Chapter 1

Setting the Stage

This chapter contains a mix of topics needed to study the remainder of this book. It has a little of metric space theory and some basics on normed vector spaces. Many readers will be familiar with some of these topics, but few will have seen them all. We start with a definition of the Riemann–Stieltjes integral, which I suspect is new to most.

1.1. Riemann–Stieltjes integrals

For a fixed closed, bounded interval $J = [a, b]$ in $\mathbb{R}$, the set of all real numbers, we want to define an extension of the usual Riemann integral from calculus. This extended integral will also assign a number to each continuous function on the interval, though later we will see how to extend it even further so we can integrate more general functions than the continuous ones. This more general integral, however, will be set in a far broader context than intervals in $\mathbb{R}$.

As in calculus, a *partition* of J is a finite, ordered subset $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$. Say that P is a *refinement* of a partition Q if $Q \subseteq P$; so P adds additional points to Q.

1.1.1. Definition. A function $\alpha : J \to \mathbb{R}$ is of *bounded variation* if there is a constant M such that for every partition $P = \{a = x_0 < \cdots < x_n = b\}$ of J,

$$\sum_{j=1}^{n} |\alpha(x_j) - \alpha(x_{j-1})| \leq M.$$

The quantity

$$\mathrm{Var}(\alpha) = \mathrm{Var}(\alpha, J) = \sup \left\{ \sum_{j=1}^{n} |\alpha(x_j) - \alpha(x_{j-1})| : P \text{ is a partition of } J \right\}$$

is called the *total variation* of α over J.

We'll see many examples of functions of bounded variation. The first below is specific and the second contains a collection of such functions.

1.1.2. Example. (a) The "mother of all" functions of bounded variation is $\alpha(x) = x$.

(b) Suppose $\alpha : [a, b] \to \mathbb{R}$ is a continuously differentiable function and M is a constant with $|\alpha'(t)| \leq M$ for all t in $[a, b]$. If $a = x_0 < x_1 < \cdots < x_n = b$, then for each j the Mean Value Theorem for derivatives says there is a point t_j in $[x_{j-1}, x_j]$ such that $\alpha(x_j) - \alpha(x_{j-1}) = \alpha'(t_j)(x_j - x_{j-1})$. Hence $\sum_j |\alpha(x_j) - \alpha(x_{j-1})| = \sum_j |\alpha'(t_j)|(x_j - x_{j-1}) \leq M(b - a)$, so that α is of bounded variation.

A function $\alpha : J \to \mathbb{R}$ is *increasing* if $\alpha(s) \leq \alpha(t)$ when $a \leq s \leq t \leq b$; the function is *decreasing* if $\alpha(s) \geq \alpha(t)$ when $a \leq s \leq t \leq b$. (We will usually avoid the terms non-decreasing and non-increasing as linguistically and psychologically awkward.) We sometimes use the terms *strictly increasing* and *strictly decreasing* when they are called for, though the reader will see this is not frequent. We note that α is a decreasing function if and only if the function $\beta : [-b, -a] \to \mathbb{R}$ defined by $\beta(s) = \alpha(-s)$ is increasing. It thus becomes possible to state and prove results for decreasing functions whenever we have a result for increasing functions. In what follows neither the separate statements for decreasing functions nor their proofs will be made explicit.

1.1.3. Example. (a) If f is a positive continuous function on $[a, b]$, then $\alpha(t) = \int_a^t f(x)dx$ is an increasing function. In fact any continuously differentiable function with a non-negative derivative is increasing.

(b) If α is a function of bounded variation and β is defined on J by $\beta(t) = \mathrm{Var}(\alpha, [a, t])$, then β is an increasing function.

If X is any set and $f, g : X \to \mathbb{R}$ are any functions, then

$$(f \vee g)(x) = \max\{f(x), g(x)\}$$

defines another function $f \vee g : X \to \mathbb{R}$. Similarly

$$(f \wedge g)(x) = \min\{f(x), g(x)\}$$

defines a function $f \wedge g : X \to \mathbb{R}$. We will see this notation frequently during the course of this book. The proofs of the next two results are routine and left to the reader.

1.1.4. Proposition. *If α, β are increasing functions on J, then so are $\alpha \vee \beta$ and $\alpha \wedge \beta$.*

For any interval $[a, b]$ let $BV[a, b]$ denote the set of all functions of bounded variation defined on $[a, b]$.

1.1.5. Proposition. (a) *Every increasing function on a bounded interval is of bounded variation.*

(b) *$BV[a, b]$ is a vector space over $\mathbb{R}$.*

Perhaps we might make explicit the definition of addition of two functions in $BV[a, b]$. If $\alpha, \beta \in BV[a, b]$, $\alpha + \beta$ is defined by $(\alpha + \beta)(t) = \alpha(t) + \beta(t)$ for all t in the interval. Similarly, if $a \in \mathbb{R}$ and $\alpha \in BV[a, b]$, $(a\alpha)(t) = a\alpha(t)$. This is referred to as defining the algebraic operations *pointwise* and will be seen repeatedly.

In light of the preceding proposition any linear combination of increasing functions is a function of bounded variation. The surprising thing is that the converse holds.

1.1.6. Proposition. *If $\alpha : [a, b] \to \mathbb{R}$ is a function of bounded variation, then we can write $\alpha = \alpha_+ - \alpha_-$, where $\alpha_\pm$ are increasing functions.*

Proof. Let $\alpha_+(t) = \frac{1}{2}[\mathrm{Var}(\alpha, [a, t]) + \alpha(t)]$ and $\alpha_-(t) = \frac{1}{2}[\mathrm{Var}(\alpha, [a, t]) - \alpha(t)]$. It is clear that $\alpha = \alpha_+ - \alpha_-$, so what we have to do is show that these functions are increasing. Let $t > s$, $\epsilon > 0$, and let $a = x_0 < \cdots < x_n = s$ such that $\sum_j |\alpha(x_j) - \alpha(x_{j-1})| > \mathrm{Var}(\alpha, [a, s]) - \epsilon$. Now it is easy to verify that

$$|\alpha(t) - \alpha(s)| \pm [\alpha(t) - \alpha(s)] \pm \alpha(s) \geq \pm\alpha(s)$$

(Note that we are not allowed to randomly make a choice of sign each time $\pm$ appears; it must be consistent.) Since $a = x_0 < \cdots < x_n < t$ is a partition of $[a, t]$, we get that

$$\begin{aligned} \mathrm{Var}(\alpha, [a, t]) \pm \alpha(t) &\geq \sum_{j=1}^{n} |\alpha(x_j) - \alpha(x_{j-1})| + |\alpha(t) - \alpha(s)| \\ &\quad \pm \big([\alpha(t) - \alpha(s)] + \alpha(s)\big) \\ &= \sum_{j=1}^{n} |\alpha(x_j) - \alpha(x_{j-1})| \\ &\quad + |\alpha(t) - \alpha(s)| \pm [\alpha(t) - \alpha(s)] \pm \alpha(s) \\ &\geq \mathrm{Var}(\alpha, [a, s]) - \epsilon \pm \alpha(s) \end{aligned}$$

Since ϵ is arbitrary, we have that $\text{Var}(\alpha, [a,t]) \pm \alpha(t) \geq \text{Var}(\alpha, [a,s]) \pm \alpha(s)$ and so the functions $\alpha_\pm$ are increasing. ■

Now that we have discussed functions of bounded variation, gotten many examples, and discovered a structure of such functions (1.1.6), we might pose a question. Why the interest? The important thing for us is that we can define integrals or averaging processes for continuous functions by using a function of bounded variation. These integrals have geometric interpretations as well as applications to the study of various problems in analysis. Let's define the integrals, where the reader will notice a close similarity with the definition of the Riemann integral. Indeed if α is the increasing function $\alpha(t) = t$, then what we do below will result in the Riemann integral over J.

If α is a function in $BV(J)$, $f : J \to \mathbb{R}$ is some function, and P is a partition, define

$$S_\alpha(f, P) = \sum_{j=1}^{n} f(t_j)[\alpha(x_j) - \alpha(x_{j-1})]$$

where the points t_j are chosen in the subinterval $[x_{j-1}, x_j]$. Yes, the notation does not reflect the dependency of this sum on the choice of the points t_j, but I am afraid we'll just have to live with that; indicating such a dependency is more awkward than any gained benefit. When the function α is the special one, $\alpha(t) = t$, let $S_\alpha(f, P) = S(f, P)$. That is,

$$S(f, P) = \sum_{j=1}^{n} f(t_j)[x_j - x_{j-1}]$$

Define the *mesh* of the partition P to be the number $\|P\| = \max\{|x_j - x_{j-1}| : 1 \leq j \leq n\}$, and for any positive number δ let $\mathcal{P}_\delta$ denote the collection of all partitions P with $\|P\| < \delta$. Recall that for the Riemann integral the fundamental result on existence is that when $f : F \to \mathbb{R}$ is a continuous function, then there is a unique number I such that for every $\epsilon > 0$ there is a $\delta > 0$ with $|I - S(f, P)| < \epsilon$ whenever $P \in \mathcal{P}_\delta$. It is precisely this number I which is the Riemann integral of f and is denoted by $I = \int_a^b f(t)dt$. We now start the process of showing that a similar existence result holds if we replace the Riemann sum $S(f, P)$ by the sum $S_\alpha(f, P)$ for an arbitrary function of bounded variation α.

Here is another bit of notation that will simplify matters and is valid for any metric space (X, d). For a function $f : X \to \mathbb{R}$, the *modulus of continuity* of f for any $\delta > 0$ is the number $\omega(f, \delta) = \sup\{|f(x) - f(y)| : d(x, y) < \delta\}$. This will be infinite for some functions, but the main place we will use it is when X is compact and f is continuous. In that case f is uniformly continuous so that we have that for any $\epsilon > 0$ there is a δ such that $\omega(f, \delta) < \epsilon$.

Just as in the definition of the Riemann integral, we want to define the integral of a function with respect to a function of bounded variation α. Here is the crucial lemma to get us to that goal.

1.1.7. Lemma. *If α is a function of bounded variation on J and $f : J \to \mathbb{R}$ is a continuous function, then for any $\epsilon > 0$ there is a $\delta > 0$ such that $|S_\alpha(f, P) - S_\alpha(f, Q)| \leq \epsilon$ whenever $P, Q \in \mathcal{P}_\delta$.*

Proof. We start by observing that when P, Q are two partitions and $P \subseteq Q$, then $S_\alpha(f, P) \leq S_\alpha(f, Q)$; this is an easy consequence of the triangle inequality. For example, suppose $P = \{a = x_0 < \cdots < x_n = b\}$ and $Q = P \cup \{x_k^*\}$ with $x_{k-1} < x_k^* < x_k$. Writing out the definition of $S_\alpha(f, P)$ and applying the triangle inequality yields the desired relation. The proof of the general case is similar.

In light of the preceding observation, it suffices to prove that there is a δ such that when $P, Q \in \mathcal{P}_\delta$ and $P \subseteq Q$, then $|S_\alpha(f, P) - S_\alpha(f, Q)| \leq \epsilon/2$. In fact if this is done and we have that P, Q are arbitrary partitions in $\mathcal{P}_\delta$, then $P \cup Q \in \mathcal{P}_\delta$ and contains both P and Q. Hence we would have that $|S_\alpha(f, P) - S_\alpha(f, Q)| \leq |S_\alpha(f, P) - S_\alpha(f, P \cup Q)| + |S_\alpha(f, P \cup Q) - S_\alpha(f, Q)| \leq \epsilon$, completing the proof.

So we assume $P \subseteq Q$ and use the uniform continuity of f to find a δ such that $\omega(f, \delta) < \frac{\epsilon}{2}\mathrm{Var}(\alpha)$. To simplify matters we will assume that Q adds only one point to P and that this point lies between x_0 and x_1. That is we assume $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ and $Q = \{a = x_0 < x_0^* < x_1 < \cdots < x_n = b\}$. Now for $2 \leq j \leq n$, let $x_{j-1} \leq t_j, s_j \leq x_j$, $x_0 \leq t_1 \leq x_1$, $x_0 \leq s_0^* \leq x_0^*, x_0^* \leq s_1^* \leq x_1$. Note that

$$
\begin{aligned}
&\left| f(t_1)[\alpha(x_1) - \alpha(x_0)] - \left\{ f(s_0^*)[\alpha(x_0^*) - \alpha(x_0)] + f(s_1^*)[\alpha(x_1) - \alpha(x_0^*)] \right\} \right| \\
&\qquad = \left| f(t_1)[\alpha(x_0^*) - \alpha(x_0) + \alpha(x_1) - \alpha(x_0^*)] \right. \\
&\qquad\qquad \left. - \left\{ f(s_0^*)[\alpha(x_0^*) - \alpha(x_0)] + f(s_1^*)[\alpha(x_1) - \alpha(x_0^*)] \right\} \right| \\
&\qquad \leq |f(t_1) - f(s_0^*)|\,|\alpha(x_0^*) - \alpha(x_0)| \\
&\qquad\qquad + |f(t_1) - f(s_1^*)|\,|\alpha(x_1) - \alpha(x_0^*)| \\
&\qquad \leq \omega(f, \delta)\,[|\alpha(x_0^*) - \alpha(x_0)| + |\alpha(x_1) - \alpha(x_0^*)|]
\end{aligned}
$$

We therefore obtain

$$
\begin{aligned}
|S_\alpha(f, P) - S_\alpha(f, Q)| &\leq \omega(f, \delta)\,[|\alpha(x_0^*) - \alpha(x_0)| + |\alpha(x_1) - \alpha(x_0^*)|] \\
&\qquad + \sum_{j=2}^{n} |f(t_j) - f(s_j)|\,|\alpha(x_j) - \alpha(x_{j-1})| \\
&\leq \omega(f, \delta)\mathrm{Var}(\alpha) \\
&< \epsilon/2
\end{aligned}
$$

An inspection of the preceding argument shows that if Q had added more than a single point to P, then the same reasoning would prevail and yielded the same result. ■

It is important to emphasize that the value of the inequality obtained in this lemma is independent of the choices of transitory points t_j in $[x_{j-1}, x_j]$ that are used to define $S(\alpha, P)$. This amply justifies not incorporating them in the notation used to denote such a sum.

1.1.8. Theorem. *If α is a function of bounded variation on J and $f : J \to \mathbb{R}$ is a continuous function, then there is a unique number I with the property that for every $\epsilon > 0$ there is a $\delta > 0$ such that when $P \in \mathcal{P}_\delta$,*

$$|S_\alpha(f, P) - I| < \epsilon$$

The number I is called the Riemann[1]–Stieltjes[2] integral of f with respect to α, is denoted by

$$I = \int_a^b f d\alpha = \int f d\alpha$$

and satisfies

$$\left| \int f d\alpha \right| \leq \operatorname{Var}(\alpha) \max\{|f(t)| : t \in J\}$$

---

[1]Georg Friedrich Bernhard Riemann was born in 1826 in Breselenz, Germany. His early schooling was closely supervised by his father. When he entered the university at Göttingen in 1846, at his father's urging he began to study theology. Later, with his father's blessing, he switched to the Faculty of Philosophy so he could study mathematics. In 1847 he transferred to Berlin where he came under the influence of Dirichlet, which was permanent. He returned to Göttingen in 1849 where he completed his doctorate in 1851, working under the direction of Gauss. He took up a lecturer position there and in 1862 he married a friend of his sister. In the autumn of that same year he contracted tuberculosis. This began a period of ill health and he went between Göttingen and Italy, where he sought to recapture his health. He died in 1866 in Selasca, Italy on the shores of beautiful Lake Maggiore. Riemann is one of the giants of analysis and geometry. He made a series of discoveries and initiated theories. There is the Riemann zeta function, Riemann surfaces, the Riemann Mapping Theorem, and many other objects and concepts named after him besides this integral.

[2]Thomas Jan Stieltjes was born in 1856 in Zwolle, The Netherlands. He attended the university at Delft, spending most of his time in the library reading mathematics rather than attending lectures. This had the effect of causing him to fail his exams three years in a row and he left the university without a degree. (This phenomenon of talented people having trouble passing exams is not unique and examples exist in the author's personal experience.) The absence of a degree plagued the progress of his career, in spite of the recognition of his mathematical talent by some of the prominent mathematicians of the day. In 1885 he was awarded membership in the Royal Academy of Sciences in Amsterdam. He received his doctorate of science in 1886 for a thesis on asymptotic series. In the same year he secured a position at the University of Toulouse in France. He did fundamental work on continued fractions and is often called the father of that subject. He extended Riemann's integral to the present setting. He died in 1894 in Toulouse, where he is buried.

Proof. According to the preceding lemma, for every integer $n \geq 1$ there is a δ_n such that if $P, Q \in \mathcal{P}_{\delta_n}$, $|S_\alpha(f,P) - S_\alpha(f,Q)| < \frac{1}{n}$. We can choose the δ_n so that they are decreasing. Let K_n be the closure of the set of numbers $\{S_\alpha(f,P) : P \in \mathcal{P}_{\delta_n}\}$. If $|f(x)| \leq M$ for all x in J, then for any partition P, $|S_\alpha(f,P)| \leq M\text{Var}(\alpha)$. So each set K_n is bounded and hence compact. Since the numbers δ_n are decreasing, for all $n \geq 1$ $\mathcal{P}_{\delta_n} \supseteq \mathcal{P}_{\delta_{n+1}}$ and so $K_n \supseteq K_{n+1}$. Finally by the choice of the δ_n, $\text{diam}\, K_n \leq n^{-1} \to 0$. Therefore by Cantor's Theorem (Theorem 1.2.1 in the next section) $\bigcap_{n=1}^\infty K_n = \{I\}$ for a single number I. It is now routine to verify that I has the stated properties; its uniqueness is guaranteed by its construction. ■

A standard example comes, of course, when $\alpha(t) = t$ for all t in J and this is the Riemann integral. The proofs of the next two results are left to the reader as a way of fixing the ideas in his/her head. These results should not come as a surprise since their counterparts for the Riemann integral are well known.

1.1.9. Proposition. *Let α and β be functions of bounded variation on J, $f, g : J \to \mathbb{R}$ continuous functions, and $s, t \in \mathbb{R}$.*

(a) $\int_a^b (sf + tg) d\alpha = s \int_a^b f d\alpha + t \int_a^b g d\alpha$.

(b) *If $f(x) \geq 0$ for all x in J and α is increasing, then* $\int_a^b f d\alpha \geq 0$.

(c) $\int_a^b f d(s\alpha + t\beta) = s \int_a^b f d\alpha + t \int_a^b f d\alpha$.

1.1.10. Proposition. *If α is a function of bounded variation on J, f is a continuous function on J, and $a < c < b$, then $\int_a^b f d\alpha = \int_a^c f d\alpha + \int_c^b f d\alpha$.*

Here is an important result that enables us to compute some Riemann–Stieltjes integrals from what we know about the Riemann integral.

1.1.11. Theorem. *If α is a function on J that has a continuous derivative at every point of J, then*

$$\int_a^b f d\alpha = \int_a^b f(x)\alpha'(x) dx$$

for any continuous function f.

Proof. We already know from Example 1.1.2 that such a function is of bounded variation, so everything makes sense. Fix a continuous function f on J, let $\epsilon > 0$, and choose δ such that simultaneously $\left|\int f d\alpha - S_\alpha(f,P)\right| < \epsilon/2$ and $\left|\int f\alpha' dx - S(f\alpha', P)\right| < \epsilon/2$ whenever $P \in \mathcal{P}_\delta$. Momentarily fix $P = \{a = x_0 < \cdots < x_n = b\}$ in $\mathcal{P}_\delta$. For each subinterval $[x_{j-1}, x_j]$ defined by P, the Mean Value Theorem for derivatives implies there is a t_j in this subinterval with $\alpha(x_j) - \alpha(x_{j-1}) = \alpha'(t_j)(x_j - x_{j-1})$. Therefore

$$\left|\int_a^b f d\alpha - \int_a^b f(x)\alpha'(x)dx\right| \leq \left|\int_a^b f d\alpha - \sum_{j=1}^n f(t_j)[\alpha(x_j) - \alpha(x_{j-1})]\right|$$
$$+ \left|\sum_{j=1}^n f(t_j)\alpha'(t_j)(x_j - x_{j-1}) - \int_a^b f(x)\alpha'(x)dx\right| < \epsilon$$

Since ϵ was arbitrary, we have the desired equality. ■

Extensions of the preceding theorem will occupy our attention later in §4.1 and beyond. Be warned, however, that there are increasing functions whose derivatives exist at many points but where nothing like the preceding result is true (2.4.9).

We introduce some special increasing functions whose roll in the general theory has significance.

1.1.12. Example. Fix s in J. When $a \leq s < b$, define

$$\alpha_s(t) = \begin{cases} 0 & \text{for } t \leq s \\ 1 & \text{for } t > s \end{cases}$$

and

$$\alpha_b(t) = \begin{cases} 0 & \text{for } t < b \\ 1 & \text{for } t = b \end{cases}$$

Each of these functions is increasing. Let $\epsilon > 0$ and choose $\delta > 0$ such that $\left|\int f d\alpha_s - S_{\alpha_s}(f, P)\right| < \epsilon/2$ whenever $P \in \mathcal{P}_\delta$ and also such that $|f(u) - f(t)| < \epsilon/2$ when $|u-t| < \delta$. Assume $s < b$. (A separate argument is required when $s = b$ and this is left to the reader. See Exercise 7.) Choose a P in $\mathcal{P}_\delta$ that contains s and let s_0 be the point in P that immediately follows s. Using s_0 as the point in $[s, s_0]$ at which to evaluate f, a moment's reflection reveals that $S_{\alpha_s}(f, P) = f(s_0)[\alpha(s_0) - \alpha(s)] = f(s_0)$. Thus $\left|f(s) - \int f d\alpha_s\right| \leq |f(s) - f(s_0)| + \left|S_{\alpha_s}(f, P) - \int f d\alpha_s\right| < \epsilon$. Since ϵ was arbitrary we have that

$$\int f d\alpha_s = f(s)$$

for every continuous function f on J. Similarly, $\int f d\alpha_b = f(b)$ for all such f.

Using the preceding example we can calculate the integrals with respect to many functions that only differ from a continuously differentiable one by having jump discontinuities. Consider the following.

1.1.13. Example. Define $\alpha : [0,1] \to \mathbb{R}$ by $\alpha(t) = t^2$ for $t \leq \frac{1}{2}$ and $\alpha(t) = t^2 + 1$ for $t > \frac{1}{2}$. What is $\int_0^1 f(t)d\alpha(t)$ for f in $C[0,1]$? Note that $\alpha(t) = t^2 + \alpha_{\frac{1}{2}}$, where $\alpha_{\frac{1}{2}}$ is defined in the preceding example. So using Proposition 1.1.9(c) and Theorem 1.1.11 we get

$$\begin{aligned}\int_0^1 f(t)d\alpha(t) &= \int_0^1 f(t)d(t^2) + \int_0^1 f(t)d\alpha_{\frac{1}{2}}(t)\\ &= 2\int_0^1 tf(t)dt + f\left(\frac{1}{2}\right)\end{aligned}$$

If α is increasing, then $\alpha(t) \leq \alpha(c)$ whenever $t < c$ and so $\alpha(c-) \equiv \lim_{t\uparrow c} \alpha(t)$ exists and $\alpha(c-) \leq \alpha(c)$; $\alpha(c-)$ is called the *left limit* of α at c. Similarly we define the *right limit* as $\alpha(c+) = \lim_{t\downarrow c} \alpha(t)$ and we have that $\alpha(c+) \geq \alpha(c)$. Note that α is continuous at c precisely when $\alpha(c-) = \alpha(c+)$. It is worthwhile pointing out that when α is discontinuous at c, then the discontinuity is called a jump discontinuity and the difference $\alpha(c+) - \alpha(c-)$ is exactly the size of the jump in values between the left and right of c. In other words, this difference is the size of the vertical jump in the graph of the function α at the point c. Another observation is that if $J = [a,b]$, then the definition of $\alpha(a-)$ is, in a sense, meaningless. We will declare, however, that $\alpha(a-) = \alpha(a)$. Similarly $\alpha(b+) = \alpha(b)$.

1.1.14. Proposition. *If $\alpha : J \to \mathbb{R}$ is a function of bounded variation, then α has at most a countable number of discontinuities.*

Proof. By Proposition 1.1.6 it suffices to show that the conclusion holds for increasing functions, so assume that α is increasing. If α is discontinuous at the points $c_n, n \geq 1$, then the sum of the jumps, $\sum_n [\alpha(c_n+) - \alpha(c_n-)]$, is at most $\alpha(b) - \alpha(a)$. If it were the case that α had an uncountable number of discontinuities, then we could find an $\epsilon > 0$ and an infinite sequence of discontinuities $\{c_n\}$ such that $\alpha(c_n+) - \alpha(c_n-) \geq \epsilon$. (Why?) In light of what we just pointed out, this would imply that $\alpha(b) - \alpha(a) = \infty$, which is nonsense. ∎

A function α is called *left-continuous* at c if the left limit $\alpha(c-)$ exists and is equal to $\alpha(c)$. The definition of *right-continuous* is analogous. So if $J = [a,b]$, every function α on J is left-continuous at a and right-continuous at b by default. Note that if $a \leq s < b$, α_s is left-continuous at s; however, α_b is not left-continuous at b.

1.1.15. Corollary. *If $\alpha : J \to \mathbb{R}$ is increasing and we define $\beta : J \to \mathbb{R}$ by $\beta(t) = \alpha(t-)$, then β is an increasing function that is left-continuous everywhere, has the same discontinuities as α on $[a,b)$, and agrees with α except possibly at its discontinuities.*

Proof. It is clear that β is increasing and left-continuous at each point of J. If c is a point of discontinuity of α, then $\beta(c) = \alpha(c-) < \alpha(c+)$. If $c < b$, then since there are points of continuity for α that approach c from the right, we have that $\beta(c+) = \alpha(c+) > \beta(c)$ and β is discontinuous at c. If $c = b$, then $\beta(b) = \alpha(b-) = \beta(b+)$, so that β is continuous at b irrespective of whether α is continuous at b. (See the function α_b defined in Example 1.1.12.) ■

Note that the integral of a continuous function with respect to a constant function is 0; hence $\int f d(\alpha + c) = \int f d\alpha$ for all continuous functions f. There are other ways that we can produce two functions of bounded variation that yield the same integral. (Compare the function defined in Exercise 4 with the functions in Example 1.1.12.) We want to examine when $\int f d\alpha = \int f d\beta$ for two fixed functions of bounded variation and for every continuous function f on J.

For a function of bounded variation α on J and $a \le t \le b$, define

1.1.16 $$\widetilde{\alpha}(t) = \alpha(t-) - \alpha(a) + [\alpha(b) - \alpha(b-)]\alpha_b$$

where α_b is defined in Example 1.1.12. Call $\widetilde{\alpha}$ the *normalization* of α. The first thing to observe is that $\alpha(t) - \widetilde{\alpha}(t) = \alpha(a)$ except at the points where α is discontinuous. Also if α is increasing, so is its normalization. Furthermore, if $\alpha = \alpha_+ - \alpha_-$, then $\widetilde{\alpha} = \widetilde{\alpha}_+ - \widetilde{\alpha}_-$. Therefore $\widetilde{\alpha}$ is also a function of bounded variation. Finally note that if α is continuous at b, then $\widetilde{\alpha}(t) = \alpha(t-) - \alpha(a)$.

1.1.17. Proposition. *If α is a function of bounded variation on J and $\widetilde{\alpha}$ is its normalization, then $\int f d\alpha = \int f d\widetilde{\alpha}$ for every continuous function f on J.*

Proof. We split this into two cases.

Case 1. α is continuous at b.

Let D be the set of points in J where α is discontinuous – a countable set. Here $\widetilde{\alpha}(t) = \alpha(t-) - \alpha(a)$; fix $\epsilon > 0$ and f in $C([a, b])$. Let $\delta > 0$ such that $\left|S_\alpha(f, P) - \int f d\alpha\right| < \epsilon/2$ and $\left|S_{\widetilde{\alpha}}(f, P) - \int f d\widetilde{\alpha}\right| < \epsilon/2$ whenever $P \in \mathcal{P}_\delta$. Since D is a countable set in $[a, b]$, we can choose $P = \{a = x_0 < \cdots < x_n = b\}$ in $\mathcal{P}_\delta$ such that $x_j \notin D$ for $0 < j \le n$. We have by definition

$$S_\alpha(f, P) = f(a)[\alpha(x_1) - \alpha(a)] + \sum_{j=2}^{n} f(x_j)[\alpha(x_j) - \alpha(x_{j-1})]$$

and

$$S_{\widetilde{\alpha}}(f, P) = f(a)[\widetilde{\alpha}(x_1) - \widetilde{\alpha}(a)] + \sum_{j=2}^{n} f(x_j)[\widetilde{\alpha}(x_j) - \widetilde{\alpha}(x_{j-1})]$$

For $0 < j \leq n$, $x_j \notin D$ and so $\alpha(x_j) = \widetilde{\alpha}(x_j) + \alpha(a)$. Since $\widetilde{\alpha}(a) = 0$, we also have $\alpha(a) = \widetilde{\alpha}(a) + \alpha(a)$; thus $\alpha(x_j) - \alpha(x_{j-1}) = \widetilde{\alpha}(x_j) - \widetilde{\alpha}(x_{j-1})$ for $1 \leq j \leq n$. That is, $S_\alpha(f,P) = S_{\widetilde{\alpha}}(f,P)$. Therefore $\left|\int f d\alpha - \int f d\widetilde{\alpha}\right| < \epsilon$. Since ϵ was arbitrary, this proves Case 1.

Case 2. α is discontinuous at b.

Here $\widetilde{\alpha}(t) = \alpha(t-) + [\alpha(b) - \alpha(b-)]\alpha_b$. Consider the function $\beta = \alpha - [\alpha(b) - \alpha(b-)]\alpha_b$. It follows that β is continuous at b since $\beta(b-) = \alpha(b-) = \beta(b)$. By Case 1, $\int f d\beta = \int f d\widetilde{\beta}$. Moreover since $\alpha_b(t-) = 0$ for all t in J, including $t = b$, it follows that $\widetilde{\beta}(t) = \alpha(t-) = \widetilde{\alpha}(t) - [\alpha(b) - \alpha(b-)]\alpha_b(t)$. Now $\int f d\alpha - [\alpha(b) - \alpha(b-)]f(b) = \int f d\beta = \int f d\widetilde{\beta} = \int f d\widetilde{\alpha} - [\alpha(b) - \alpha(b-)]f(b)$. After canceling we get that $\int f d\alpha = \int f d\widetilde{\alpha}$. ■

Later in this book we will prove the converse of the above result. To be precise, Proposition 4.5.3 shows that if α and β are two functions of bounded variation on the interval J, then $\int f d\alpha = \int f d\beta$ for every continuous function f on J if and only if $\widetilde{\alpha} = \widetilde{\beta}$.

Exercises. We continue to assume that $J = [a,b]$ unless the interval is otherwise specified.

(1) (This first exercise has rather little to do with this section and could have been assigned in your course on advanced calculus, but maybe it wasn't.) For $n \geq 0$ define $f_n : [0,1] \to \mathbb{R}$ by $f_n(t) = t^n \sin\left(\frac{1}{t}\right)$ when $t \neq 0$ and $f_n(0) = 0$. (a) Show that f_0 is not continuous at 0. (b) Show that f_1 is continuous on $[0,1]$ but is not differentiable at 0. (c) Show that f_2 is differentiable at every point of $[0,1]$, but f_2' is not continuous at 0 and so f_2 is not twice differentiable at 0. (d) For $k \geq 1$ show that f_{2k} is $k-1$ differentiable on $[0,1]$, but $f_{2k}^{(k-1)}$ is not continuous at 0 and hence f_{2k} is not k-times differentiable on the unit interval. (e) Show that if $f(t) = e^{-t^{-2}} \sin\left(t^{-1}\right)$ for $t \neq 0$ and zero at $t = 0$, then f is infinitely differentiable on the unit interval.

(2) Show that a function of bounded variation is a bounded function; that is, there is a constant M with $|\alpha(t)| \leq M$ for all t in J.

(3) Show that the function f_1 in Exercise 1 is not of bounded variation even though it is continuous.

(4) Define α on the unit interval $[0,1]$ by $\alpha(\frac{1}{2}) = 1$ and $\alpha(t) = 0$ when $t \neq \frac{1}{2}$. Observe that α is neither left nor right-continuous at $\frac{1}{2}$. Show that α is of bounded variation and find increasing functions $\alpha_\pm$ such that $\alpha = \alpha_+ - \alpha_-$. Compute $\int f d\alpha$ for an arbitrary continuous function f on the unit interval. Are you surprised?

(5) If α is an increasing function on J such that $\int f d\alpha = 0$ for every continuous function f on J, show that α is constant. Contrast this with Exercise 4. (Hint: First show that you can assume that $\alpha(a) = 0$, then show that α must be continuous at each point of J. Use Example 1.1.12. Now show that α is identically 0.)

(6) Give the details of the proof of Proposition 1.1.9

(7) If α_b is defined as in Example 1.1.12, show that $\int f d\alpha_b = f(b)$ for every continuous function f on J.

(8) Suppose a left-continuous increasing function $\alpha : J \to \mathbb{R}$ has a discontinuity at t_0 and $a_0 = \alpha(t_0+) - \alpha(t_0-)$. Let α_{t_0} be the increasing function defined as in Example 1.1.12: $\alpha_{t_0}(t) = 0$ for $t \le t_0$ and $\alpha_{t_0}(t) = 1$ for $t > t_0$. (Once again if t_0 is the right hand end point of J, a separate argument is required.) (a) Show that $\alpha - a_0\alpha_{t_0}$ is an increasing function that is continuous at t_0. (b) Show that any left-continuous increasing function α on J can be written as $\alpha = \beta + \gamma$, where both β and γ are increasing, β is continuous, and γ has the property that if γ is continuous on the open subinterval (c, d), then γ is constant there.

(9) If γ is an increasing function with the property of γ in Exercise 8, calculate $\int f d\gamma$ for any continuous function f on J.

(10) If A is any countable subset of J, show that there is an increasing function α on J such that A is precisely the set of discontinuities of α. (So, in particular, there is an increasing function on J with discontinuities at all the rational numbers in J.)

(11) Let $\{r_n\}$ denote the set of all rational numbers in the interval J and define $\alpha : J \to \mathbb{R}$ by $\alpha(t) = \sum \frac{1}{2^n}$ where the sum is taken over all n such that $r_n < t$. (a) Show that α is a strictly increasing function that is left-continuous and satisfies $\alpha(a) = 0$ and $\alpha(b) = 1$. (b) Show that α is continuous at each irrational number and discontinuous at all the rational numbers in J.

(12) If α is an increasing function, discuss the possibility of defining $\int f d\alpha$ when f has a discontinuity.

(13) Suppose $\alpha : [0, \infty) \to \mathbb{R}$ is an increasing function. Is it possible to define $\int_0^\infty f d\alpha$ for a continuous function $f : [0, \infty) \to \mathbb{R}$?

1.2. Metric spaces redux

As I stated in the Preface, the reader is expected to know the basics of metric spaces. My experience is that the students for whom this book is intended usually do, though there is an occasional student who seems to lack some of it. Such students, however, usually know the concepts as they apply to Euclidean spaces. So I am not going to review this material. I will include, however, two basic results we will use very often. The first is Cantor's Theorem, already used in the preceding section; the second is a list of equivalent formulations of compactness. The different statements of compactness will be used so often that I want to be sure that all readers are on the same page. Cantor's Theorem, besides already being used, will be used in the proof of the second result and can also be used to prove that a continuous function on a compact metric space is uniformly continuous (Exercise 2). Besides, its proof is short. Then in the remainder of this section we will see some results about continuous functions on a metric space that aren't encountered in all courses on the subject, and we'll explore a few other ideas that might not be so familiar to the reader.

The Baire Category Theorem is a topic from metric spaces that may not be familiar to all readers. This result will not be used in this section but will be seen in a significant way later in this book. In case a reader is not familiar with this, (s)he can see it together with a proof in §A.1.

Throughout this section (X, d) is a metric space. If $x \in X$ and $r > 0$, we let $B(x; r) = \{y \in X : d(x, y) < r\}$. For any subset A of X, $\operatorname{cl} A$ denotes the closure of A and $\operatorname{int} A$ denotes its interior. Also recall that a metric space is complete if every Cauchy sequence converges.

1.2.1. Theorem (Cantor's[3] Theorem). *A metric space (X, d) is complete if and only if for every sequence of closed subsets $\{F_n\}$ such that $F_{n+1} \subseteq F_n$ for each $n \geq 1$ and* $\operatorname{diam} F_n \equiv \sup\{d(x, y) : x, y \in F_n\} \to 0$*, there is a point x_0 in X such that*

$$\bigcap_{n=1}^{\infty} F_n = \{x_0\}$$

Proof. Assume X is complete and $\{F_n\}$ is as stated in the theorem. If $x_n \in F_n$ and $m \geq n$, then $d(x_n, x_m) \leq \operatorname{diam} F_n$; hence $\{x_n\}$ is a Cauchy sequence. Since X is complete, there is an x_0 in X such that $x_n \to x_0$. It follows that $x_0 \in \bigcap_{n=1}^{\infty} F_n$. If there is another point x in this intersection, then $d(x, x_0) \leq \operatorname{diam} F_n$ for every $n \geq 1$; hence $x = x_0$.

[3]Georg Cantor was the child of an international family. His father was born in Denmark and his mother was Russian; he himself was born in 1845 in St. Petersburg where his father was a successful merchant and stock broker. He is recognized as the father of set theory, having invented cardinal and ordinal numbers and proved that the irrational numbers are uncountable. He received his doctorate from the University of Berlin in 1867 and spent most of his career at the University of Halle. His work was a watershed event in mathematics but was condemned by many prominent mathematicians at the time. The work was simply too radical with counterintuitive results such as $\mathbb{R}$ and $\mathbb{R}^d$ having the same number of points. He began to suffer from depression around 1884. This progressed and plagued him the rest of his life. He died in a sanatorium in Halle in 1918.

Now assume that X satisfies the stated condition and let's prove that X is complete. If $\{x_n\}$ is a Cauchy sequence, put $F_n = \text{cl}\{x_m : m \geq n\}$. So F_n is a closed set, $F_n \subseteq F_{n+1}$, and $\text{diam}\, F_n \leq \sup\{d(x_n, x_m) : m \geq n\}$. Because we have a Cauchy sequence it follows that $\text{diam}\, F_n \to 0$. Thus $\bigcap_{n=1}^{\infty} F_n = \{x_0\}$ for some point x_0. But then $d(x_0, x_n) \leq \text{diam}\, F_n$, so $x_n \to x_0$. ■

The following is the basic result on compactness, whose definition we take as the property that every open cover has a finite subcover.

1.2.2. Theorem. *The following statements are equivalent for a closed subset K of a metric space (X, d).*

(a) *K is compact.*

(b) *If $\mathcal{F}$ is a collection of closed subsets of K having the property that every finite subcollection has non-empty intersection, then $\bigcap_{F \in \mathcal{F}} F \neq \emptyset$.*

(c) *Every infinite set in K has a limit point.*

(d) *Every sequence in K has a convergent subsequence.*

(e) *(K, d) is a complete metric space that is totally bounded. That is, if $r > 0$, then there are points $x_1, \ldots, x_n$ such that $X = \bigcup_{k=1}^{n} B(x_k; r)$.*

Proof. (a) *implies* (b). Let $\mathcal{F}$ be a collection of closed subsets of K having the property that every finite subcollection has non-empty intersection. Suppose $\bigcap\{F : F \in \mathcal{F}\} = \emptyset$. If $\mathcal{G} = \{X \backslash F : F \in \mathcal{F}\}$, then it follows that $\mathcal{G}$ is an open cover of X and therefore of K. By (a), there are $F_1, \ldots, F_n$ in $\mathcal{F}$ such that $K \subseteq \bigcup_{j=1}^{n} (X \backslash F_j) = X \backslash \left[\bigcap_{j=1}^{n} F_j\right]$. But since each F_j is a subset of K this implies $\bigcap_{j=1}^{n} F_j = \emptyset$, contradicting the assumption about $\mathcal{F}$.

(b) *implies* (c). Assume that (c) is false. So there is an infinite subset S of K with no limit point; it follows that there is an infinite sequence $\{x_n\}$ of distinct points in S with no limit point. Thus for each $n \geq 1$, $F_n = \{x_k : k \geq n\}$ is a closed set (it contains all its limit points) and $\bigcap_{n=1}^{\infty} F_n = \emptyset$. But each finite subcollection of $\{F_1, F_2, \ldots\}$ has non-empty intersection, contradicting (b).

(c) *implies* (d). Assume $\{x_n\}$ is a sequence of distinct points. By (c), $\{x_n\}$ has a limit point. We are tempted here to say that there is a sequence in the set $\{x_n\}$ that converges to x, but we have to manufacture an actual subsequence of the original sequence. This takes a little bit of effort that is left to the interested reader.

(d) *implies* (e). If $\{x_n\}$ is a Cauchy sequence in K, then (d) implies it is a convergent sequence; by Exercise 1 the original sequence converges.

Thus (K, d) is complete. Now fix an $r > 0$. Let $x_1 \in K$; if $K \subseteq B(x_1; r)$, we are done. If not, then there is a point x_2 in $K \backslash B(x_1; r)$. Once again, if $K \subseteq B(x_1; r) \cup B(x_2; r)$, we are done; otherwise pick an x_3 in $K \backslash [B(x_1; r) \cup B(x_2; r)]$. Continue. If this process does not stop after a finite number of steps, we produce an infinite sequence $\{x_n\}$ in K with $d(x_n, x_m) \geq r$ whenever $n \neq m$. But this implies that this sequence can have no convergent subsequence, contradicting (d).

(e) *implies* (d). Fix an infinite sequence $\{x_n\}$ in K and let $\{\epsilon_n\}$ be a decreasing sequence of positive numbers such that $\epsilon_n \to 0$. By (e) there is a covering of K by a finite number of balls of radius ϵ_1. Thus there is a ball $B(y_1; \epsilon_1)$ that contains an infinite number of points from $\{x_n\}$; let $\mathbb{N}_1 = \{n \in \mathbb{N} : d(x_n, y_1) < \epsilon_1\}$. Now consider the sequence $\{x_n : n \in \mathbb{N}_1\}$ and balls of radius ϵ_2. As we just did, there is a point y_2 in K such that $\mathbb{N}_2 = \{n \in \mathbb{N}_1 : d(y_2, x_n) < \epsilon_2\}$ is an infinite set. Using induction we can show that for each $k \geq 1$ we get a point y_k in K and an infinite set of positive integers $\mathbb{N}_k$ such that $\mathbb{N}_{k+1} \subseteq \mathbb{N}_k$ and $\{x_n : n \in \mathbb{N}_k\} \subseteq B(y_k; \epsilon_k)$. If $F_k = \text{cl}\,\{x_n : n \in \mathbb{N}_k\}$, then $F_{k+1} \subseteq F_k$ and $\operatorname{diam} F_k \leq 2\epsilon_k$. Since K is complete, Cantor's Theorem implies that $\bigcap_{k=1}^{\infty} F_k = \{x\}$ for some point x in X. Now pick integers n_k in $\mathbb{N}_k$ such that $n_k < n_{k+1}$. It follows that $x_{n_k} \to x$.

(e) *implies* (a). We first prove the following.

1.2.3. Claim. If X satisfies (d) and $\mathcal{G}$ be an open cover of X, then there is an $r > 0$ such that for each x in X there is a G in $\mathcal{G}$ such that $B(x; r) \subseteq G$.

Let $\mathcal{G}$ be an open cover of X and suppose the claim is false; so for every $n \geq 1$ there is an x_n in X such that $B(x_n; n^{-1})$ is not contained in any set G in $\mathcal{G}$. By (d) there is an x in X and a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to x$. Since $\mathcal{G}$ is a cover, there is a G in $\mathcal{G}$ such that $x \in G$; choose a positive ϵ such that $B(x; \epsilon) \subseteq G$. Let $n_k > 2\epsilon^{-1}$ such that $x_{n_k} \in B(x; \epsilon/2)$. If $y \in B(x_{n_k}; n_k^{-1})$, then $d(x, y) \leq d(x, x_{n_k}) + d(x_{n_k}, y) < \epsilon/2 + n_k^{-1} < \epsilon$, so that $y \in B(x; \epsilon) \subseteq G$. That is $B(x_{n_k}; n_k^{-1}) \subseteq G$, contradicting the restriction imposed on x_{n_k}. This establishes the claim.

From here it is easy to complete the proof. We know that (e) implies (d), so for an open cover $\mathcal{G}$ let $r > 0$ be the number guaranteed by (1.2.3). Now let $x_1, \ldots, x_n \in X$ such that $X = \bigcup_{k=1}^{n} B(x_k; r)$ and for $1 \leq k \leq n$ let $G_k \in \mathcal{G}$ such that $B(x_k; r) \subseteq G_k$. $\{G_1, \ldots, G_n\}$ is the sought after finite subcover. ■

We might note that the statement in Claim 1.2.3 is sometimes called the Lebesgue Covering Lemma. It follows from part (d) of the last theorem that every compact metric space is separable. See Exercise 6.

We will in the course of this book look at functions that are real-valued as well as complex-valued. When we see a result that applies to both, we use $\mathbb{F}$ to denote that we are dealing either with $\mathbb{R}$ or $\mathbb{C}$. In the later chapters we will focus on complex-valued objects.

1.2.4. Definition. $C(X)$ denotes the space of all continuous functions $f : X \to \mathbb{F}$ and $C_b(X)$ denotes the space of all bounded continuous functions in $C(X)$. That is $f \in C_b(X)$ if and only if $f \in C(X)$ and there is a constant M such that $|f(x)| \leq M$ for all x in X. When we want to emphasize the underlying field of scalars, we will write $C(X, \mathbb{R}), C(X, \mathbb{C}), C_b(X, \mathbb{R})$, or $C_b(X, \mathbb{C})$. This, however, will not be frequent.

Of course when X is compact, $C(X) = C_b(X)$. $C(X)$ has a natural algebraic structure where the algebraic operations are defined pointwise as we did for $BV[a, b]$. If $f, g \in C(X)$, $(f+g) : X \to \mathbb{F}$ is defined by $(f+g)(x) = f(x) + g(x)$ and we define $(fg) : X \to \mathbb{F}$ by $(fg)(x) = f(x)g(x)$. Similarly we can define af when $a \in \mathbb{F}$ and $f \in C(X)$. In these cases $f + g, fg$, and af belong to $C(X)$. With such definitions, $C(X)$ is an *algebra*. That is, it is a vector space over $\mathbb{F}$ that has a multiplicative structure and all the usual distributive laws are valid. In the same way $C_b(X)$ is an algebra. As vector spaces $C(X)$ and $C_b(X)$ are finite dimensional if and only if X is a finite metric space (Exercise 3). This will become apparent as soon as we see how to manufacture examples of continuous functions.

Consider $C(X) = C(X, \mathbb{R})$ as a vector space over $\mathbb{R}$ and impose an order structure on $C(X)$ pointwise as follows: if $f, g \in C(X)$, say $f \leq g$ if $f(x) \leq g(x)$ for all x in X. The notation $g \geq f$ means that $f \leq g$. Define the *positive cone* of $C(X)$ to be

$$C(X)_+ = \{f \in C(X) : f \geq 0\}$$

Similarly put this order structure on $C_b(X)$ and define the positive cone of $C_b(X)$ as $C_b(X)_+ = C(X)_+ \cap C_b(X)$. It is also possible to put an order structure on $C(X, \mathbb{C})$, though it is not as neat. Namely for f, g in $C(X, \mathbb{C})$ say that $f \leq g$ if both f and g are real-valued and $f(x) \leq g(x)$ for all x in X.

The proofs of the next two propositions are routine.

1.2.5. Proposition. *Let $f, g, h, k \in C(X)$.*

(a) *$f \leq g$ if and only if $g - f \geq 0$.*

(b) *If $f \leq g$ and $g \leq h$, then $f \leq h$.*

(c) *If $f \leq g$, then $f + h \leq g + h$.*

(d) *If $f \leq g$ and $h \leq k$, then $f + h \leq g + k$.*

(e) *If* $f \leq g$ *and* $a \in [0, \infty)$*, then* $af \leq ag$*; if* $a \in (-\infty, 0]$*,* $af \geq ag$*.*
Similar results hold for $C_b(X)$*.*

Recall the definitions of $f \vee g$ and $f \wedge g$ from §1.1.

1.2.6. Proposition. *If* $f, g \in C(X)$*, then the following hold.*

(a) $f \vee g$ *belongs to* $C(X)$*.*

(b) $f \wedge g$ *belongs to* $C(X)$*.*

(c) $|f|$*, defined by* $|f|(x) = |f(x)|$*, belongs to* $C(X)$*.*

(d) *If* $f_+ = f \vee 0$ *and* $f_- = -f \wedge 0$*, then* $f = f_+ - f_-$*,* $f_+ f_- = 0$*, and* $|f| = f_+ + f_-$*.*
Similar results hold for $C_b(X)$*.*

Part (d) of this last proposition can be interpreted as saying that the positive cone $C(X)_+$ spans $C(X)$.

For any subset A of X, $\operatorname{dist}(x, A) = \inf\{d(x, a) : a \in A\}$, the distance from x to A. Note that $\operatorname{dist}(x, A) = \operatorname{dist}(x, \operatorname{cl} A)$.

1.2.7. Proposition. *If* F *is a closed subset of* X*,* $|\operatorname{dist}(x, F) - \operatorname{dist}(y, F)| \leq d(x, y)$*. Consequently the function* $f : X \to \mathbb{R}$ *defined by* $f(x) = \operatorname{dist}(x, F)$ *is uniformly continuous.*

Proof. If $z \in F$, then $\operatorname{dist}(x, F) \leq d(x, z) \leq d(x, y) + d(y, z)$. Taking the infimum over all z in F we get $\operatorname{dist}(x, F) \leq d(x, y) + \operatorname{dist}(y, F)$ or $\operatorname{dist}(x, F) - \operatorname{dist}(y, F) \leq d(x, y)$. Interchanging the roles of x and y we get $\operatorname{dist}(y, F) - \operatorname{dist}(x, F) \leq d(x, y)$, whence the result. ■

Urysohn's Lemma is usually proved in a first course on point set topology, but students don't seem to realize how much easier it is to prove in the metric space setting than in that of a normal topological space.

1.2.8. Theorem (Urysohn's[4] Lemma)**.** *If* A *and* B *are two disjoint closed subsets of* X*, then there is a continuous function* $f : X \to \mathbb{R}$ *having the following properties:*

(a) $0 \leq f(x) \leq 1$ *for all* x *in* X*;*

[4]Pavel Samuilovich Urysohn was born in 1898 in Odessa, Ukraine. He was awarded his habilitation in June 1921 from the University of Moscow, where he remained as an instructor. He began his work in analysis but switched to topology where he made several important contributions, especially in developing a theory of dimension. His work attracted attention from the mathematicians of the day, and in 1924 he set out for a tour of the major universities in Germany, Holland, and France, meeting with Hausdorff, Hilbert, and others. That same year, while swimming off the coast of Brittany, France, he drowned. He is buried in Batz-sur-Mer in Brittany. In just three years he left his mark on mathematics.

(b) $f(x) = 0$ *for all* x *in* A;

(c) $f(x) = 1$ *for all* x *in* B.

Proof. Define $f : X \to \mathbb{R}$ by

$$f(x) = \frac{\text{dist}\,(x, A)}{\text{dist}\,(x, A) + \text{dist}\,(x, B)}$$

which is well defined since the denominator never vanishes. It is easy to check that f has the desired properties. ■

1.2.9. Corollary. *If F is a closed subset of X and G is an open set containing F, then there is a continuous function $f : X \to \mathbb{R}$ such that $0 \leq f(x) \leq 1$ for all x in X, $f(x) = 1$ when $x \in F$, and $f(x) = 0$ when $x \notin G$.*

Proof. In Urysohn's Lemma, take A to be the complement of G and $B = F$. ■

This corollary can be thought of as the "local" version of Urysohn's Lemma, though this comment will only be seen to make sense once we have used it. Indeed, later in this book we will need to construct sequences of continuous functions that have various properties, and Urysohn's Lemma and the preceding corollary are the keys. Here is one such example but first a word about notation.

In this book the difference of two sets will be denoted by $A \backslash B$ rather than $A - B$. The reason for this is that often we will deal with vector spaces and if A and B are subsets of a vector space $\mathcal{X}$, the notation $A - B$ will mean $\{a - b : a \in A, b \in B\}$, the set of all differences of vectors from the two sets. So we use $A \backslash B$ to avoid confusion.

1.2.10. Proposition. *If G is an open subset and F is a closed subset of X such that $F \subseteq G$, then the following hold.*

(a) *There is a sequence of continuous functions $\{f_n\}$ such that for all $n \geq 1$, $0 \leq f_n(x) \leq 1$, $f_n(x) = 1$ when $x \in F$, $f_n(x) = 0$ when $x \notin G$, $f_n \leq f_{n+1}$, and $f_n(x) \nearrow 1$ when $x \in G$.*

(b) *There is a sequence of continuous functions $\{g_n\}$ such that $0 \leq g_n(x) \leq 1$, $g_n(x) = 1$ when $x \in F$, $g_n(x) = 0$ when $x \notin G$, $g_n \geq g_{n+1}$ for all $n \geq 1$, and $g_n(x) \searrow 0$ when $x \notin F$.*

Proof. (a) Let $F_n = F \cup \{x : \text{dist}\,(x, X \backslash G) \geq \frac{1}{n}\}$. So each F_n is closed, $F \subseteq F_n \subseteq F_{n+1}$, and $\bigcup_{n=1}^{\infty} F_n = G$. Use Urysohn's Lemma to find a continuous function $h_n : X \to [0, 1]$ such that $h_n(x) = 1$ when $x \in F_n$ and $h_n(x) = 0$ when $x \notin G$. Put $f_n = \max\{h_1, \ldots, h_n\}$; so f_n is continuous by Proposition 1.2.6. Clearly this sequence of functions is increasing, $f_n(x) = 1$ when $x \in F_n$, which includes F, and $f_n(x) = 0$ when $x \notin G$. If $x \in G$, then there is an N with x in F_N; since $F_N \subseteq F_n$ when $n \geq N$, we have that $f_n(x) = 1$ for all $n \geq N$.

(b) Put $H = X \backslash F$ and $K = X \backslash G$. So K is closed, H is open, and $K \subseteq H$. By part (a) there is a sequence $\{f_n\}$ in $C(X)$ with $0 \leq f_n(x) \leq 1$ for all x in X, $f_n(x) = 1$ when $x \in K$, $f_n(x) = 0$ when $x \notin H$, $f_n \leq f_{n+1}$, and $f_n(x) \nearrow 1$ when $x \in H$. Let $g_n = 1 - f_n$ and verify that $\{g_n\}$ has the desired properties. ■

It is worthwhile for what we will do later in this book to take the following point of view in the preceding proposition. For any set E in X, define the *characteristic function* of E, $\chi_E : X \to \mathbb{F}$, by

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases}$$

The functions f_n in Proposition 1.2.10(a) can be said to satisfy $\chi_F \leq f_n \leq f_{n+1} \leq \chi_G$ and $f_n(x) \to \chi_G(x)$ for every x, even though the functions χ_F and χ_G may not be continuous. (See Exercise 10.) In fact, the conditions just given on $\{f_n\}$ capture all the properties given in Proposition 1.2.10. Similarly the functions g_n in part (b) of the preceding proposition satisfy $\chi_G \geq g_n \geq g_{n+1} \geq \chi_F$ and $g_n(x) \to \chi_F(x)$ for every x.

1.2.11. Theorem (Partition of unity)**.** *If $\{G_1, \ldots, G_n\}$ is an open cover of X, then there are continuous functions $\phi_1, \ldots, \phi_n$ on X with the following properties:*

(a) $0 \leq \phi_k(x) \leq 1$ *for all x and $1 \leq k \leq n$;*

(b) $\phi_k(x) = 0$ *when $x \notin G_k$ and $1 \leq k \leq n$;*

(c) $\sum_{k=1}^{n} \phi_k(x) = 1$ *for all x in X.*

Proof. We begin by letting $g_k(x) = \text{dist}\,(x, X \backslash G_k)$. Note that when $x \in G_k$, $g_k(x) > 0$. Since the sets $\{G_k\}$ cover X, $g(x) = \sum_{k=1}^{n} g_k(x) > 0$ for all x in X. Put $\phi_k = g_k/g$. Clearly (a) and (c) hold. Since $g_k(x) = 0$ for $x \notin G_k$, the same holds for $\phi_k(x)$ and so (b) is also satisfied. ■

For the open cover $\{G_1, \ldots, G_n\}$ and the functions $\phi_1, \ldots, \phi_n$ as in the preceding theorem, we say that these functions are a *partition of unity subordinate to the cover*. Using characteristic functions we can restate conditions (a) and (b) in this theorem as $0 \leq \phi_k \leq \chi_{G_k}$.

1.2.12. Corollary. *If K is a closed subset of X and $\{G_1, \ldots, G_n\}$ is an open cover of K, then there are continuous functions $\phi_1, \ldots, \phi_n$ on X with the following properties:*

(a) $0 \leq \phi_k \leq \chi_{G_k}$ *for $1 \leq k \leq n$;*

(b) $\sum_{k=1}^{n} \phi_k(x) = 1$ *for all* x *in* K*;*

(c) $\sum_{k=1}^{n} \phi_k(x) \leq 1$ *for all* x *in* X*.*

Proof. Note that if we put $G_{n+1} = X \backslash K$, then $\{G_1, \ldots, G_{n+1}\}$ is an open cover of X. Let $\{\phi_1, \ldots, \phi_{n+1}\}$ be a partition of unity subordinate to this cover. The reader can check that the functions $\phi_1, \ldots, \phi_n$ satisfy (a) and (b) since $\phi_{n+1}(x) = 0$ for x in K. For any x in X, $\sum_{k=1}^{n} \phi_k(x) = 1 - \phi_{n+1}(x) \leq 1$, giving (c). ■

Partitions of unity are a way of putting together local results to get a global result. This comment will not make complete sense to the reader but will become evident later. The typical situation will be that we know how to do something in an open neighborhood like G_k, such as manufacturing a function f_k with certain properties. Using the partition of unity $\{\phi_k\}$ and writing $f = \sum_k f_k \phi_k$, we will obtain a function f defined on the entire space that reflects the local properties enjoyed by the functions f_k. Again, this is meant as a rough idea of how partitions of unity will be used, and the reader can be forgiven if the picture remains murky. We will use partitions of unity frequently in this book and the fog will then lift.

This section concludes with a result from topology, the Tietze Extension Theorem. This is valid for normal spaces, but only is stated here for metric spaces. No proof is given here since the author knows no proof for metric spaces that is simpler than the one for normal spaces. (Anyone who has such a proof is invited to communicate it to the author.)

1.2.13. Theorem (Tietze[5] Extension Theorem)**.** *If* X *is a metric space,* Y *is a closed subset of* X*, and* $f : Y \to \mathbb{F}$ *is a continuous function with* $|f(y)| \leq M$ *for all* y *in* Y*, then there is a continuous function* ϕ *on* X *such that* $\phi(y) = f(y)$ *for* y *in* Y *and* $|\phi(x)| \leq M$ *for all* x *in* X*.*

Exercises. For these exercises (X, d) is always a metric space.

(1) Show that if $\{x_n\}$ is a Cauchy sequence in X and there is a subsequence $\{x_{n_k}\}$ that converges to a point x in X, then $x_n \to x$.

(2) Use (1.2.3) to show that if X is compact, Y is another metric space, and $f : X \to Y$ is a continuous function, then f is uniformly continuous.

[5] Heinrich Franz Friedrich Tietze was born in 1880 in Schleinz, Austria. In 1898 he entered the Technische Hochschule in Vienna. He continued his studies in Vienna and received his doctorate in 1904 and his habilitation in 1908, with a thesis in topology. His academic career was interrupted by service in the Austrian army in World War I; just before this he obtained the present theorem. After the war he had a position first at Erlangen and then at Munich, where he remained until his retirement. In addition to this theorem he made other contributions to topology and did significant work in combinatorial group theory, a field in which he was one of the pioneers. He had 12 PhD students, all at Munich, where he died in 1964.

(3) Show that $C_b(X)$ is a finite-dimensional vector space over $\mathbb{F}$ if and only if X is a finite metric space.

(4) If G is an open set and K is a compact set with $K \subseteq G$, then there is a $\delta > 0$ such that $\{x : \text{dist}\,(x, K) < \delta\} \subseteq G$. Find an example of an open set G in a metric space X and a closed, non-compact subset F of G such that there is no $\delta > 0$ with $\{x : \text{dist}\,(x, F) < \delta\} \subseteq G$.

(5) Recall that a metric space is *separable* if there is a countable dense subset. The following exercise will be used often in this book and so it is included here to be sure the reader is familiar with it. If (X, d) is separable, $E \subseteq X$, $r > 0$, and the balls $\{B(x; r) : x \in E\}$ are pairwise disjoint, then E is countable.

(6) Use Theorem 1.2.2(d) to show that a compact metric space is separable.

(7) Let I be any non-empty set and for each i in I, let X_i be a copy of $\mathbb{R}$ with the metric $d_i(x, y) = |x - y|$. Let X be the disjoint union of the sets X_i. That's a verbal description that can be used in any circumstance, but if you want precision you can say $X = \mathbb{R} \times I$, the Cartesian product where I has the discrete topology. Define a metric on X by letting d agree with d_i on X_i; when $x \in X_i$, $y \in X_j$, where $i \neq j$, then $d(x, y) = 1$. (a) Show that d is indeed a metric on X. (b) Show that $\{X_i : i \in I\}$ is the collection of components of X and each of these components is an open subset of X. (c) Show that (X, d) is separable if and only if I is a countable set.

(8) Show that a subset of $\mathbb{R}$ is connected if and only if it is an interval.

(9) For two subsets A and B of X, define the distance from A to B by $\text{dist}\,(A, B) = \inf\{d(a, b) : a \in A, b \in B\}$. (a) Show that $\text{dist}\,(A, B) = \text{dist}\,(B, A) = \text{dist}\,(\text{cl}\, A, \text{cl}\, B)$. (b) If A and B are two disjoint closed subsets of X such that B is compact, then $\text{dist}\,(A, B) > 0$. (c) Give an example of two disjoint closed subsets A and B of the plane $\mathbb{R}^2$ such that $\text{dist}\,(A, B) = 0$.

(10) If $E \subseteq X$, show that the characteristic function χ_E is a continuous function on X if and only if E is simultaneously an open and a closed set.

1.3. Normed spaces

The next concept plays a central role in analysis.

1.3.1. Definition. If $\mathcal{X}$ is a vector space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, a *norm* on $\mathcal{X}$ is a function $\|\cdot\| : \mathcal{X} \to [0, \infty)$ having the following properties for a scalar a and vectors x, y in $\mathcal{X}$:

(a) $\|x\| = 0$ if and only if $x = 0$;

(b) $\|ax\| = |a|\|x\|$;

(c) $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality).

A *normed space* is a pair $(\mathcal{X}, \|\cdot\|)$ consisting of a vector space $\mathcal{X}$ and a norm.

Most of the objects that are studied in analysis are connected to normed spaces or their cousins.

1.3.2. Example. (a) Let $\mathcal{X} = \mathbb{F}^d$ for some natural number d, and for $x = (x_1, \ldots, x_d)$ in $\mathbb{F}^d$ define $\|x\| = \left[\sum_{k=1}^d |x_k|^2\right]^{\frac{1}{2}}$. This is a norm but showing that the triangle inequality holds involves a bit of development. You may take this on faith until §3.1 below. Two other norms on $\mathbb{F}^d$ are $\|x\|_1 = \sum_{k=1}^d |x_k|$ and $\|x\|_\infty = \max\{|x_k| : 1 \leq k \leq d\}$.

(b) Let ℓ^1 denote the set of all sequences $\{a_n\}$ of numbers in $\mathbb{F}$ such that $\sum_{n=1}^\infty |a_n| < \infty$. Define addition and scalar multiplication for elements of ℓ^1 by: $\{a_n\}+\{b_n\} = \{a_n+b_n\}$ and $a\{a_n\} = \{aa_n\}$. The reader is asked to show that this transforms ℓ^1 into a vector space and that $\|\{a_n\}\| = \sum_{n=1}^\infty |a_n|$ defines a norm on ℓ^1.

(c) Let c_0 denote the set of all sequences $\{a_n\}$ of numbers from $\mathbb{F}$ such that $a_n \to 0$. If addition and scalar multiplication are defined entrywise on c_0 as they were on ℓ^1, then c_0 is a vector space. Moreover $\|\{a_n\}\| = \sup_n |a_n|$ defines a norm on c_0.

(d) Let ℓ^∞ denote the set of all bounded sequences of numbers $\{a_n\}$ from $\mathbb{F}$ and define addition and scalar multiplication entrywise as in the previous two examples; this makes ℓ^∞ into a vector space. If we set $\|\{a_n\}\| = \sup_n |a_n|$ for each $\{a_n\}$ in ℓ^∞, this is a norm.

(e) Let c_{00} denote the vector space of all finite sequences. There are many choices of a norm for c_{00}; for example, either of the norms used for c_0 or ℓ^1 can be used.

(f) Let $J = [a, b]$ be a bounded interval in the real line and let $BV(J)$ denote the functions of bounded variation on J. If we define $\|\alpha\| = |\alpha(a)| + \mathrm{Var}(\alpha)$ for α in $BV(J)$, then this is a norm on $BV(J)$.

(g) If X is a metric space and $f \in C_b(X)$, $\|f\| = \sup\{|f(x)| : x \in X\}$ is a norm on $C_b(X)$. To show that the triangle inequality holds observe that $|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\| + \|g\|$. Taking the supremum over all x gives the inequality. Verifying that the remaining properties of a norm are satisfied is straightforward. Actually if we consider $\mathbb{N}$ as a metric space, where the metric is the discrete or trivial metric ($d(x, y) = 1$ for $x \neq y$ and $d(x, x) = 0$), then ℓ^∞ in part (d) is seen to be identical with $C_b(\mathbb{N})$.

If $\mathcal{X}$ is a normed space, let

$$\text{ball}\,\mathcal{X} = \{X \in \mathcal{X} : \|x\| \leq 1\}$$

This closed unit ball will play a prominent role as our study of normed spaces progresses.

1.3.3. Proposition. *If $\mathcal{X}$ is a normed space and $x, y \in \mathcal{X}$, then $\big|\|x\| - \|y\|\big| \leq \|x - y\|$. Moreover if $d : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ is defined by $d(x, y) = \|x - y\|$, then $(\mathcal{X}, d)$ is a metric space.*

Proof. We have $\|x\| = \|(x - y) + y\| \leq \|x - y\| + \|y\|$; hence $\|x\| - \|y\| \leq \|x - y\|$. If the rolls of x and y are interchanged in the preceding inequality and this new relation is combined with the preceding one, we get $\big|\|x\| - \|y\|\big| \leq \|x - y\|$. The proof that $d(x, y)$ is a metric is routine. ■

When discussing a metric or any metric space concepts on a normed space it is always assumed that the above metric is the one under discussion.

1.3.4. Definition. A *Banach*[6] *space* is a normed space that is complete with respect to its metric. That is, every Cauchy sequence converges.

All the examples in (1.3.2) except for (e) are Banach spaces. The proof that c_0 is a Banach space takes a little work, but should be within the ability of the reader. The proof that ℓ^1 is complete takes a rather technical argument that will be encountered again and so we give this below.

1.3.5. Proposition. *ℓ^1 is a Banach space.*

Proof. Again the proof that ℓ^1 is a normed space is routine. For each $n \geq 1$, let $x^n = \{x_k^n\} \in \ell^1$ and assume that $\{x^n\}$ is a Cauchy sequence. If $k \geq 1$, then $|x_k^n - x_k^m| \leq \|x^n - x^m\|$, and so for each $k \geq 1$ we have that $\{x_k^n\}$ is a

[6]Stefan Banach was born in 1892 in Krakow, presently in Poland but then part of Austria-Hungary (Polish history is complicated). In 1922 the university in Lvov (a city that during its history has belonged to at least three countries and is presently in Ukraine) awarded Banach his habilitation for a thesis on measure theory. His earlier thesis on Integral Equations is sometimes said to mark the birth of functional analysis. He was one of the stars in a particularly brilliant constellation of Polish mathematicians during this period. Banach and his colleague Hugo Steinhaus (see Theorem 6.7.5 for a biographical note) in Lvov as well as other mathematicians in Warsaw began publishing a series of mathematical monographs. The first to appear in 1931 was Banach's *Théorie des Opérations Linéaires*, which had an enormous impact on analysis and continues to hold its place as a classic still worth reading. Banach was one of the founders of functional analysis, which had been brewing in academic circles for sometime. You will see Banach's name appear often as this book progresses. He died in Lvov in 1945 just after the end of World War II.

Cauchy sequence in $\mathbb{F}$; put $x_k = \lim_n x_k^n$. For any $K \geq 1$ and any $n \geq 1$,

$$\begin{aligned}\left(\sum_{k=1}^{K} |x_k|\right) &\leq \left(\sum_{k=1}^{K} |x_k - x_k^n|\right) + \left(\sum_{k=1}^{K} |x_k^n|\right) \\ &\leq \left(\sum_{k=1}^{K} |x_k - x_k^n|\right) + \|x^n\|\end{aligned}$$

Choose n sufficiently large that $|x_k - x_k^n| < 1/K$ for $1 \leq k \leq K$. Since every Cauchy sequence is bounded (see Exercise 1), $\left(\sum_{k=1}^{K} |x_k|\right) \leq 1 + C$, where $C \geq \|x^n\|$ for all $n \geq 1$. Therefore $x = \{x_k\} \in \ell^1$. Now we will show that $x^n \to x$ in ℓ^1.

Let $\epsilon > 0$ and choose N_1 such that $\|x^n - x^m\| < \epsilon/3$ when $n, m \geq N_1$. Fix an $m \geq N_1$, and for this m, since $x - x^m \in \ell^1$, we can choose L such that $\sum_{k=L+1}^{\infty} |x_k - x_k^m| < \epsilon/3$. Now choose N_2 such that $|x_k - x_k^n| < \epsilon/3L$ for $1 \leq k \leq L$ when $n \geq N_2$. Put $N = \max\{N_1, N_2\}$. Adopt the notation that $T_L(x - x^n)$ is the truncation of the sequence $x - x^n$ after the L-th term; that is, $T_L(x - x^n)$ is the sequence that is $x_k - x_k^n$ for $1 \leq k \leq L$ and then is identically 0 for $k > L$. Observe that for $n \geq N_2$, $\|T_L(x - x^n)\| < L\frac{\epsilon}{3L} = \epsilon/3$. Let $R_L(x - x^n)$ be the remainder sequence, with a 0 in the first L places and then agreeing with $x - x^n$. If n is an arbitrary integer larger than N and m is that integer larger than N_1 that was fixed above, then

$$\begin{aligned}\|x - x^n\| &\leq \|T_L(x - x^n)\| + \|R_L(x - x^n)\| \\ &< \frac{\epsilon}{3} + \|R_L(x - x^m)\| + \|R_L(x^m - x^n)\| \\ &< \frac{2\epsilon}{3} + \|x^m - x^n\| \\ &< \epsilon\end{aligned}$$

■

Here is another important example of a Banach space, one that will occupy a considerable amount of our attention.

1.3.6. Proposition. *$C_b(X)$ is a Banach space. Moreover a sequence $\{f_n\}$ in $C_b(X)$ converges to f if and only if $f_n(x) \to f(x)$ uniformly for x in X.*

Proof. We prove the last part first. Assume $\|f_n - f\| \to 0$ and let $\epsilon > 0$. So there is an integer N such that $\|f_n - f\| < \epsilon$ for $n \geq N$. Thus for any x in X and $n \geq N$, $|f_n(x) - f(x)| \leq \|f_n - f\| < \epsilon$. This says that $f_n(x) \to f(x)$ uniformly for x in X. Conversely, assume $f_n(x) \to f(x)$ uniformly for x in X. So if $\epsilon > 0$ there is an integer N such that for $n \geq N$, $|f_n(x) - f(x)| \leq \epsilon$ for all x in X. Thus $\|f_n - f\| = \sup\{|f_n(x) - f(x)| : x \in X\} \leq \epsilon$ when $n \geq N$.

To establish the completeness of $C_b(X)$, assume that $\{f_n\}$ is a Cauchy sequence in $C_b(X)$. Since $|f_n(x) - f_m(x)| \leq \|f_n - f_m\|$, we have, as in the preceding paragraph, that $\{f_n(x)\}$ is uniformly Cauchy in x. As is standard in metric space theory, this implies there is a continuous function f on X such that $f_n(x) \to f(x)$ uniformly in x. Choose N_1 such that $\|f_n - f_m\| < 1$ for $n, m \geq N_1$. Thus for $n \geq N_1$, $\|f_n\| \leq \|f_n - f_{N_1}\| + \|f_{N_1}\| \leq 1 + \|f_{N_1}\|$. So if $M = 1 + \max\{\|f_1\|, \ldots, \|f_{N_1}\|\}$, we have that $\|f_n\| \leq M$ for all $n \geq 1$. It follows that $|f(x)| \leq M$ for all x in X; that is, $f \in C_b(X)$. Since $f_n(x) \to f(x)$ uniformly on X, $f_n \to f$ in $C_b(X)$ from the first paragraph in this proof. ■

The reader should now begin to undergo a change of point of view about functions. Namely, as indicated by the last result as well as the examples in (1.3.2), we want to start thinking of continuous functions as well as sequences as points in a larger space. The author remembers having some difficulty making this change when he was a graduate student, but it comes shortly with study and working with the functions as points. Indeed much of the power of modern analysis rests on adopting such a point of view and this should become apparent as we progress.

To prove one of the main results of this section we need the following.

1.3.7. Lemma. *If X is a metric space and S is a dense subset of X such that every Cauchy sequence of elements of S has a limit in X, then X is complete.*

Proof. Let $\{x_n\}$ be a Cauchy sequence in X and for every n let $s_n \in S$ such that $d(x_n, s_n) < n^{-1}$. We want to show that $\{s_n\}$ is a Cauchy sequence. If $\epsilon > 0$, choose N_1 such that $d(x_n, x_m) < \epsilon/3$ when $n, m \geq N_1$. Let $N \geq N_1$ such that $n^{-1} < \epsilon/3$ when $n \geq N$. Thus for $n, m \geq N$, $d(s_n, s_m) \leq d(s_n, x_n) + d(x_n, x_m) + d(x_m, s_m) < \epsilon$. Hence, by hypothesis, there is an x in X such that $s_n \to x$. Therefore $d(x_n, x) \leq n^{-1} + d(s_n, x) \to 0$. ■

The next result concerns normed spaces. There is a similar result about metric spaces that some readers might have seen and will be stated after we prove this theorem, but the normed space version is really the one we want to concentrate on. One way for the reader to have a picture of what is happening in this proof is to keep in mind the process of defining the real numbers once we have the rational numbers. One way to do this is by means of a device know as a Dedekind cut, but another is by saying a real number is an equivalence class of Cauchy sequences of rational numbers. It is this last one that is analogous to what we undertake in the proof of the next theorem.

1.3.8. Theorem. *If $\mathcal{X}$ is a normed space, then there is a Banach space $\widehat{\mathcal{X}}$ and a linear isometry $U : \mathcal{X} \to \widehat{\mathcal{X}}$ such that $U(\mathcal{X})$ is dense in $\widehat{\mathcal{X}}$. Moreover $\widehat{\mathcal{X}}$ is unique in the sense that if $\mathcal{Y}$ is another Banach space with a linear isometry $W : \mathcal{X} \to \mathcal{Y}$ such that $W(\mathcal{X})$ is dense in $\mathcal{Y}$, then there is a surjective isometry $V : \widehat{\mathcal{X}} \to \mathcal{Y}$ such that $VU = W$.*

Proof. Let $\mathcal{X}_0$ denote the collection of all Cauchy sequences in $\mathcal{X}$. Note that $\mathcal{X}_0$ can be made into a vector space by defining $a\{x_n\} + \{y_n\} = \{ax_n + y_n\}$. Also observe that for $\{x_n\}$ in $\mathcal{X}_0$, $\{\|x_n\|\}$ is a Cauchy sequence in $\mathbb{R}$ by Proposition 1.3.3; so $\|\{x_n\}\| \equiv \lim \|x_n\|$ exists. It is routine to verify that this satisfies the properties of a norm on $\mathcal{X}_0$ except that it is possible that $\|\{x_n\}\| = 0$ without $\{x_n\}$ being the identically 0 sequence. (For example $\{x, 0, 0, \dots\} \in \mathcal{X}_0$ and $\|(x, 0, 0, \dots)\| = 0$.) We therefore introduce a relation on $\mathcal{X}_0$ by saying that $\{x_n\} \sim \{y_n\}$ if $\|x_n - y_n\| \to 0$; equivalently, if $\|\{x_n - y_n\}\| = 0$. It is left to the reader to check, using the various properties of the norm and Cauchy sequences, that $\sim$ is an equivalence relation on $\mathcal{X}_0$. Let $\widehat{\mathcal{X}}$ denote the set of equivalence classes for $\sim$. (This is a rather common event in mathematics. We have a set $\mathcal{X}$ and an equivalence relation. We examine the collection of equivalence classes $\widehat{\mathcal{X}}$ and want to impose a structure on $\widehat{\mathcal{X}}$ that reflects the structure of the original set $\mathcal{X}$. In this case we want to show that $\widehat{\mathcal{X}}$ is a normed space and, eventually, that it is a Banach space.)

To show that $\widehat{\mathcal{X}}$ is a vector space, let $\xi, \eta \in \widehat{\mathcal{X}}$ and let $\{x_n\}, \{y_n\}$ be representatives of ξ, η, respectively. We want to define $\xi + \eta$ as the equivalence class of the sequence $\{x_n + y_n\}$, which is easily seen to belong to $\mathcal{X}_0$. To be sure that $\xi + \eta$ is well defined, we must check that the definition is independent of the representatives chosen. That is, suppose $\{x_n'\} \sim \{x_n\}$ and $\{y_n'\} \sim \{y_n\}$. We must check that $\{x_n' + y_n'\} \sim \{x_n + y_n\}$. We leave the details to the reader. Similarly if $\xi \in \widehat{\mathcal{X}}$ with representative $\{x_n\}$ and $a \in \mathbb{F}$, we define $a\xi$ to be the equivalence class of the sequence $\{ax_n\}$. Once this is done it is easily checked that $\widehat{\mathcal{X}}$ satisfies the axioms of a vector space.

Now we put a norm on $\widehat{\mathcal{X}}$ by letting $\|\xi\| = \|\{x_n\}\|$ for any representative of ξ. Here Proposition 1.3.3 implies that if $\{x_n\} \sim \{y_n\}$, then $\|\{x_n\}\| = \|\{y_n\}\|$ so that this is well defined. Now note that if $\|\xi\| = 0$, then $\{x_n\} \sim \{0, 0, \dots\}$, the zero of $\widehat{\mathcal{X}}$, so that ξ is the zero of $\widehat{\mathcal{X}}$. Let $\{x_n\} \in \xi$ and $\{y_n\} \in \eta$. So $\|\xi + \eta\| = \lim \|x_n + y_n\| \le \lim \|x_n\| + \lim \|y_n\| = \|\xi\| + \|\eta\|$. That is $\|\cdot\|$ satisfies the triangle inequality on $\widehat{\mathcal{X}}$. The fact that $\|a\xi\| = |a|\|\xi\|$ for every ξ in $\widehat{\mathcal{X}}$ is straightforward. Thus we have that $\widehat{\mathcal{X}}$ is a normed space.

If $x \in \mathcal{X}$, let ξ_x be the equivalence class of the sequence $\{x, x, \dots\}$. Clearly $\xi_x \in \widehat{\mathcal{X}}$. If we define $U : \mathcal{X} \to \widehat{\mathcal{X}}$ by $U(x) = \xi_x$, then it is easily verified that $U : \mathcal{X} \to \widehat{\mathcal{X}}$ is a linear isometry. We want to show that the range of U is dense in $\widehat{\mathcal{X}}$ and that every Cauchy sequence from $U(\mathcal{X})$ has a limit in $\widehat{\mathcal{X}}$. Once this is done, the preceding lemma implies that $\widehat{\mathcal{X}}$ is complete and we have finished the proof of the existence part of the theorem.

To show the density of $U(\mathcal{X})$ let $\{x_n\} \in \xi \in \widehat{\mathcal{X}}$ and let $\epsilon > 0$. So there is an integer N such that $\|x_n - x_m\| < \epsilon/2$ for $n, m \geq N$. Thus $\|U(x_N) - \xi\| = \lim_n \|x_N - x_n\| < \epsilon$. Therefore the range of U is dense in $\widehat{\mathcal{X}}$. Now let $\{U(x_n\}$ be a Cauchy sequence in $U(\mathcal{X})$. Since U is an isometry, $\{x_n\}$ is a Cauchy sequence in $\mathcal{X}$; let ξ be its equivalence class in $\widehat{\mathcal{X}}$. We will show that $U(x_n) \to \xi$ in $\widehat{\mathcal{X}}$. In fact, if $\epsilon > 0$, there is an N such that $\|x_n - x_m\| < \epsilon/2$ for all $n, m \geq N$. Thus if $m \geq N$, $\|U(x_m) - \xi\| = \lim_n \|x_m - x_n\| < \epsilon$.

Now for the proof of uniqueness. Let $\mathcal{Y}$ and W be as in the statement of the theorem. If $\{x_n\} \in \xi \in \widehat{\mathcal{X}}$, the fact that W is an isometry implies that $\{W(x_n)\}$ is a Cauchy sequence in $\mathcal{Y}$. Thus there is a y in $\mathcal{Y}$ such that $W(x_n) \to y$. Let $V(\xi) = y$. The reader can check that V is well defined and linear. Also $\|y\| = \lim \|W(x_n)\| = \lim \|x_n\| = \|\xi\|$, so that V is an isometry. Moreover for any x in $\mathcal{X}$, $VU(x) = V(\xi_x) = W(x)$ by the definition of V. Thus V is a linear isometry with dense range and it must therefore be surjective. ■

The unique Banach space $\widehat{\mathcal{X}}$ obtained in the preceding theorem is called the *completion* of $\mathcal{X}$. Using this same proof we can establish a similar result about general metric spaces, though we do not have to discuss any linearity conditions.

1.3.9. Theorem. *If X is a metric space then there is a complete metric space $\widehat{X}$ and an isometry $\tau : X \to \widehat{X}$ such that $\tau(X)$ is dense in $\widehat{X}$. Moreover $\widehat{X}$ is unique in the sense that if Y is a complete metric space such that there is an isometry $\phi : X \to Y$ with $\phi(X)$ dense in Y, then there is a surjective isometry $\psi : \widehat{X} \to Y$ with $\psi\tau = \phi$.*

As a practical point, it is easier to think of $\mathcal{X}$ as contained in $\widehat{\mathcal{X}}$ rather than work through the isometry U. When we actually complete a specific normed space, this is what happens. For example, if $\mathcal{X} = c_{00}$ as in Example 1.3.2(e) and it is given the supremum norm, then $\widehat{c_{00}} = c_0$.

Exercises.

(1) Show that if $\mathcal{X}$ is a normed space and $\{x_n\}$ is a Cauchy sequence in $\mathcal{X}$, then there is a constant C such that $\|x_n\| \leq C$ for all n.

(2) Let A be a subset of the normed space $\mathcal{X}$, and denote by $\bigvee A$ the intersection of all closed linear subspaces of $\mathcal{X}$ that contain A; this is called the *closed linear span of A*. Prove the following. (a) $\bigvee A$ is a closed linear subspace of $\mathcal{X}$; (b) $\bigvee A$ is the smallest closed linear

subspace of $\mathcal{X}$ that contains A. (c) $\bigvee A$ is the closure of

$$\left\{\sum_{k=1}^{n} a_k x_k : n \geq 1, \text{ for } 1 \leq k \leq n, a_k \in \mathbb{F}, x_k \in A\right\}$$

(3) If $\mathcal{X}$ is a normed space, show that $\mathcal{X}$ is complete if and only if whenever $\{x_n\}$ is a sequence in $\mathcal{X}$ such that $\sum_{n=1}^{\infty} \|x_n\| < \infty$, then $\sum_{n=1}^{\infty} x_n$ converges in $\mathcal{X}$. (Note: To say that an infinite series $\sum_{n=1}^{\infty} x_n$ converges in a normed space means that the sequence of partial sums $\{\sum_{k=1}^{n} x_k\}$ converges. What else, right?)

(4) (a) Show that the space c_0 in Example 1.3.2(c) is a Banach space. (Hint: If $x^k = \{x_n^k\}$ is a Cauchy sequence in c_0, show that for every $\epsilon > 0$ there is an integer N such that $|x_n^k| < \epsilon$ for $n \geq N$ and all $k \geq 1$.) (b) Let c be all the sequences $\{x_n\}$ of numbers from $\mathbb{F}$ such that $\lim_n x_n$ exists, and show that c is a Banach space with the supremum norm $\|\{x_n\}\| = \sup_n |x_n|$. (c) If $e = \{1, 1, \dots\}$, show that $c = c_0 + \mathbb{F}e$.

(5) Show that c_{00} is dense in both c_0 and ℓ^1.

(6) Let $w = \{w_n\}$ be a sequence of strictly positive real numbers and define $\ell^1(w)$ to be all the sequences $\{x_n\}$ of numbers from $\mathbb{F}$ such that $\sum_{n=1}^{\infty} w_n|x_n| < \infty$. Define a norm on $\ell^1(w)$ by $\|\{x_n\}\| = \sum_{n=1}^{\infty} w_n|x_n|$ and show that $\ell^1(w)$ is a Banach space.

(7) Let w be as in the preceding exercise and define $c_0(w)$ to be all the sequences $\{x_n\}$ of numbers from $\mathbb{F}$ such that $\lim_n w_n x_n = 0$. If we define a norm on this space by $\|\{x_n\}\| = \sup_n w_n|x_n|$, show that $c_0(w)$ is a Banach space.

(8) Let J be a bounded interval in $\mathbb{R}$ and denote by $BV(J)$ the set of all functions of bounded variation on J. Show that with the norm defined in Example 1.3.2(f), $BV(J)$ is a Banach space.

(9) Let $\mathcal{X} = C([0,1])$ but give it the norm $\|f\| = \int_0^1 |f(t)|dt$. Show that $\mathcal{X}$ with this norm is not complete.

(10) Find a sequence $\{f_n\}$ in $C([0,1])$ such that when $n \neq m$, $\|f_n - f_m\| \geq 1$ and that this implies that $\operatorname{ball} C([0,1])$ is not compact. Can you generalize this to $C_b(X)$ for any infinite metric space X? What happens when X is a finite metric space?

(11) Show that the sequence $\{f_n\}$ found in Exercise 10 is a linearly independent set.

(12) (a) Show that if $\mathcal{X}$ is a normed space, then $\mathcal{X}$ is separable if and only if $\operatorname{ball} \mathcal{X}$ is separable. (b) Show that c_0 is a separable Banach space. (c) Show that ℓ^∞ is not separable. (Hint: You know that the collection of all subsets of $\mathbb{N}$ is uncountable. Use these sets to

define an uncountable set of functions that will enable you to use Exercise 1.2.5.)

(13) This continues the preceding exercise. Recall that a compact metric space is separable. If (X, d) is a compact metric space, fix a countable dense subset D in X. Let $\mathcal{G}$ be the collection of all finite open covers of X by balls of the form $B(x; \frac{1}{n})$, where $x \in D$ and $n \geq 1$. (a) Show that $\mathcal{G}$ is countable. (b) For each $\gamma = \{G_1, \ldots, G_n\}$ in $\mathcal{G}$ fix a partition of unity $\Phi_\gamma = \{\phi_1, \ldots, \phi_n\}$ subordinate to γ and let $\mathcal{X}_\gamma$ denote the collection of all linear combinations with rational coefficients of the functions in Φ_γ. Show that $\bigcup_{\gamma \in \mathcal{G}} \mathcal{X}_\gamma$ is a countable dense subset in $C(X)$ and, hence, that $C(X)$ is separable.

(14) Let $\mathcal{X}$ be a normed space and let G be an open subset of $\mathcal{X}$. Show that G is a connected set if and only if for any two points a and b in G there are points $a = x_0, x_1, \ldots, x_n = b$ in G such that for $1 \leq j \leq n$ the line segment $[x_{j-1}, x_j] \equiv \{tx_j + (1-t)x_{j-1} : 0 \leq t \leq 1\} \subseteq G$.

1.4. Locally compact spaces

We want to extend the concept of a compact metric space. It is likely that at least some of the readers have seen the following definition, which can be made in the setting of an arbitrary topological space, not just a metric space as is done here.

1.4.1. Definition. A metric space (X, d) is *locally compact* if for every point x in X there is a radius $r > 0$ such that $B(x; r)$ has compact closure.

1.4.2. Example. (a) Euclidean space is a locally compact space.

(b) If X is any set and d is the discrete metric ($d(x, y) = 1$ when $x \neq y$ and $d(x, x) = 0$), then (X, d) is locally compact.

(c) Every compact metric space is locally compact, and here is a way to get additional examples of locally compact spaces once we are given a compact metric space. Let (Y, d) be a compact metric space, fix a point y_0 in Y, and let $X = Y \backslash \{y_0\}$. (X, d) is locally compact. Can it be compact? Concrete examples occur by letting $Y = [0, 1]$ and $y_0 = \{1\}$ or letting $Y = [0, 1] \cup \{2\}$ and $y_0 = \{2\}$.

(d) It is not necessarily the case that all locally compact metric spaces have the property that the closure of every ball of finite radius is compact, as is the case with $\mathbb{F}^d$. For example, let Y, X, and y_0 be as in the previous example. If y is a point of Y distinct from y_0 and $r > d(y, y_0)$, then the ball in X of radius r and centered at y may not have a compact closure in X.

The main interest we have in locally compact spaces is to do analysis on them. This is possible even for non-metric spaces. However in this book we limit ourselves to the metric space case for the sake of easier exposition. Note that with Euclidean spaces being locally compact, we do not lack examples where analysis on a locally compact space might be profitable. One of our principal tools will be to embed the locally compact metric space inside a compact metric space. This is not always possible, but let's set the stage. We need to introduce a special space of continuous functions on a locally compact space.

1.4.3. Definition. If X is a locally compact metric space, say that a continuous function $f : X \to \mathbb{F}$ *vanishes at infinity* if for every $\epsilon > 0$ the set $\{x \in X : |f(x)| \geq \epsilon\}$ is compact. Let $C_0(X)$ denote the set of all continuous functions on X that vanish at infinity. For any continuous function $f : X \to \mathbb{F}$, define the *support* of f to be the set $\operatorname{cl}\{x : f(x) \neq 0\}$ and denote this by $\operatorname{spt}(f)$. Say that f has *compact support* if $\operatorname{spt}(f)$ is a compact subset of X. Let $C_c(X)$ denote the set of all continuous functions $f : X \to \mathbb{F}$ having compact support.

1.4.4. Example. (a) If $\mathbb{N}$ has the discrete metric, then $C_b(\mathbb{N}) = \ell^\infty$, the space of bounded sequences; $C_0(\mathbb{N}) = c_0$, the space of sequences that converge to 0; $C_c(\mathbb{N}) = c_{00}$, the space of all finitely non-zero sequences.

(b) Consider Example 1.4.2(c), where $X = Y \backslash \{y_0\}$. Here we can identify $C_0(X)$ with those functions in $C(Y)$ that vanish at y_0.

Say that a locally compact space X is *σ-compact* if $X = \bigcup_{n=1}^\infty K_n$ where each K_n is a compact subset. It is left as an exercise for the reader to show that if X is σ-compact, then we can write X as the union of compact sets K_n such that $K_n \subseteq \operatorname{int} K_{n+1}$. (See Exercise 2.) Note that Euclidean space is σ-compact as is the locally compact space given in Example 1.4.2(c). In this last case we can take $K_n = \{y \in Y : d(y, y_0) \geq \frac{1}{n}\}$, which is a compact subset of X.

1.4.5. Proposition. *Let X be a locally compact space.*

(a) *Both $C_c(X)$ and $C_0(X)$ are subalgebras of $C_b(X)$ and $C_c(X) \subseteq C_0(X)$. Moreover $C_0(X)$ is closed in $C_b(X)$. Hence $C_0(X)$ is a Banach space.*

(b) *$C_c(X)$ is dense in $C_0(X)$.*

(c) *If $f \in C_0(X)$, then f is uniformly continuous.*

(d) *If X is σ-compact, there is a sequence of functions $\{\phi_n\}$ in $C_c(X)$ such that $0 \leq \phi_n \leq \phi_{n+1} \leq 1$ for all $n \geq 1$ and for every f in $C_0(X)$, $\|f\phi_n - f\| \to 0$ as $n \to \infty$.*

Proof. (a) The fact that $C_c(X)$ and $C_0(X)$ are subalgebras of $C_b(X)$ and $C_c(X) \subseteq C_0(X)$ is routinely proven. Let $\{f_n\}$ be a sequence in $C_0(X)$ and assume that $f_n \to f$ in $C_b(X)$. That is $\|f_n - f\| = \sup\{|f_n(x) - f(x)| : x \in X\} \to 0$. Let $\epsilon > 0$ and put $K = \{x : |f(x)| \geq \epsilon\}$; we want to show that K is compact. Clearly K is closed. Let $N \geq 1$ such that $\|f_n - f\| < \epsilon/2$ for all $n \geq N$; fix an $n \geq N$. If $y \in K$, then $\epsilon \leq |f(y) - f_n(y)| + |f_n(y)| \leq \epsilon/2 + |f_n(y)|$, and so $y \in \{x : |f_n(x)| \geq \epsilon/2\}$. That is, K is a closed subset of $\{x : |f_n(x)| \geq \epsilon/2\}$, a compact set. Thus K is compact and so $f \in C_0(X)$.

(d) As in Exercise 2 we write $K = \bigcup_{n=1}^{\infty} K_n$ where each K_n is compact and $K_n \subseteq \operatorname{int} K_{n+1}$ for all $n \geq 1$. By Urysohn's Lemma we can find a ϕ_n in $C(X)$ such that $0 \leq \phi_n \leq 1$, $\phi_n(x) = 1$ for all x in K_n, and $\phi_n(x) = 0$ for all x in $X \backslash \operatorname{int} K_{n+1}$. It follows that $\phi_n C_c(X)$ and $0 \leq \phi_n \leq \phi_{n+1} \leq 1$. If $f \in C_0(X)$ and $\epsilon > 0$, let $K = \{x : |f(x)| \geq \epsilon\}$. Since K is compact, there is an integer N such that $K \subseteq \operatorname{int} K_n$ for all $n \geq N$. It is easy to see that $|f(x) - f(x)\phi_n(x)| = 0$ when $x \in K_n$. Hence for $n \geq N$, $\|f - f\phi_n\| < \epsilon$.

(b) If $f \in C_0(X)$ and $\{\phi_n\}$ is a sequence in $C_c(X)$ as in (d), then $f\phi_n \in C_c(X)$ for all $n \geq 1$ and $f\phi_n \to f$.

(c) Let $\epsilon > 0$ and put $L = \{x : |f(x)| \geq \epsilon/2\}$; so L is compact. Since X is locally compact, for every x in L there is an $r_x > 0$ such that $\operatorname{cl} B(x; r_x)$ is compact. Let $x_1, \ldots, x_m \in L$ such that $L \subseteq \bigcup_{j=1}^{m} B(x_j; r_{x_j})$. Choose $\gamma > 0$ such that $\operatorname{dist}(x, L) \leq \gamma$ implies $x \in \bigcup_{j=1}^{m} B(x_j; r_{x_j})$ and set $K = \{x : \operatorname{dist}(x, L) \leq \gamma\}$. Note that K is compact since it is contained in $\bigcup_{j=1}^{m} \operatorname{cl} B(x_j; r_{x_j})$. If we only consider f as a function on K, it is uniformly continuous there; so there exists δ such that $0 < \delta < \gamma$ and if $x, y \in K$ and $d(x, y) < \delta$, then $|f(x) - f(y)| < \epsilon$. Let $x, y \in X$ such that $d(x, y) < \delta$. If $x, y \in L$, then $|f(x) - f(y)| < \epsilon$. If $x \in L$ but $y \notin L$, then the fact that $d(x, y) < \delta < \gamma$ implies that $x, y \in K$; hence $|f(x) - f(y)| < \epsilon$. If neither point belongs to L, then $|f(x) - f(y)| \leq |f(x)| + |f(y)| < \epsilon/2 + \epsilon/2 = \epsilon$. ■

Fix a locally compact metric space (X, d). The objective that will occupy us for the remainder of the section is to embed X in a compact metric space (X_∞, ρ) in a "congenial" way. In fact we want to have X_∞ differ from X by a single point that we denote by ∞, the point at infinity. For those who have studied abstract topological spaces this is called the *one-point compactification* of X. We will not assume that the reader has this background and will give a self-contained exposition. It is possible that even the reader who has seen this topic has not addressed the question of when X_∞ is metrizable.

As a preview, consider the locally compact space X obtained from the compact space Y as in Example 1.4.2(c). Clearly the one-point compactification we seek is the space Y, where the point y_0 plays the role of the point

∞. The reader might keep this example in mind when we prove the next theorem.

Note that if Z is a set and d_1 and d_2 are two metrics on Z, then we say that the two metrics are *equivalent metrics* provided that the open sets defined by one metric coincide with those defined by the other. This is equivalent to the requirement that for a sequence $\{z_n\}$ in Z and a point z, $d_1(z_n, z) \to 0$ if and only if $d_2(z_n, z) \to 0$. See Exercise 6.

1.4.6. Theorem. *If (X, d) is a locally compact metric space, the following statements are equivalent.*

(a) *There is a metric space (X_∞, ρ) such that $X_\infty = X \cup \{\infty\}$, X_∞ is compact, the metric d on X is equivalent to the restriction of ρ to X, and a sequence $\{x_n\}$ in X satisfies $\rho(x_n, \infty) \to 0$ if and only if for every compact subset K of X, there is an integer N such that $x_n \notin K$ for all $n \geq N$.*

(b) *X is σ-compact.*

(c) *The metric space $C_0(X)$ is separable.*

Before we begin the proof, let's observe a few things so as to better understand the theorem. Note that the metric ρ obtained for X_∞ is not required to agree with the original metric for X but only to be equivalent to it. In fact if $X = \mathbb{R}$, there is no way that the metric on the compactification will be the same as the metric on $\mathbb{R}$ since one is bounded and the other is not. Second, again if $X = \mathbb{R}$, then the space X_∞ differs from $\mathbb{R}$ by a single point; we have not added the two points $\pm\infty$. Finally note that if $f \in C_0(X)$ and we extend f to be defined on X_∞ by letting $f(\infty) = 0$, then this extension is continuous. In fact if $x_n \to \infty$ and $\epsilon > 0$, let $K = \{x \in X : |f(x)| \geq \epsilon\}$, which is compact since $f \in C_0(X)$. Thus there is an integer N such that $x_n \notin K$ for all $n \geq N$. Thus $|f(x_n)| < \epsilon$ for $n \geq N$. Finally let us observe that the last requirement of the metric ρ contained in (a) is equivalent to the requirement that for every $\epsilon > 0$, $\{x \in X : \rho(x, \infty) \geq \epsilon\}$ is compact. In fact this last statement is that the function $f : X \to [0, \infty)$ defined by $f(x) = \rho(x, \infty)$ vanishes at infinity. So the condition is seen to be the same as the condition that $f(x_n) \to 0$ whenever $\{x_n\}$ escapes every compact subset of X.

Proof. (a) *implies* (b). If ρ is the metric on X_∞ and $K_n = \{x \in X : \rho(x, \infty) \geq \frac{1}{n}\}$, then each K_n is compact and their union is all of X.

(b) *implies* (c). (The reader may want to first look at Exercise 1.3.13, which outlines a proof that $C(X)$ is separable when X is a compact metric space. The proof given here is basically the same but technically more complex.) We write $X = \bigcup_{n=1}^{\infty} K_n$, where each K_n is compact and $K_n \subseteq \operatorname{int} K_{n+1}$ for all $n \geq 1$ (see Exercise 2). Find a decreasing sequence of positive numbers $\{\delta_n\}$ such that $\{x : \operatorname{dist}(x, K_n) < \delta_n\} \subseteq \operatorname{int} K_{n+1}$. (How?) For each n and each $k \geq n$, let $\{B(a^j_{nk}; \delta_k) : 1 \leq j \leq m_{nk}\}$ be open disks with a^j_{nk} in K_n such that $K_n \subseteq \bigcup_{j=1}^{m_{nk}} B(a^j_{nk}; \delta_k)$. Note that this union of open disks is contained in $\operatorname{int} K_{n+1}$. As in (1.2.12) for each $n \geq 1$ and $k \geq n$, let $\{\phi^j_{nk} : 1 \leq j \leq m_{nk}\}$ be continuous functions with $0 \leq \phi^j_{nk} \leq 1$, $\phi^j_{nk}(x) = 0$ for $x \notin B(a^j_{nk}; \delta_k)$, $\sum_{j=1}^{m_{nk}} \phi^j_{nk}(x) = 1$ when $x \in K_n$, and $\sum_{j=1}^{m_{nk}} \phi^j_{nk} \leq 1$. Let $\mathcal{M}$ be the linear span of the functions $\{\phi^j_{nk} : n \geq 1, k \geq n$, and $1 \leq j \leq m_{nk}\}$ with coefficients from the rational numbers $\mathbb{Q}$. Note that $\mathcal{M}$ is a countable subset of $C_c(X)$. We will show that $\mathcal{M}$ is dense in $C_c(X)$ and hence in $C_0(X)$ (1.4.5).

Fix f in $C_c(X)$ and let $\epsilon > 0$; so there is an integer n such that $f(x) = 0$ when $x \notin K_n$. Since f is uniformly continuous there is a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon/2$ whenever $d(x, y) < \delta$. Pick $k \geq n$ such that $\delta_k < \delta$. For $1 \leq j \leq m_{nk}$ let $q^j_{nk} \in \mathbb{Q}$ such that $|q^j_{nk} - f(a^j_{nk})| < \epsilon/2$. Hence $g = \sum_{j=1}^{m_{nk}} q^j_{nk}\phi^j_{nk} \in \mathcal{M}$. Now fix x in K_n. Thus

$$
\begin{aligned}
|f(x) - g(x)| &= \left|\sum_{j=1}^{m_{nk}} [f(x) - q^j_{nk}]\phi^j_{nk}(x)\right| \\
&\leq \sum_{j=1}^{m_{nk}} \left|f(x) - f(a^j_{nk})\right| \phi^j_{nk}(x) + \sum_{j=1}^{m_{nk}} \left|f(a^j_{nk}) - q^j_{nk}\right| \phi^j_{nk}(x) \\
&\leq \sum_{j=1}^{m_{nk}} \left|f(x) - f(a^j_{nk})\right| \phi^j_{nk}(x) + \epsilon/2
\end{aligned}
$$

Now when $\phi^j_{nk}(x) \neq 0$, $x \in B(a^j_{nk}; \delta_k)$ and so $\left|f(x) - f(a^j_{nk})\right| < \epsilon/2$. Putting this inequality into the preceding one yields that $|f(x) - g(x)| < \epsilon$ when $x \in K_n$. Suppose now that $x \in \left[\bigcup_{j=1}^{m_{nk}} B(a^j_{nk}; \delta_k)\right] \backslash K_n$. So $f(x) = 0$. If $x \in B(a^j_{nk}; \delta_k)$, then $\left|f(a^j_{nk})\right| = \left|f(a^j_{nk}) - f(x)\right| < \epsilon/2$. Therefore

$$
\begin{aligned}
|f(x) - g(x)| &= |g(x)| \\
&\leq \sum_{j=1}^{m_{nk}} \left|q^j_{nk}\right| \phi^j_{nk}(x) \\
&\leq \sum_{j=1}^{m_{nk}} \left[\left|q^j_{nk} - f(a^j_{nk})\right| + \left|f(a^j_{nk})\right|\right] \phi^j_{nk}(x) \\
&< (\epsilon/2 + \epsilon/2) \sum_{j=1}^{m_{nk}} \phi^j_{nk}(x) \\
&\leq \epsilon
\end{aligned}
$$

Finally if $x \notin \bigcup_{j=1}^{m_{nk}} B(a_{nk}^j;\delta_k)$, then $f(x) = 0 = g(x)$. Therefore we have that $\|f - g\| < \epsilon$ and so $\mathcal{M}$ is dense in $C_0(X)$.

(c) *implies* (a). Here is an outline of the proof. First note that when $f \in C_0(X)$, $\rho_f(x,y) = |f(x) - f(y)|$ is symmetric ($\rho_f(x,y) = \rho_f(y,x)$) and satisfies the triangle inequality. ρ_f is called a semimetric. Extending f to X_∞ by setting $f(\infty) = 0$ enables us to see that ρ_f defines a semimetric on X_∞. Now using the fact that $C_0(X)$ is separable we can use a countable dense subset of $\operatorname{ball} C_0(X)$ to generate a sequence of such semimetrics and sum them up to get a true metric on X_∞. The details follow.

Let $\{f_n\}$ be a countable dense sequence in the unit ball of $C_0(X)$ and define $\rho(x,y) = \sum_{n=1}^\infty \frac{1}{2^n}|f_n(x) - f_n(y)|$ for all x, y in X_∞. Note that for x in X, $\rho(x,\infty) = \sum_{n=1}^\infty \frac{1}{2^n}|f_n(x)|$ since $f_n(\infty) = 0$ for all n. Clearly $\rho(x,y) = \rho(y,x)$. If $\rho(x,y) = 0$, then $f_n(x) = f_n(y)$ for all $n \geq 1$. Now if $x \neq y$ there is a function f in $C_c(X)$ such that $f(x) = 1$ and $f(y) = 0$, assuming $x \neq \infty$. (Why?) Let $n \geq 1$ such that $\|f - f_n\| < \frac{1}{2}$. It follows that $|f_n(x)| > \frac{1}{2}$ and $|f_n(y)| < \frac{1}{2}$. Thus it cannot be that $\rho(x,y) = 0$. We leave it to the reader to verify that the triangle inequality holds for ρ so that it is a metric on X.

It remains to prove that the inclusion map $(X,d) \to (X_\infty,\rho)$ is a homeomorphism onto its image and that $\{x : \rho(x,\infty) \geq \epsilon\}$ is compact for every $\epsilon > 0$.

Claim. A sequence $\{x_n\}$ converges to x in (X_∞,ρ) if and only if $f_k(x_n) \to f_k(x)$ for all $k \geq 1$.

If $x_n \to x$ in (X_∞,ρ), then the fact that $2^{-k}|f_k(x_n) - f_k(x)| \leq \rho(x_n,x)$ implies $f_k(x_n) \to f_k(x)$. Now assume that $f_k(x_n) \to f_k(x)$ for all $k \geq 1$ and let $\epsilon > 0$. Choose m such that $\sum_{k=m}^\infty \frac{1}{2^k} < \epsilon/2$ and choose N such that for $n \geq N$ and $1 \leq k \leq m$, $|f_k(x_n) - f_k(x)| < \epsilon/2m$. Thus for $n \geq N$,

$$\rho(x_n,x) < \epsilon/2 + \sum_{k=1}^m \frac{1}{2^k}\frac{\epsilon}{2m} < \epsilon$$

proving the claim.

Claim. The inclusion map $(X,d) \to (X_\infty,\rho)$ is a homeomorphism.

The first claim proves that if $x_n \to x$ in X, then $\rho(x_n,x) \to 0$; that is, the inclusion map $(X,d) \to (X_\infty,\rho)$ is continuous. For the converse, suppose that $\rho(x_n,x) \to 0$, where x and x_n belong to X for all $n \geq 1$. Suppose that $\{d(x_n,x)\}$ does not converge to 0; then there is an $\epsilon > 0$ and a subsequence $\{x_{n_j}\}$ such that $d(x_{n_j},x) \geq \epsilon$ for all n_j. We can assume that ϵ is small, say $\epsilon < \frac{1}{2}$. Using Urysohn's Lemma there is a function f in $C_c(X)$ such that $0 \leq f \leq 1$, $f(x) = 1$, and $f(y) = 0$ when $d(y,x) \geq \epsilon/2$.

Since $\{f_k\}$ is dense in ball $C_0(X)$, there is an f_k such that $\|f_k - f\| < \epsilon/2$. Hence $|f_k(x_{n_j})| = |f_k(x_{n_j}) - f(x_{n_j})| < \epsilon/2$. On the other hand, $1 = f(x) \leq |f(x) - f_k(x)| + |f_k(x)| < \epsilon/2 + |f_k(x)|$, and so $|f_k(x)| > 1 - \epsilon/2 > \epsilon/2$. This contradicts the assumption that $f_k(x_{n_j}) \to f_k(x)$, which, by the first claim, contradicts the assumption that $\rho(x_n, x) \to 0$.

Claim. $\{x : \rho(x, \infty) \geq \epsilon\}$ is compact for every $\epsilon > 0$.

Let $\epsilon > 0$ and put $K = \{x \in X : \rho(x, \infty) \geq \epsilon\}$. Suppose K is not compact. For each $n \geq 1$, put $K_n = \{x \in X : |f_k(x)| \geq \frac{1}{n} \text{ for } 1 \leq k \leq n\}$. Since K is closed and each K_n is compact, it cannot be that $K \subseteq K_n$. Therefore there is a point x_n in K such that $x_n \notin K_n$. That is $|f_k(x_n)| < \frac{1}{n}$ for $1 \leq k \leq n$. This says that for every $k \geq 1$, $\lim_n f_k(x_n) = 0$. According to the first claim this implies that $x_n \to \infty$ or that $\rho(x_n, \infty) \to 0$. Since each $x_n \in K$, this is a contradiction. Therefore K must be compact.

It remains to show that (X_∞, ρ) is compact. So let $\{x_n\}$ be a sequence in X_∞ and let's show it has a convergent subsequence. One of two things holds; either there is a compact set K such that $x_n \in K$ for infinitely many values of n, or every compact subset of X contains an x_n for only a finite number of values of n. In the first of these cases there is automatically a convergent subsequence. So assume the latter case is valid. Write $X = \bigcup_{j=1}^\infty K_j$, where each K_j is compact and $K_j \subseteq \operatorname{int} K_{j+1}$ for all $j \geq 1$. Thus we can construct a subsequence $\{x_{n_j}\}$ such that for all $j \geq 1$, $x_{n_j} \notin K_j$. But if K is any compact subset of X, there is a j_0 such that $K \subseteq K_{j_0}$. Therefore $x_{n_j} \notin K$ whenever $j \geq j_0$. From the preceding paragraph we know that this implies that $x_{n_j} \to \infty$. ∎

Here is a sample of the way we can use the one-point compactification to parlay results about compact spaces into results about σ-compact locally compact spaces.

1.4.7. Proposition. *Let X be σ-compact locally compact metric space with a compact subset K. If $K \subseteq \bigcup_{k=1}^n U_k$ where each U_k is open, then there are functions $\phi_1, \ldots, \phi_n$ in $C_c(X)$ such that the following hold.*

(a) *For $1 \leq k \leq n$, $0 \leq \phi_k \leq \chi_{U_k}$.*

(b) *$\sum_{k=1}^n \phi_k(x) = 1$ for all x in K.*

(c) *$\sum_{k=1}^n \phi_k(x) \leq 1$ for all x in X.*

Proof. Start by finding open sets $G_1, \ldots, G_n$ such that for $1 \leq k \leq n$, $\operatorname{cl} G_k \subseteq U_k$ and $\operatorname{cl} G_k$ is compact. Now consider X_∞ and apply Corollary 1.2.12. ∎

It should be pointed out that we do not need Corollary 1.2.12 to prove the preceding proposition and, in fact, the conclusion is valid without the assumption that X is σ-compact. However under those circumstances the proof requires more effort.

Exercises.

(1) Is $\mathbb{Q}$ locally compact?

(2) If X is a locally compact metric space that is σ-compact, show that we can write X as the union of compact sets K_n such that $K_n \subseteq \operatorname{int} K_{n+1}$.

(3) If (X_k, d_k) is a locally compact metric space for $1 \leq k \leq n$, show that $X = \prod_{k=1}^n X_k$ with the metric $d(\{x_k\}, \{y_k\}) = \max_k d_k(x_k, y_k)$ is a locally compact metric space.

(4) It is known and usually proved in an elementary course on metric spaces that if (X_k, d_k) is a metric space for all $k \geq 1$, then $X = \prod_{k=1}^\infty X_k$ is a metric space. Give a necessary and sufficient condition that X is locally compact. (See the preceding exercise.)

(5) Show that every σ-compact metric space is separable.

(6) If Z is a set and d_1 and d_2 are two metrics on Z, then prove the following statements are equivalent. (a) If $\{z_n\}$ is a sequence in Z and $z \in Z$, then $d_1(z_n, z) \to 0$ if and only if $d_2(z_n, z) \to 0$. (b) A set G is open in (Z, d_1) if and only if it is open in (Z, d_2). (c) A set F is closed in (Z, d_1) if and only if it is closed in (Z, d_2).

(7) Show that every open subset of Euclidean space is σ-compact. More generally, show that if X is a σ-compact metric space and if G is an open subset, then G is σ-compact.

(8) Show that the one-point compactification of $\mathbb{R}$ is homeomorphic to the circle.

(9) If X is a locally compact metric space and $f, g \in C_0(X)$, show that $f \vee g$ and $f \wedge g \in C_0(X)$. Similarly, if $f, g \in C_c(X)$, show that $f \vee g$ and $f \wedge g \in C_c(X)$.

(10) Consider the metric space (X, d) defined in Exercise 1.2.7. (a) Show that (X, d) is locally compact. (b) Show that (X, d) is not σ-compact and hence $\widehat{X}$ is non-metrizable. (c) Find an infinite subset A of X such that no sequence of points converges to ∞ in the one-point compactification.

(11) If X is locally compact and σ-compact and $\phi \in C_0(X)$ such that $\{x : \phi(x) = 0\} = \emptyset$, show that $\phi C_0(X)$ is dense in $C_0(X)$. (Hint: use Proposition 1.4.5(b).) If X is not assumed to be σ-compact and ϕ is any function in $C_0(X)$, what is the closure of $\phi C_0(X)$?

(12) If X is a locally compact metric space, is the condition that X is σ-compact equivalent to the condition that X is separable? (See Exercise 5.)

1.5. Linear functionals

1.5.1. Definition. If $\mathcal{X}$ is a vector space, a *linear functional* on $\mathcal{X}$ is a function $L : \mathcal{X} \to \mathbb{F}$ satisfying $L(ax + by) = aL(x) + bL(y)$ for all x, y in $\mathcal{X}$ and all scalars a and b. If $\mathcal{X}$ is a normed space, say that a linear functional L is *bounded* if there is a constant M such that $|L(x)| \leq M\|x\|$ for all x in $\mathcal{X}$.

The concept of a linear functional applies to any vector space, so it is unlikely to be of much value in the study of Banach space. The idea of a bounded linear functional, however, connects the concept to the norm and makes it more relevant.

1.5.2. Example. (a) Give $\mathcal{X} = \mathbb{F}^d$ the norm $\|x\| = \sum_{k=1}^d |x_k|$; if $y_1, \ldots, y_d \in \mathbb{F}$ and we define $L : \mathcal{X} \to \mathbb{F}$ by $L(x) = \sum_{k=1}^d x_k y_k$, then L is a bounded linear functional. To see that it is bounded, let $M = \max\{|y_k| : 1 \leq k \leq d\}$ and note that $|L(x)| \leq M\|x\|$.

(b) If X is a compact metric space, $x \in X$, and $L : C_b(X) \to \mathbb{F}$ is defined by $L(f) = f(x)$, then L is a bounded linear functional, where the constant M can be taken to be 1.

(c) Let $\mathcal{X} = \ell^1$ as in Example 1.3.2(b). If $\{b_n\} \in \ell^\infty$ as in Example 1.3.2(d), define $L : \ell^1 \to \mathbb{F}$ by

$$L(\{a_n\}) = \sum_{n=1}^{\infty} a_n b_n$$

It follows that L is a bounded linear functional on ℓ^1 with

$$|L(\{a_n\})| \leq \|\{b_n\}\|_\infty \|\{a_n\}\|$$

for all $\{a_n\}$ in ℓ^1.

(d) If J is a bounded interval in $\mathbb{R}$ and α is a function of bounded variation on J, then Theorem 1.1.8 says that $L : C(J) \to \mathbb{R}$ defined by $L(f) = \int f d\alpha$ is a bounded linear functional with $|L(f)| \leq \mathrm{Var}(\alpha)\|f\|$.

The reader can consult Exercises 3 and 4 for examples of unbounded linear functionals, though it might be underlined that in both these exercises the normed spaces are not Banach spaces. Getting an example of a Banach space and an unbounded linear functional on it requires the Axiom of Choice. See Exercise 5. The next result demonstrates that the boundedness of a linear functional is intimately connected to the topology defined on a normed space.

1.5.3. Proposition. *If $\mathcal{X}$ is a normed space and $L : \mathcal{X} \to \mathbb{F}$ is a linear functional, then the following statements are equivalent.*

(a) *L is bounded.*

(b) *L is a continuous function.*

(c) *L is a continuous function at* 0.

(d) *L is continuous at some point of $\mathcal{X}$.*

Proof. (a) *implies* (b). Let M be a constant such that $|L(x)| \leq M\|x\|$ for all x in $\mathcal{X}$. In fact $|L(x) - L(y)| = |L(x-y)| \leq M\|x-y\|$, from which it follows that not only is L continuous but it is uniformly continuous.

It is trivial that (b) implies (c) and that (c) implies (d).

(d) *implies* (a). Suppose L is continuous at x_0; so there is a $\delta > 0$ with $|L(x-x_0)| = |L(x) - L(x_0)| < 1$ whenever $\|x - x_0\| < \delta$. If x is an arbitrary non-zero vector in $\mathcal{X}$, then $\left\|\left[\frac{\delta}{2\|x\|}x + x_0\right] - x_0\right\| = \frac{\delta}{2} < \delta$. Hence $\left|L\left(\frac{\delta}{2\|x\|}x\right)\right| < 1$, and this implies that $|L(x)| \leq \frac{2}{\delta}\|x\|$. Since x was an arbitrary non-zero vector and the preceding inequality also holds for $x = 0$, we have that L is bounded where the constant M can be taken to be $\frac{2}{\delta}$. ■

If L is a bounded linear functional on $\mathcal{X}$, we define

$$\|L\| = \sup\{|L(x)| : x \in \operatorname{ball}\mathcal{X}\}$$

Clearly $\|L\| < \infty$; this is called the *norm* of L.

1.5.4. Proposition. *If L is a bounded linear functional on $\mathcal{X}$, then $|L(x)| \leq \|L\|\|x\|$ for all x in $\mathcal{X}$ and*

$$\begin{aligned}\|L\| &= \sup\{|L(x)|/\|x\| : x \neq 0\} \\ &= \inf\{M : |L(x)| \leq M\|x\| \text{ for all } x \in \mathcal{X}\}\end{aligned}$$

Proof. If $x \in \mathcal{X}$, then $x/\|x\| \in \operatorname{ball}\mathcal{X}$ and so $|L(x/\|x\|)| \leq M$, demonstrating the inequality. Since the first equation in the display is trivial, we will only prove the second, which is not a lot more difficult. Let α denote the infimum in question. We just saw that $\|L\|$ is a possible such constant, so $\alpha \leq \|L\|$. On the other hand if M is any of the constants in question, then the definition of $\|L\|$ shows that $\|L\| \leq M$. Hence $\|L\| \leq \alpha$. ■

1.5.5. Definition. If $\mathcal{X}$ is a normed space and $\mathcal{X}^*$ denotes the collection of all bounded linear functionals on $\mathcal{X}$, then $\mathcal{X}^*$ is called the *dual space* of $\mathcal{X}$.

For a variety of reasons it is often convenient to denote the elements of $\mathcal{X}^*$ by x^* rather than something like L.

1.5.6. Proposition. *If $\mathcal{X}$ is a normed space, then $\mathcal{X}^*$ with its norm is a Banach space, where the operations of addition and scalar multiplication are defined pointwise: $(x^* + y^*)(x) = x^*(x) + y^*(x)$ and $(ax^*)(x) = ax^*(x)$.*

Proof. It is routine to show that $\mathcal{X}^*$ is a vector space and that its norm satisfies all the axioms of a norm on this vector space. We establish the completeness of $\mathcal{X}^*$. Let $\{x_n^*\}$ be a Cauchy sequence in $\mathcal{X}^*$. It follows that for every x in $\mathcal{X}$, $|x_n^*(x) - x_m^*(x)| = |(x_n^* - x_m^*)(x)| \leq \|x_n^* - x_m^*\|\|x\|$. Thus $\{x_n^*(x)\}$ is a Cauchy sequence in $\mathbb{F}$ and so $L(x) = \lim_n x_n^*(x)$ exists. The reader can show that this defines a linear functional $L : \mathcal{X} \to \mathbb{F}$; we want to show that L is bounded and that $\|x_n^* - L\| \to 0$. Since any Cauchy sequence in a normed space is uniformly bounded (Exercise 1.3.1.), $M = \sup_n \|x_n^*\| < \infty$. Therefore for any x in $\mathcal{X}$ we have that $|x_n^*(x)| \leq M\|x\|$ for all $n \geq 1$; it follows that $|L(x)| \leq M\|x\|$ and so $L \in \mathcal{X}^*$. Also if $\epsilon > 0$ and $x \in \operatorname{ball}\mathcal{X}$, then for all $n, m \geq 1$, $|L(x) - x_n^*(x)| \leq |L(x) - x_m^*(x)| + |x_m^*(x) - x_n^*(x)| \leq |L(x) - x_m^*(x)| + \|x_m^* - x_n^*\|$. There is an integer N such that for $n, m \geq N$, $\|x_m^* - x_n^*\| < \epsilon$ and hence $|L(x) - x_n^*(x)| \leq |L(x) - x_m^*(x)| + \epsilon$ for any $m \geq N$. Letting $m \to \infty$ shows that $|L(x) - x_n^*(x)| \leq \epsilon$ for all $n \geq N$ and all x in $\operatorname{ball}\mathcal{X}$; that is $\|L - x_n^*\| \leq \epsilon$ for $n \geq N$. ■

It is worth emphasizing that the preceding proposition shows that even though $\mathcal{X}$ is not assumed to be a Banach space, its dual space is.

We will spend considerable time and effort in this book determining $\mathcal{X}^*$ for a variety of normed spaces. Why? As it turns out there is a rich theory about dual spaces that says that under certain assumptions on a subset A of a normed space $\mathcal{X}$, if we prove results about the action of every bounded linear functional x^* on A, then we can conclude that the results hold uniformly over A – that is, in relation to the topology defined by the norm. That is a powerful technique, analogous to a statement that if something holds pointwise it holds uniformly; though in this abstract formulation, such a principle surely seems quite distant.

Exercises.

(1) If X is a compact metric space and $L : C_b(X) \to \mathbb{F}$ is defined as in Example 1.5.2(b), show that $\|L\| = 1$.

(2) If $L : \ell^1 \to \mathbb{F}$ is defined as in Example 1.5.2(c), show that $\|L\| = \sup_n |b_n|$.

(3) Let $\mathcal{X} = C([0,1])$ but give it the norm $\|f\| = \int_0^1 |f(t)|dt$. (See Exercise 1.3.9.) Define $L : \mathcal{X} \to \mathbb{R}$ by $L(f) = f(\frac{1}{2})$ and show that L is an unbounded linear functional.

(4) Let $\mathcal{P}$ be the vector space of all polynomials with real coefficients and define $\|p\| = \max\{|p(x)| : 0 \leq x \leq 1\}$. (a) Show that $\|\cdot\|$ is a norm on $\mathcal{P}$. (b) If $L : \mathcal{P} \to \mathbb{R}$ is defined by $L(p) = p'(0)$, show that L is a linear functional on $\mathcal{P}$ that is not bounded.

(5) This exercise outlines the construction of an unbounded linear functional on a Banach space. Consider the Banach space c_0 and for each $n \geq 1$ let e_n denote the sequence with 1 in the n-th place and zeros elsewhere. Let $\{s_j : j \in J\}$ be vectors in c_0 such that $\mathcal{S} = \{e_n : n \geq 1\} \cup \{s_j : j \in J\}$ is an algebraic basis for c_0. That is, every element of c_0 is the linear combination of a finite number of elements from $\mathcal{S}$. Define $L : c_0 \to \mathbb{F}$ by

$$L\left(\sum_n a_n e_n + \sum_{j \in J} b_j s_j\right) = \sum_n n a_n e_n$$

for all collections of finitely non-zero scalars $\{a_n\} \cup \{b_j\}$. Show that L is an unbounded linear functional on c_0.

(6) Let $\mathcal{X}$ be a normed space and let $\mathcal{M}$ be a dense vector subspace of $\mathcal{X}$. If $L : \mathcal{M} \to \mathbb{F}$ is a bounded linear functional, show that there is a unique bounded linear functional $\widetilde{L} : \mathcal{X} \to \mathbb{F}$ such that $\widetilde{L}(x) = L(x)$ for all x in $\mathcal{M}$ and show that $\|\widetilde{L}\| = \|L\|$.

Chapter 3

A Hilbert Space Interlude

Here we just peek behind the curtain of Hilbert space theory, though in this chapter's last section we will see a fundamental and crucial result of the subject, the Riesz Representation Theorem. In fact it is this theorem that necessitates the insertion of this chapter this early in the book. The idea here is to do just enough to facilitate the completion of measure theory that occurs in the following chapter. Later in Chapter 5 we'll draw back that curtain and get a more complete view of the subject.

3.1. Introduction to Hilbert space

Here we encounter the first place where we must discuss complex numbers. We will make a short excursion through this topic here, but we will increasingly be involved with the complex numbers as we go on in this book – complex measures, functions, and normed spaces. Be prepared. The reader should feel free to consult Wikipedia or the first section of my book on complex variables [**7**] for a discussion of $\mathbb{C}$ if (s)he is not conversant with the topic. What follows is just a short excursion.

We remind the reader that a complex number is $z = x + iy$, where x and y are real numbers, called the *real part* and *imaginary part* of z, respectively. When z is so expressed, we define the complex conjugate of z as $\bar{z} = x - iy$. It follows that $x = \frac{1}{2}(z + \bar{z})$ and $y = \frac{1}{2i}(z - \bar{z})$. Denote this by $x = \operatorname{Re} z$ and $y = \operatorname{Im} y$. We also know that if $z \in \mathbb{C}$, $z\bar{z} = |z|^2 = x^2 + y^2$. When $z \neq 0$, define $\operatorname{sign} z = z/|z|$; let $\operatorname{sign} 0 = 1$. (Note that this is consistent with the definition of sign(f) in §2.7.) We have that

$$|\operatorname{sign} z| = 1, \quad z = |z| \operatorname{sign} z, \quad z\,\overline{\operatorname{sign} z} = |z|$$

We will want to treat the scalar field $\mathbb{F}$, which could be the real or complex numbers. For a in $\mathbb{F}$ we define $\bar{a} = a$ when $\mathbb{F} = \mathbb{R}$ and $\bar{a}$ is the complex conjugate of a when $\mathbb{F} = \mathbb{C}$. In this section we will need to incorporate this "conjugate" into the concepts we introduce, and it is convenient to do this even when the underlying scalar field is the real numbers. Indeed, we will see later in §5.4 some results that hold when the scalars are complex but fail otherwise. This will start a process where we migrate to always having our scalars be complex.

3.1.1. Definition. For a vector space $\mathcal{H}$ over $\mathbb{F}$ an *inner product* is a function $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{F}$ satisfying the following for x, y, z belonging to $\mathcal{H}$ and a, b in $\mathbb{F}$:

(a) $\langle ax + by, z \rangle = a\langle x, z\rangle + b\langle y, z\rangle$;

(b) $\langle x, ay + bz\rangle = \bar{a}\langle x, y\rangle + \bar{b}\langle x, z\rangle$;

(c) $\langle x, x\rangle \geq 0$ with equality if and only if $x = 0$;

(d) $\langle x, y\rangle = \overline{\langle y, x\rangle}$.

We might also make it clear that when $\mathbb{F} = \mathbb{C}$ and we say that an element a in $\mathbb{C}$ satisfies $a \geq 0$, we mean that $a \in \mathbb{R} \subseteq \mathbb{C}$ and $a \geq 0$. Note that by taking $a = b = 0$ in (b) we get that for all x, $\langle x, 0\rangle = 0$; similarly $\langle 0, x\rangle = 0$ for all x.

3.1.2. Example. (a) For $x, y \in \mathbb{F}^d$, if $\langle x, y\rangle = \sum_{k=1}^d x_k \overline{y_k}$, then this is an inner product on $\mathbb{F}^d$. We might note that if $a_1, \ldots, a_d$ are strictly positive real numbers, then $\langle x, y\rangle = \sum_{k=1}^d a_k x_k \overline{y_k}$ also defines an inner product on $\mathbb{F}^d$.

(b) Let $\mathcal{H}$ denote the collection of all finitely non-zero sequences of scalars in $\mathbb{F}$; that is, $\{x_n\} \in \mathcal{H}$ if $x_n \in \mathbb{F}$ for all n and $x_n = 0$ except for a finite number of coordinates n. We can make $\mathcal{H}$ into a vector space by defining addition and scalar multiplication coordinatewise. If we define $\langle x, y\rangle = \sum_{n=1}^\infty x_n \overline{y_n}$, then this is an inner product on $\mathcal{H}$.

(c) If $(X, \mathcal{A}, \mu)$ is a measure space and $\mathcal{H} = L^2(\mu)$, define $\langle f, g\rangle = \int f\bar{g}d\mu$ for f, g in $L^2(\mu)$. It is easy to check that this defines an inner product.

(d) Let X be a compact metric space, let $\mathcal{H} = C(X)$, and for a positive regular Borel measure on X define $\langle f, g\rangle = \int f(x)\overline{g(x)}d\mu(x)$; then this is an inner product on $C(X)$.

The next result will be referred to as the CBS Inequality.

3.1.3. Theorem (Cauchy–Bunyakovsky–Schwarz Inequality[1]). *If $\mathcal{H}$ is a vector space with an inner product, then*

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle$$

for all x, y in $\mathcal{H}$. Moreover equality occurs if and only if x, y are linearly dependent.

Proof. Notice that for x, y in $\mathcal{H}$ and a in $\mathbb{F}$ we have that

$$\begin{aligned} 0 &\leq \langle x - ay, x - ay\rangle \\ &= \langle x, x\rangle - a\langle y, x\rangle - \bar{a}\langle x, y\rangle + |a|^2\langle y, y\rangle \end{aligned}$$

Consider the polar representation of the number $\langle y, x\rangle$: $\langle y, x\rangle = \beta e^{i\theta}$ with $\beta \geq 0$. If t is an arbitrary real number and we substitute $a = e^{-i\theta}t$ in the above inequality, then we get

$$\begin{aligned} 0 &\leq \langle x, x\rangle - e^{-i\theta}t\beta e^{i\theta} - e^{i\theta}t\beta e^{-i\theta} + t^2\langle y, y\rangle \\ &= \langle x, x\rangle - 2\beta t + t^2\langle y, y\rangle \\ &= \gamma - 2\beta t + \alpha t^2 \equiv q(t) \end{aligned}$$

where $\gamma = \langle x, x\rangle$ and $\alpha = \langle y, y\rangle$. Thus $q(t)$ is a quadratic polynomial with real coefficients and $q(t) \geq 0$ for all $t \geq 0$. So the graph of $q(t)$ stays above the x-axis except that it might be tangent at a single point; that is,

[1]Augustin Louis Cauchy was born in Paris in August 1789, a month after the storming of the Bastille. He was educated in engineering and his first job was in 1810 working on the port facilities at Cherbourg in preparation for Napoleon's contemplated invasion of England. In 1812 he returned to Paris and his energies shifted toward mathematics. His contributions were monumental, with a plethora of results in analysis bearing his name. His collected works fill 27 published volumes. As a human being he left much to be desired. He was highly religious with a totally dogmatic personality, often treating others with dismissive rudeness. Two famous examples were his treatment of Abel and Galois, where he refused to consider their monumental works. Both suffered an early death. Perhaps better treatment by Cauchy would have given them the recognition that would have resulted in a longer life and a productive career to the betterment of mathematics; we'll never know. He had two doctoral students, one of which was Bunyakovsky. Cauchy died in 1857 in Sceaux near Paris.

Viktor Yakovlevich Bunyakovsky was born in 1804 in what is presently Vinnitsa in the Ukraine. In 1825 he received a doctorate in Paris working under Cauchy. The next year he went to St Petersburg, where he made his career. In 1859 this result appeared in his monograph on integral inequalities; he seems to be the first to have discovered this, 25 years before Schwarz. He made contributions to number theory, geometry, and applied mathematics. He died in St Petersburg in 1889.

Hermann Amandus Schwarz was born in 1843 in Hermsdorf, Silesia, presently in Poland. He began his studies at Berlin in Chemistry, but switched to mathematics and received his doctorate in 1864 under the direction of Weierstrass. He held positions at Halle, Zurich, Göttingen, and Berlin. His work centered on various geometry problems that were deeply connected to analysis. This included work on surfaces and conformal mappings in analytic function theory, any student of which will see his name in prominence. He died in Berlin in 1921.

$q(t) = 0$ has at most one real root. From the quadratic formula we get that $0 \geq 4\beta^2 - 4\alpha\gamma$. Therefore

$$0 \geq \beta^2 - \alpha\gamma = |\langle x, y\rangle|^2 - \langle x, x\rangle\langle y, y\rangle$$

proving the inequality.

The proof of the necessary and sufficient condition for equality consists of a careful analysis of the above inequalities to see the implications of such an equality. The reader is urged to carry out this argument. ■

Observe that when we examine the space $L^2(\mu)$, the CBS Inequality is the same as Hölder's Inequality for $p = 2 = q$. Also note that in the proof of the CBS Inequality the property of an inner product that $\langle x, x\rangle = 0$ implies that $x = 0$ is used only in deriving the condition for equality. A function that has all the properties of an inner product save this one is called a *semi-inner product.* Therefore when we have a semi-inner product the CBS Inequality holds and a necessary and sufficient condition for equality is that there exist scalars a, b, not both 0, such that $\langle ax + by, ax + by\rangle = 0$.

3.1.4. Corollary. *If $\langle \cdot, \cdot \rangle$ is an inner product on $\mathcal{H}$, then $\|x\| = \sqrt{\langle x, x\rangle}$ defines a norm on $\mathcal{H}$.*

Proof. If $x, y \in \mathcal{H}$, then using the CBS Inequality we get

$$\begin{aligned}
\|x + y\|^2 &= \langle x + y, x + y\rangle \\
&= \|x\|^2 + \langle y, x\rangle + \langle x, y\rangle + \|y\|^2 \\
&= \|x\|^2 + 2\mathrm{Re}\,\langle x, y\rangle + \|y\|^2 \\
&\leq \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \\
&\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\
&= (\|x\| + \|y\|)^2
\end{aligned}$$

Therefore the triangle inequality holds. The remainder of the proof that $\|\cdot\|$ defines a norm is straightforward. ■

Let's underline that in the course of the preceding proof we established that for all x, y in $\mathcal{H}$ we have

3.1.5 $$\|x + y\|^2 = \|x\|^2 + 2\mathrm{Re}\,\langle x, y\rangle + \|y\|^2$$

We will find this useful in the future; it is called the *polar identity.*

When we have that $\mathcal{H} = L^2(\mu)$ as in Example 3.1.2(c), the norm defined by the inner product is exactly $\|\cdot\|_2$ as defined in (2.7.2) for $p = 2$. Now that we have a normed space we can use what we know about that subject, though because of the way this norm arises we will see it will have extra structure.

3.1.6. Definition. A *Hilbert*[2] *space* is a vector space with an inner product such that with respect to the norm defined by this inner product it is a complete metric space.

So every Hilbert space is a Banach space. For the remainder of this chapter $\mathcal{H}$ is a Hilbert space.

3.1.7. Example. (a) $\mathbb{F}^d$ with any of the inner products defined in Example 3.1.2(a) is a Hilbert space.

(b) $L^2(\mu)$ is a Hilbert space.

(c) We define ℓ^2 as the space of all sequences $\{x_n\}$ such that $\sum_{n=1}^{\infty} |x_n|^2 < \infty$. If we consider the measure space $(\mathbb{N}, 2^{\mathbb{N}}, \mu)$, where μ is counting measure, then ℓ^2 is precisely $L^2(\mu)$, and so ℓ^2 is another example of a Hilbert space.

(d) The inner product spaces in parts (b) and (d) of (3.1.2) are not Hilbert spaces since they fail to be complete.

When we are presented with an incomplete inner product space, we can complete it just as we did for a normed space in Theorem 1.3.8. This result is stated here, though its proof will not be given as it follows routinely by mimicking the proof of (1.3.8).

3.1.8. Theorem. *If $\mathcal{X}$ is a vector space over $\mathbb{F}$ that has an inner product and $\mathcal{H}$ is its completion as a normed space, then there is an inner product on $\mathcal{H}$ such that $\langle x, y\rangle_{\mathcal{H}} = \langle x, y\rangle_{\mathcal{X}}$ for all x, y in $\mathcal{X}$ and the norm on $\mathcal{H}$ is defined by this inner product.*

What is the completion of the spaces in Examples 3.1.2(b) and 3.1.2(d)? The completion of the first of these is exactly ℓ^2. The completion of the space in Example 3.1.2(d) is $L^2(\mu)$.

---

[2]David Hilbert was born in 1862 in Königsberg, Prussia, now Kaliningrad, Russia. He received his doctorate from the University of Königsberg in 1885. He continued there on the faculty until 1895 when he was appointed to a chair in mathematics at Göttingen, where he remained for the rest of his life. His first major work at Göttingen was the proof of the basis theorem in a part of algebra called invariant theory. Previously the world's leader in this subject, Paul Albert Gordon, had proved a special case by horrendous computations. Gordon was highly critical of what he called Hilbert's formal approach. Needless to say, history smiled on this formal approach and Gordon receded into a historical note. (Except that Emmy Noether was his student.) This so-called formal approach turned out to characterize Hilbert's method in mathematics. He encountered a problem, stripped it down to its essence, recast it in his own framework, and proceeded to advance the entire subject in terms that resonate with the present generation of mathematicians. One of his most famous contributions was the presentation of 23 problems at the 1900 International Congress of Mathematicians in Paris. Many remain unsolved and those that were solved brought instant fame to the solver. They set the research direction for much of twentieth century mathematics and their influence continues. Hilbert contributed to many areas of mathematics besides algebra, including several parts of analysis and geometry. He was one of the subject's titans. In a 1930 address in Königsberg to accept being made an honorary citizen he said, "We must know, we shall know." A fitting epitaph. Hilbert died in 1943 in Göttingen.

Exercises.

(1) Carry out the details of the proof of the necessary and sufficient condition for equality in the CBS Inequality.

(2) Let w be a sequence of strictly positive real numbers and define $\ell^2(w)$ to be the set of all sequences $\{x_n\}$ of numbers from $\mathbb{F}$ such that $\sum_{n=1}^{\infty} w_n|x_n|^2 < \infty$. Show that $\langle\{x_n\},\{y_n\}\rangle = \sum_{n=1}^{\infty} w_n x_n \overline{y_n}$ defines an inner product on $\ell^2(w)$ and that this is a Hilbert space. Is there a measure μ such that $\ell^2(w)$ is naturally identified with $L^2(\mu)$?

(3) Generalize the definition of ℓ^2 as follows. Let I be any set and for a function $x : I \to \mathbb{F}$ say that $\sum_{i\in I} x(i) = a$ if for every $\epsilon > 0$ there is a finite subset J_0 of I such that $|a - \sum_{i\in J} x(i)| < \epsilon$ for every finite set J containing J_0. (a) Show that if such an x converges, there are at most a countable number of i in I such that $x(i) \neq 0$. Define $\ell^2(I)$ to be the set of all functions $x : I \to \mathbb{F}$ such that $\sum_{i\in I} |x(i)|^2 < \infty$. For x, y in $\ell^2(I)$ let $\langle x, y\rangle = \sum_{i\in I} x(i)\overline{y(i)}$. (b) Show that this defines an inner product. (c) Find a measure space $(X, \mathcal{A}, \mu)$ such that $\ell^2(I)$ is naturally identified with $L^2(\mu)$.

3.2. Orthogonality

The subject of this section is the very thing that makes the structure of a Hilbert space so transparent; it has a concept of angle. We can define the angle between two non-zero vectors x and y as $\arccos[\langle x, y\rangle/\|x\|\|y\|]$. The ambiguity of the function arccos of course lends ambiguity to this process. However we will not pursue this concept of the angle and instead focus on the concept of two vectors forming a right angle.

3.2.1. Definition. If $x, y \in \mathcal{H}$, say that x and y are *orthogonal*, in symbols $x \perp y$, if $\langle x, y\rangle = 0$. If A and B are two subsets of $\mathcal{H}$, say that these sets are orthogonal if $x \perp y$ for any x in A and any y in B. If we have a set of vectors S in $\mathcal{H}$, we say that the vectors in S are *pairwise orthogonal* if $x \perp y$ for any two distinct elements of S.

Note that in Euclidean space (3.1.2(a)) this coincides with the traditional definition of orthogonality.

3.2.2. Example. Let $\mathcal{H} = L^2([0,1],\lambda)$, where λ is Lebesgue measure. For n in $\mathbb{Z}$ let $e_n(t) = \exp(2\pi i n t)$. It follows that $e_n \perp e_m$ whenever $n \neq m$.

3.2.3. Theorem (Pythagorean Theorem)**.** *If $x_1, \dots, x_n$ are pairwise orthogonal vectors in $\mathcal{H}$, then*

$$\|x_1 + \cdots + x_n\|^2 = \|x_1\|^2 + \cdots + \|x_n\|^2$$

Proof. We prove the case where $n = 2$; the general case proceeds by induction. If $x \perp y$, then the polar identity (3.1.5) implies that $\|x + y\|^2 = \|x\|^2 + 2\text{Re}\,\langle x, y\rangle + \|y\|^2 = \|x\|^2 + \|y\|^2$. ■

We note that if $x \perp y$, then $x \perp -y$ and so we also have that $\|x - y\|^2 = \|x\|^2 + \|y\|^2$.

3.2.4. Theorem (Parallelogram Law)**.** *If $x, y \in \mathcal{H}$, then*

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

Proof. The proof consists in looking at the polar identity in two different ways and then adding both:

$$\|x + y\|^2 = \|x\|^2 + 2\text{Re}\,\langle x, y\rangle + \|y\|^2$$
$$\|x - y\|^2 = \|x\|^2 - 2\text{Re}\,\langle x, y\rangle + \|y\|^2$$

Now add. ■

3.2.5. Definition. If $\mathcal{X}$ is any vector space and $A \subseteq \mathcal{X}$, say that A is *convex* if $tx + (1 - t)y \in A$ whenever $x, y \in A$ and $0 \leq t \leq 1$.

Since $\{tx + (1 - t)y : 0 \leq t \leq 1\}$ is the straight line segment joining x and y, this means that A is convex when the line joining any two points in A is wholly contained in A. It is easy to draw pictures of subsets of the plane that are and are not convex. Also observe that the convex subsets of $\mathbb{R}$ are precisely the intervals.

Any open or closed ball in a normed space is convex. If $\mathcal{X}$ is any vector space, then any linear subspace is convex and the intersection of any collection of convex subsets is again convex. So if $A \subseteq \mathcal{X}$, we can define the *convex hull* of A as the intersection of all convex subsets of $\mathcal{X}$ that contain A; denote this by $\text{co}(A)$. Similarly, when $\mathcal{X}$ is a normed space we define the *closed convex hull* of A, denoted by $\overline{\text{co}}(A)$, as the closure of $\text{co}(A)$. See Exercise 7.

The next result is one the salient facts in the structure of Hilbert space. It says that the distance from a point to a closed convex subset of $\mathcal{H}$ is always attained at a unique point.

3.2.6. Theorem. *If $x \in \mathcal{H}$ and K is a closed convex subset of $\mathcal{H}$, then there is a unique x_0 in K such that*

$$\|x - x_0\| = \text{dist}\,(x, K) = \inf\{\|x - y\| : y \in K\}$$

Proof. Replacing K by $K - x$, we see that without loss of generality we may assume that $x = 0$. (Verify!) By definition there is a sequence $\{y_n\}$ in

K such that $\|y_n\| \to \text{dist}\,(0, K) = d$. The Parallelogram Law implies that

$$\left\|\frac{y_n - y_m}{2}\right\|^2 = \frac{1}{2}(\|y_n\|^2 + \|y_m\|^2) - \left\|\frac{y_n + y_m}{2}\right\|^2$$

Now we use the convexity of K to conclude that $\frac{1}{2}(y_n + y_m) \in K$ and so that $\|\frac{1}{2}(y_n + y_m)\|^2 \geq d^2$. If $\epsilon > 0$ we can choose N such that for $n \geq N$, $\|y_n\|^2 < d^2 + \frac{1}{4}\epsilon^2$. Therefore for $n, m \geq N$

$$\left\|\frac{y_n - y_m}{2}\right\|^2 < \frac{1}{2}\left(2d^2 + \frac{1}{2}\epsilon^2\right) - d^2 = \frac{1}{4}\epsilon^2$$

That is, $\|y_n - y_m\| < \epsilon$ for $n, m \geq N$; thus $\{y_n\}$ is a Cauchy sequence. Since K is closed there is an x_0 in K such that $y_n \to x_0$. Also $d \leq \|x_0\| \leq \|x_0 - y_n\| + \|y_n\| \to d$. This proves existence.

To establish uniqueness of the point x_0, suppose there is another point y_0 in K with $\|y_0\| = d$. Since $\frac{1}{2}(x_0 + y_0) \in K$, $d \leq \|\frac{1}{2}(x_0 + y_0)\| \leq \frac{1}{2}(\|x_0\| + \|y_0\|) = d$, so that $d = \|\frac{1}{2}(x_0+y_0)\|$. From here we get that the Parallelogram Law implies that $d^2 = \|\frac{1}{2}(x_0+y_0)\|^2 = d^2 - \|\frac{1}{2}(x_0-y_0)\|$, whence $x_0 = y_0$. ■

Now let's see what happens when the convex set in the preceding theorem is a linear subspace.

3.2.7. Theorem. *If $\mathcal{M}$ is a closed linear subspace of $\mathcal{H}$, $x \in \mathcal{H}$, and x_0 is the unique point in $\mathcal{M}$ such that $\|x - x_0\| = \text{dist}\,(x, \mathcal{M})$, then $x - x_0 \perp \mathcal{M}$. Conversely if $x_0 \in \mathcal{M}$ such that $x - x_0 \perp \mathcal{M}$, then x_0 is the unique point in $\mathcal{M}$ with $\|x - x_0\| = \text{dist}\,(x, \mathcal{M})$.*

Proof. Let $x_0 \in \mathcal{M}$ such that $\|x - x_0\| = \text{dist}\,(x, \mathcal{M})$. If y is an arbitrary vector in $\mathcal{M}$, then $x_0 + y \in \mathcal{M}$ and so $\|x - x_0\|^2 \leq \|x - (x_0 + y)\|^2 = \|(x - x_0) - y\|^2 = \|x - x_0\|^2 - 2\text{Re}\,\langle x - x_0, y\rangle + \|y\|^2$. Thus

$$2\text{Re}\,\langle x - x_0, y\rangle \leq \|y\|^2$$

for any y in $\mathcal{M}$. Now we perform a standard trick. Fix a y in $\mathcal{M}$ and suppose $\langle x - x_0, y\rangle = re^{i\theta}$ with $r \geq 0$. Substitute $te^{i\theta}y$ for y in the preceding inequality. This gives us $2\text{Re}\,[te^{-i\theta}re^{i\theta}] \leq t^2\|y\|^2$ or $2tr \leq t^2\|y\|^2$; this is valid for all t. Letting $t \to 0$ shows that $r = 0$; that is, $\langle x - x_0, y\rangle = 0$ so that $x - x_0 \perp y$.

For the converse assume that $x_0 \in \mathcal{M}$ such that $x - x_0 \perp \mathcal{M}$. Thus for any y in $\mathcal{M}$ we have that $x - x_0 \perp x_0 - y$ and so $\|x - y\|^2 = \|(x - x_0) + (x_0 - y)\|^2 = \|x - x_0\|^2 + \|x_0 - y\|^2 \geq \|x - x_0\|^2$. Therefore $\|x - x_0\| = \text{dist}\,(x, \mathcal{M})$. ■

If $A \subseteq \mathcal{H}$ and $A \neq \emptyset$, let $A^\perp = \{x \in \mathcal{H} : x \perp \mathcal{A}\}$. It is easy to see that no matter what properties the set A has, $A^\perp$ is always a closed linear subspace of $\mathcal{H}$. Also let's introduce a bit of notation that could have been introduced earlier in this book. If $\mathcal{X}$ is any normed space, the notation $\mathcal{M} \leq \mathcal{X}$ will mean that $\mathcal{M}$ is a closed linear subspace of $\mathcal{X}$. We need to make a distinction between vector subspaces of $\mathcal{X}$ that are closed and those that are not. A *linear manifold* or *submanifold* is a vector subspace of $\mathcal{X}$ that is not necessarily closed; a *subspace* is a closed linear manifold.

In light of Theorem 3.2.7, whenever $\mathcal{M} \leq \mathcal{H}$ we can define a function $P : \mathcal{H} \to \mathcal{H}$ as follows: for any x in $\mathcal{H}$, Px is the unique vector in $\mathcal{M}$ with $x - Px \perp \mathcal{M}$. Note that when $x \in \mathcal{M}$, $Px = x$. This function will be called the *orthogonal projection* of $\mathcal{H}$ onto $\mathcal{M}$, or simply the *projection* of $\mathcal{H}$ onto $\mathcal{M}$. Sometimes this is designated by $P = P_{\mathcal{M}}$.

Whenever $\mathcal{X}$ and $\mathcal{Y}$ are vector spaces and $T : \mathcal{X} \to \mathcal{Y}$ is a linear transformation, the *kernel* and *range* of T are defined by

$$\ker T = \{x \in \mathcal{X} : Tx = 0\}$$
$$\operatorname{ran} T = \{Tx : x \in \mathcal{X}\}$$

We note that $\ker T$ is a linear manifold in $\mathcal{X}$ and $\operatorname{ran} T$ is a linear manifold in $\mathcal{Y}$.

3.2.8. Theorem. *If $\mathcal{M} \leq \mathcal{H}$ and P is the orthogonal projection of $\mathcal{H}$ onto $\mathcal{M}$, then the following statements are true.*

(a) *P is a linear transformation.*

(b) *$\|Px\| \leq \|x\|$ for all x in $\mathcal{H}$.*

(c) *$P^2 = P$ (here $P^2 = P \circ P$, the composition of P with itself).*

(d) *$\ker P = \mathcal{M}^\perp$ and $\operatorname{ran} P = \mathcal{M}$.*

Proof. Keep in mind that for every x in $\mathcal{H}$, $x - Px \perp \mathcal{M}$ and $\|x - Px\| = \operatorname{dist}(x, \mathcal{M})$.

(a) Let $x_1, x_2 \in \mathcal{H}$ and $a_1, a_2 \in \mathbb{F}$. If $y \in \mathcal{M}$, then $\langle [a_1x_1 + a_2x_2] - [a_1Px_1 + a_2Px_2], y\rangle = a_1\langle x_1 - Px_1, y\rangle + a_2\langle x_2 - Px_2, y\rangle = 0$. Since $a_1Px_1 + a_2Px_2 \in \mathcal{M}$, the uniqueness statement in Theorem 3.2.7 implies $P(a_1x_1 + a_2x_2) = a_1Px_1 + a_2Px_2$ and so P is linear.

(b) If $x \in \mathcal{H}$, then $x - Px \perp Px$ since $x - Px \in \mathcal{M}^\perp$ while $Px \in \mathcal{M}$. Therefore $\|x\|^2 = \|x - Px\|^2 + \|Px\|^2 \geq \|Px\|^2$.

(c) Since $Py = y$ for any y in $\mathcal{M}$, for any x in $\mathcal{H}$ we have that $P(Px) = Px$.

(d) If $Px = 0$, then $x = x - Px \in \mathcal{M}^\perp$. Conversely if $x \in \mathcal{M}^\perp$, then 0 is the unique point in $\mathcal{M}$ such that $x - 0 = x \perp \mathcal{M}$; thus $Px = 0$. Thus $\ker P = \mathcal{M}^\perp$. The fact that $P\mathcal{H} = \mathcal{M}$ is clear. ■

3.2.9. Corollary. *If $\mathcal{M} \leq \mathcal{H}$, then $\mathcal{M} = (\mathcal{M}^\perp)^\perp$.*

Proof. For this proof we use Exercise 4: if $P = P_{\mathcal{M}}$, then $I - P$ is the projection onto $\mathcal{M}^{\perp}$, where I denotes the identity linear transformation. By the preceding theorem, $(\mathcal{M}^{\perp})^{\perp} = \ker(I - P)$. But $0 = (I - P)x$ if and only if $Px = x$; that is, if and only if $x \in \operatorname{ran} P$. Thus $(\mathcal{M}^{\perp})^{\perp} = \operatorname{ran} P = \mathcal{M}$. ■

3.2.10. Corollary. *If $A \subseteq \mathcal{H}$, then $(A^{\perp})^{\perp}$ is the closed linear span of A in $\mathcal{H}$.*

For a discussion of the closed linear span of a set, the reader should look at Exercise 1.3.2. The proof of this corollary as well as the next one is straightforward.

3.2.11. Corollary. *If $\mathcal{Y}$ is a linear submanifold of $\mathcal{H}$, then $\mathcal{Y}$ is dense in $\mathcal{H}$ if and only if $\mathcal{Y}^{\perp} = (0)$.*

The preceding corollaries are quite powerful tools. For example, this last corollary says something very deep: that a topological property, the density of $\mathcal{Y}$, is equivalent to an algebraic one, $\mathcal{Y}^{\perp} = (0)$. Whenever two such disparate conditions are equivalent, the force is with us.

Exercises.

(1) Let λ be normalized area measure on the closed unit disk $\operatorname{cl}\mathbb{D}$ in the complex plane. That is, λ is the Radon measure on $\operatorname{cl}\mathbb{D}$ associated with the positive linear functional $f \mapsto \pi^{-1}\int\int_{\mathbb{D}} f(x+iy)dxdy$ and $\lambda(\mathbb{D}) = 1$. For $n, m \geq 0$, let $f_{nm}(z) = z^n \bar{z}^m$. When is $f_{nm} \perp f_{jk}$? (Hint: Use polar coordinates.)

(2) Let x, y be two linearly independent vectors in $\mathcal{H}$ with $1 = \|x\| = \|y\|$. Show that for $0 < t < 1$, $\|ty + (1-t)x\| < 1$. What does this say about the geometry of the boundary of $\operatorname{ball}\mathcal{H}$? What happens if instead of the Hilbert space $\mathcal{H}$, we look at $\mathbb{F}^2$ with the norm $\|(x, y)\| = |x| + |y|$? (Also see the next exercise.)

(3) Consider $\mathbb{F}^2$ with the norm $\|(x, y)\| = |x| + |y|$. (a) Use a compactness argument to show that if K is any closed subset of $\mathbb{F}^2$ and $x \in \mathbb{F}^2$, then there is a point x_0 in K with $\|x - x_0\| = \operatorname{dist}(x, K)$. (b) Give an example of a closed convex subset K of $\mathbb{F}^2$ and a point x in $\mathbb{F}^2$ such that the distance from x to K is not attained at a unique point. (Also see the end of the preceding exercise.)

(4) If $\mathcal{M} \leq \mathcal{H}$ and P is the projection onto $\mathcal{M}$, show that $I - P$ is the projection onto $\mathcal{M}^{\perp}$.

(5) Let $\mathcal{M}, \mathcal{N} \leq \mathcal{H}$, $P = P_{\mathcal{M}}$, $Q = P_{\mathcal{N}}$. If $PQ = QP$, show that $PQ = P_{\mathcal{M}\cap\mathcal{N}}$. In this case, what is $\ker PQ$?

(6) Let X be a compact metric space and give $C(X)$ its usual supremum norm. Show that this norm is not given by an inner product

(so that $C(X)$ is not a Hilbert space) by showing that the norm does not satisfy the Parallelogram Law.

(7) If $\mathcal{X}$ is a normed space, show that $\overline{\text{co}}(A)$ is the intersection of all closed convex subsets of $\mathcal{X}$ that contain A.

(8) If F is a closed subset of the normed space $\mathcal{X}$, show that F is convex if and only if for any x, y in F, $\frac{1}{2}(x+y) \in F$. Give a counterexample to this statement if F is not assumed to be closed.

(9) Show that if $\mathcal{H} = \mathbb{F}^d$ with the usual inner product and if K is any closed subset of $\mathcal{H}$, then for any x in $\mathcal{H}$ there is a point x_0 in K with $\|x - x_0\| = \text{dist}\,(x, K)$. (Later we'll show that this is true in any Hilbert space. Of course since we have not assumed that K is convex, the point x_0 is not necessarily unique.)

(10) Is ℓ^1 a Hilbert space? Why?

3.3. The Riesz Representation Theorem

For Hilbert spaces it is possible to give a simple representation of all the bounded linear functionals on it and the proof of this representation is rather simple. As we will see this is not true of all Banach spaces. For example, when X is a compact metric space, the representation of the elements of $C(X)^*$ will occupy a lot of our energy. In a sense it is our motivation for establishing measure theory. (This motivation is not historically true – measure theory preceded a consideration of bounded linear functionals. The original motivation was to be able to fully explore the Fourier transform and trigonometric series.)

I am certain that at this point the student will not fully appreciate the importance of representing linear functionals on a Banach or Hilbert space. Why should (s)he? So I'll have to ask the reader to be patient until we see the power of this illustrated by applications. Perhaps the reader's patience can be increased if (s)he reflects on the name of one of the subjects that is partially covered in this book: Functional Analysis. In fact one application of the next theorem on representing linear functionals will appear in the first section of the next chapter. Another application of linear functionals in a Banach space setting will take place in §8.5 when we use them to prove the Stone–Weierstrass Theorem, which generalizes the result that the polynomials are uniformly dense in $C[0, 1]$.

Start by realizing that if $x_0 \in \mathcal{H}$ and $L : \mathcal{H} \to \mathbb{F}$ is defined by $L(x) = \langle x, x_0 \rangle$, then L is a bounded linear functional on $\mathcal{H}$. The fact that it is linear is a direct consequence of the definition of an inner product; the fact that it is bounded follows from the CBS Inequality: $|L(x)| = |\langle x, x_0 \rangle| \leq \|x\|\|x_0\|$.

So $\|L\| \leq \|x_0\|$. In fact, if $x = \|x_0\|^{-1}x_0$, then $\|x\| = 1$ and $L(x) = \|x_0\|$; hence $\|L\| = \|x_0\|$.

3.3.1. Theorem (Riesz[3] Representation Theorem)**.** *If $\mathcal{H}$ is a Hilbert space and $L : \mathcal{H} \to \mathbb{F}$ is a continuous linear functional, then there is a unique vector x_0 in $\mathcal{H}$ such that $L(x) = \langle x, x_0 \rangle$ for all x in $\mathcal{H}$ and $\|L\| = \|x_0\|$.*

Proof. The theorem is trivially true when $L \equiv 0$; so assume that L is not the zero linear functional and let $\mathcal{M} = \ker L$. Since L is continuous, its kernel is a closed linear subspace of $\mathcal{H}$; since $L \neq 0$, $\mathcal{M} \neq \mathcal{H}$. Thus $\mathcal{M}^\perp \neq (0)$ and we can find a vector y_0 in $\mathcal{M}^\perp$ such that $L(y_0) = 1$. If $x \in \mathcal{H}$ and $a = L(x)$, then $L(x - ay_0) = 0$; that is, $x - L(x)y_0 \in \mathcal{M}$. Therefore $0 = \langle x - L(x)y_0, y_0 \rangle = \langle x, y_0 \rangle - L(x)\|y_0\|^2$. Since x was arbitrary, we see that for $x_0 = \|y_0\|^{-2}y_0$, $L(x) = \langle x, x_0 \rangle$ for all x in $\mathcal{H}$. Since the proof of the norm equality was done before the statement of the theorem, we are done. ■

3.3.2. Corollary. *If $(X, \mathcal{A}, \mu)$ is a measure space and $L : L^2(\mu) \to \mathbb{F}$ is a bounded linear functional, then there is a function h in $L^2(\mu)$ such that*

$$L(f) = \langle f, h \rangle = \int f\bar{h}\, d\mu$$

for all f in $L^2(\mu)$ and $\|h\|_2 = \|L\|$.

3.3.3. Corollary. *If $L : \ell^2 \to \mathbb{F}$ is a bounded linear functional, there is a unique sequence $\{b_n\}$ in ℓ^2 such that $L(\{a_n\}) = \sum_{n=1}^\infty a_n\overline{b_n}$ for all $\{a_n\}$ in ℓ^2 and $\|L\| = \left(\sum_{n=1}^\infty |b_n|^2\right)^{\frac{1}{2}}$.*

Exercises.

(1) Suppose $\mathcal{H}$ is a Hilbert space and $\mathcal{M}$ is a closed subspace. Show that if $L : \mathcal{M} \to \mathbb{F}$ is a bounded linear functional, then there is a bounded linear functional $\widetilde{L} : \mathcal{H} \to \mathbb{F}$ such that $\widetilde{L}(x) = L(x)$ for every x in $\mathcal{M}$ and $\|\widetilde{L}\| = \|L\|$. (Later we will prove this same result in any normed space. It is called the Hahn–Banach Theorem.)

(2) If $\mathcal{H}$ is a Hilbert space, $\mathcal{M}$ is a closed subspace of $\mathcal{H}$, and $x \in \mathcal{H}$ such that $x \notin \mathcal{M}$, show that there is a bounded linear functional

[3] Frigyes Riesz was born in 1880 in Györ in present day Hungary. His younger brother, Marcel Riesz, was also a famous mathematician; indeed there is an important theorem in complex/harmonic analysis called the F. and M. Riesz Theorem. F. Riesz obtained his doctorate from the University of Budapest in 1902. He is widely considered as one of the founders of functional analysis and has additional representation theorems named after him besides the one in this section. For example see Theorem 4.3.8. He made fundamental contributions to Fourier analysis and ergodic theory and was a driving force in establishing modern mathematics in Hungary, including the founding of the János Bolyai Mathematical Institute in Szeged in 1922. He brought a high degree of artistry and elegance to mathematics. In 1945 he accepted the chair of mathematics at the University of Budapest, where he remained until his death in 1956.

L on $\mathcal{H}$ such that $L(y) = 0$ for every y in $\mathcal{M}$, $L(x) = 1$, and $\|L\| = [\operatorname{dist}(x, \mathcal{M})]^{-1}$.

(3) If $\mathcal{H}$ is a Hilbert space and $\mathcal{M}$ is a closed linear subspace, show that a necessary and sufficient condition for there to be a bounded linear functional L on $\mathcal{H}$ with $\mathcal{M} = \ker L$ is that there is a single vector x in $\mathcal{H}$ with $\mathcal{H} = \mathcal{M} \vee \mathbb{C}x = \bigvee\{\mathcal{M} \cup \{x\}\}$. (This condition can be expressed by saying that $\mathcal{M}$ has codimension 1.)