=1}^{\infty} |x_i y_i| \le \|x\|_p \, \|y\|_q$.*

PROOF. We may suppose that $\|x\|_p > 0$ and $\|y\|_q > 0$ (since, otherwise, there is nothing to show). Now, for $n \ge 1$ we use Young's inequality to estimate:

$$\sum_{i=1}^{n} \left| \frac{x_i y_i}{\|x\|_p \, \|y\|_q} \right| \le \frac{1}{p} \sum_{i=1}^{n} \left| \frac{x_i}{\|x\|_p} \right|^p + \frac{1}{q} \sum_{i=1}^{n} \left| \frac{y_i}{\|y\|_q} \right|^q \le \frac{1}{p} + \frac{1}{q} = 1.$$

Thus, $\sum_{i=1}^{n} |x_i y_i| \le \|x\|_p \, \|y\|_q$ for any $n \ge 1$, and the result follows. □

Our proof of the triangle inequality will be made easier if we first isolate one of the key calculations. Notice that if $x \in \ell_p$, then the sequence $(\,|x_n|^{p-1})_{n=1}^{\infty} \in \ell_q$, because $(p-1)q = p$. Moreover,

$$\|(\,|x_n|^{p-1})\|_q = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/q} = \|x\|_p^{p-1}.$$

Theorem 3.8. (Minkowski's Inequality) *Let $1 < p < \infty$. If $x, y \in \ell_p$, then $x + y \in \ell_p$ and $\|x + y\|_p \le \|x\|_p + \|y\|_p$.*

PROOF. We have already shown that $x + y \in \ell_p$ (Lemma 3.5). To prove the triangle inequality, we once again let q be defined by $1/p + 1/q = 1$, and we now use Hölder's inequality to estimate:

$$\begin{aligned}
\sum_{i=1}^{\infty} |x_i + y_i|^p &= \sum_{i=1}^{\infty} |x_i + y_i| \cdot |x_i + y_i|^{p-1} \\
&\le \sum_{i=1}^{\infty} |x_i| \cdot |x_i + y_i|^{p-1} + \sum_{i=1}^{\infty} |y_i| \cdot |x_i + y_i|^{p-1} \\
&\le \|x\|_p \cdot \|(\,|x_n + y_n|^{p-1})\|_q + \|y\|_p \cdot \|(\,|x_n + y_n|^{p-1})\|_q \\
&= \|x + y\|_p^{p-1} \left(\|x\|_p + \|y\|_p \right).
\end{aligned}$$

That is, $\|x + y\|_p^p \le \|x + y\|_p^{p-1} \left(\|x\|_p + \|y\|_p \right)$, and the triangle inequality follows. □

EXERCISES

24. The conclusion of Lemma 3.7 also holds in the case $p = 1$ and $q = \infty$. Why?

25. The same techniques can be used to show that $\|f\|_p = \left(\int_0^1 |f(t)|^p \, dt \right)^{1/p}$ defines a norm on $C[0, 1]$ for any $1 < p < \infty$. State and prove the analogues of Lemma 3.7 and Theorem 3.8 in this case. (Does Lemma 3.7 still hold in this setting for $p = 1$ and $q = \infty$?)

26. Given $a, b > 0$, show that $\lim_{p\to\infty}(a^p + b^p)^{1/p} = \max\{a, b\}$. [Hint: If $a < b$ and $r = a/b$, show that $(1/p)\log(1 + r^p) \to 0$ as $p \to \infty$.] What happens as $p \to 0$? as $p \to -1$? as $p \to -\infty$?

Limits in Metric Spaces

Now that we have generalized the notion of distance, we can easily define the notions of convergence and continuity in metric spaces. It will help a bit, though, if we first generate some notation for "small" sets. Throughout this section, unless otherwise specified, we will assume that we are always dealing with a generic metric space (M, d).

Given $x \in M$ and $r > 0$, the set $B_r(x) = \{y \in M : d(x, y) < r\}$ is called the **open ball** about x of radius r. If we also need to refer to the metric d, then we write $B_r^d(x)$. We may occasionally refer to the set $\{y \in M : d(x, y) \leq r\}$ as the *closed* ball about x of radius r, but we will not bother with any special notation for closed balls.

Examples 3.9

(a) In $\mathbb{R}$ we have $B_r(x) = (x - r, x + r)$, the open interval of radius r about x, while in $\mathbb{R}^2$ the set $B_r(x)$ is the open disk of radius r centered at x.

(b) In a *discrete* space $B_1(x) = \{x\}$ and $B_2(x) = M$.

(c) In a normed vector space $(V, \|\cdot\|)$ the balls centered at 0 play a special role (see Exercise 32); in this setting $B_r(0) = \{x : \|x\| < r\}$.

A subset A of M is said to be **bounded** if it is contained in some ball, that is, if $A \subset B_r(x)$ for some $x \in M$ and some $r > 0$. But exactly which x and r does not much matter. In fact, A is bounded if and only if for *any* $x \in M$ we have $\sup_{a\in A} d(x, a) < \infty$. (Why?) Related to this is the **diameter** of A, defined by $\text{diam}(A) = \sup\{d(a, b) : a, b \in A\}$. The diameter of A is a convenient measure of size because it does not refer to points outside of A.

EXERCISES

Each of the following exercises is set in a generic metric space (M, d).

27. Show that $\text{diam}(B_r(x)) \leq 2r$, and give an example where strict inequality occurs.

28. If $\text{diam}(A) < r$, show that $A \subset B_r(a)$ for some $a \in A$.

▷ **29.** Prove that A is bounded if and only if $\text{diam}(A) < \infty$.

▷ **30.** If $A \subset B$, show that $\text{diam}(A) \leq \text{diam}(B)$.

31. Give an example where $\text{diam}(A \cup B) > \text{diam}(A) + \text{diam}(B)$. If $A \cap B \neq \emptyset$, show that $\text{diam}(A \cup B) \leq \text{diam}(A) + \text{diam}(B)$.

▷ **32.** In a normed vector space $(V, \|\cdot\|)$ show that $B_r(x) = x + B_r(0) = \{x + y : \|y\| < r\}$ and that $B_r(0) = rB_1(0) = \{rx : \|x\| < 1\}$.

A **neighborhood** of x is any set containing an open ball about x. You should think of a neighborhood of x as a "thick" set of points near x. We will use this new terminology to streamline our definition of convergence.

We say that a sequence of points (x_n) in M **converges** to a point $x \in M$ if $d(x_n, x) \to 0$. Now, since this definition is stated in terms of the sequence of *real numbers* $\left(d(x_n, x)\right)_{n=1}^{\infty}$, we can easily derive the following equivalent reformulations:

$$\begin{cases} (x_n) \text{ converges to } x \text{ if and only if, given any } \varepsilon > 0, \text{ there is} \\ \text{an integer } N \geq 1 \text{ such that } d(x_n, x) < \varepsilon \text{ whenever } n \geq N, \end{cases}$$

or

$$\begin{cases} (x_n) \text{ converges to } x \text{ if and only if, given any } \varepsilon > 0, \text{ there is} \\ \text{an integer } N \geq 1 \text{ such that } \{x_n : n \geq N\} \subset B_\varepsilon(x). \end{cases}$$

If it should happen that $\{x_n : n \geq N\} \subset A$ for some N, we say that the sequence (x_n) is **eventually** in A. Thus, our last formulation can be written

$$\begin{cases} (x_n) \text{ converges to } x \text{ if and only if, given any } \varepsilon > 0, \\ \text{the sequence } (x_n) \text{ is eventually in } B_\varepsilon(x) \end{cases}$$

or, in yet another incarnation,

$$\begin{cases} (x_n) \text{ converges to } x \text{ if and only if the sequence} \\ (x_n) \text{ is eventually in every neighborhood of } x. \end{cases}$$

This final version is blessed by a total lack of Ns and εs! In any event, just as with real sequences, we usually settle for the shorthand $x_n \to x$ in place of the phrase (x_n) converges to x. On occasion we will want to display the set M, or d, or both, and so we may also write $x_n \xrightarrow{d} x$ or $x_n \to x$ in (M, d). We also define Cauchy (or d-Cauchy, if we need to specify d) in the obvious way: A sequence (x_n) is **Cauchy** in (M, d) if, given any $\varepsilon > 0$, there is an integer $N \geq 1$ such that $d(x_m, x_n) < \varepsilon$ whenever $m, n \geq N$. We can reword this just a bit to read: (x_n) is Cauchy if and only if, given $\varepsilon > 0$, there is an integer $N \geq 1$ such that $\operatorname{diam}(\{x_n : n \geq N\}) \leq \varepsilon$. (How?)

Much of what we already know about sequences of real numbers will carry over to this new setting – but not everything! The reader is strongly encouraged to test the limits of this transition by supplying proofs for the following easy results.

EXERCISES

Each of the following exercises is set in a metric space M with metric d.

33. Limits are unique. [Hint: $d(x, y) \leq d(x, x_n) + d(x_n, y)$.]

▷ **34.** If $x_n \to x$ in (M, d), show that $d(x_n, y) \to d(x, y)$ for any $y \in M$. More generally, if $x_n \to x$ and $y_n \to y$, show that $d(x_n, y_n) \to d(x, y)$.

35. If $x_n \to x$, then $x_{n_k} \to x$ for any subsequence (x_{n_k}) of (x_n).

▷ **36.** A convergent sequence is Cauchy, and a Cauchy sequence is bounded (that is, the set $\{x_n : n \geq 1\}$ is bounded).

▷ **37.** A Cauchy sequence with a convergent subsequence converges.

38. A sequence (x_n) has a Cauchy subsequence if and only if it has a subsequence (x_{n_k}) for which $d\left(x_{n_k}, x_{n_{k+1}}\right) < 2^{-k}$ for all k.

▷ **39.** If every subsequence of (x_n) has a *further* subsequence that converges to x, then (x_n) converges to x.

Now, while several familiar results about sequences in $\mathbb{R}$ have carried over successfully to the "abstract" setting of metric spaces, at least a few will not survive the journey. Two especially fragile cases are: Cauchy sequences *need not* converge and bounded sequences *need not* have convergent subsequences. A few specific examples might help your appreciation of their delicacy.

Examples 3.10

(a) Consider the sequence $(1/n)_{n=1}^{\infty}$ living in the space $M = (0, 1]$ under its usual metric. Then, $(1/n)$ is Cauchy but, annoyingly, does not converge to any point in M. (Why?) Notice too that $(1/n)$ is a bounded sequence with no convergent subsequence.

(b) Consider $M = \mathbb{R}$ supplied with the discrete metric. Then, $(n)_{n=1}^{\infty}$ is a bounded sequence with no Cauchy subsequence!

(c) At least one good thing happens in any discrete space: Cauchy sequences always converge. But for a simple reason. In a discrete space, a sequence (x_n) is Cauchy if and only if it is *eventually constant*; that is, if and only if $x_n = x$ for some (fixed) x and all n sufficiently large. (Why?)

Let's take a closer look at $\mathbb{R}^n$ (with its usual metric). Since $d(x, y) = \|x - y\|_2 = \left(\sum_{i=1}^{n} |x_i - y_i|^2\right)^{1/2} \geq |x_j - y_j|$ for any $j = 1, \ldots, n$, it follows that a sequence of vectors $x^{(k)} = (x_1^k, \ldots, x_n^k)$ in $\mathbb{R}^n$ converges (is Cauchy) if and only if each of the coordinate sequences $(x_j^k)_{k=1}^{\infty}$ converges (is Cauchy) in $\mathbb{R}$. (Why?) Thus, nearly every fact about convergent sequences in $\mathbb{R}$ "lifts" successfully to $\mathbb{R}^n$. For example, any Cauchy sequence in $\mathbb{R}^n$ converges in $\mathbb{R}^n$, and any bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

How much of this has to do with the particular metric that we chose for $\mathbb{R}^n$? And will this same result "lift" to the spaces ℓ_1, ℓ_2, or ℓ_∞, for example? We cannot hope for much, but each of these spaces shares at least one thing in common with $\mathbb{R}^n$. Since all three of the norms $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ satisfy $\|x\| \geq |x_j|$ for any j, it follows that convergence in ℓ_1, ℓ_2, or ℓ_∞ will imply "coordinatewise" convergence. That is, if $x^{(k)} = (x_n^k)_{n=1}^{\infty}$, $k = 1, 2, \ldots$, is a sequence (of sequences!) in, say, ℓ_1, and if $x^{(k)} \to x$ in ℓ_1, then we must have $x_n^k \to x_n$ (as $k \to \infty$) for each $n = 1, 2, \ldots$. A simple example will convince you that the converse does not hold, in general, in this new setting. The sequence $e^{(k)} = (0, \ldots, 0, 1, 0, \ldots)$, where the kth entry is 1 and the rest are 0s, converges "coordinatewise" to $0 = (0, 0, \ldots)$, but $(e^{(k)})$ does *not* converge to 0 in any of the metric spaces ℓ_1, ℓ_2, or ℓ_∞. Why? Because in each of the three spaces we have $d(e^{(k)}, 0) = \|e^{(k)}\| = 1$. In fact, $(e^{(k)})$ is not even Cauchy because in each case we also have $\|e^{(k)} - e^{(m)}\| \geq 1$ for any $k \neq m$.

EXERCISES

40. Here is a positive result about ℓ_1 that may restore your faith in intuition. Given any (fixed) element $x \in \ell_1$, show that the sequence $x^{(k)} = (x_1, \ldots, x_k, 0, \ldots) \in \ell_1$ (i.e., the first k terms of x followed by all 0s) converges to x in ℓ_1-norm. Show that the same holds true in ℓ_2, but give an example showing that it fails (in general) in ℓ_∞.

41. Given $x, y \in \ell_2$, recall that $\langle x, y\rangle = \sum_{i=1}^{\infty} x_i y_i$. Show that if $x^{(k)} \to x$ and $y^{(k)} \to y$ in ℓ_2, then $\langle x^{(k)}, y^{(k)}\rangle \to \langle x, y\rangle$.

▷ **42.** Two metrics d and ρ on a set M are said to be **equivalent** if they generate the same convergent sequences; that is, $d(x_n, x) \to 0$ if and only if $\rho(x_n, x) \to 0$. If d is any metric on M, show that the metrics ρ, σ, and τ, defined in Exercise 6, are all equivalent to d.

▷ **43.** Show that the usual metric on $\mathbb{N}$ is equivalent to the discrete metric. Show that any metric on a *finite* set is equivalent to the discrete metric.

▷ **44.** Show that the metrics induced by $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ on $\mathbb{R}^n$ are all equivalent. [Hint: See Exercise 18.]

45. We say that two norms on the same vector space X are equivalent if the metrics they induce are equivalent. Show that $\|\cdot\|$ and $|||\cdot|||$ are equivalent on X if and only if they generate the same sequences tending to 0; that is, $\|x_n\| \to 0$ if and only if $|||x_n||| \to 0$.

▷ **46.** Given two metric spaces (M, d) and (N, ρ), we can define a metric on the product $M \times N$ in a variety of ways. Our only requirement is that a sequence of pairs (a_n, x_n) in $M \times N$ should converge precisely when both coordinate sequences (a_n) and (x_n) converge (in (M, d) and (N, ρ), respectively). Show that each of the following define metrics on $M \times N$ that enjoy this property and that all three are equivalent:

$$d_1\big((a, x), (b, y)\big) = d(a, b) + \rho(x, y),$$
$$d_2\big((a, x), (b, y)\big) = \big(d(a, b)^2 + \rho(x, y)^2\big)^{1/2},$$
$$d_\infty\big((a, x), (b, y)\big) = \max\{d(a, b), \rho(x, y)\}.$$

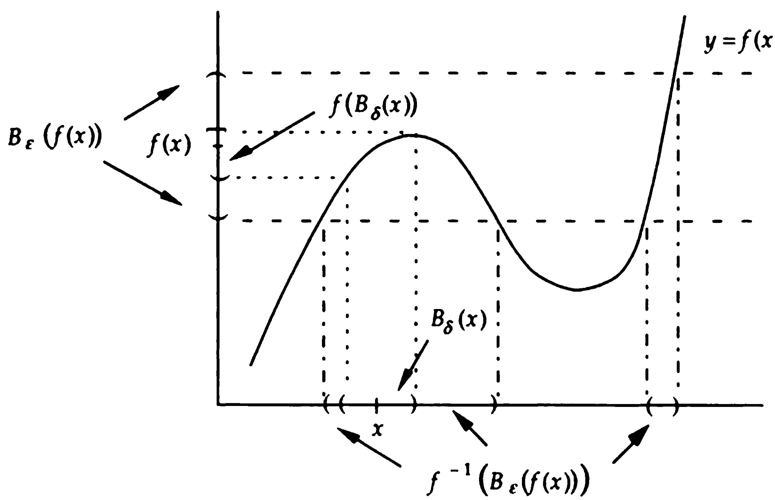
Henceforth, any implicit reference to "the" metric on $M \times N$, sometimes called **the product metric**, will mean one of d_1, d_2, or d_∞. Any one of them will serve equally well; use whichever looks most convenient for the argument at hand.

While we are not yet ready for an all-out attack on continuity, it couldn't hurt to give a hint as to what is ahead. Given a function $f : (M, d) \to (N, \rho)$ between two metric spaces, and given a point $x \in M$, we have at least two plausible sounding definitions for the continuity of f at x. Each definition is derived from its obvious counterpart for real-valued functions by replacing absolute values with an appropriate metric.

For example, we might say that f is continuous at x if $\rho(f(x_n), f(x)) \to 0$ whenever $d(x_n, x) \to 0$. That is, f should send sequences converging to x into sequences converging to $f(x)$. This says that f "commutes" with limits: $f(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} f(x_n)$. Sounds like a good choice.

Or we might try doctoring the familiar ε-δ definition from a first course in calculus. In this case we would say that f is continuous at x if, given any $\varepsilon > 0$, there always exists a $\delta > 0$ such that $\rho(f(x), f(y)) < \varepsilon$ whenever $d(x, y) < \delta$. Written in slightly different terms, this definition requires that $f\left(B_\delta^d(x)\right) \subset B_\varepsilon^\rho(f(x))$. That is, f maps a sufficiently small neighborhood of x into a given neighborhood of $f(x)$.

We will rewrite the definition once more, but this time we will use an inverse image. Recall that the *inverse image* of a set A, under a function $f : X \to Y$, is defined to be the *set* $\{x \in X : f(x) \in A\}$ and is usually written $f^{-1}(A)$. (The inverse image of any set under any function always makes sense. Although the notation is similar, inverse images have nothing whatever to do with inverse *functions*, which don't always make sense.) Stated in terms of an inverse image, our condition reads: $B_\delta^d(x) \subset f^{-1}\left(B_\varepsilon^\rho(f(x))\right)$. Look a bit imposing? Well, it actually tells us quite a bit. It says that the inverse image of a "thick" set containing $f(x)$ must still be "thick" near x. Curious. Figure 3.2 may help you with these new definitions. Better still, draw a few pictures of your own!



This sets the stage for what is ahead. Each of the two possible definitions for continuity seems perfectly reasonable. Certainly we would hope that the two turn out to be equivalent. But what do convergent sequences have to do with "thick" sets? And just what is a "thick" set anyway?

Notes and Remarks

The quotation at the start of this chapter is taken from Fréchet [1950]; his thesis appears in Fréchet [1906]. His book, Fréchet [1928], was published as one of the volumes in a series of monographs edited by Émile Borel. The authors in this series include every "name" French mathematician of that time: Baire, Borel, Lebesgue, Lévy, de La Vallée Poussin, and many others. The full title of Fréchet's book, including subtitle, is enlightening: *Les espaces abstraits et leur théorie considérée comme introduction a l'analyse générale* (Abstract spaces and their theory considered as an introduction to general analysis). The paper by Riesz mentioned in the introductory passage is Riesz [1906].

It was Hausdorff who gave us the name "metric space." Indeed, his classic work *Grundzüge der Mengenlehre*, Leipzig, 1914, is the source for much of our terminology regarding abstract sets and abstract spaces. An English translation of Hausdorff's book is available as *Set Theory* (Hausdorff [1937]). If we had left it up to Fréchet, we would be calling metric spaces "spaces of type (D)."

For more on metric spaces, normed spaces, and $\mathbb{R}^n$, see Copson [1968], Goffman and Pedrick [1965], Goldberg [1976], Hoffman [1975], Kaplansky [1977], Kasriel [1971], Kolmogorov and Fomin [1970], and Kuller [1969]. For a look at modern applications of metric space notions, see Barnsley [1988] and Edgar [1990].

Normed vector spaces were around for some time before anyone bothered to formalize their definition. Quite often you will see the great Polish mathematician Stefan Banach mentioned as the originator of normed vector spaces, but this is only partly true. In any case, it is fair to say that Banach gave the first *thorough* treatment of normed vector spaces, beginning with his thesis (Banach [1922]). We will have cause to mention Banach's name frequently in these notes.

The several "name" inequalities that we saw in this chapter are, for the most part, older than the study of norms and metrics. Most fall into the category of "mean values" (various types of averages). An excellent source of information on inequalities and mean values of every shape and size is a dense little book with the apt title *Inequalities*, by Hardy, Littlewood, and Pólya [1952]. Beckenbach and Bellman [1961] provide an elementary introduction to inequalities, including a few applications. For a very slick, yet elementary proof of the inequalities of Hölder and Minkowski, see Maligranda [1995].

Certain applications to numerical analysis and computational mathematics have caused a renewed interest in mean values. For a brief introduction to this exciting area, see the selection "On the arithmetic-geometric mean and similar iterative algorithms" in Schoenberg [1982], and the articles by Almkvist and Berndt [1988], Carlson [1971], and Miel [1983]. For a discussion of some of the computational practicalities, see D. H. Bailey [1988].