6

Further Topics

In this chapter we present a smorgasbord of treats. The sections of this chapter are, for the most part, independent of each other (the exception is that the third section makes use of the main theorem of the second section). The sections are not uniform in length or level of difficulty. They may be added as topics for lectures, or used as sources of student projects.

The first section gives a proof of the important Weierstrass approximation theorem and of its generalization due to Marshall Stone (1903–1989; USA). We offer a proof of the latter that is relatively recent [26] and has not appeared, to our knowledge, in any text. The second section presents a theorem of René Baire and gives an application to real analysis. This material is standard in a real analysis text; we include it because we like it and because we use Baire's theorem in the third section, where we prove three fundamental results of functional analysis. In the fourth section we prove the existence of a set of real numbers that is not Lebesgue measurable. The fifth section investigates contraction mappings and how these special maps can be used to solve problems in differential equations. In the penultimate section we study the algebraic structure of the space of continuous functions. In the last section we give a very brief introduction to how some of the ideas in the text are used in quantum mechanics. Some of these topics certainly belong in a functional analysis text; some probably do not. Enjoy!

6.1 The Classical Weierstrass Approximation Theorem and the Generalized Stone–Weierstrass Theorem

The classical Weierstrass approximation theorem was first proved by Weierstrass in 1885. Several different proofs have appeared since Weierstrass's (see [54], page 266). We present the proof of it due to the analyst and probabalist Sergei Bernstein (1880–1968; Ukraine) [18].[1] His proof gives a clever, and perhaps surprising, application of probability. Marshall Stone's generalization of Weierstrass's famous theorem appeared a quarter of a century later, in [118].

Recall that $C([a,b];\mathbb{R})$, or $C([a,b])$, denotes the Banach space of all continuous real-valued functions with norm $\|f\|_\infty = \sup\{|f(x)| \,\big|\, a \le x \le b\}$.

Theorem 6.1 (The Weierstrass Approximation Theorem). *The polynomials are dense in $C([a,b])$. That is, given any $f \in C([a,b])$ and an $\epsilon > 0$ there exists a polynomial $p \in C([a,b])$ such that $\|f - p\|_\infty < \epsilon$.*

Proof. First we will show that the result is true for $a = 0$ and $b = 1$. Thus, we consider $f \in C([0,1])$ and proceed to describe a polynomial that is close to f (with respect to the supremum norm). The polynomial we will use is a so-called *Bernstein polynomial*. The nth-degree Bernstein polynomial associated to f is defined by

$$p_n(x) = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k},$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is the binomial coefficient.

The idea for this definition comes from probability. Imagine a coin with probability x of getting heads. In n tosses, the probability of getting exactly k heads is thus

$$\binom{n}{k} x^k (1-x)^{n-k}.$$

If $f\left(\frac{k}{n}\right)$ dollars are paid when exactly k heads are thrown in n tosses, then the average dollar amount (after throwing n tosses very many times) paid when n tosses are made is

$$\sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

This expression is what we called $p_n(x)$.

---

[1] As an incidental historical remark, Bernstein's Ph.D. dissertation contained the first solution to Hilbert's 19th problem on elliptic differential equations.

We now show that given an $\epsilon > 0$ there exists n large enough so that $\|p_n - f\|_\infty < \epsilon$.

This should seem plausible: If n is very large, then we expect $\frac{k}{n}$ to be very close to x. We thus expect the average dollar amount paid, $p_n(x)$, to be very close to $f(x)$.

To prove that $\|p_n - f\|_\infty < \epsilon$ for sufficiently large n, we recall the binomial theorem:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}. \tag{6.1}$$

If we differentiate this with respect to x and multiply both sides by x, we get

$$nx(x+y)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k x^k y^{n-k}. \tag{6.2}$$

If instead, we differentiate twice and multiply both sides by x^2, we get

$$n(n-1)x^2(x+y)^{n-2} = \sum_{k=0}^{n} \binom{n}{k} k(k-1) x^k y^{n-k}. \tag{6.3}$$

Equations (6.1)–(6.3), with $y = 1 - x$, read

$$1 = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k}, \tag{6.4}$$

$$nx = \sum_{k=0}^{n} \binom{n}{k} k x^k (1-x)^{n-k}, \tag{6.5}$$

and

$$n(n-1)x^2 = \sum_{k=0}^{n} \binom{n}{k} k(k-1) x^k (1-x)^{n-k}. \tag{6.6}$$

Therefore,

$$\begin{aligned}
\sum_{k=0}^{n} (k-nx)^2 \binom{n}{k} x^k (1-x)^{n-k} &= \sum_{k=0}^{n} k^2 \binom{n}{k} x^k (1-x)^{n-k} \\
&\quad - 2\sum_{k=0}^{n} nkx \binom{n}{k} x^k (1-x)^{n-k} \\
&\quad + \sum_{k=0}^{n} n^2x^2 \binom{n}{k} x^k (1-x)^{n-k} \\
&= [nx + n(n-1)x^2] - 2nx \cdot nx + n^2x^2 \\
&= nx(1-x).
\end{aligned} \tag{6.7}$$

At this stage, you may not see where the proof is headed. Fear not, and forge ahead!

For our given f, choose $M > 0$ such that $|f(x)| \leq M$ for all $x \in [0, 1]$. Since f is continuous on $[a, b]$, it is uniformly continuous, and therefore there exists a

$\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$ (this ϵ is the ϵ fixed at the beginning). Then

$$\begin{aligned}
|f(x) - p_n(x)| &= \left| f(x) - \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k} \right| \\
&= \Big| \sum_{k=0}^{n} \left(f(x) - f\left(\frac{k}{n}\right)\right)\binom{n}{k}x^k(1-x)^{n-k} \Big| \\
&= \Big| \sum_{|k-nx|<\delta n} \left(f(x) - f\left(\frac{k}{n}\right)\right)\binom{n}{k}x^k(1-x)^{n-k} \\
&\qquad + \sum_{|k-nx|\ge\delta n} \left(f(x) - f\left(\frac{k}{n}\right)\right)\binom{n}{k}x^k(1-x)^{n-k} \Big| \\
&\le \Big| \sum_{|k-nx|<\delta n} \left(f(x) - f\left(\frac{k}{n}\right)\right)\binom{n}{k}x^k(1-x)^{n-k} \Big| \\
&\qquad + \Big| \sum_{|k-nx|\ge\delta n} \left(f(x) - f\left(\frac{k}{n}\right)\right)\binom{n}{k}x^k(1-x)^{n-k} \Big|.
\end{aligned}$$

If $|k - nx| < \delta n$, then $|x - \frac{k}{n}| < \delta$, so that $|f(x) - f\left(\frac{k}{n}\right)| < \epsilon$. Then

$$\begin{aligned}
&\Big| \sum_{|k-nx|<\delta n} \left(f(x) - f\left(\frac{k}{n}\right)\right)\binom{n}{k}x^k(1-x)^{n-k} \Big| \\
&\quad \le \sum_{|k-nx|<\delta n} \left| f(x) - f\left(\frac{k}{n}\right)\right| \binom{n}{k}x^k(1-x)^{n-k} \\
&\quad < \epsilon \cdot \left(\sum_{|k-nx|<\delta n} \binom{n}{k}x^k(1-x)^{n-k} \right) \\
&\quad \le \epsilon \cdot \left(\sum_{k=0}^{n} \binom{n}{k}x^k(1-x)^{n-k} \right) = \epsilon.
\end{aligned}$$

If $|k - nx| \ge \delta n$, then

$$\begin{aligned}
&\Big| \sum_{|k-nx|\ge\delta n} \left(f(x) - f\left(\frac{k}{n}\right)\right)\binom{n}{k}x^k(1-x)^{n-k} \Big| \\
&\quad \le \sum_{|k-nx|\ge\delta n} \left| f(x) - f\left(\frac{k}{n}\right)\right| \binom{n}{k}x^k(1-x)^{n-k} \\
&\quad \le \sum_{|k-nx|\ge\delta n} \left(\left|f(x)\right| + \left|f\left(\frac{k}{n}\right)\right| \right) \binom{n}{k}x^k(1-x)^{n-k} \\
&\quad \le 2M \cdot \left(\sum_{|k-nx|\ge\delta n} \binom{n}{k}x^k(1-x)^{n-k} \right)
\end{aligned}$$

$$\leq \frac{2M}{n^2\delta^2} \cdot \left(\sum_{k=0}^{n} (k - nx)^2 \binom{n}{k} x^k (1-x)^{n-k} \right),$$

since $\dfrac{k - nx}{n\delta} \geq 1$ for these terms. By (7) the last expression is equal to

$$\frac{2M}{n^2\delta^2} nx(1-x).$$

And since $x(1-x) \leq \frac{1}{4}$ for each value of $x \in [0, 1]$, this is less than or equal to

$$\frac{M}{2n\delta^2}.$$

We have now shown that $|f(x) - p_n(x)|$ is is less than or equal to

$$\Big| \sum_{|k-nx|<\delta n} \left(f(x) - f\left(\frac{k}{n}\right) \right) \binom{n}{k} x^k (1-x)^{n-k} \Big|$$
$$+ \Big| \sum_{|k-nx|\geq\delta n} \left(f(x) - f\left(\frac{k}{n}\right) \right) \binom{n}{k} x^k (1-x)^{n-k} \Big|,$$

and that this, in turn, is less than

$$\epsilon + \frac{M}{2n\delta^2}.$$

If n is now chosen larger than $\dfrac{M}{2\delta^2\epsilon}$, we have

$$\|f - p_n\|_\infty < \epsilon + \frac{M}{2\frac{M}{2\delta^2\epsilon}\delta^2} = 2\epsilon.$$

We have now proved the Weierstrass approximation theorem for the interval $[0, 1]$ and are ready to extend this argument to an arbitrary interval $[a, b]$. The method employed here to generalize from $[0, 1]$ to $[a, b]$ is useful, and should be kept in mind. We consider any $f \in C([a, b])$, let $\epsilon > 0$ be arbitrary, and note that a and b are any two real numbers satisfying $a < b$. Define

$$g(x) = f\left(x(b-a) + a\right), \qquad x \in [0, 1].$$

Note that $g \in C([0, 1])$, $g(0) = f(a)$, and $g(1) = f(b)$. By the preceding argument there exists a (Bernstein) polynomial $p \in C([0, 1])$ such that $\|p - g\|_\infty < \epsilon$. Define

$$q(x) = p\left(\frac{x-a}{b-a}\right), \qquad x \in [a, b].$$

Note that q is a polynomial in $C([a, b])$, $q(a) = p(0)$, and $q(b) = p(1)$. Then

$$|f(x) - q(x)| = \left| g\left(\frac{x-a}{b-a}\right) - p\left(\frac{x-a}{b-a}\right) \right|,$$

and thus

$$\|f - q\|_\infty = \sup_{x \in [a,b]} \Big\{ |f(x) - q(x)| \Big\}$$

$$= \sup_{x\in[a,b]} \left\{ \left| g\left(\frac{x-a}{b-a}\right) - p\left(\frac{x-a}{b-a}\right) \right| \right\}$$
$$= \sup_{x\in[0,1]} \left\{ |g(x) - p(x)| \right\} = \|g - p\|_\infty < \epsilon.$$

This completes the proof. □

Marshall Stone recognized that the interval, $[a, b]$ in the real line that Weierstrass used could be replaced by a more general subset of a more general metric space. In fact, he realized that the set $[a, b]$ could be replaced by any compact subset of any Hausdorff topological space. Since this book does not assume knowledge of topological spaces, we will give a proof for a compact subset of a metric space (a metric space is an example of a Hausdorff topological space). The proof goes over, verbatim, if the set X in the theorem is considered as a subset of a Hausdorff topological space.

In the theorem, we write $C(X;\mathbb{R})$ to emphasize that the functions are *real*valued. After the proof we will make a comment about the theorem for complex-valued functions.

Most proofs of the Stone–Weierstrass theorem make use of the Weierstrass approximation theorem; one attractive feature of the proof presented here is that the Weierstrass approximation theorem is not used to deduce the more general result. Hence, the Weierstrass approximation theorem is subsumed by the Stone–Weierstrass theorem. The proof we give is due to Brosowski and Deutsch [26]; we follow their presentation closely. Following the proof, we will say more about standard proofs, and also about further generalizations.

The trouble, in a general metric space, is that "polynomials" might not make sense. We observe that polynomials are exactly the functions that can be obtained from the two functions 1 and x by multiplication by a scalar, by addition, and/or by multiplication. This characterization of the polynomials is captured by the three hypotheses of the theorem.

Theorem 6.2 **(**The Stone–Weierstrass Theorem**).** *Let X be a compact metric space, and assume that $A \subseteq C(X;\mathbb{R})$ satisfies the following conditions:*

- **(a)** *A is an algebra: If $f, g \in A$ and $\alpha \in \mathbb{R}$, then $f + g$, $f \cdot g$, and αf are all in A.*
- **(b)** *The constant function $x \to 1$ is in A (and hence A contains all constant functions).*
- **(c)** *A separates points: For $x \neq y \in X$ there exists an $f \in A$ such that $f(x) \neq f(y)$.*

Then A is dense in $C(X;\mathbb{R})$.

Proof. Brosowski and Deutsch break their proof into three parts. They prove two preliminary lemmas, and then prove the Stone–Weierstrass theorem.

The first lemma states:

Consider any point $x_0 \in X$ and any open set U_0 in X containing x_0. Then there exists an open set $V_0 \subseteq U_0$ containing x_0 such that for each $\epsilon > 0$ there exists $g \in A$ satisfying

(i) $0 \leq g(x) \leq 1, x \in X$;
(ii) $g(x) < \epsilon,\ x \in V_0$;
(iii) $g(x) > 1 - \epsilon,\ x \in X \setminus V_0$.

To prove this lemma we first make use of hypothesis (c) to deduce the existence of a function $g_x \in A$ with $g_x(x_0) \neq g_x(x)$ for each $x_0 \in X \setminus U_0$. The function $h_x = g_x - g_x(x_0)$ is in A and satisfies $0 = h_x(x_0) \neq h_x(x)$. The function

$$p_x = \Big(\frac{1}{\|h_x\|_\infty^2}\Big) h_x^2$$

is also in A and satisfies $p_x(x_0) = 0$, $p_x(x) > 0$, and $0 \leq p_x \leq 1$.

Let $U_x = \{y \in X : p_x(y) > 0\}$. Then U_x is an open set and contains x. Since $A \setminus U_0$ is compact (by Exercise 2.1.13), it contains a finite collection of points $x_1, x_2, \ldots, x_m$ such that

$$X \setminus U_0 \subseteq \bigcup_{i=1}^{m} U_{x_i}.$$

The function

$$p = \Big(\frac{1}{m}\Big) \sum_{i=1}^{m} p_{x_i}$$

is in A and satisfies $0 \leq p \leq 1$, $p(x_0) = 0$, and $p > 0$ on $X \setminus U_0$. Since $X \setminus U_0$ is compact, there exists $0 < \delta < 1$ such that $p \geq \delta$ on $X \setminus U_0$ (see Exercise 2.1.14). The set

$$V_0 = \Big\{x \in X \;\Big|\, p(x) < \frac{\delta}{2}\Big\}$$

is an open set. Further, V_0 contains the point x_0 and is contained in the set U_0.

Let k be the smallest integer greater than $\frac{1}{\delta}$. Then $k - 1 \leq \frac{1}{\delta}$, and so $k < \frac{2}{\delta}$. Therefore, $1 < k\delta < 2$. Define functions

$$q_n(x) = \Big(1 - p^n(x)\Big)^{k^n}$$

for $n = 1, 2, \ldots$. Then $q_n \in A$, $0 \leq q_n \leq 1$, and $q_n(x_0) = 1$.

If $x \in V_0$, then

$$kp(x) \leq k \cdot \frac{\delta}{2} < 1.$$

The inequality proved in Exercise 6.1.2 implies that

$$q_n(x) \geq 1 - \Big(kp(x)\Big)^n \geq 1 - \Big(k \cdot \frac{\delta}{2}\Big)^n.$$

The last expression goes to 1 as $n \to \infty$.

If $x \in X \setminus U_0$, then

$$kp(x) \geq k\delta > 1,$$

and the same inequality from the exercise implies that

$$\begin{aligned} q_n(x) &= \frac{1}{k^n p^n(x)} \Big(1 - p^n(x)\Big)^{k^n} k^n p^n(x) \\ &\leq \frac{1}{k^n p^n(x)} \Big(1 - p^n(x)\Big)^{k^n} \Big(1 + k^n p^n(x)\Big) \\ &\leq \frac{1}{k^n p^n(x)} \Big(1 - p^n(x)\Big)^{k^n} \Big(1 + p^n(x)\Big)^{k^n} \\ &= \frac{1}{k^n p^n(x)} \Big(1 - p^{2n}(x)\Big)^{k^n} \\ &\leq \frac{1}{(k\delta)^n}. \end{aligned}$$

The last expression goes to 0 as $n \to \infty$.

We can therefore choose n large enough so that q_n satisfies $0 \leq q_n \leq 1$, $q_n(x) < \epsilon$ for each $x \in X \setminus U_0$, and $q_n(x) > 1 - \epsilon$ for each $x \in V_0$. Define $g = 1 - q_n$. It is left as an exercise (Exercise 6.1.3(a)) to show that g satisfies (i)–(iii).

The second lemma states:
Consider disjoint closed subsets Y and Z of X. For each $0 < \epsilon < 1$ there exists $g \in A$ satisfying

(i) $0 \leq g(x) \leq 1, x \in X$;
(ii) $g(x) < \epsilon, x \in Y$;
(iii) $g(x) > 1 - \epsilon, x \in Z$.

To prove this lemma we begin by considering the open set $U = X \setminus Z$ in X. If $x \in Y$, then $x \in U$, and the first lemma gives an open set V_x of X containing x with certain properties. Since X is compact, X contains a finite collection of points $x_1, x_2, \ldots, x_m$ such that

$$X \subseteq \bigcup_{i=1}^{m} V_{x_i}.$$

Let g_i be the function associated to V_{x_i} as given in the first lemma satisfying $0 \leq g_i \leq 1$, $g_i(x) < \frac{\epsilon}{m}$ for each $x \in V_{x_i}$, and $g_i(x) > 1 - \frac{\epsilon}{m}$ for each $x \in X \setminus U = Z$. Define $g = g_1 g_2 \cdots g_m$. It is left as an exercise (Exercise 6.1.3(b)) to show that this function satisfies (i)–(iii).

We now return to the proof of the Stone–Weierstrass theorem. Consider $f \in C(X; \mathbb{R})$. We aim to show that corresponding to any given $\epsilon > 0$ there exists an element $g \in A$ satisfying $\|f - g\|_\infty < \epsilon$. In fact, we will show that there exists an element $g \in A$ satisfying $\|f - g\|_\infty < 2\epsilon$.

Replacing f by $f + \|f\|_\infty$, we can assume that $f \geq 0$. We also assume that $\epsilon < \frac{1}{3}$. Start by choosing an integer n such that $(n-1)\epsilon \geq \|f\|_\infty$ and define sets

$X_i, Y_i,\ i = 0, 1, \ldots, n$, by

$$X_i = \left\{x \in X \,\middle|\, f(x) \leq \left(i - \frac{1}{3}\right)\epsilon\right\}$$

and

$$Y_i = \left\{x \in X \,\middle|\, f(x) \geq \left(i + \frac{1}{3}\right)\epsilon\right\}.$$

Then we see that

$$\begin{aligned} &X_i \cap Y_i = \emptyset, \\ &\emptyset \subseteq X_0 \subseteq X_1 \cdots \subseteq X_n = X, \end{aligned}$$

and

$$Y_0 \supseteq Y_1 \supseteq \cdots \supseteq Y_n = \emptyset.$$

From the second lemma we have, corresponding to each $i = 0, 1, \ldots, n$, a function $g_i \in A$ satisfying $0 \leq g_i \leq 1$, $g_i(x) < \frac{\epsilon}{n}$ for each $x \in X_i$, and $g_i(x) > 1 - \frac{\epsilon}{n}$ for each $x \in Y_i$.

Define

$$g = \epsilon \sum_{i=0}^{n} g_i.$$

This function is in A. Consider an arbitrary element $x \in X$. From the chain of subsets $\emptyset \subseteq X_0 \subseteq X_1 \cdots \subseteq X_n = X$, we see that there is an $i \geq 1$ such that $x \in X_i \setminus X_{i-1}$. For this value of i,

$$\left(i - \frac{4}{3}\right)\epsilon < f(x) < \left(i - \frac{1}{3}\right)\epsilon$$

and

$$g_j(x) < \frac{\epsilon}{n} \text{ for every } j \geq i.$$

Note also that $x \in Y_j$ for every value of $j \leq i - 2$ and thus

$$g_j(x) > 1 - \frac{\epsilon}{n} \text{ for every } j \leq i - 2.$$

These last two inequalities yield

$$\begin{aligned} g(x) &= \epsilon \sum_{j=0}^{i-1} g_j(x) + \epsilon \sum_{j=i}^{n} g_j(x) \\ &\leq i\epsilon + \epsilon(n - i + 1)\frac{\epsilon}{n} \\ &\leq i\epsilon + \epsilon^2 \\ &< \left(i + \frac{1}{3}\right)\epsilon, \end{aligned}$$

and, for $i \geq 2$,

$$\begin{aligned} g(x) &\geq \epsilon \sum_{j=0}^{i-2} g_j(x) \\ &\geq (i-1)\epsilon\Big(1-\frac{\epsilon}{n}\Big) = (i-1)\epsilon - \frac{i-1}{n}\epsilon^2 > (i-1)\epsilon - \epsilon^2 \\ &> \Big(i-\frac{4}{3}\Big)\epsilon. \end{aligned}$$

This last equality, $g(x) > \Big(i - \frac{4}{3}\Big)\epsilon$, is proved for $i \geq 2$; for $i = 1$ it is straightforward. Therefore,

$$|f(x) - g(x)| \leq \Big(i + \frac{1}{3}\Big)\epsilon - \Big(i - \frac{4}{3}\Big)\epsilon < 2\epsilon,$$

and the Stone–Weierstrass theorem is proved. □

The conditions (a), (b), and (c) are usually rather easy to check if you are given a subset A of $C(X;\mathbb{R})$. Also, we point out that if $\mathbb{R}$ is replaced by $\mathbb{C}$ and the following condition (d) is added to the hypotheses, then the conclusion of the theorem still holds:

(d) If $f \in A$ then $\overline{f} \in A$.

Our proof of the Stone–Weierstrass theorem is not found in other texts. In many texts ([30] or [112], for example), a proof is given that was published in 1961 by Errett Bishop (1928–1983; USA) [19]. Bishop deduces the conclusion of the Stone–Weierstrass theorem from a powerful result now called "Bishop's theorem."

As mentioned in the paragraphs preceding Brosowski and Deutsch's proof, X can be replaced by a Hausdorff space. In what follows, we will use this more general language; if it makes you more comfortable, you may continue to think of X as a metric space. Consider a compact Hausdorff space X and a closed unital subalgebra A of $C(X;\mathbb{C})$. A subset S of X is said to be *A-symmetric* if every $h \in A$ that is real-valued on S is actually constant on S. Assume that $f \in C(X;\mathbb{C})$ and to each A-symmetric subset S of X there is a function $g_S \in A$ such that $g_S(x) = f(x)$ for all $x \in S$. Bishop's theorem asserts that with these hypotheses, f must be in A. Why is the Stone–Weierstrass theorem a consequence of Bishop's theorem? With notation as in the the statement of the Stone–Weierstrass theorem, consider $\overline{A}$, a closed unital subalgebra of $C(X;\mathbb{C})$. Let $f \in C(X;\mathbb{C})$ be arbitrary, and let S be an $\overline{A}$-symmetric subset of X. By Exercise 6.1.7, $S = \{s\}$ for some $s \in X$. Since $\overline{A}$ contains all constant functions, $\overline{A}$ contains the constant function g defined by $g(x) = f(s)$ for all $x \in X$. Clearly, then, $g(x) = f(x)$ for all $x \in S$. Bishop's theorem now implies that $f \in \overline{A}$. Since f was arbitrary, we conclude that $\overline{A}$ must contain all of $C(X;\mathbb{C})$, which is the conclusion of the Stone–Weierstrass theorem.

The next natural question is, How can one prove Bishop's theorem? The standard proof of this theorem uses ideas of Louis de Branges [24]. This approach requires sophisticated machinery (including the Hahn–Banach theorem, which will be proved later in this chapter) that we are not in a position to develop here. In 1977, Silvio Machado (1932–1981) offered a more elementary proof of Bishop's Theorem [85] (and hence of the Stone–Weierstrass theorem). In 1984, Thomas Ransford (born 1958; England) incorporated Brosowski and Deutsch's ideas and gave a shortened, simplified version of Machado's proof [102]. The only ingredient in Ransford's proof that can be thought of as "nonelementary" is Zorn's lemma. We have not yet encountered Zorn's lemma but we point out that the Hahn–Banach theorem, and hence de Branges's approach, also uses this lemma. Zorn's lemma will be discussed in Section 3 of this chapter. Reading and presenting Ransford's paper would be a nice student project.

We end with a few closing remarks. Subalgebras of $C(X;\mathbb{C})$ that separate the points of X and contain the constant function 1 (and hence all constant functions) are called *uniform algebras*. These algebras need not be closed under complex conjugation. The theory of uniform algebras is described in [48].

The classical Weierstrass approximation theorem asserts that on certain subsets of the real line, the polynomials are dense in the continuous functions. Stone's generalization aims to capture the essence of the collection of polynomials (in order to replace $\mathbb{R}$ with some more general space on which polynomials may not make sense). Another interesting way to generalize the classical theorem is as follows. Replace subsets of $\mathbb{R}$ with subsets of $\mathbb{C}$. Then polynomials make sense, and one can ask to characterize the subsets of $\mathbb{C}$ on which the polynomials are dense in the continuous functions. There are theorems, such as a famous one due to Carle Runge (1856–1927; Germany), that address this question. Runge's theorem is really a theorem of complex analysis; see page 198 of [30] for a proof. In the present context, it is interesting that the standard proof (as in [30]) can be replaced by a proof using functional analysis. This functional analytic proof is elegant, but (again!) requires the rather powerful Hahn–Banach theorem (see Section 3). For more on the connection between functional analysis and Runge's theorem, see Chapters 13 and 20 of [111]. The proof using the Hahn–Banach theorem uses the observation that the collection of (complex) differentiable functions on a compact subset of $\mathbb{C}$ forms a Banach space. There is a third proof, which uses the observation that this Banach space is, in fact, a Banach algebra. This is a proof that is more elementary in the sense that it does not require the Hahn–Banach theorem, and was first given in a nice article [50] by Sandy Grabiner (born 1939; USA).

Marshall Harvey Stone was born on April 8, 1903, in New York City (Figure 6.1). His father, Harlan Fiske Stone, was a distinguished lawyer who served on the U.S. Supreme Court, including service for five years as chief justice.

Stone attended public schools in New Jersey and graduated from Harvard University in 1922. His Ph.D., also done at Harvard, was awarded in 1926. He spent the majority of the first part of his professional life at Harvard, and then most of his career at the University of Chicago. He left Harvard to become the head of the mathematics department at Chicago. As head, he was largely responsible for turning Chicago's mathematics department into what many consider the strongest

FIGURE 6.1. Marshall Stone in 1982.

mathematics department in the United States at that time. During World War II, he was involved in secret work for the United States government.

Stone's Ph.D. thesis, ***Ordinary linear homogeneous differential equations of order*** *n* ***and the related expansion problems***, was written under the direction of George David Birkhoff (1884–1944; U.S.A.). Over the next few years he continued this work, by studying the eigenfunctions of differential operators. This then led him to work with Hermitian operators on Hilbert spaces. Much of his work in this area was motivated by the quantum-mechanical theories developing at the same time. In particular, Stone was interested in extending Hilbert's spectral theory from bounded to unbounded operators. His work, at many times, paralleled the work of von Neumann. In 1932, he published his classic book ***Linear Transformations in Hilbert Space and their Applications to Analysis***, and he credits von Neumann and Riesz as the two primary sources of ideas for his work. It should be noted that this is the same publication year as Banach's treatise [11]; the two books are quite different in their aims and styles. Some other particularly important contributions of Stone's include his celebrated extension of Weierstrass's theorem on polynomial approximation (as discussed in Section 6.1) and his work on rings of continuous functions. The latter can be viewed as early work on commutative Banach algebras. Stone's work is characterized by brilliant combined use of ideas from analysis, algebra, and topology.

Stone had exceptional talent as a writer, which is demonstrated by his writings on many different topics. For example, his paper "The generalized Weierstrass approximation theorem" [119] and his book mentioned in the preceding paragraph are very enjoyable to read. He was interested in many things, especially in education and travel, and he wrote about these things as well as about mathematics. He authored, for example, a paper on mid twentieth century mathematics in China [120], uniting two of his interests. He traveled to many different countries, and was even shipwrecked in Antarctica. He died on January 9, 1989, shortly after becoming ill, in Madras, India.

6.2 The Baire Category Theorem with an Application to Real Analysis

After giving necessary definitions, we prove an important theorem of René Baire. We proceed to deduce from it that it is impossible for a function to be continuous at each rational point and discontinuous at each irrational point of the interval (0, 1). Finally, we give another argument that no such function can exist, an argument due to Volterra. Volterra gave his proof about twenty years before the first appearance of Baire's theorem.

Throughout this section $M = (M, d)$ will denote a metric space.

A subset X in a metric space M is *nowhere dense* if $M \setminus \overline{X}$ is dense in M. Any subset of M that is a countable union of nowhere dense subsets of M is said to be of *first category* (in M). A subset of M that is not of first category is said to be of *second category*.

We use the following lemma to prove the Baire category theorem. In fact, the lemma and Baire's theorem are equivalent statements.

Lemma 6.3. *If $\{U_n\}_{n=1}^{\infty}$ is a sequence of open dense subsets of a complete metric space M, then $\bigcap_{n=1}^{\infty} U_n$ is dense in M.*

Proof. Consider $x \in M$ and $\epsilon > 0$. We aim to show that there exists an element $y \in \bigcap_{n=1}^{\infty} U_n$ such that $y \in B_\epsilon(x)$.

Since $\overline{U_1} = M$, there exists $y_1 \in U_1 \cap B_\epsilon(x)$. Since $U_1 \cap B_\epsilon(x)$ is open, there exists an open ball

$$B_{\epsilon_1}(y_1) \subseteq U_1 \cap B_\epsilon(x).$$

Let $\delta_1 = \min\{\frac{\epsilon_1}{2}, 1\}$.

Since $\overline{U_2} = M$, there exists $y_2 \in U_2 \cap B_{\delta_1}(y_1)$. Since $U_2 \cap B_{\delta_1}(y_1)$ is open, there exists an open ball

$$B_{\epsilon_2}(y_2) \subseteq U_2 \cap B_{\delta_1}(y_1).$$

Let $\delta_2 = \min\{\frac{\epsilon_2}{2}, \frac{1}{2}\}$.

Since $\overline{U_3} = M$, there exists $y_3 \in U_3 \cap B_{\delta_2}(y_2)$. Since $U_3 \cap B_{\delta_2}(y_2)$ is open, there exists an open ball

$$B_{\epsilon_3}(y_3) \subseteq U_3 \cap B_{\delta_2}(y_2).$$

Let $\delta_3 = \min\{\frac{\epsilon_3}{2}, \frac{1}{3}\}$.

Continuing in this way, we create a sequence $\{y_n\}_{n=1}^{\infty}$ and open balls $\{B_{\epsilon_n}(y_n)\}_{n=1}^{\infty}$ and $\{B_{\delta_n}(y_n)\}_{n=1}^{\infty}$ with $\delta_n = \min\{\frac{\epsilon_n}{2}, \frac{1}{n}\}$ satisfying

$$B_{\epsilon_n}(y_n) \subseteq U_n \cap B_{\delta_{n-1}}(y_{n-1})$$

for $n = 1, 2, \ldots$. If $m > n$, then this implies that $y_m \in B_{\delta_{n-1}}(y_n)$ and thus

$$d(y_m, y_n) < \delta_n \leq \frac{1}{n}.$$

Hence $\{y_n\}_{n=1}^{\infty}$ is a Cauchy sequence. Since M is a complete metric space, there exists a $y \in M$ such that $d(y_n, y) \to 0$ as $n \to \infty$. From $y_m \in B_{\delta_n}(y_n)$ for all

$m > n$, it follows that $y \in \overline{B_{\delta_n}(y_n)}$ for all $n = 1, 2, \ldots$. Therefore,

$$y \in \overline{B_{\delta_n}(y_n)} \subseteq \overline{B_{\frac{\epsilon_n}{2}}\left(y_n\right)} \subseteq \overline{B_{\epsilon_n}(y_n)} \subseteq U_n$$

for all $n = 1, 2, \ldots$. From this string of inclusions we can conclude both

$$y \in \bigcap_{n=1}^{\infty} U_n$$

and $y \in B_{\epsilon_1}(y_1) \subseteq B_\epsilon(x)$, as desired. $\square$

Theorem 6.4 (The Baire Category Theorem)**.** *Any nonempty, complete metric space is of second category.*

Proof. Let M be a nonempty complete metric space that is of first category. Then M can be written as a countable union of nowhere dense subsets, $M = \bigcup_{n=1}^{\infty} A_n$. Then $M = \bigcup_{n=1}^{\infty} \overline{A_n}$ also, and from De Morgan's law we have $\emptyset = \bigcap_{n=1}^{\infty}(M \setminus \overline{A_n})$. But each set $M \setminus \overline{A_n}$ is open and dense (the latter from the definition of nowhere dense), and the lemma thus gives that $\bigcap_{n=1}^{\infty}(M \setminus \overline{A_n})$ is dense in M. This is impossible, and thus M must be of second category. $\square$

There are many applications of Baire's theorem in analysis. Our main use of it will appear in the next section, where we use it to establish fundamental results about linear operators between Banach spaces. We now give one application of it to the theory of real functions.

The function defined on (0, 1) by

$$g(x) = \begin{cases} \dfrac{1}{q} & \text{if } x = \dfrac{p}{q} \text{ in reduced form,} \\ 0 & \text{if } x \text{ is irrational,} \end{cases}$$

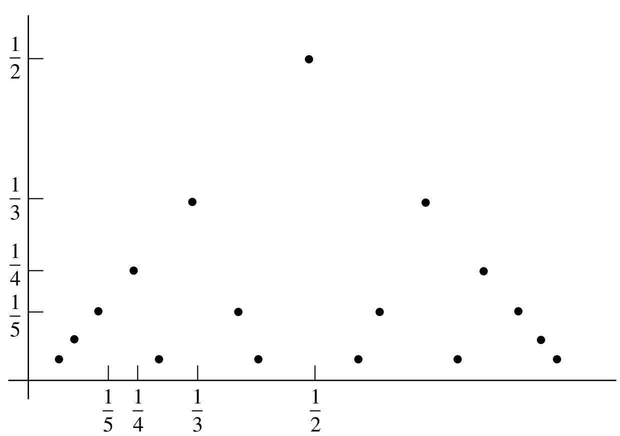

FIGURE 6.2. The graph of $g(x)$.

is a standard example of a function that is continuous at each irrational point of (0, 1) and discontinuous at each rational point of (0, 1) (Exercise 6.2.5). See Figure 6.2. It is natural to wonder whether there is a function defined on (0, 1) that is continuous at each rational point of (0, 1) and discontinuous at each irrational point of (0, 1). It may seem somewhat surprising that no such function can exist. The proof of this usually given makes use of Baire's theorem, and we now supply this proof.

We start by defining, for any bounded real-valued function f defined on an open interval I, the *oscillation of f on I* by

$$\omega_f(I) = \sup\{f(x) \,\big|\, x \in I\} - \inf\{f(x) \,\big|\, x \in I\}.$$

For $a \in I$ define the *oscillation of f at a* by

$$\omega_f(a) = \inf\{\omega_f(J) \,\big|\, J \subseteq I \text{ is an open interval containing } a\}.$$

The connection between continuity and oscillation is made precise in the following straightforward lemma.

Lemma 6.5. *A bounded real-valued function f defined on an open interval I is continuous at $a \in I$ if and only if $\omega_f(a) = 0$ and the set $\{x \in I \,\big|\, \omega_f(x) < \epsilon\}$ is an open set for each $\epsilon > 0$.*

Proof. If f is continuous at a, then given $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(a)| < \epsilon$ whenever $|x - a| < \delta$. Therefore,

$$\omega_f(a) \leq \omega_f\Big((a - \delta, a + \delta)\Big) < \epsilon.$$

Since this holds for every $\epsilon > 0$, we have $\omega_f(a) = 0$.

Conversely, if $\omega_f(a) = 0$, then for any given $\epsilon > 0$ there exists an open interval $J \subseteq I$ containing a such that $\omega_f(J) < \epsilon$. Since J is open, there exists a $\delta > 0$ such that the interval $(a - \delta, a + \delta)$ is contained in J. Hence $|x - a| < \delta$ implies that

$$|f(x) - f(a)| \leq \omega_f\Big((a - \delta, a + \delta)\Big) \leq \omega_f(J) < \epsilon,$$

as desired.

To prove the second assertion of the lemma consider $\epsilon > 0$ and $x_0 \in \{x \in I : \omega_f(x) < \epsilon\}$. Let J be an open interval containing x_0 and satisfying $\omega_f(J) < \epsilon$. For any $y \in J$, $\omega_f(y) \leq \omega_f(J) < \epsilon$, and thus $J \subseteq \{x \in I : \omega_f(x) < \epsilon\}$, proving that $\{x \in I : \omega_f(x) < \epsilon\}$ is open. □

Theorem 6.6. *There is no function defined on* (0, 1) *that is continuous at each rational point of* (0, 1) *and discontinuous at each irrational point of* (0, 1)*.*

Proof. Suppose, to the contrary, that such an f exists. By the lemma, the set

$$U_n = \left\{x \in (0, 1) : \omega_f(x) < \frac{1}{n}\right\}$$

is open for each $n = 1, 2, \ldots$. By the first part of the lemma, the set $\bigcap_{n=1}^{\infty} U_n$ is equal to $\mathbb{Q} \cap (0, 1)$. Since the rational numbers are dense in (0, 1), each U_n is also dense in $(0, 1)$. Set $V_n = (0, 1) \setminus U_n$, $n = 1, 2, \ldots$. Then $(0, 1) \setminus \mathbb{Q} = \bigcup_{n=1}^{\infty} V_n$, and each V_n is nowhere dense in $(0, 1)$. If $\{r_1, r_2, \ldots\}$ is an enumeration of the rational numbers in $(0, 1)$, then

$$(0, 1) = V_1 \cup \{r_1\} \cup V_2 \cup \{r_2\} \cup \cdots,$$

and so

$$[0, 1] = \{0\} \cup \{1\} \cup V_1 \cup \{r_1\} \cup V_2 \cup \{r_2\} \cup \cdots.$$

Each set in this union is nowhere dense, and thus $[0, 1]$ is of first category, contradicting the Baire category theorem and completing our proof. □

As already mentioned, the nonexistence of a function continuous exactly on the rational numbers is usually deduced as a corollary to Baire's theorem. Baire's theorem appeared in 1899 [9]. Two decades before the appearance of this paper the Italian mathematician Vito Volterra gave a proof of the nonexistence of such a function; he gave this proof while he was still a student. Volterra also argues by contradiction, but he avoids altogether the somewhat sophisticated ideas of category. We first encountered Volterra's proof in William Dunham's wonderful article [37].

Following Volterra's proof, we will assume that such an f exists, and let g be defined by

$$g(x) = \begin{cases} \dfrac{1}{q} & \text{if } x = \dfrac{p}{q} \text{ in reduced form,} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Consider any rational point x_0 in the open interval $(0, 1)$. By continuity of f at x_0 there exists $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subseteq (0, 1)$ and $|f(x) - f(x_0)| < \frac{1}{2}$ whenever $|x - x_0| < \delta$. Choose a_1 and b_1 such that $[a_1, b_1] \subseteq (x_0 - \delta, x_0 + \delta)$. Then

$$|f(x) - f(y)| \leq |f(x) - f(x_0)| + |f(x_0) - f(y)| < \frac{1}{2} + \frac{1}{2} = 1$$

for all $x, y \in [a_1, b_1]$. Next, we choose an irrational point in the open interval (a_1, b_1). By the same argument, there exist points a_1' and b_1' such that $[a_1', b_1'] \subseteq (a_1, b_1)$ and $|g(x) - g(y)| < 1$ for all $x, y \in [a_1', b_1']$. Thus, for all $x, y \in [a_1', b_1']$, we have both $|f(x) - f(y)| < 1$ and $|g(x) - g(y)| < 1$.

Repeat this argument starting with the open interval (a_1', b_1') in place of $(0, 1)$ to construct a closed interval $[a_2', b_2'] \subseteq (a_1', b_1')$ such that for all $x, y \in [a_2', b_2']$ we have both $|f(x) - f(y)| < \frac{1}{2}$ and $|g(x) - g(y)| < \frac{1}{2}$.

Keep repeating this argument to construct intervals

$$(0, 1) \supseteq [a_1', b_1'] \supseteq [a_2', b_2'] \supseteq \cdots \supseteq [a_n', b_n'] \supseteq \cdots$$

and such that for all $x, y \in [a_n', b_n']$ we have both $|f(x) - f(y)| < \frac{1}{2^n}$ and $|g(x) - g(y)| < \frac{1}{2^n}$. By the nested interval theorem[2] there exists exactly one point contained in all intervals $[a_n', b_n']$. It follows that both f and g are continuous at this point, and hence that this point is simultaneously rational and irrational. Since this is impossible, we are done.

6.3 Three Classical Theorems from Functional Analysis

In this section we present the open mapping theorem, the Banach–Steinhaus theorem, and the Hahn–Banach theorem. We also discuss their history and applications. The first was proved by Banach, the second jointly by Banach and Steinhaus, and the third was proved, independently, by Hans Hahn (1879–1934; German) and Banach. These three results are fundamental theorems of functional analysis, and it may be argued that any book purporting to be a functional analysis text must include them.

The first two can be viewed as consequences of the Baire category theorem. These two theorems have to do with linear operators between normed linear spaces. The Hahn–Banach theorem is about "linear functionals": linear mappings from a linear space into the underlying field. It is different in flavor from the other two theorems of the section, but is put in this section because these three theorems are often thought of as the "bread and butter" theorems of elementary functional analysis.

Suppose that X is a linear space, α is a scalar, and $A \subseteq X$. We will use the notation αA to denote the set $\{\alpha x \,|\, x \in A\}$. In particular, we note the equality of open sets: $\alpha B_\beta(x) = B_{\alpha\beta}(x)$.

Theorem 6.7 (The Open Mapping Theorem)**.** *Consider Banach spaces X and Y and an element $T \in \mathcal{B}(X, Y)$. If T is onto, then $T(U)$ is open in Y whenever U is open in X.*

Proof. We split the proof into three steps:

- **(i)** There exists $\epsilon > 0$ such that $B_\epsilon(0) \subseteq \overline{T(B_{\frac{1}{2}}(0))}$.
- **(ii)** For the $\epsilon > 0$ found in (i), we have $B_\epsilon(0) \subseteq T(B_1(0))$.
- **(iii)** $T(U)$ is open in Y whenever U is open in X.

To prove (i), we begin by observing that since T is onto,

$$Y = T(X) = T\Big(\bigcup_{n=1}^{\infty} B_{\frac{n}{2}}(0)\Big) = \bigcup_{n=1}^{\infty} T\Big(B_{\frac{n}{2}}(0)\Big).$$

The Baire category theorem implies that there is an integer N such that one of the sets $T\Big(B_{\frac{N}{2}}(0)\Big)$ is not nowhere dense in Y. By Exercise 6.2.1 the interior of

[2]This is a standard theorem from a first real analysis course and is an immediate consequence of the completeness of the real numbers.

$\overline{T\left(B_{\frac{N}{2}}(0)\right)}$ is not empty. We can thus find, by the definition of interior, a $y_0 \in Y$ and an $r > 0$ such that $B_r(y_0) \subseteq \overline{T\left(B_{\frac{N}{2}}(0)\right)}$.

Let $\epsilon = \frac{r}{2N}$. Note that

$$\frac{y_0}{N} + y \in \overline{T\left(B_{\frac{1}{2}}(0)\right)}$$

for each $y \in B_{2\epsilon}(0)$. Now consider any $y \in B_\epsilon(0)$. Since both $\frac{y_0}{N}$ and $-\frac{y_0}{N}$ are in $\overline{T\left(B_{\frac{1}{2}}(0)\right)}$, we have

$$\begin{aligned} y &= \frac{-y_0}{2N} + \left(\frac{y_0}{2N} + y\right) \\ &= \frac{-y_0}{2N} + \frac{1}{2}\left(\frac{y_0}{N} + 2y\right) \\ &\in \overline{T\left(B_{\frac{1}{4}}(0)\right)} + \overline{T\left(B_{\frac{1}{4}}(0)\right)} \\ &\subseteq \overline{T\left(B_{\frac{1}{2}}(0)\right)}. \end{aligned}$$

This proves (i).

To prove (ii) we consider any $y \in B_\epsilon(0)$. By (i) we can choose $x_1 \in B_{\frac{1}{2}}(0)$ such that Tx_1 and y are as close to each other as we please; choose x_1 to satisfy

$$\|y - Tx_1\|_Y < \frac{\epsilon}{2}.$$

That is, $y - Tx_1 \in B_{\frac{\epsilon}{2}}(0) = \frac{1}{2}B_\epsilon(0) \subseteq \frac{1}{2}\overline{T(B_{\frac{1}{2}}(0))} = \overline{T(B_{\frac{1}{4}}(0))}$. Now we can choose x_2 in $B_{\frac{1}{4}}(0)$ such that Tx_2 and $y - Tx_1$ are as close to each other as we please; choose x_2 to satisfy

$$\|(y - Tx_1) - Tx_2\|_Y < \frac{\epsilon}{4}.$$

Continue in this way, creating a sequence

$$x_n \in B_{2^{-n}}(0),$$
$$\left\| y - \sum_{k=1}^{n} Tx_k \right\|_Y < \frac{\epsilon}{2^n}.$$

It follows from Lemma 3.21 that $\sum_{k=1}^{\infty} x_k$ converges in X. Let x denote this infinite sum. Then

$$\|x\|_X \leq \sum_{k=1}^{\infty} \|x_k\|_X < \sum_{k=1}^{\infty} 2^{-k} = 1,$$

showing that $x \in B_1(0)$. Finally, it follows from the continuity of T that $Tx = y$.

We now move on to the proof of the third part. Consider an element Tx in $T(U)$. Since x is in the open set U, there is a $\delta > 0$ such that $B_\delta(x) \subseteq U$. We will be done when we show that $B_{\delta\epsilon}(Tx) \subseteq T(U)$. To see this, let $y \in B_{\delta\epsilon}(Tx)$ and write y as $Tx + y_1$, for some $y_1 \in B_{\delta\epsilon}(0)$. Then $y_2 = \frac{y_1}{\delta}$ is in $B_\epsilon(0) \subseteq T(B_1(0))$. Write $y_2 = Tx_2$ for some $x_2 \in B_1(0)$. Then $y = Tx + \delta Tx_2 = T(x + \delta x_2)$, with $x + \delta x_2 \in U$. This completes the proof. $\square$

The next result, the Banach–Steinhaus theorem, is sometimes referred to as the *uniform boundedness principle*. In fact, there are different versions of this principle, as a perusal of functional analysis texts shows. These principles give conditions on a collection of operators under which each operator in the collection is bounded (in norm) by a single (finite) number.

Theorem 6.8 (The Banach–Steinhaus Theorem)**.** *Consider a Banach space X and a normed linear space Y. If $\mathcal{A} \subseteq \mathcal{B}(X, Y)$ is such that*

$$\sup\{\|Tx\|_Y \;\big|\, T \in \mathcal{A}\} < \infty$$

for each $x \in X$, then

$$\sup\{\|T\|_{B(X,Y)} \;\big|\, T \in \mathcal{A}\} < \infty.$$

Proof. Define sets

$$E_n = \{x \in X \;\big|\, \|Tx\|_Y \le n \text{ for all } T \text{ in } \mathcal{A}\} = \bigcap_{T \in \mathcal{A}} \{x \in X \;\big|\, \|Tx\|_Y \le n\}.$$

Each of these sets E_n is closed, and their union $\bigcup_{n=1}^{\infty} E_n$ is all of X. Using Exercise 6.2.1 just as in the proof of the open mapping theorem, the Baire category theorem implies that $(E_N)^\circ \ne \emptyset$, for some integer N. By definition of interior, there exists a point x_0 and a positive number r such that $B_r(x_0) \subseteq E_N$. That is, $\|Tx\|_Y \le N$ for each x in $B_r(x_0)$ and each $T \in \mathcal{A}$. We aim to show that there is a number K such that $\|Tx\|_Y \le K$ for every element $x \in X$ of norm 1 and each $T \in \mathcal{A}$. For an arbitrary element x in X of norm 1, we consider the element $y = \frac{r}{2}x + x_0$. Then $y \in B_r(x_0)$, and so $\|Ty\|_Y \le N$. Therefore,

$$\frac{r}{2}\|Tx\|_Y - \|Tx_0\|_Y \le \left\|\frac{r}{2}Tx + Tx_0\right\|_Y = \|Ty\|_Y \le N,$$

and hence

$$\|Tx\|_Y \le \frac{2}{r}\Big(N + \|Tx_0\|_Y\Big).$$

Since the number $K = \frac{2}{r}(N + \|Tx_0\|_Y)$ is independent of x, we are done. $\square$

The Banach–Steinhaus theorem is very useful as a theoretical tool in functional analysis, and for more "concrete" applications in other areas. We will not discuss these applications.

We now move on to one of the single most important results in functional analysis: the Hahn–Banach theorem.

Let X be a linear space over $\mathbb{R}$ (respectively $\mathbb{C}$). A linear operator from X into $\mathbb{R}$ (respectively $\mathbb{C}$) is called a *linear functional*. Assume now that X is a normed linear space. The collection of all continuous linear functionals[3] on X is called the *dual*

[3]Recall that in this context the words *continuous* and *bounded* are interchangeable.

space of X; this space is very important, and we will only touch on its properties. The notation X^* is used to denote the dual space of X, so that X^* is shorthand for $\mathcal{B}(X, \mathbb{R})$ (respectively $\mathcal{B}(X, \mathbb{C})$). For example, the next project an interested student might take on would be to identify the dual spaces of his/her favorite Banach spaces (this material is standard in most first-year graduate functional analysis texts). The early work on dual spaces led to the idea of the "adjoint" of a linear operator, an idea that has proved extremely useful in the theory of operators on Hilbert space.

If M is a proper subspace of X, and λ is a linear functional on M taking values in the appropriate field $\mathbb{R}$ or $\mathbb{C}$, then a linear functional Λ on X is called an *extension* of λ if $\lambda(x) = \Lambda(x)$ for each $x \in M$. The Hahn–Banach theorem guarantees the extension of bounded linear functionals in a norm-preserving fashion, and it is this latter assertion — about norm preservation — that is the power of the theorem.

There are many versions of the Hahn–Banach theorem. The family of these existence theorems enjoy many, and varied, applications. For an account of the history and applications of these theorems, [96] is warmly recommended. As stated at the beginning of this section, the theorem is credited, independently, to Hahn and Banach. However, this is one of those situations in mathematics where there is another person who has not received proper credit. In this case, Eduard Helly (1884–1943; Austria) should perhaps be recognized as the originator of the theorem (see [65], [96]). Briefly, Helly proved a version of the Hahn–Banach theorem [62], roughly fifteen years before the publication of the proofs of Hahn and Banach. Helly then enlisted in the army, and was severely injured in World War I. Eventually, he returned to Vienna but then was forced to flee in 1938 to avoid persecution by the Nazis. These many years outside of the academic setting damaged his career in mathematics, and caused his earlier work to remain obscure. Incidentally, the Banach–Steinhaus theorem also appeared in Helly's 1912 paper.

The proof given here of the Hahn–Banach theorem uses the axiom of choice. The relationship between the Hahn–Banach theorem and the axiom of choice is discussed in [96]. The axiom of choice is, as the name suggests, an *axiom* of set theory. That is, it is a statement that cannot be deduced from the usual axioms of set theory. Its history is rich; for a discussion of the axiom of choice see [59], page 59. The axiom of choice is stated in the next section, where we use it to prove the existence of a nonmeasurable set. As you read, note that both of our applications of the axiom of choice are assertions of existence of something: an extension of a linear functional in the Hahn–Banach theorem and a nonmeasurable set in the next section.

As it turns out, the axiom of choice has several equivalent formulations, and it is one of these other forms that we will find most useful in our proof of the Hahn–Banach theorem. We will use a form known as Zorn's lemma. To state this lemma, we need some preliminary language. A *partially ordered set* is a nonempty set S together with a relation "$\preceq$" satisfying two conditions:

(i) $x \preceq x$ for each $x \in S$;
(ii) If $x \preceq y$ and $y \preceq z$, then $x \preceq z$.

If for any x and y in a partially ordered set S either $x \preceq y$ or $y \preceq x$, we say that S is a *totally ordered set*. Consider a subset T of a partially ordered set S. An element $x \in S$ is an *upper bound* for T if $y \preceq x$ for each $y \in T$. An element $x \in S$ is a *maximal* element if $x \preceq y$ implies $y \preceq x$. *Zorn's lemma* asserts that every partially ordered set S in which every totally ordered subset has an upper bound must contain a maximal element.

We are now ready to prove a version of the Hahn–Banach theorem having to do with *real-valued* linear functionals.

Theorem 6.9 (The Hahn–Banach Theorem (Real Case))**.** *Let X be a real normed linear space and let M be a subspace. If $\lambda \in M^*$, then there exists $\Lambda \in X^*$ such that $\Lambda = \lambda$ on M and $\|\Lambda\| = \|\lambda\|$.*

(It is important to keep norms straight. For example, the two norms appearing in this equality are different; the equality, more precisely written, reads: $\|\Lambda\|_{\mathcal{B}(X,\mathbb{R})} = \|\lambda\|_{\mathcal{B}(M,\mathbb{R})}$. We will indulge in the common practice of not writing these cumbersome subscripts.)

Proof. We begin by considering the real-valued function

$$p(x) = \|\lambda\| \cdot \|x\|$$

defined on all of X. Observe that

$$p(x + y) \leq p(x) + p(y) \qquad \text{and} \qquad p(\alpha x) \leq \alpha p(x)$$

for all $x, y \in X$ and $\alpha \geq 0$. Also observe that $\lambda(x) \leq p(x)$ for all $x \in M$.

Next, consider a fixed $z \in X \setminus M$. For all $x, y \in M$ we have that

$$\begin{aligned} \lambda(x) - \lambda(y) &= \lambda(x - y) \\ &\leq p(x - y) = p((x + z) + (-z - y)) \leq p(x + z) + p(-z - y). \end{aligned}$$

Hence,

$$-p(-z - y) - \lambda(y) \leq p(x + z) - \lambda(x)$$

for all $x, y \in M$. Therefore, $y \in M$ implies that

$$s = \sup_{y \in M}\{-p(-z - y) - \lambda(y)\} \leq p(x + z) - \lambda(x),$$

and hence that

$$s \leq \inf_{x \in M}\{p(x + z) - \lambda(x)\}.$$

For the z specified above, define the subspace M_z of X to be the subspace generated by M and z:

$$M_z = \{x + \alpha z \mid x \in M, \alpha \in \mathbb{R}\}.$$

Notice that the representation $w = x + \alpha z$ is unique for $w \in M_z$. Define $\overline{\lambda}(w) = \lambda(x) + \alpha s$ on M_z. Then $\overline{\lambda}$ is linear, and $\overline{\lambda}(x) = \lambda(x)$ for each $x \in M$. We have thus extended λ from M to a bigger subspace M_z of X. If M_z actually equals X, we are done (and without using Zorn's lemma!). Since $\frac{x}{\alpha} \in M$ for every $\alpha \neq 0$, we see that

$$-p(-z - \frac{x}{\alpha}) - \lambda(\frac{x}{\alpha}) \leq s \leq p(\frac{x}{\alpha} + z) - \lambda(\frac{x}{\alpha}), \qquad \alpha \neq 0.$$

From this we can deduce that $\overline{\lambda}(w) \leq p(w)$ for each $w \in M_z$ (using the first inequality if $\alpha < 0$ and the second if $\alpha > 0$).

If $M_z \neq X$, we still have work to do. Recall that we are attempting to extend λ from M to X. We could repeat the process described above, extending λ from M_z to a bigger subspace of X, but chances are, we will never reach all of X in this way. To get around this difficulty we will need to employ Zorn's lemma. Denote by S the set of all pairs (M', λ') where M' is a subspace of X containing M, and λ' is an extension of λ from M to M' satisfying $\lambda' \leq p$ on M'. Define a relation "$\preceq$" on S by

$$(M', \lambda') \preceq (M'', \lambda'')$$

if M' is a proper subspace of M'' and $\lambda' = \lambda''$ on M'. This defines a partial ordering on S. Let $T = \{(M_a, \lambda_a)\}_{a \in A}$ be a totally ordered subset of S. The pair $(\bigcup_{a \in A} M_a, \tilde{\lambda})$, where $\tilde{\lambda}(x) = \lambda_a(x)$ for $x \in M_a$, is an element of S, and is seen to be an upper bound for T. Since T was an arbitrary totally ordered subset of S, Zorn's lemma now implies that S has a maximal element, which we will call $(M_\infty, \lambda_\infty)$. Observe that λ_∞ is an extension of λ from M to M_∞ that satisfies $\lambda_\infty \leq p$ on M_∞. We aim to show that M_∞ is, in fact, all of X. If it is not, then we apply the process of extension used in passing from M to M_z to create the element $(M_{\infty z}, \lambda_{\infty z})$ of S. This element satisfies

$$(M_\infty, \lambda_\infty) \preceq (M_{\infty z}, \lambda_{\infty z}).$$

Since $(M_\infty, \lambda_\infty)$ is maximal in S, we must have that $M_\infty = M_{\infty z}$, contradicting the definition of "$\preceq$". If we now let $\Lambda = \lambda_\infty$, we have an extension of λ to all of X satisfying $\Lambda(x) \leq p(x)$ for all $x \in X$. Replacing x by $-x$ gives

$$|\Lambda(x)| \leq p(x) = \|\lambda\| \cdot \|x\|$$

for all $x \in X$, showing that $\|\Lambda\| \leq \|\lambda\|$. The only thing left to do is to show that this inequality is actually an equality. To do this, for each $\epsilon > 0$ choose $x \in M$ such that $\|x\| \leq 1$ and $|\lambda(x)| > \|\lambda\| - \epsilon$. Then $|\Lambda(x)| > \|\lambda\| - \epsilon$ also and, consequently, $\|\Lambda\| \geq \|\lambda\|$, completing the proof. □

We now prove a complex version of the Hahn–Banach theorem. The proof of the complex case was first given in 1938 [20]. The proof follows from the real case. In fact, most of the work is already done in the real case, and it seems surprising that over a decade passed between the appearance of the proofs of the two cases, especially in light of the explosive development of functional analysis at the time. We are now considering a complex linear space X, a subspace M of X, and a complex-valued bounded linear functional λ defined on M. We wish to extend this functional to one defined on all of X in a way that does not force the norm of the functional to grow. To see how this can be done, define real-valued functions λ_1 and λ_2 by

$$\lambda(x) = \lambda_1(x) + i\lambda_2(x)$$

for $x \in M$. You should check that λ_1 and λ_2 are real linear, and note that λ_1 satisfies

$$|\lambda_1(x)| \leq |\lambda(x)| \leq \|\lambda\| \cdot \|x\|$$

for each $x \in X$. If we view X as a linear space over $\mathbb{R}$, the first version of the Hahn–Banach theorem tells us that there is a real-valued bounded linear functional Λ_1 on X satisfying $\Lambda_1(x) = \lambda_1(x)$ for each $x \in M$ and $\|\Lambda_1\| \leq \|\lambda\|$. Next, define Λ on X by

$$\Lambda(x) = \Lambda_1(x) - i\Lambda_1(ix).$$

This is our desired extension, and it will not be hard to show that it has all the desired properties. First, you should check that it is a complex linear map on X. Since for $x \in M$,

$$\lambda_1(ix) + i\lambda_2(ix) = \lambda(ix) = i\lambda(x) = -\lambda_2(x) + i\lambda_1(x)$$

and λ_1 and λ_2 are real-valued, we have that

$$-\lambda_2(x) = \lambda_1(ix).$$

Thus,

$$\begin{aligned}
\lambda(x) &= \lambda_1(x) + i\lambda_2(x) \\
&= \lambda_1(x) - i\lambda_1(ix) \\
&= \Lambda_1(x) - i\Lambda_1(ix) \\
&= \Lambda(x)
\end{aligned}$$

for each $x \in M$, showing that Λ extends λ. It should be clear that $\|\lambda\| \leq \|\Lambda\|$, and we will be done when we show that this is actually an equality. For $x \in X$ write the complex number $\Lambda(x)$ in polar form $re^{i\theta}$ with nonnegative r and real θ. Then

$$|\Lambda(x)| = e^{-i\theta}\Lambda(x) = \Lambda(e^{-i\theta}x),$$

showing that $\Lambda(e^{-i\theta}x)$ is real and hence equal to its real part, $\Lambda_1(e^{-i\theta}x)$. Therefore,

$$|\Lambda(x)| = \Lambda_1(e^{-i\theta}x) \leq \|\Lambda_1\| \cdot \|e^{-i\theta}x\| \leq \|\lambda\| \cdot \|x\|,$$

implying that $\|\Lambda\| \leq \|\lambda\|$, as desired.

The Hahn–Banach theorem has many applications, both in functional analysis and in other areas of mathematics. Its applications within functional analysis focus on separation properties, and the interested reader should follow up by reading any one of the functional analysis texts in the References. We end this section with a brief discussion of the theorem's use in another area of mathematics.

One of the most important problems in an area of mathematics called potential theory is the so-called Dirichlet problem. Potential theory is a branch of partial differential equations and has its roots in problems of the 1700s such as the problem of determining gravitational forces exerted by bodies of various shapes (it was already known that the earth is some sort of ellipsoid). See Section 22.4 of [76] for more of this history.

We consider an open, bounded, *connected* (that is, it cannot be written as the disjoint union of two open subsets) set U in $\mathbb{R}^n$. The *boundary* ∂U of U is the set $\overline{U} \cap (\overline{\mathbb{R} \setminus U})$. Given a continuous function f defined on ∂U, the Dirichlet problem is to find a continuous function u defined on all of $\overline{U}$ that is a solution to the Laplacian equation

$$\frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2} = 0$$

(on U) and meets the additional requirement that u take on the same values that f does on ∂U. Usually, if such a u exists, it is not too hard to show that it is the unique solution to the Dirichlet problem (for a specified U and f). Showing that a solution exists at all is harder (much harder), and various techniques can be used. One way to show this existence is to use the Hahn–Banach theorem. Further details can be found on page 155 of [47].

6.4 The Existence of a Nonmeasurable Set

The primary goal of this section is to show that there is a subset of $\mathbb{R}$ that is not Lebesgue measurable. The example can easily be adapted to give a subset of $\mathbb{R}^n$ that is not Lebesgue measurable. The example given here is a modification of one given by Giuseppe Vitali (1875–1932; Italy) [126].

The facts we need about Lebesgue measure are the following:

(i) It is *translation invariant*. That is, $m(E) = m(x + E)$ for each $x \in \mathbb{R}$ and $E \in \mathcal{M}$. Here, the set $x + E$ is defined by

$$x + E = \{x + y \mid y \in E\}.$$

(ii) $m([0, 1)) = 1$.

(iii) It is, as is every measure, countably additive.

The second follows from the definition of m; the third is proved in Theorem 3.6. The first has not yet been mentioned. We point out that a set $E \subseteq \mathbb{R}$ is measurable if and only if $x + E$ is measurable, and now (i) follows from the next theorem.

Theorem 6.10. *Lebesgue outer measure m^* is translation invariant on $2^{\mathbb{R}}$.*

PROOF. For an interval $I = (a, b)$, $(a, b]$, $[a, b]$, or $[a, b)$, $m^*(I) = m(I) = b - a$ by definition, and it should be clear that $m^*(x + I) = m^*(I)$ for each $x \in \mathbb{R}$. Let $E \subseteq \mathbb{R}$ be arbitrary and cover E with a countable number of intervals I_n,

$$E \subseteq \bigcup_{n=1}^{\infty} I_n.$$

Then

$$x + E \subseteq \bigcup_{n=1}^{\infty}(x + I_n),$$

and we have

$$m^*(x + E) \leq \sum_{n=1}^{\infty} m(x + I_n) = \sum_{n=1}^{\infty} m(I_n).$$

Taking the infimum over all such covers of E, we get

$$m^*(x + E) \leq m^*(E).$$

Then also

$$m^*(E) \leq m^*(-x + (x + E)) \leq m^*(x + E). \qquad \square$$

We also will use the *axiom of choice*: Let I be any nonempty set. If $\{A_i : i \in I\}$ is a nonempty family of pairwise disjoint sets such that $A_i \neq \emptyset$ for each $i \in I$, then there exists a set $E \subseteq \bigcup_{i \in I} A_i$ such that $E \cap A_i$ consists of exactly one element for each $i \in I$. The axiom of choice is discussed in further detail in the preceding section. If you have not yet read that section, you should read the discussion on the axiom of choice (and the equivalent Zorn's lemma) found there before proceeding.

We now turn our attention to showing that there exists a set $E \subseteq \mathbb{R}$ that is not Lebesgue measurable. We start by defining an equivalence relation $\sim$ on $(0, 1)$ by $x \sim y$ if $x - y \in \mathbb{Q}$. The equivalence classes play the role of the A_i's in the axiom of choice as stated above. Therefore, we conclude the existence of a set E in $(0, 1)$ consisting of exactly one element of each equivalence class.

We will show that E cannot be Lebesgue measurable. Assume, to the contrary, that E is Lebesgue measurable. Let $r_1, r_2, \ldots$ be an enumeration of the rational numbers in $(-1, 1)$. Define sets

$$E_n = \{x + r_n \,|\, x \in E\}, \qquad n = 1, 2, \ldots.$$

Since m is translation invariant, $m(E_n) = m(E)$ for each n. Note that each E_n is contained in the interval $(-1, 2)$, and thus

$$\bigcup_{n=1}^{\infty} E_n \subseteq (-1, 2).$$

Also, notice that

$$(0, 1) \subseteq \bigcup_{n=1}^{\infty} E_n.$$

To see that this is the case choose any $x \in [0, 1)$ and let y be the unique element of E that is equivalent to x. Then $x - y \in \mathbb{Q} \cap (-1, 1)$ and thus must be one of the rationals r_k for some k. In this case, $x = y + r_k \in E_k \subseteq \bigcup_{n=1}^{\infty} E_n$. Finally, we note that $E_n \cap E_m = \emptyset$ for $n \neq m$. To see that this is the case, consider an element $x \in E_n \cap E_m$. Then $x = y + r_n = z + r_m$ for some $y, z \in E$. Then

$y - z = r_m - r_n \in \mathbb{Q}$, which shows that $y \sim z$ and $y \neq z$ (since $n \neq m$). This is impossible, since E contains precisely one element from each equivalence class.

Combining the observations of the previous paragraph with (i)–(iii) above, we get

$$3 = m((-1, 2)) \geq m\Big(\bigcup_{n=1}^{\infty} E_n\Big) = \sum_{n=1}^{\infty} m(E_n) = \sum_{n=1}^{\infty} m(E),$$

and therefore $m(E) = 0$. On the other hand,

$$1 = m((0, 1)) \leq m\Big(\bigcup_{n=1}^{\infty} E_n\Big) = \sum_{n=1}^{\infty} m(E_n) = \sum_{n=1}^{\infty} m(E),$$

implying that $m(E) > 0$. Clearly, we cannot have both, and we must thus conclude that E is not Lebesgue measurable.

This proof uses the axiom of choice. Interestingly, one *must* use the axiom of choice to prove the existence of sets that are not Lebesgue measurable. This follows from results of Robert Solovay [115].

6.5 Contraction Mappings

Fixed point theorems have many applications in mathematics and are also used in other areas, such as in mathematical economics (see, for example, [22], [92]). We mentioned the Schauder fixed point theorem in the proof of Theorem 5.22 about invariant subspaces. Most theorems that ensure the existence of solutions of differential, integral, and operator equations can be reduced to fixed point theorems. The theory behind these theorems belongs to topology and makes use of ideas such as continuity and compactness. Two of the most important names associated to this broad area are Henri Poincaré (1854–1912; France) and Luitzen Brouwer (1881–1966; Netherlands). In this section we will prove a fixed point theorem known as Banach's contraction mapping principle and study applications of it to differential equations. This type of application is of an aesthetically appealing nature: The result may be known and may be provable via the methods of "hard analysis," but the techniques of functional analysis reveal a beautifully enlightening proof. The proof of our theorem is a generalization of an analytic technique due to (Charles) Emile Picard (1856–1941; France).

We begin with a differential equation subject to a boundary condition:

$$\begin{aligned} \frac{dy}{dx} &= \phi(x, y), \\ y(x_0) &= y_0. \end{aligned}$$

"Finding a solution" means to construct a (necessarily continuous) function $y(x)$ that passes through the point (x_0, y_0) and has slope $\phi(x, y)$ near x_0. This solution, if it exists, will thus be an element of $C([a, b])$ for some closed interval $[a, b]$ containing x_0.

Given the above system, how do we know whether a solution exists at all? If one exists, can we tell whether it is the only one? Consider the following "easy" example:

$$\frac{dy}{dx} = y^{\frac{2}{3}},$$
$$y(0) = 0.$$

This, as you can easily check, has solutions

$$y_1(x) = 0 \quad \text{and} \quad y_2(x) = \frac{1}{27}x^3.$$

Therefore, uniqueness does not always follow from existence. We shall later see conditions ensuring uniqueness.

In 1820, Cauchy proved the first uniqueness and existence theorems for a system of type

$$\frac{dy}{dx} = \phi(x, y),$$
$$y(x_0) = y_0.$$

However, he imposed severe restrictions on ϕ, and the proof was unnecessarily complicated. There subsequently followed improvements on Cauchy's theorem and proof, including an improvement of the proof due to Picard. The result we present next is a general fixed point theorem, the proof of which uses Picard's iterative method that he employed in his version of Cauchy's theorem. We will then rephrase the problem about differential equations into the language of the fixed point theorem. Rephrasing the differential equations problem in the language of functional analysis yields a remarkably simple result with a powerful conclusion. This rephrasing in a more general setting also greatly increases the scope of applications.

At this juncture we must remember that a solution to our system is an element of $C([a, b])$, and that $C([a, b])$ endowed with the supremum norm

$$\|f\|_\infty = \sup\{|f(x)| \,\big|\, a \le x \le b\}$$

is a complete metric space.

Let $X = (X, d)$ be a metric space. A map $T : X \to X$ is a *contraction* if there exists $M \in [0, 1)$ such that $d(Tx, Ty) \le Md(x, y)$ for all $x, y \in X$.

Here is the main result of this section:

Theorem 6.11 (The Banach Contraction Mapping Principle)**.** *Let X be a complete metric space and T a contraction $X \to X$. Then there exists a unique point $x \in X$ with $Tx = x$.*

Proof. Choose any $x_0 \in X$. Put

$$x_1 = Tx_0,$$
$$x_2 = Tx_1 (= T^2 x_0),$$

$$\vdots$$
$$x_n = T^n x_0.$$

We will show that this defines a Cauchy sequence $\{x_n\}$. Then, since X is complete, we know that this sequence converges, say to x. We will finish by showing that x is a fixed point of T and that it is the only fixed point of T.

Let M be as in the definition above and note that

$$d(x_{n+1}, x_n) \leq M^n d(x_1, x_0).$$

This is left as Exercise 6.5.1.
Then, if $m > n$,

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \cdots + d(x_{n+1}, x_n) \\ &\leq (M^{m-1} + M^{m-2} + \cdots + M^n) d(x_1, x_0) \\ &\leq \left(\frac{M^n}{1-M}\right) d(x_1, x_0). \end{aligned}$$

Since

$$\frac{M^n}{1-M} \to 0$$

as $n \to \infty$, we see that $\{x_n\}_{n=1}^{\infty}$ is Cauchy. Let $x = \lim_{n\to\infty} x_n$ and notice that

$$d(Tx, x) \leq d(Tx, Tx_n) + d(Tx_n, x) \leq Md(x, x_n) + d(x_{n+1}, x).$$

Since both

$$d(x, x_n) \to 0 \quad \text{and} \quad d(x_{n+1}, x) \to 0$$

as $n \to \infty$, we see that $d(Tx, x) = 0$. In other words, $Tx = x$. To complete the proof it remains to be shown that x is the only fixed point of T. Suppose that $Ty = y$ and that $x \neq y$. Then $d(x, y) > 0$, and thus

$$d(x, y) = d(Tx, Ty) \leq Md(x, y) < d(x, y),$$

a contradiction. Therefore, it must be the case that $x = y$. This completes the proof. □

We now return to our differential equation with boundary condition:

$$\begin{aligned} \frac{dy}{dx} &= \phi(x, y), \\ y(x_0) &= y_0. \end{aligned}$$

As you should verify, this is equivalent to the integral equation

$$y(x) = y_0 + \int_{x_0}^{x} \phi(t, y(t))dt.$$

Define a map

$$T : C([a, b]) \to C([a, b])$$

by $Tf = g$, where

$$g(x) = y_0 + \int_{x_0}^{x} \phi(t, f(t))dt.$$

Any solution of the original differential system is a solution to the integral equation, which in turn is a fixed point of the map T. Let us now assume that ϕ satisfies a "Lipschitz condition" in the second variable.[4] That is, we assume that there exists a positive number K such that

$$|\phi(x, y) - \phi(x, z)| \leq K|y - z|$$

for all $x \in [a, b]$ and all $y, z \in \mathbb{R}$. In this case,

$$\begin{aligned}
|(Tf)(x) - (Tg)(x)| &= \left| \int_{x_0}^{x} [\phi(t, f(t)) - \phi(t, g(t))]dt \right| \\
&\leq \int_{x_0}^{x} |\phi(t, f(t)) - \phi(t, g(t))|dt \\
&\leq \int_{x_0}^{x} K|f(t) - g(t)|dt \\
&\leq \int_{x_0}^{x} Kd(f, g)dt \\
&\leq K(b - a)d(f, g).
\end{aligned}$$

Since this holds for all $x \in [a, b]$, we have

$$d(Tf, Tg) \leq K(b - a)d(f, g)$$

for all $f, g \in C([a, b])$. From this, we see that the map T is a contraction as long as $K(b - a) < 1$. We thus have the following corollary to Banach's theorem:

Theorem 6.12. *Let notation be as in the preceding discussion. If there exists a* $K > 0$ *with* $K(b - a) < 1$*, then there exists a unique* $f \in C([a, b])$ *such that* $f(x_0) = y_0$ *and* $f'(x) = \phi(x, f(x))$ *for all other* $x \in [a, b]$.

In some simple situations, Picard's method of successive approximation can actually be used to construct the unique solution to a differential equation with boundary condition. For example, consider

$$\begin{aligned}
f'(x) &= x + f(x), \\
f(0) &= 0.
\end{aligned}$$

This system is equivalent to the integral equation

$$f(x) = \int_{0}^{x} (t + f(t))dt.$$

[4] Named in honor of Rudolf Otto Sigismund Lipschitz (1832–1903; Königsberg, Prussia, now Kaliningrad, Russia).

Thus we put

$$(Tf)(x) = y_0 + \int_{x_0}^{x} \phi(t, f(t))dt,$$

where $y_0 = 0$, $x_0 = 0$, and $\phi(t, f(t)) = t + f(t)$. Does ϕ satisfy a Lipschitz condition? We have

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

which is less than or equal to $K|y-z|$ if we put $K = 1$. Since we need $K(b-a) < 1$ in order to apply our theorem, we just need to restrict attention to an interval $[a, b]$ of length less than one containing $x_0 = 0$. Then our theorem implies the existence and uniqueness of a solution; what is this solution? Let us choose, quite naively, $f_0(x) = 0$. Then

$$f_1(x) = (Tf_0)(x) = \int_0^x (t + f_0(t))dt = \frac{1}{2}x^2,$$

$$f_2(x) = \frac{1}{2}x^2 + \frac{1}{6}x^3,$$

$$\vdots$$

$$f_n(x) = \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{(n+1)!}x^{n+1}.$$

From this we see that $f_n(x) \to f(x) = e^x - x - 1$, and sure enough, $e^x - x - 1$ solves the original system.

There are problems with using Picard's iteration scheme. First, it may be the case that the iterates cannot be solved for with elementary functions. Or even if they can theoretically be solved for, they may be too hard to calculate. Second, even if we "have" the iterates, to figure out what they converge to may be very difficult. The power of this theorem, even though it is a "constructive" proof, is for ensuring existence and uniqueness. The theorem is also useful when one is interested in using a computer to get numerical approximations to a solution. The computational side of this problem is one that we will not go into at all, but is worthy of the interested reader's further investigation (see [92]).

6.6 The Function Space $C([a, b])$ as a Ring, and its Maximal Ideals

In this short section we show that the maximal ideals of the ring $C([a, b])$ are in one-to-one correspondence with the points of $[a, b]$. We start with definitions of rings and ideals. The intent of this section is to study the *algebraic structure* of function spaces. Though this may be of more interest to students who have studied abstract algebra, the material found here is self-contained and requires no background not already found in this text.

A *ring* is a nonempty set $\mathcal{R}$ together with two binary operations, "+" and "·", such that for all a, b, c in $\mathcal{R}$:

1. $a + b$ is in $\mathcal{R}$;
2. $a + b = b + a$;
3. $(a + b) + c = a + (b + c)$;
4. There is an element 0 in $\mathcal{R}$ such that $a + 0 = 0 + a = a$;
5. There is an element $-a$ in $\mathcal{R}$ such that $a + (-a) = 0$;
6. $a \cdot b$ is in $\mathcal{R}$;
7. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$;
8. $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$.

If, in addition, there is an element 1 in $\mathcal{R}$ such that $a \cdot 1 = 1 \cdot a = a$ for each $a \in \mathcal{R}$, $\mathcal{R}$ is said to be a *ring with identity* (or *ring with unit*). The use of the word ring in this section is different than the usage in the context of measure theory.

We really have no need for this definition, except to notice that $C([a, b])$ is a ring with identity. You are asked to prove this in Exercise 6.6.1.

A nonempty subset $\mathcal{J}$ of a ring $\mathcal{R}$ is called an *ideal* of $\mathcal{R}$ if

1. a and b both in $\mathcal{J}$ imply $a + b$ also in $\mathcal{J}$;
2. a in $\mathcal{J}$ and r in $\mathcal{R}$ imply that both ar and ra are in $\mathcal{J}$.

An ideal $\mathcal{J}$ of a ring $\mathcal{R}$ is a *proper* ideal if $\mathcal{J} \neq \mathcal{R}$, and is a *maximal* ideal if $\mathcal{J} = \mathcal{M}$ for every proper ideal $\mathcal{M}$ satisfying $\mathcal{J} \subseteq \mathcal{M}$. Recall from the Section 5.3 what a Banach algebra is, and that $C([a, b])$ is one. An ideal in a Banach algebra is, among other things, a subspace, and it makes sense to ask whether a given ideal is closed in the algebra.

Theorem 6.13. *A maximal ideal in a Banach algebra is always a closed ideal.*

PROOF. Let $\mathcal{J}$ be a maximal ideal. We first observe that the set of invertible elements in the Banach algebra is open (see the discussion following the proof of Theorem 5.6). Since $\mathcal{J}$ must have empty intersection with the set of invertible elements (Exercise 6.6.2), $\mathcal{J}$ cannot be dense. Therefore, $\overline{\mathcal{J}}$ is not the entire algebra. Also, it is straightforward to check that $\overline{\mathcal{J}}$ is an ideal. Since $\mathcal{J}$ is maximal, it must be the case that $\mathcal{J} = \overline{\mathcal{J}}$, as desired. □

In general, it can be quite a hard problem to identify the closed ideals of a Banach algebra. However, it is sometimes feasible to characterize the closed ideals that are maximal ideals. Our next theorem does just that for the Banach algebra $C([a, b])$. It must be said that this example is just the tip of a huge iceberg. First, the interval $[a, b]$ can be replaced by a much more general type of topological space (a compact Hausdorff space X). The proof in this case depends on a theorem, Urysohn's lemma,[5] from topology that is outside of the scope of this text. But even this result about $C(X)$ is far from the whole story; one might start by considering, for

[5]Due to Pavel Urysohn (1898–1924; Ukraine).

example, other types of functions in place of the continuous functions considered here. See [124], Sections VII.3–VII.5 for much more on this.

We define, for each $x \in [a, b]$, the set

$$\mathcal{J}_x = \{f \in C([a, b]) \mid f(x) = 0\}.$$

We are now ready to prove the main theorem of this section.

Theorem 6.14. *Each ideal $\mathcal{J}_x$ is a maximal ideal of $C([a, b])$, and moreover, every maximal ideal of $C([a, b])$ is of this form. Finally, $\mathcal{J}_x = \mathcal{J}_y$ if and only if $x = y$.*

Proof. Exercise 6.6.4 shows that $\mathcal{J}_x$ is a proper ideal of $C([a, b])$ for each $x \in [a, b]$. Suppose that there is a proper ideal $\mathcal{J}$ satisfying $\mathcal{J}_x \subseteq \mathcal{J}$. For $x \in [a, b]$ define $\lambda_x : C([a, b]) \to \mathbb{R}$ by $\lambda_x(f) = f(x)$. Since

$$\lambda_x(f + \alpha g) = (f + \alpha g)(x) = f(x) + \alpha g(x) = \lambda_x(f) + \alpha\lambda_x(g)$$

for all $f, g \in C([a, b])$ and each real number α, it follows that λ_x is a linear functional. Since λ_x is not identically zero, $\mathcal{J}_x = \ker \lambda_x$ is a proper subspace of $C([a, b])$. Choose any element $f \in C([a, b]) \setminus \mathcal{J}_x$. For any $g \in C([a, b])$, the element

$$g - \frac{\lambda_x(g)}{\lambda_x(f)} f$$

is in $\mathcal{J}_x$. So g is in the subspace generated by $\mathcal{J}_x$ and f; since g was arbitrary, the subspace generated by $\mathcal{J}_x$ and f is all of $C([a, b])$. This shows that $\mathcal{J}_x$ is a maximal subspace and hence must be a maximal ideal (any ideal is a subspace, so if $\mathcal{J}_x$ cannot even fit inside a proper subspace, there is no hope that it might fit inside a proper ideal).

We have now shown that each ideal of the form $\mathcal{J}_x$ is a maximal ideal. We aim to show that in fact, these are the only maximal ideals. Assume that $\mathcal{J}$ is a proper ideal. We want to show that there exists $x \in [a, b]$ such that $\mathcal{J} \subseteq \mathcal{J}_x$. Suppose, to the contrary, that $\mathcal{J} \nsubseteq \mathcal{J}_x$ for every $x \in [a, b]$. Then, for every $x \in [a, b]$, there is a function $f_x \in \mathcal{J}$ with $f_x(x) \neq 0$. Each of these functions is continuous, and thus there exist open intervals U_x on which f_x^2 is a strictly positive function. Notice that these open intervals form an open cover for the compact set $[a, b]$. Thus there exists a finite number of points $x_1, \dots, x_n$ in $[a, b]$ such that the function

$$f = f_{x_1}^2 + \cdots + f_{x_n}^2$$

is strictly positive on all of $[a, b]$. By construction, $f \in \mathcal{J}$ and f is invertible, contradicting the result of Exercise 6.6.2. Thus, $\mathcal{J} \subseteq \mathcal{J}_x$ for some $x \in [a, b]$, and if $\mathcal{J}$ is maximal, $\mathcal{J} = \mathcal{J}_x$.

The final assertion of the theorem is left as Exercise 6.6.5. $\square$

6.7 Hilbert Space Methods in Quantum Mechanics

Quantum mechanics attempts to describe and account for the properties of molecules and atoms and their constituents. The first attempts, in large part due to Niels Bohr (1885–1962; Denmark), had limited success and are now known as the "old quantum theory." The "new quantum theory" was developed around 1925 by Werner Heisenberg (1901–1976; Germany) and Erwin Schrödinger (1887–1961; Austria), and later was extended by Paul Dirac (1902–1984; England). The story of the history and development of quantum theory is extremely interesting, for many different reasons, and includes deep philosophical questioning, personal and political intrigue, as well as fascinating mathematics and physics. There are many very good books on the subject that address these different facets, and it would be difficult to give a short recommended reading list. It is the aim of this section to give some inkling of how Hilbert space and operator theory can be used in quantum mechanics. Our treatment is not rigorous. Good references that contain more details on what we introduce here include [39], [43], [113].

For simplicity, we restrict ourselves to a consideration of one-dimensional motion. That is, we consider a particle (such as an electron) moving along a straight line, so that there is a function $f(x, t)$ of position x (on that straight line) and time t such that the probability that the particle is in the interval $[a, b]$ at time t is given by

$$\int_a^b |f(x, t)|^2 dx.$$

The (complex-valued) function f is called the *state function* for the particle. It should be clear that we expect

$$\int_{-\infty}^{\infty} |f(x, t)|^2 dx = 1,$$

for each fixed value of t. We now consider t to be fixed, and write $f(x)$ in place of $f(x, t)$. In summary, the state of a quantum particle in one dimension is a function f in the Hilbert space $L^2(\mathbb{R})$ satisfying $\|f\|^2 = \int_{-\infty}^{\infty} |f(x)|^2 dx = 1$.

The particle's position, x, is an example of an *observable*, that is, a quantity that can be measured. Other observables useful in quantum mechanics include the particle's momentum and energy. We will study position and momentum, and discuss Heisenberg's uncertainty principle. Throughout this section we are taking *Planck's constant*, $\hbar$, to be 1.

The *Fourier transform*[6] of any $L^2(\mathbb{R})$ function f is another square integrable function $\hat{f}$, given by

$$\hat{f}(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-iwt} dt.$$

[6]The Fourier transform is a very useful object and not beyond the level of this book. A strong temptation to include more on this topic was resisted. Further investigation of the Fourier transform's properties and applications would make a nice project.

It is not obvious that this makes sense for all $L^2(\mathbb{R})$ functions; we refer the interested reader to Chapter 7 of [43]. In addition to the fact that the Fourier transform makes sense, we will use one of its most fundamental properties, without supplying a proof, that

$$\|f\|_2 = \|\hat{f}\|_2$$

for each f. This equation is called the *Plancherel identity*, named in honor of Michel Plancherel (1885–1967; Switzerland).

If w denotes the momentum of the particle, then the Fourier transform of the state function can be used to give the probability that w is in the interval $[a, b]$. The probability that the momentum of the particle is in $[a, b]$ is given by

$$\int_a^b |\hat{f}(w)|^2 dw.$$

If $\overline{x}, \overline{w}$ denote the average, or *mean*, values of position x and momentum w, respectively, then

$$\overline{x} = \int_{-\infty}^{\infty} x|f(x)|^2 dx, \qquad \overline{w} = \int_{-\infty}^{\infty} w|\hat{f}(w)|^2 dw,$$

and the *variance* of each is given by

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \overline{x})^2 |f(x)|^2 dx, \qquad \sigma_w^2 = \int_{-\infty}^{\infty} (w - \overline{w})^2 |\hat{f}(w)|^2 dw.$$

Figure 6.3 gives some illustration of what the size of the variance tells us.

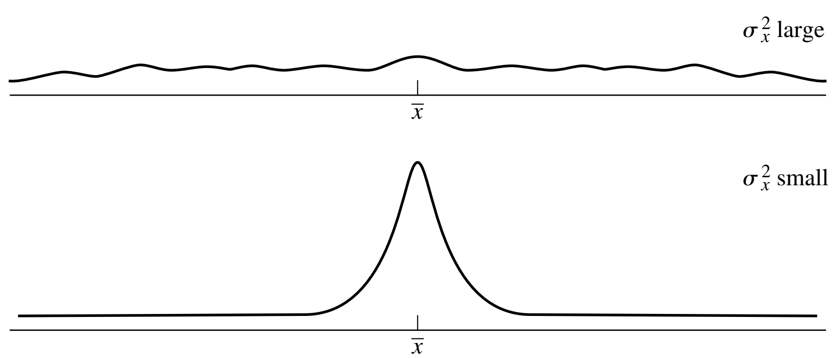

FIGURE 6.3. $|f(x)|^2$ will resemble the top graph if σ_x^2 is big, and the bottom graph if σ_x^2 is small.

Heisenberg's uncertainty principle says in this context that σ_x^2 and σ_w^2 cannot be "small" simultaneously. Specifically,

$$\sigma_x^2 \cdot \sigma_w^2 \geq \frac{1}{4}.$$

From Figure 6.3, this means, informally, that position and momentum cannot be "localized" simultaneously.

We justify this inequality for $\overline{x} = 0 = \overline{w}$ (this assumption is not very restrictive). Define operators M (for "multiplication") and D (for "differentiation") on $L^2(\mathbb{R})$ by

$$Mf(x) = xf(x) \qquad \text{and} \qquad Df(x) = f'(x).$$

We point out that these are not defined on all of $L^2(\mathbb{R})$, but for our superficial treatment we will not worry about this and will always take for granted that f is a member of the "right" domain. It should be clear that

$$\sigma_x^2 = \|Mf\|^2.$$

Also,

$$\sigma_w^2 = \int_{-\infty}^{\infty} w^2 |\hat{f}(w)|^2 dw = \int_{-\infty}^{\infty} |\widehat{Df}(w)|^2 dw = \|Df\|^2$$

(in this, the first equality should be clear; a justification of the second is asked for in Exercise 6.7.3; the third follows from the Plancherel identity). Since

$$(xf(x))' = f(x) + xf'(x),$$

we have that

$$D(Mf) = f + M(Df),$$

or, in "operator" form,

$$DM - MD = I$$

on the intersection of the subspaces of $L^2(\mathbb{R})$ on which M and D are defined. It is straightforward to see that M is Hermitian, that is, $\langle Mf, g\rangle = \langle f, Mg\rangle$, for all $f, g \in L^2(\mathbb{R})$. It is also true, but not as easy to show, that $\langle DMf, f\rangle = -\langle Mf, Df\rangle$ for all $f \in L^2(\mathbb{R})$. As you can check, this boils down to showing that

$$\int_{-\infty}^{\infty} x(|f(x)|^2)' dx = -\int_{-\infty}^{\infty} |f(x)|^2 dx,$$

and you can find a proof of this equality in Section 2.8 of [39]. Then, using the Cauchy–Schwarz inequality (Theorem 1.3),

$$\begin{aligned} 1 &= \|f\|^2 = \langle f, f\rangle \\ &= \Big\langle (DM - MD)f, f \Big\rangle \\ &= \langle DMf, f\rangle - \langle MDf, f\rangle \\ &= -\langle Mf, Df\rangle - \langle Df, Mf\rangle \end{aligned}$$

$$\begin{aligned}&\leq 2\Big|\langle Mf, Df\rangle\Big|\\&\leq 2\|Mf\|\cdot\|Df\|\\&= 2\sigma_x\cdot\sigma_w,\end{aligned}$$

which yields the desired result that $1 \leq 4\sigma_x^2\sigma_w^2$.

We pointed out above that the operators of quantum mechanics we have discussed are not defined on all elements of the Hilbert space $L^2(\mathbb{R})$. In fact, the setting of bounded operators on Hilbert space is *not* appropriate for this context. That the operators satisfy $DM - MD = I$ is critically important, and this next result tells us that this could not happen in the $B(H)$ setting.

Theorem 6.15. *There do not exist bounded linear operators S and T on any Hilbert space that satisfy $ST - TS = I$.*

Proof. Suppose, to the contrary, that there is a Hilbert space H and operators $S, T \in B(H)$ that satisfy $ST - TS = I$. We will prove, using induction, that

$$nT^{n-1} = ST^n - T^nS$$

for each positive integer n. The case $n = 1$ is our hypothesis. Assume that it holds for $n > 1$. Then

$$\begin{aligned}(n+1)T^n &= nT^{n-1}T + T^nI\\&= (ST^n - T^nS)T + T^n(ST - TS)\\&= ST^{n+1} - T^nST + T^nST - T^{n+1}S\\&= ST^{n+1} - T^{n+1}S.\end{aligned}$$

Therefore,

$$nT^{n-1} = ST^n - T^nS$$

holds for each positive integer n. Recalling that the operator norm is submultiplicative, an application of the triangle inequality yields

$$n\|T^{n-1}\| \leq 2\|S\|\cdot\|T\|\cdot\|T^{n-1}\|$$

for each n. This tells us that either $\|T^{n-1}\| = 0$ for some n, or that $n \leq 2\|S\|\cdot\|T\|$ for all n. Since the latter cannot happen, we have that $\|T^{n-1}\| = 0$ for some value of n. Therefore, $T^{n-1} = 0$. Since

$$nT^{n-2} = ST^{n-1} - T^{n-1}S,$$

we deduce that $T^{n-2} = 0$. We repeat this argument n times and ultimately deduce that $I = 0$, a clear contradiction. □

In finite dimensions there is an alternative, and basic, proof for this theorem: the trace of any matrix of form $ST - TS$ is zero, while the trace of the $n \times n$ identity matrix is n. There are, necessarily infinite, matrices that satisfy the equation

$ST - TS = I$; for example, take

$$S = \begin{pmatrix} 0 & 1 & 0 & 0\dots \\ 0 & 0 & 2 & 0\dots \\ 0 & 0 & 0 & 3\dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \text{ and } T = \begin{pmatrix} 0 & 0 & 0 & 0\dots \\ 1 & 0 & 0 & 0\dots \\ 0 & 1 & 0 & 0\dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Notice, though, that S does not define a bounded operator on the sequence space ℓ^2. See [31] for more on the equation $ST - TS = I$ and its role in the matrix mechanics of Heisenberg, Born, and Jordan, and for more on algebraic structures in quantum mechanics.

The uncertainty principle has other interpretations. For example, $f(t)$ might represent the amplitude of a signal (like a sound wave) at time t. We have changed the name of the independent variable from x to t for what we hope is an obvious reason.

Example 1. Fix real numbers θ and a, and let

$$f(t) = \begin{cases} \frac{1}{\sqrt{2a}} e^{i\theta t} & \text{if } t \in (-a, a), \\ 0 & \text{otherwise.} \end{cases}$$

In Exercise 6.7.2 you are asked to show that $\int_{-\infty}^{\infty} |f(t)|^2 dt = \int_{-a}^{a} |f(t)|^2 dt = 1$ and that

$$\hat{f}(w) = \frac{1}{\sqrt{\pi a}} \frac{\sin a(w - \theta)}{w - \theta}$$

(Figure 6.4).

Since $\int_{-\infty}^{\infty} |\hat{f}(w)|^2 dw$ must be 1, we observe that if a is large (that is, if the time duration of $f(t)$ is big), then the frequencies of f are near to θ. Likewise, if a is small, then the frequencies are spread out (Figure 6.5).

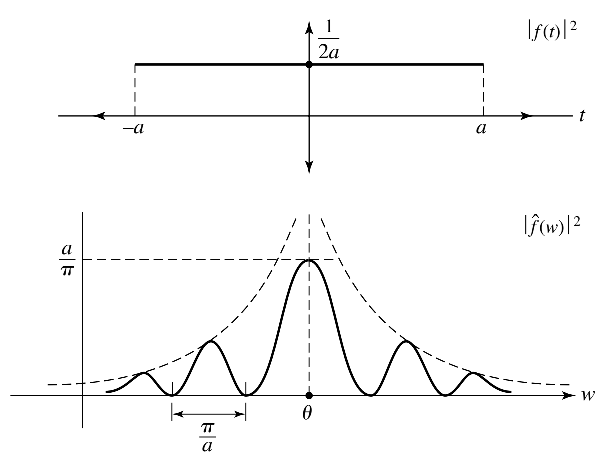

FIGURE 6.4. The area under each curve must be 1.

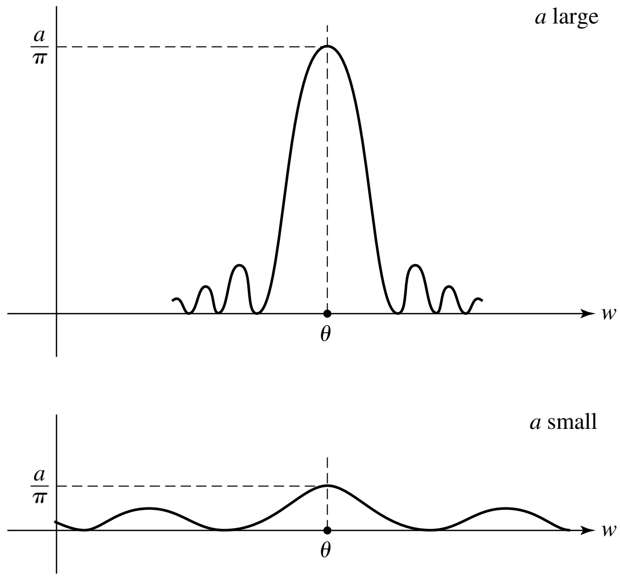

FIGURE 6.5. $|\hat{f}(w)|^2$, for large and small values of a.

We now return to the general case where $f(t)$ represents the amplitude of a signal at time t. Let a and b be positive numbers and set

$$\alpha^2 = \int_{-a}^{a} |f(t)|^2 dt$$

and

$$\beta^2 = \int_{-b}^{b} |\hat{f}(w)|^2 dw.$$

Observe that the ordered pair (α, β) is in the unit square $[0, 1] \times [0, 1]$. If $\alpha = 1$ (as is the case in Example 1), then the signal is "time-limited" (the signal is confined to the interval $[-a, a]$); $\beta = 1$ means that the signal is "band-limited." Can the ordered pair (α, β) be anywhere in the unit square? As it turns out, there is a positive number $\lambda_1 < 1$ such that

$$\arccos \alpha + \arccos \beta \geq \arccos \sqrt{\lambda_1}. \tag{6.8}$$

This assertion is an "uncertainty principle." It tells us that there are some ordered pairs (α, β) in the square that are not allowed. For example, since $\lambda_1 < 1, \alpha = \beta = 1$ cannot be achieved (for any values of a and b). This version of the uncertainty principle was proved during the years 1961–1962 by three mathematicians working at Bell Labs (see Section 2.9 of [39] for a complete reference to this theorem and for its proof). Some questions are apparent. What is this number λ_1 (and why the funny name λ_1)? As you will read, λ_1 is a function of a and b. For a given pair a and b, and hence a specified λ_1, why must α and β satisfy (6.8)? And can we identify the portion of the unit square that the points (α, β) fill in? To begin to address these questions, consider a and b as fixed positive numbers and consider the two closed subspaces M and N of $L^2(\mathbb{R})$ defined by

$$M = \{f \in L^2(\mathbb{R}) \,\big|\, f(t) = 0 \text{ for } t \notin [-a, a]\}$$

and

$$N = \{f \in L^2(\mathbb{R}) \,\big|\, \hat{f}(w) = 0 \text{ for } w \notin [-b, b]\}.$$

M is the class of time-limited functions, while N is the class of band-limited functions. Next, we consider two linear operators $T_{t\ell}$ and $T_{b\ell}$ on $L^2(\mathbb{R})$ defined by

$$T_{t\ell} f(t) = \begin{cases} f(t) & \text{if } t \in [-a, a], \\ 0 & \text{otherwise,} \end{cases}$$

and

$$T_{b\ell} f(t) = \frac{1}{\sqrt{2\pi}} \int_{-b}^{b} \hat{f}(w) e^{iwt} dw.$$

These operators are the projections from $L^2(\mathbb{R})$ onto M and N, respectively. We are interested in the operator

$$T = T_{b\ell} T_{t\ell}.$$

This operator is given by the formula (Exercise 6.7.4)

$$T f(t) = \int_{-a}^{a} \frac{\sin b(t-s)}{\pi(t-s)} f(s) ds.$$

As it turns out, T has a countable number of real eigenvalues. The number λ_1 in (6.8) is the largest, and these eigenvalues satisfy

$$1 > \lambda_1 \geq \lambda_2 \geq \cdots \geq 0.$$

The number λ_1 is a function of the values of a and b only (because $T_{t\ell}$ and $T_{b\ell}$ are) and, in fact, depends only on the value of the product ab. As this product gets large, λ_1 approaches 1, as in Figure 6.6. As $ab \to \infty$, (6.8) thus implies that more and more points (α, β) are allowable; this is as one might expect from the definitions of α and β.

We have now described what λ_1 is. Next, let us think a bit about why (6.8) might be true. A consequence of (6.8) is that as one of α or β nears 1, the other must get smaller. Certainly, the restriction

$$\alpha^2 + \beta^2 \leq 1$$

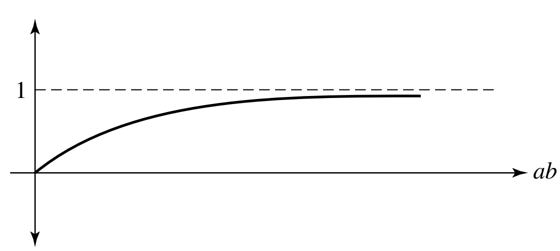

FIGURE 6.6. λ_1 as a function of ab.

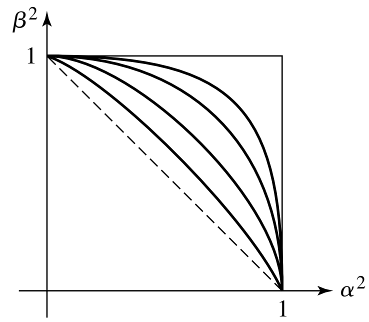

FIGURE 6.7. The curve $\arccos\alpha + \arccos\beta = \arccos\sqrt{\lambda_1}$, for various values of ab. As $ab \to 0$ this curve approaches the diagonal. Points beneath the curve satisfy (6.8).

would keep α and β from being near 1 simultaneously. In Exercise 6.7.5 you are asked to show that (6.8) is true whenever $\alpha^2 + \beta^2 \le 1$. Are there any allowable points (α, β) inside the unit square but outside the circle $\alpha^2 + \beta^2 = 1$? That is, can one find $f \in L^2(\mathbb{R})$ that gives (α, β) outside the circle? The answer is yes if and only if (α, β) satisfies (6.8). The rest of the proof of the Bell Labs result involves showing that for each given a and b (so that λ_1 is determined), $f \in L^2(\mathbb{R})$ implies that (6.8) holds, and conversely, if α and β satisfy (6.8), then there exists f in $L^2(\mathbb{R})$ such that

$$\alpha^2 = \int_{-a}^{a} |f(t)|^2 dt \qquad \text{and} \qquad \beta^2 = \int_{-b}^{b} |\hat{f}(w)|^2 dw.$$

Finally, Figure 6.7 shows the curve $\arccos\alpha + \arccos\beta = \arccos\sqrt{\lambda_1}$, for various values of the product ab.

Of all the individuals profiled in this book, **John von Neumann** probably enjoys the greatest degree of name recognition (Figure 6.8). In fact, I am fairly confident that he is the only one for whom an obituary appeared in *Life* magazine (February 1957). Also, an interview with his (second) wife appeared in *Good Housekeeping* (September 1956).

He was born on December 28, 1903, in Budapest, Hungary. His father was a banker in a well-to-do Jewish family. Von Neumann's brilliance was apparent even at a very young age, and there are many stories about his precociousness and the exceptional mental capabilities that remained with him throughout his life. It is said that he had a photographic memory. While this is impressive and interesting, it is not what he is remembered for. Throughout his life he could grasp very difficult concepts extremely quickly. In addition to being very sharp of mind, von Neumann worked tirelessly. He published approximately 60 articles in pure mathematics and 20 in physics. Altogether he published over 150 papers, most of the rest on applications to economics and computer science.

Trained as a chemist as well as a mathematician, von Neumann was well prepared for the scientific career that he

FIGURE 6.8. John von Neumann.

FIGURE 6.9. Stamp in honor of von Neumann.

would ultimately have. He attended good schools, and was awarded the Ph.D. from the University of Budapest in 1926. His thesis was about set theory. He worked in Germany until 1930, working mostly on the new quantum-mechanical theory and operator theory. He extended Hilbert's spectral theory from bounded to unbounded operators. This work paralleled, in large part, Stone's work at the time, but the two worked independently. Von Neumann published his great book uniting quantum mechanics and operator theory, ***Mathematische Grundlagen der Quantenmechanik***, in 1932. Notice that in this same year both Banach's and Stone's books also appeared.

In 1930 von Neumann came to the United States, becoming one of the original six members of the Institute for Advanced Study at Princeton. As mentioned already, his early work focused on set theory, quantum mechanics, and operator theory. His famous proof of the "ergodic theorem" came in the early 1930s. The techniques that he developed in this context served him later when he studied rings of operators, which became his focus later in the 1930s. "Rings of operators" are now called "operator algebras"; an important subclass of these are the "von Neumann algebras." In 1933, building on Haar's work on measures, von Neumann solved an important special case of the fifth of Hilbert's 23 problems.[7]

Around 1940, von Neumann changed the focus of his work from pure to applied mathematics. During World War II he did much work for a variety of government and civil agencies. He wrote extensively on topical subjects, including ballistics, shock waves, and aerodynamics. In addition to the over 150 published papers, he wrote many more that remain unpublished for security reasons. Von Neumann held strong political views, and was very much involved with the controversial scientific politics of World War II and the subsequent Cold War. He was one of the key players in the creation of the atomic bomb at the Los Alamos Scientific Laboratory, and later served on the Atomic Energy Commission (appointed by President Eisenhower in 1955). We will not go into his politics here; there exist extensive accounts of this in the literature. The book by Macrae [87] and its bibliography make a good starting point.

[7] These problems are described in the biographical material on Hilbert.

Also post-1940, he worked on mathematical economics. He is credited with the first application of game theory to economics. His theory of mathematical economics is based on his *minimax theorem*, which he had proved much earlier, in 1928 [97]. This theory was laid out in ***The Theory of Games and Economic Behavior***, written jointly with Oskar Morgenstern (1902–1977; Germany) and published in 1944. Their book is now a classic.

In the years following World War II, von Neumann devoted considerable attention to the development of the modern computer. He was interested in every aspect of computing. He made significant contributions in several different areas, including parallel processing and errors involved with large computations (specifically, inverting large matrices and Monte Carlo methods). He also worked on questions about weather forecasting, and his work has had an impact on this field. In a more philosophical area, he drew an analogy between the computer and the human nervous system.

John von Neumann died on February 8, 1957, in Washington, D.C.

Exercises for Chapter 6

Section 6.1

6.1.1 Describe the Bernstein polynomials for the two functions $f(x) = x$ and $g(x) = x^2$ on $[0, 1]$.

6.1.2 Prove that $(1 + x)^n \geq 1 + nx$ for all $x \geq -1$ and each positive integer n.

6.1.3 **(a)** Complete the proof of the first lemma used in the proof of the Stone–Weierstrass theorem by showing that the described g satisfies (i)–(iii).

(b) Complete the proof of the second lemma used in the same theorem.

6.1.4 In the Weierstrass approximation theorem, it is crucial that the interval is compact, as addressed in Stone's generalization. One of our favorite noncompact subsets of $\mathbb{R}$ is $\mathbb{R}$ itself. Show that uniform polynomial approximation is not always guaranteed on $\mathbb{R}$. Here are two approaches: (i) Come up with a function in $C(\mathbb{R})$ that cannot be uniformly approximated by polynomials; (ii) show that the uniform limit of polynomials $\mathbb{R} \to \mathbb{R}$ is still a polynomial.

6.1.5 Deduce the Weierstrass approximation theorem from the Stone–Weierstrass theorem (thus showing that the former is indeed a special case of the latter).

6.1.6 Consider the collection $\mathcal{P}_e$ of all even polynomials.

(a) Use the Stone–Weierstrass theorem to show that $\mathcal{P}_e$ is dense in $C([0, 1])$.

(b) Explain why the Stone–Weierstrass theorem cannot be used to show that $\mathcal{P}_e$ is dense in $C([-1, 1])$.

(c) The following question remains: Is $\mathcal{P}_e$ dense in $C([-1, 1])$? Answer this question, and prove your assertion.

6.1.7 Let A be as in the hypotheses of Bishop's theorem.

(a) Prove that functions of the form $f + \overline{f}$, for $f \in A$, are real-valued and separate the points of X.
(b) Deduce that an $\overline{A}$-symmetric subset of X must be a singleton.

6.1.8 Find (in another book) another proof of the Weierstrass approximation theorem. Write up your own account of this proof. For example, if you have read the final section of the chapter, you may want to find the proof that makes use of the Fourier transform.

Section 6.2

6.2.1 Show that X is nowhere dense in M if and only if its closure has empty interior, $(\overline{X})^\circ = \emptyset$.

6.2.2 Show that any finite subset of $\mathbb{R}$ is nowhere dense in $\mathbb{R}$. Give examples to show that a countable set can be nowhere dense in $\mathbb{R}$, but that this is not always the case.

6.2.3 Let $f(x) = \sin\left(\frac{1}{x}\right)$ for $x \neq 0$, and $f(0) = 0$. Compute $\omega_f(0)$. How does this tie in with what you know about the continuity of this function?

6.2.4 First category sets are in some sense "small," while second category sets are "large." Another notion of set size that we have discussed is "measure zero": "Small" sets are of measure zero, while "large" sets are not. Is there a connection between these two ways of describing the size of a set? The answer is no. Please show this by doing two things:

(a) Describe a set that is of first category, but not of measure zero.
(b) Describe a set that is of measure zero, but not of first category.

6.2.5 Prove that the function defined on $(0, 1)$ by

$$g(x) = \begin{cases} \frac{1}{q} & \text{if } x = \frac{p}{q} \text{ in reduced form,} \\ 0 & \text{if } x \text{ is irrational,} \end{cases}$$

is continuous at each irrational point of $(0, 1)$ and discontinuous at each rational point of $(0, 1)$.

6.2.6 The goal of this exercise is to show that "most" continuous functions are nowhere differentiable. Let E_n denote the set of all $f \in C([0, 1])$ for which there exists an $x_f \in [0, 1]$ satisfying

$$|f(x) - f(x_f)| \leq n|x - x_f|$$

for every $x \in [0, 1]$.

(a) Show that E_n is nowhere dense in $C([0, 1])$. To do this, approximate $f \in C([0, 1])$ by a piecewise linear function g whose pieces each have slope $\pm 2n$. Then, if $\|h - g\|_\infty$ is sufficiently small, the function h cannot be in E_n.
(b) Deduce from (a) that the nowhere differentiable functions are of second category in $C([0, 1])$.

Section 6.3

6.3.1 Let X and Y be Banach spaces. By considering the right shift on ℓ^2 into itself defined by $S(x_1, x_2, \ldots) = (0, x_1, x_2, \ldots)$, we see that the property of being one-to-one is not enough to guarantee that a bounded linear operator be invertible (compare what happens in the finite-dimensional case). However, if a one-to-one bounded linear operator is also onto, then it must be invertible. Use the open mapping theorem to prove that the inverse is also bounded.

6.3.2 Let X be a Banach space in the two different norms $\|\cdot\|_1$ and $\|\cdot\|_2$.

(a) Show that if there exists a constant M satisfying $\|x\|_1 \le M\|x\|_2$ for all $x \in X$, then the two norms are equivalent.
(b) Does the result of (a) contradict the result of Exercise 5.2.2? Explain your answer to this question.

Section 6.4

6.4.1 Show that $x \sim y$ if $x - y \in \mathbb{Q}$ defines an equivalence relation on the open interval $(0, 1)$.

Section 6.5

6.5.1 Give an inductive proof of the inequality $d(x_{n+1}, x_n) \le M^n d(x_1, x_0)$, $n = 1, 2, \ldots$, and thereby complete the proof of the contraction mapping theorem.

6.5.2 Show that the method of successive approximations applied to the differential equation $f' = f$ with $f(0) = 1$ yields the usual formula for e^x.

6.5.3 For each of the following sets give an example of a continuous mapping of the set into itself that has no fixed point:

(a) $\mathbb{R}$.
(b) $(0, 1]$.
(c) $[-10, -7] \cup [2, 4]$.

6.5.4 Give an example of a mapping of $[0, 1]$ into itself that is not continuous and has no fixed points.

6.5.5 Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} x + e^{\frac{-x}{2}} & \text{if } x \ge 0, \\ e^{\frac{x}{2}} & \text{if } x < 0. \end{cases}$$

Show that $|f(x) - f(y)| < |x - y|$ for all x and y, yet f has no fixed point. Does this contradict the contraction mapping principle? Explain.

Section 6.6

6.6.1 For f and g in $C([a,b])$ define $f+g$ and $f \cdot g$ by

$$(f+g)(x) = f(x) + g(x) \qquad \text{and} \qquad (f \cdot g)(x) = f(x) \cdot g(x).$$

Show that $C([a,b])$ is a ring with identity.

6.6.2 Prove that a proper ideal of a ring with identity contains no invertible elements.

6.6.3 If $\mathcal{J}$ is an ideal in a Banach algebra, then $\overline{\mathcal{J}}$ is also an ideal. (In this statement, "Banach algebra" can be replaced by "normed ring.")

6.6.4 Prove that $\mathcal{J}_x$ is a proper ideal of $C([a,b])$.

6.6.5 Prove that for $x, y \in [a,b]$, $\mathcal{J}_x = \mathcal{J}_y$ if and only if $x = y$. (In proving this, you may find yourself assuming that $x \neq y$ and constructing an actual function f that is in one of the ideals but is not in the either. It is this step that requires Urysohn's lemma when $[a,b]$ is replaced by a compact Hausdorff space; see the paragraph preceding Theorem 6.14. In fact, in that case f is not actually constructed but only its existence implied.)

Section 6.7

6.7.1 Fix a positive real number a and let $\chi_{[-a,a]}$ denote the characteristic function

$$\chi_{[-a,a]}(t) = \begin{cases} 1 & \text{if } t \in [-a,a], \\ 0 & \text{otherwise.} \end{cases}$$

Show that the Fourier transform of $\chi_{[-a,a]}$ is given by

$$\widehat{\chi}_{[-a,a]}(w) = \frac{2\sin(aw)}{w}.$$

6.7.2 As in Example 1, fix real numbers θ and a, and let

$$f(t) = \begin{cases} \frac{1}{\sqrt{2a}} e^{i\theta t} & \text{if } t \in (-a,a), \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\int_{-\infty}^{\infty} |f(t)|^2 dt = 1$ and

$$\hat{f}(w) = \frac{1}{\sqrt{\pi a}} \frac{\sin a(w-\theta)}{w-\theta}.$$

Use that $e^{i\theta} = \cos\theta + i\sin\theta$ for every real number θ.

6.7.3 In the section we claimed that

$$\sigma_w^2 = \int_{-\infty}^{\infty} w^2 |\hat{f}(w)|^2 dw = \int_{-\infty}^{\infty} |\widehat{Df}(w)|^2 dw = \|Df\|^2.$$

The point of this exercise is to verify the middle equality, and hence all three equalities. Note that

$$\widehat{Df}(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(t) e^{-iwt} dt.$$

Use integration by parts to show that this equals $iw\hat{f}(w)$, and hence that the equation holds.

6.7.4 Verify the formula

$$Tf(t) = \int_{-a}^{a} \frac{\sin b(t-s)}{\pi(t-s)} f(s)ds$$

for $T = T_{b\ell}T_{t\ell}$, as used in the section.

6.7.5 Show that (6.8) (in the text of the section) is true whenever $\alpha^2 + \beta^2 \leq 1$.

http://www.springer.com/978-0-387-95224-6

Beginning Functional Analysis
Saxe, K.
2002, XI, 197 p., Hardcover
ISBN: 978-0-387-95224-6