Chapter 6

Sequences and Series of Functions

6.1 Discussion: The Power of Power Series

In 1689, Jakob Bernoulli published his *Tractatus de seriebus infinitis* summarizing what was known about infinite series toward the end of the 17th century. Full of clever calculations and conclusions, this publication was also notable for one particular question that it didn't answer; namely, what is the precise value of the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots .$$

Bernoulli convincingly argued that $\sum 1/n^2$ converged to something less than 2 (see Example 2.4.4) but he was unable to find an explicit expression for the limit. Generally speaking, it is much harder to sum a series than it is to determine whether or not it converges. In fact, being able to find the sum of a convergent series is the exception rather than the rule. In this case, however, the series $\sum 1/n^2$ seemed so elementary; more elementary than, say, $\sum_{n=1}^{\infty} n^2/2^n$ or $\sum_{n=1}^{\infty} 1/n(n+1)$, both of which Bernoulli was able to handle. "If anyone finds and communicates to us that which has so far eluded our efforts," Bernoulli wrote, "great will be our gratitude."[1]

Geometric series are the most prominent class of examples that can be readily summed. In Example 2.7.5 we proved that

$$(1) \qquad \frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$$

[1]As quoted in [12], which contains a much more thorough account of this story.

S. Abbott, *Understanding Analysis*, Undergraduate Texts in Mathematics, DOI 10.1007/978-1-4939-2712-8_6

for all $|x| < 1$. Thus, for example, $\sum_{n=0}^{\infty} 1/2^n = 2$ and $\sum_{n=0}^{\infty}(-1/3)^n = 3/4$. Geometric series were part of mathematical folklore long before Bernoulli; however, what was relatively novel in Bernoulli's time was the idea of operating on infinite series such as (1) with tools from the budding theory of calculus. For instance, what happens if we take the derivative on each side of equation (1)? The left side is easy enough—we just get $1/(1-x)^2$. But what about the right side? Adopting a 17th century mindset, a natural way to proceed is to treat the infinite series as a polynomial, albeit of infinite degree. Differentiation across equation (1) in this fashion gives

$$\text{(2)} \qquad \frac{1}{(1-x)^2} = 0 + 1 + 2x + 3x^2 + 4x^3 + \cdots .$$

Is this a valid formula, at least for values of x in $(-1, 1)$? Empirical evidence suggests it is. Setting $x = 1/2$ we get

$$4 = \sum_{n=1}^{\infty} \frac{n}{2^{n-1}} = 1 + 1 + \frac{3}{4} + \frac{4}{8} + \frac{5}{16} + \cdots ,$$

which feels plausible, and is in fact true. Although not Bernoulli's requested series, this does suggest a possible new line of attack.

Manipulations of this sort can be used to create a wide assortment of new series representations for familiar functions. Substituting $-x^2$ for x in (1) gives

$$\text{(3)} \qquad \frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + x^8 - \cdots ,$$

for all $x \in (-1, 1)$.

Once again closing our eyes to the potential danger of treating an infinite series as though it were a polynomial, let's see what happens when we take antiderivatives. Using the fact that

$$(\arctan(x))' = \frac{1}{1+x^2} \qquad \text{and} \qquad \arctan(0) = 0,$$

equation (3) becomes

$$\text{(4)} \qquad \arctan(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots .$$

Plugging $x = 1$ into equation (4) yields the striking relationship

$$\text{(5)} \qquad \frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \cdots .$$

The constant π, which arises from the geometry of circles, has somehow found its way into an equation involving the reciprocals of the odd integers. Is this a valid formula? Can we really treat the infinite series in (3) like a finite polynomial? Even if the answer is yes there is still another mystery to solve in this example.

Plugging $x = 1$ into equations (1), (2), or (3) yields mathematical gibberish, so is it prudent to anticipate something meaningful arising from equation (4) at this same value? Will any of these ideas get us closer to computing $\sum_{n=1}^{\infty} 1/n^2$?

As it turned out, Bernoulli's plea for help was answered in an unexpected way by Leonard Euler. At a young age, Euler was a student of Jakob Bernoulli's brother Johann, and the stellar pupil quickly rose to become the preeminent mathematician of his age. Euler's solution is impossible to anticipate. In 1735, he announced that

$$1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots = \frac{\pi^2}{6},$$

a provocative formula that, even more than equation (5), hints at deep connections between geometry, number theory and analysis. Euler's argument is quite short, but it needs to be viewed in the context of the time in which it was created. The "infinite polynomials" in this discussion are examples of *power series*, and a major catalyst for the expanding power of calculus in the 17th and 18th centuries was a proliferation of techniques like the ones used to generate formulas (2), (3), and (4). The machinations of both algebra and calculus are relatively straightforward when restricted to the class of polynomials. So, if in fact power series could be treated more or less like unending polynomials, then there was a great incentive to try to find power series representations for familiar functions like e^x, $\sqrt{1+x}$, or $\sin(x)$.

The appearance of $\arctan(x)$ in (4) is an encouraging sign that this might indeed always be possible. One of Isaac Newton's more significant achievements was to produce a generalization of the binomial formula. If $n \in \mathbf{N}$, then old-fashioned finite algebra leads to the formula

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

Through a process of experimentation and intuition Newton realized that for $r \notin \mathbf{N}$, the infinite series

$$(1+x)^r = 1 + rx + \frac{r(r-1)}{2!}x^2 + \frac{r(r-1)(r-2)}{3!}x^3 + \cdots$$

was meaningful, at least for $x \in (-1, 1)$. Setting $r = -1$, for example, yields

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 - \cdots,$$

which is easily seen to be equivalent to equation (1). Setting $r = 1/2$ we get

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{2^2 2!}x^2 + \frac{3}{2^3 3!}x^3 - \frac{3 \cdot 5}{2^4 4!}x^4 + \cdots.$$

One way to lend a little credence to this formula for $\sqrt{1+x}$ is to focus on the first few terms and square the series:

$$\begin{aligned}\left(\sqrt{1+x}\right)^2 &= \left(1+\frac{1}{2}x-\frac{1}{8}x^2+\cdots\right)\left(1+\frac{1}{2}x-\frac{1}{8}x^2+\cdots\right)\\ &= 1+\left(\frac{1}{2}+\frac{1}{2}\right)x+\left(-\frac{1}{8}+\frac{1}{4}-\frac{1}{8}\right)x^2+\cdots\\ &= 1+x+0x^2+0x^3+\cdots.\end{aligned}$$

Amid all of the unfounded assumptions we are making about infinity, calculations like this induce a feeling of optimism about the legitimacy of our search for power series representations.

Newton's binomial series is the starting point for a modern proof of Euler's famous sum, which is sketched out in detail in Section 8.3. Euler's original 1735 argument, however, started from the power series representation for $\sin(x)$. The formula

$$\sin x = x - \frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots$$

was known to Newton, Bernoulli, and Euler alike. In contrast to equation (1), we will see that this formula is valid for all $x \in \mathbf{R}$. Factoring out x and dividing yields a power series with leading coefficient equal to 1:

$$\frac{\sin x}{x} = 1-\frac{x^2}{3!}+\frac{x^4}{5!}-\frac{x^6}{7!}+\cdots. \tag{6}$$

Euler's idea was to continue factoring the power series in (6), and his strategy for doing this was very much in keeping with what we have seen so far—treat the power series as though it were a polynomial and then extend the pattern to infinity.

Factoring a polynomial of, say, degree three is straightforward if we know its roots. If $p(x) = 1+ax+bx^2+cx^3$ has roots r_1, r_2, and r_3, then

$$p(x) = \left(1-\frac{x}{r_1}\right)\left(1-\frac{x}{r_2}\right)\left(1-\frac{x}{r_3}\right).$$

To see this just directly substitute to get $p(0)=1$ and $p(r_1)=p(r_2)=p(r_3)=0$.

The roots of the power series in (6) are the nonzero roots of $\sin x$, or $x = \pm\pi, \pm 2\pi, \pm 3\pi$, and so on. All right then—relying on his fabled intuition, Euler surmised that

$$\begin{aligned}&1-\frac{x^2}{3!}+\frac{x^4}{5!}-\frac{x^6}{7!}+\cdots\\ &\quad= \left(1-\frac{x}{\pi}\right)\left(1+\frac{x}{\pi}\right)\left(1-\frac{x}{2\pi}\right)\left(1+\frac{x}{2\pi}\right)\left(1-\frac{x}{3\pi}\right)\left(1+\frac{x}{3\pi}\right)\cdots\\ &\quad= \left(1-\frac{x^2}{\pi^2}\right)\left(1-\frac{x^2}{4\pi^2}\right)\left(1-\frac{x^2}{9\pi^2}\right)\cdots,\end{aligned} \tag{7}$$

where in the last step adjacent pairs of factors have been multiplied together. What happens if we continue to multiply out the factors on the right? Well, the constant term comes out to be 1 which happily matches the constant term on the left. The magic comes when we compare the x^2 term on each side of (7). Multiplying out the infinite number of factors on the right (using our imagination as necessary) and collecting like powers of x, equation (7) becomes

$$1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots$$
$$= 1 + \left(-\frac{1}{\pi^2} - \frac{1}{4\pi^2} - \frac{1}{9\pi^2} - \cdots \right) x^2 + \left(\frac{1}{4\pi^4} + \frac{1}{9\pi^4} + \cdots \right) x^4 + \cdots .$$

Equating the coefficients of x^2 on each side yields

$$-\frac{1}{3!} = -\frac{1}{\pi^2} - \frac{1}{4\pi^2} - \frac{1}{9\pi^2} - \cdots ,$$

which when we multiply by $-\pi^2$ becomes

$$\frac{\pi^2}{6} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots .$$

Numerical approximations of each side of this equation confirmed for Euler that, despite the audacious leaps in his argument, he had landed on solid ground. By our standards, this derivation falls well short of being a proper proof, and we will have to tend to this in the upcoming chapters. The takeaway of this discussion is that the hard work ahead is worth the effort. Infinite series representations of functions are both useful and surprisingly elegant, and can lead to remarkable conclusions when they are properly handled.

The evidence so far suggests power series are quite robust when treated as if they were finite in nature. Term-by-term differentiation produced a valid conclusion in equation (2), and taking antiderivatives fared similarly well in (4). We will see that these manipulations are *not* always justified for infinite series of more general types of functions. What is it about power series in particular that makes them so impervious to the dangers of the infinite? Of the many unanswered questions in this discussion, this last one is probably the most central, and the most important to understanding series of functions in general.

6.2 Uniform Convergence of a Sequence of Functions

Adopting the same strategy we used in Chapter 2, we will initially concern ourselves with the behavior and properties of converging *sequences* of functions. Because convergence of infinite series is defined in terms of the associated sequence of partial sums, the results from our study of sequences will be immediately applicable to the questions we have raised about both power series and more general infinite series of functions.

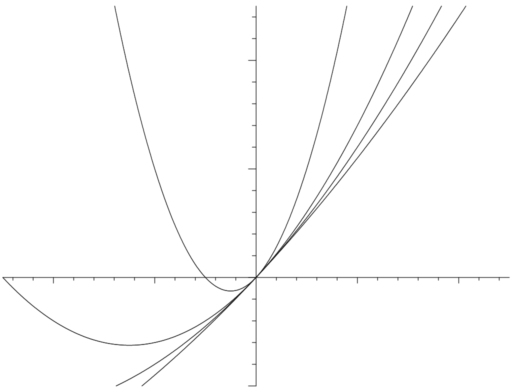

Figure 6.1: f_1, f_5, f_{10}, AND f_{20} WHERE $f_n = (x^2 + nx)/n$.

Pointwise Convergence

Definition 6.2.1. For each $n \in \mathbf{N}$, let f_n be a function defined on a set $A \subseteq \mathbf{R}$. The sequence (f_n) of functions *converges pointwise on* A to a function f if, for all $x \in A$, the sequence of real numbers $f_n(x)$ converges to $f(x)$.

In this case, we write $f_n \to f$, $\lim f_n = f$, or $\lim_{n\to\infty} f_n(x) = f(x)$. This last expression is helpful if there is any confusion as to whether x or n is the limiting variable.

Example 6.2.2. (i) Consider

$$f_n(x) = (x^2 + nx)/n$$

on all of $\mathbf{R}$. Graphs of f_1, f_5, f_{10}, and f_{20} (Fig. 6.1) give an indication of what is happening as n gets larger. Algebraically, we can compute

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \frac{x^2 + nx}{n} = \lim_{n\to\infty} \frac{x^2}{n} + x = x.$$

Thus, (f_n) converges pointwise to $f(x) = x$ on $\mathbf{R}$.

(ii) Let $g_n(x) = x^n$ on the set $[0, 1]$, and consider what happens as n tends to infinity (Fig. 6.2). If $0 \le x < 1$, then we have seen that $x^n \to 0$. On the other hand, if $x = 1$, then $x^n \to 1$. It follows that $g_n \to g$ pointwise on $[0, 1]$, where

$$g(x) = \begin{cases} 0 & \text{for } 0 \le x < 1 \\ 1 & \text{for } x = 1. \end{cases}$$

(iii) Consider $h_n(x) = x^{1+\frac{1}{2n-1}}$ on the set $[-1, 1]$ (Fig. 6.3). For a fixed $x \in [-1, 1]$ we have

$$\lim_{n\to\infty} h_n(x) = x \lim_{n\to\infty} x^{\frac{1}{2n-1}} = |x|.$$

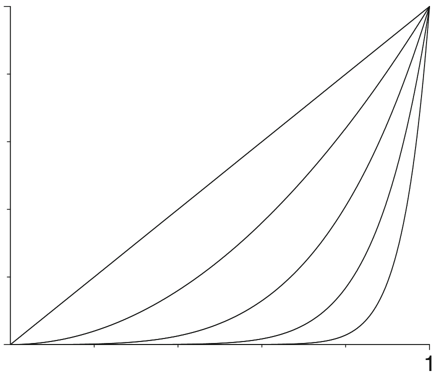

Figure 6.2: $g(x) = \lim_{n\to\infty} x^n$ IS NOT CONTINUOUS ON $[0, 1]$.

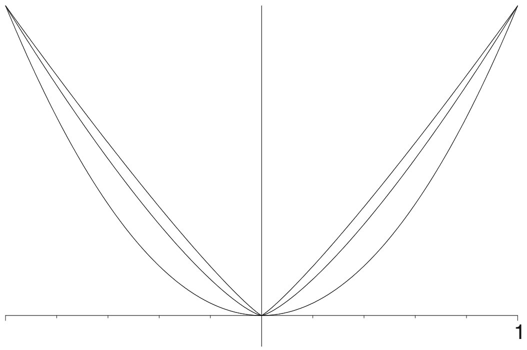

Figure 6.3: $h_n \to |x|$ ON $[-1, 1]$; LIMIT IS NOT DIFFERENTIABLE.

Examples 6.2.2 (ii) and (iii) are our first indication that there is some difficult work ahead of us. The central theme of this chapter is analyzing which properties the limit function inherits from the approximating sequence. In Example 6.2.2 (iii) we have a sequence of differentiable functions converging pointwise to a limit that is not differentiable at the origin. In Example 6.2.2 (ii), we see an even more fundamental problem of a sequence of continuous functions converging to a limit that is not continuous.

Continuity of the Limit Function

With Example 6.2.2 (ii) firmly in mind, we begin this discussion with a doomed attempt to prove that the pointwise limit of continuous functions is continuous. Upon discovering the problem in the argument, we will be in a better position to understand the need for a stronger notion of convergence for sequences of functions.

Assume (f_n) is a sequence of continuous functions on a set $A \subseteq \mathbf{R}$, and assume (f_n) converges pointwise to a limit f. To argue that f is continuous, fix a point $c \in A$, and let $\epsilon > 0$. We need to find a $\delta > 0$ such that

$$|x - c| < \delta \quad \text{implies} \quad |f(x) - f(c)| < \epsilon.$$

By the triangle inequality,

$$\begin{aligned} |f(x) - f(c)| &= |f(x) - f_n(x) + f_n(x) - f_n(c) + f_n(c) - f(c)| \\ &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(c)| + |f_n(c) - f(c)|. \end{aligned}$$

Our first, optimistic impression is that each term in the sum on the right-hand side can be made small—the first and third by the fact that $f_n \to f$, and the middle term by the continuity of f_n. In order to use the continuity of f_n, we must first establish which particular f_n we are talking about. Because $c \in A$ is fixed, choose $N \in \mathbf{N}$ so that

$$|f_N(c) - f(c)| < \frac{\epsilon}{3}.$$

Now that N is chosen, the continuity of f_N implies that there exists a $\delta > 0$ such that

$$|f_N(x) - f_N(c)| < \frac{\epsilon}{3}$$

for all x satisfying $|x - c| < \delta$.

But here is the problem. We also need

$$|f_N(x) - f(x)| < \frac{\epsilon}{3} \quad \textit{for all } x \text{ satisfying } |x - c| < \delta.$$

The values of x depend on δ, which depends on the choice of N. Thus, we cannot go back and simply choose a different N. More to the point, the variable x is not fixed the way c is in this discussion but represents any point in the interval $(c-\delta, c+\delta)$. Pointwise convergence implies that we can make $|f_n(x)-f(x)| < \epsilon/3$ for large enough values of n, but *the value of n depends on the point x.* It is possible that different values for x will result in the need for different—larger—choices for n. This phenomenon is apparent in Example 6.2.2 (ii). To achieve the inequality

$$|g_n(1/2) - g(1/2)| < \frac{1}{3},$$

we need $n \geq 2$, whereas

$$|g_n(9/10) - g(9/10)| < \frac{1}{3}$$

is true only after $n \geq 11$.

Uniform Convergence

To resolve this dilemma, we define a new, stronger notion of convergence of functions.

Definition 6.2.3 (Uniform Convergence). Let (f_n) be a sequence of functions defined on a set $A \subseteq \mathbf{R}$. Then, (f_n) *converges uniformly on* A to a limit function f defined on A if, for every $\epsilon > 0$, there exists an $N \in \mathbf{N}$ such that $|f_n(x) - f(x)| < \epsilon$ whenever $n \geq N$ and $x \in A$.

To emphasize the difference between uniform convergence and pointwise convergence, we restate Definition 6.2.1, being more explicit about the relationship between ϵ, N, and x. In particular, notice where the domain point x is referenced in each definition and consequently how the choice of N then does or does not depend on this value.

Definition 6.2.1B. Let (f_n) be a sequence of functions defined on a set $A \subseteq \mathbf{R}$. Then, (f_n) *converges pointwise on* A to a limit f defined on A if, for every $\epsilon > 0$ and $x \in A$, there exists an $N \in \mathbf{N}$ (perhaps dependent on x) such that $|f_n(x) - f(x)| < \epsilon$ whenever $n \geq N$.

The use of the adverb *uniformly* here should be reminiscent of its use in the phrase "uniformly continuous" from Chapter 4. In both cases, the term "uniformly" is employed to express the fact that the response (δ or N) to a prescribed ϵ can be chosen to work simultaneously for all values of x in the relevant domain.

Example 6.2.4. (i) Let

$$g_n(x) = \frac{1}{n(1+x^2)}.$$

For any fixed $x \in \mathbf{R}$, we can see that $\lim g_n(x) = 0$ so that $g(x) = 0$ is the pointwise limit of the sequence (g_n) on $\mathbf{R}$. Is this convergence uniform? The observation that $1/(1+x^2) \leq 1$ for all $x \in \mathbf{R}$ implies that

$$|g_n(x) - g(x)| = \left| \frac{1}{n(1+x^2)} - 0 \right| \leq \frac{1}{n}.$$

Thus, given $\epsilon > 0$, we can choose $N > 1/\epsilon$ (which does not depend on x), and it follows that

$$n \geq N \quad \text{implies} \quad |g_n(x) - g(x)| < \epsilon$$

for all $x \in \mathbf{R}$. By Definition 6.2.3, $g_n \to 0$ uniformly on $\mathbf{R}$.

(ii) Look back at Example 6.2.2 (i), where we saw that $f_n(x) = (x^2 + nx)/n$ converges pointwise on $\mathbf{R}$ to $f(x) = x$. On $\mathbf{R}$, the convergence is not uniform. To see this write

$$|f_n(x) - f(x)| = \left| \frac{x^2 + nx}{n} - x \right| = \frac{x^2}{n},$$

and notice that in order to force $|f_n(x) - f(x)| < \epsilon$, we are going to have to choose

$$N > \frac{x^2}{\epsilon}.$$

Although this is possible to do for each $x \in \mathbf{R}$, there is no way to choose a single value of N that will work for all values of x at the same time.

On the other hand, we can show that $f_n \to f$ uniformly on the set $[-b, b]$. By restricting our attention to a bounded interval, we may now assert that

$$\frac{x^2}{n} \leq \frac{b^2}{n}.$$

Given $\epsilon > 0$, then, we can choose

$$N > \frac{b^2}{\epsilon}$$

independently of $x \in [-b, b]$.

Graphically speaking, the uniform convergence of f_n to a limit f on a set A can be visualized by constructing a band of radius $\pm\epsilon$ around the limit function f. If $f_n \to f$ uniformly, then there exists a point in the sequence after which each f_n is *completely* contained in this ϵ-strip (Fig. 6.4). This image should be compared with the graphs in Figures 6.1–6.2 from Example 6.2.2 and the one in Figure 6.5.

Cauchy Criterion

Recall that the Cauchy Criterion for convergent sequences of real numbers was an equivalent characterization of convergence which, unlike the definition, did not make explicit mention of the limit. The usefulness of the Cauchy Criterion suggests the need for an analogous characterization of uniformly convergent sequences of functions. As with all statements about uniformity, pay special attention to the relationship between the response variable ($N \in \mathbf{N}$) and the domain variable ($x \in A$).

Theorem 6.2.5 (Cauchy Criterion for Uniform Convergence). *A sequence of functions (f_n) defined on a set $A \subseteq \mathbf{R}$ converges uniformly on A if and only if for every $\epsilon > 0$ there exists an $N \in \mathbf{N}$ such that $|f_n(x) - f_m(x)| < \epsilon$ whenever $m, n \geq N$ and $x \in A$.*

Proof. Exercise 6.2.5. □

Continuity Revisited

The stronger assumption of uniform convergence is precisely what is required to remove the flaws from our attempted proof that the limit of continuous functions is continuous.

Theorem 6.2.6 (Continuous Limit Theorem). *Let (f_n) be a sequence of functions defined on $A \subseteq \mathbf{R}$ that converges uniformly on A to a function f. If each f_n is continuous at $c \in A$, then f is continuous at c.*

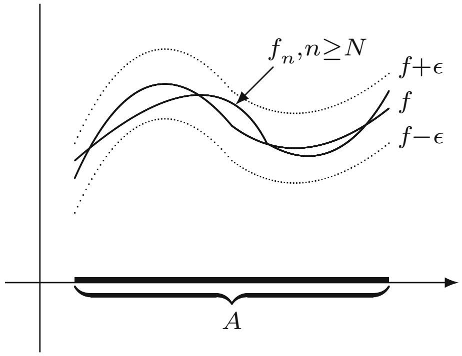

Figure 6.4: $f_n \to f$ UNIFORMLY ON A.

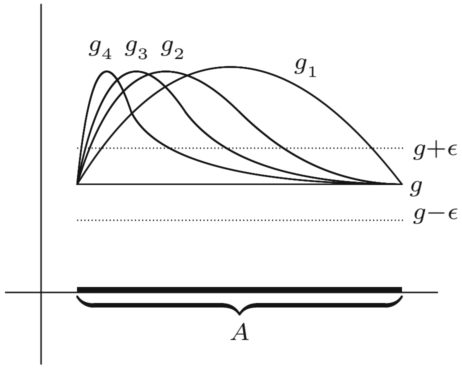

Figure 6.5: $g_n \to g$ POINTWISE, BUT NOT UNIFORMLY.

Proof. Fix $c \in A$ and let $\epsilon > 0$. Choose N so that

$$|f_N(x) - f(x)| < \frac{\epsilon}{3}$$

for all $x \in A$. Because f_N is continuous, there exists a $\delta > 0$ for which

$$|f_N(x) - f_N(c)| < \frac{\epsilon}{3}$$

is true whenever $|x - c| < \delta$. But this implies

$$\begin{aligned} |f(x) - f(c)| &= |f(x) - f_N(x) + f_N(x) - f_N(c) + f_N(c) - f(c)| \\ &\leq |f(x) - f_N(x)| + |f_N(x) - f_N(c)| + |f_N(c) - f(c)| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Thus, f is continuous at $c \in A$. □

Exercises

Exercise 6.2.1. Let

$$f_n(x) = \frac{nx}{1+nx^2}.$$

(a) Find the pointwise limit of (f_n) for all $x \in (0, \infty)$.

(b) Is the convergence uniform on $(0, \infty)$?

(c) Is the convergence uniform on $(0, 1)$?

(d) Is the convergence uniform on $(1, \infty)$?

Exercise 6.2.2. (a) Define a sequence of functions on $\mathbf{R}$ by

$$f_n(x) = \begin{cases} 1 & \text{if } x = 1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

and let f be the pointwise limit of f_n.

Is each f_n continuous at zero? Does $f_n \to f$ uniformly on $\mathbf{R}$? Is f continuous at zero?

(b) Repeat this exercise using the sequence of functions

$$g_n(x) = \begin{cases} x & \text{if } x = 1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n} \\ 0 & \text{otherwise.} \end{cases}$$

(c) Repeat the exercise once more with the sequence

$$h_n(x) = \begin{cases} 1 & \text{if } x = \frac{1}{n} \\ x & \text{if } x = 1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n-1} \\ 0 & \text{otherwise.} \end{cases}$$

In each case, explain how the results are consistent with the content of the Continuous Limit Theorem (Theorem 6.2.6).

Exercise 6.2.3. For each $n \in \mathbf{N}$ and $x \in [0, \infty)$, let

$$g_n(x) = \frac{x}{1+x^n} \quad \text{and} \quad h_n(x) = \begin{cases} 1 & \text{if } x \geq 1/n \\ nx & \text{if } 0 \leq x < 1/n. \end{cases}$$

Answer the following questions for the sequences (g_n) and (h_n):

(a) Find the pointwise limit on $[0, \infty)$.

(b) Explain how we know that the convergence *cannot* be uniform on $[0, \infty)$.

(c) Choose a smaller set over which the convergence is uniform and supply an argument to show that this is indeed the case.

Exercise 6.2.4. Review Exercise 5.2.8 which includes the definition for a uniformly differentiable function. Use the results discussed in Section 6.2 to show that if f is uniformly differentiable, then f' is continuous.

Exercise 6.2.5. Using the Cauchy Criterion for convergent sequences of real numbers (Theorem 2.6.4), supply a proof for Theorem 6.2.5. (First, define a candidate for $f(x)$, and then argue that $f_n \to f$ uniformly.)

Exercise 6.2.6. Assume $f_n \to f$ on a set A. Theorem 6.2.6 is an example of a typical type of question which asks whether a trait possessed by each f_n is inherited by the limit function. Provide an example to show that *all* of the following propositions are false if the convergence is only assumed to be pointwise on A. Then go back and decide which are true under the stronger hypothesis of uniform convergence.

(a) If each f_n is uniformly continuous, then f is uniformly continuous.

(b) If each f_n is bounded, then f is bounded.

(c) If each f_n has a finite number of discontinuities, then f has a finite number of discontinuities.

(d) If each f_n has fewer than M discontinuities (where $M \in \mathbf{N}$ is fixed), then f has fewer than M discontinuities.

(e) If each f_n has at most a countable number of discontinuities, then f has at most a countable number of discontinuities.

Exercise 6.2.7. Let f be uniformly continuous on all of $\mathbf{R}$, and define a sequence of functions by $f_n(x) = f(x + \frac{1}{n})$. Show that $f_n \to f$ uniformly. Give an example to show that this proposition fails if f is only assumed to be continuous and not uniformly continuous on $\mathbf{R}$.

Exercise 6.2.8. Let (g_n) be a sequence of continuous functions that converges uniformly to g on a compact set K. If $g(x) \neq 0$ on K, show $(1/g_n)$ converges uniformly on K to $1/g$.

Exercise 6.2.9. Assume (f_n) and (g_n) are uniformly convergent sequences of functions.

(a) Show that $(f_n + g_n)$ is a uniformly convergent sequence of functions.

(b) Give an example to show that the product $(f_n g_n)$ may not converge uniformly.

(c) Prove that if there exists an $M > 0$ such that $|f_n| \leq M$ and $|g_n| \leq M$ for all $n \in \mathbf{N}$, then $(f_n g_n)$ does converge uniformly.

Exercise 6.2.10. This exercise and the next explore partial converses of the Continuous Limit Theorem (Theorem 6.2.6). Assume $f_n \to f$ pointwise on $[a, b]$ and the limit function f is continuous on $[a, b]$. If each f_n is increasing (but not necessarily continuous), show $f_n \to f$ uniformly.

Exercise 6.2.11 (Dini's Theorem). Assume $f_n \to f$ pointwise on a compact set K and assume that for each $x \in K$ the sequence $f_n(x)$ is increasing. Follow these steps to show that if f_n and f are continuous on K, then the convergence is uniform.

(a) Set $g_n = f - f_n$ and translate the preceding hypothesis into statements about the sequence (g_n).

(b) Let $\epsilon > 0$ be arbitrary, and define $K_n = \{x \in K : g_n(x) \geq \epsilon\}$. Argue that $K_1 \supseteq K_2 \supseteq K_3 \supseteq \cdots$, and use this observation to finish the argument.

Exercise 6.2.12 (Cantor Function). Review the construction of the Cantor set $C \subseteq [0,1]$ from Section 3.1. This exercise makes use of results and notation from this discussion.

(a) Define $f_0(x) = x$ for all $x \in [0,1]$. Now, let

$$f_1(x) = \begin{cases} (3/2)x & \text{for } 0 \leq x \leq 1/3 \\ 1/2 & \text{for } 1/3 < x < 2/3 \\ (3/2)x - 1/2 & \text{for } 2/3 \leq x \leq 1. \end{cases}$$

Sketch f_0 and f_1 over $[0,1]$ and observe that f_1 is continuous, increasing, and constant on the middle third $(1/3, 2/3) = [0,1]\backslash C_1$.

(b) Construct f_2 by imitating this process of flattening out the middle third of each nonconstant segment of f_1. Specifically, let

$$f_2(x) = \begin{cases} (1/2)f_1(3x) & \text{for } 0 \leq x \leq 1/3 \\ f_1(x) & \text{for } 1/3 < x < 2/3 \\ (1/2)f_1(3x-2) + 1/2 & \text{for } 2/3 \leq x \leq 1. \end{cases}$$

If we continue this process, show that the resulting sequence (f_n) converges uniformly on $[0,1]$.

(c) Let $f = \lim f_n$. Prove that f is a continuous, increasing function on $[0,1]$ with $f(0) = 0$ and $f(1) = 1$ that satisfies $f'(x) = 0$ for all x in the open set $[0,1]\backslash C$. Recall that the "length" of the Cantor set C is 0. Somehow, f manages to increase from 0 to 1 while remaining constant on a set of "length 1."

Exercise 6.2.13. Recall that the Bolzano–Weierstrass Theorem (Theorem 2.5.5) states that every bounded sequence of real numbers has a convergent subsequence. An analogous statement for bounded sequences of functions is not true in general, but under stronger hypotheses several different conclusions are possible. One avenue is to assume the common domain for all of the functions in the sequence is countable. (Another is explored in the next two exercises.)

Let $A = \{x_1, x_2, x_3, \ldots\}$ be a countable set. For each $n \in \mathbf{N}$, let f_n be defined on A and assume there exists an $M > 0$ such that $|f_n(x)| \leq M$ for all $n \in \mathbf{N}$ and $x \in A$. Follow these steps to show that there exists a subsequence of (f_n) that converges pointwise on A.

(a) Why does the sequence of real numbers $f_n(x_1)$ necessarily contain a convergent subsequence (f_{n_k})? To indicate that the subsequence of functions (f_{n_k}) is generated by considering the values of the functions at x_1, we will use the notation $f_{n_k} = f_{1,k}$.

(b) Now, explain why the sequence $f_{1,k}(x_2)$ contains a convergent subsequence.

(c) Carefully construct a nested family of subsequences $(f_{m,k})$, and show how this can be used to produce a single subsequence of (f_n) that converges at every point of A.

Exercise 6.2.14. A sequence of functions (f_n) defined on a set $E \subseteq \mathbf{R}$ is called *equicontinuous* if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $|f_n(x)-f_n(y)| < \epsilon$ for all $n \in \mathbf{N}$ and $|x - y| < \delta$ in E.

(a) What is the difference between saying that a sequence of functions (f_n) is equicontinuous and just asserting that each f_n in the sequence is individually uniformly continuous?

(b) Give a qualitative explanation for why the sequence $g_n(x) = x^n$ is not equicontinuous on $[0, 1]$. Is each g_n uniformly continuous on $[0, 1]$?

Exercise 6.2.15 (Arzela–Ascoli Theorem). For each $n \in \mathbf{N}$, let f_n be a function defined on $[0, 1]$. If (f_n) is bounded on $[0, 1]$—that is, there exists an $M > 0$ such that $|f_n(x)| \leq M$ for all $n \in \mathbf{N}$ and $x \in [0, 1]$—and if the collection of functions (f_n) is equicontinuous (Exercise 6.2.14), follow these steps to show that (f_n) contains a uniformly convergent subsequence.

(a) Use Exercise 6.2.13 to produce a subsequence (f_{n_k}) that converges at every rational point in $[0, 1]$. To simplify the notation, set $g_k = f_{n_k}$. It remains to show that (g_k) converges uniformly on all of $[0, 1]$.

(b) Let $\epsilon > 0$. By equicontinuity, there exists a $\delta > 0$ such that

$$|g_k(x) - g_k(y)| < \frac{\epsilon}{3}$$

for all $|x - y| < \delta$ and $k \in \mathbf{N}$. Using this δ, let $r_1, r_2, \ldots, r_m$ be a *finite* collection of rational points with the property that the union of the neighborhoods $V_\delta(r_i)$ contains [0,1].

Explain why there must exist an $N \in \mathbf{N}$ such that

$$|g_s(r_i) - g_t(r_i)| < \frac{\epsilon}{3}$$

for all $s, t \geq N$ and r_i in the finite subset of $[0, 1]$ just described. Why does having the set $\{r_1, r_2, \ldots, r_m\}$ be finite matter?

(c) Finish the argument by showing that, for an arbitrary $x \in [0, 1]$,

$$|g_s(x) - g_t(x)| < \epsilon$$

for all $s, t \geq N$.

6.3 Uniform Convergence and Differentiation

Example 6.2.2 (iii) imposes some significant restrictions on what we might hope to be true regarding differentiation and uniform convergence. If $h_n \to h$ uniformly and each h_n is differentiable, we should not anticipate that $h'_n \to h'$ because in this example $h'(x)$ does not even exist at $x = 0$. There are also examples (see Exercise 6.3.4) where $f_n \to f$ uniformly with (f_n) and f all differentiable, but the sequence (f'_n) diverges at every point of the domain.

The key assumption necessary to be able to prove any facts about the derivative of the limit function is that the *sequence of derivatives* be uniformly convergent. This may sound as though we are assuming what it is we would like to prove, and there is some validity to this complaint. The more hypotheses a proposition has, the more difficult it is to apply. The content of the next theorem is that if we are given a pointwise convergent sequence of differentiable functions, and if we know that the sequence of derivatives converges uniformly *to something*, then the limit of the derivatives is indeed the derivative of the limit.

Theorem 6.3.1 (Differentiable Limit Theorem). *Let $f_n \to f$ pointwise on the closed interval $[a, b]$, and assume that each f_n is differentiable. If (f'_n) converges uniformly on $[a, b]$ to a function g, then the function f is differentiable and $f' = g$.*

Proof. Fix $c \in [a, b]$ and let $\epsilon > 0$. We want to argue that $f'(c)$ exists and equals $g(c)$. Because f' is defined by the limit

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c},$$

our task is to produce a $\delta > 0$ so that

$$\left| \frac{f(x) - f(c)}{x - c} - g(c) \right| < \epsilon$$

whenever $0 < |x - c| < \delta$.

To motivate the strategy of the proof, observe that for all $x \neq c$ and all $n \in \mathbf{N}$, the triangle inequality implies

$$\begin{aligned} \left| \frac{f(x) - f(c)}{x - c} - g(c) \right| \quad \le \quad & \left| \frac{f(x) - f(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} \right| \\ & + \left| \frac{f_n(x) - f_n(c)}{x - c} - f'_n(c) \right| + \left| f'_n(c) - g(c) \right|. \end{aligned}$$

Our intent is to first find an f_n that forces the first and third terms on the right-hand side to be less than $\epsilon/3$. Once we establish which f_n we want, we can then use the differentiability of f_n to produce a δ that makes the middle term less than $\epsilon/3$ for all x satisfying $0 < |x - c| < \delta$.

Let's start by choosing an N_1 such that

$$|f'_m(c) - g(c)| < \frac{\epsilon}{3} \tag{1}$$

for all $m \geq N_1$. We now invoke the uniform convergence of (f'_n) to assert (via Theorem 6.2.5) that there exists an N_2 such that $m, n \geq N_2$ implies

$$|f'_m(x) - f'_n(x)| < \frac{\epsilon}{3} \quad \text{for all } x \in [a, b].$$

Set $N = \max\{N_1, N_2\}$.

The function f_N is differentiable at c, and so there exists a $\delta > 0$ for which

$$\left| \frac{f_N(x) - f_N(c)}{x - c} - f'_N(c) \right| < \frac{\epsilon}{3} \tag{2}$$

whenever $0 < |x - c| < \delta$. This is our sought after δ, but it takes some effort to show that it has the desired property.

Fix an x satisfying $0 < |x - c| < \delta$, let $m \geq N$, and apply the Mean Value Theorem to $f_m - f_N$ on the interval $[c, x]$, (If $x < c$ the argument is the same.) By MVT, there exists an $\alpha \in (c, x)$ such that

$$f'_m(\alpha) - f'_N(\alpha) = \frac{(f_m(x) - f_N(x)) - (f_m(c) - f_N(c))}{x - c}.$$

Recall that our choice of N implies

$$|f'_m(\alpha) - f'_N(\alpha)| < \frac{\epsilon}{3},$$

and so it follows that

$$\left| \frac{f_m(x) - f_m(c)}{x - c} - \frac{f_N(x) - f_N(c)}{x - c} \right| < \frac{\epsilon}{3}.$$

Because $f_m \to f$ we can take the limit as $m \to \infty$, and the Order Limit Theorem (Theorem 2.3.4) asserts that

$$\left| \frac{f(x) - f(c)}{x - c} - \frac{f_N(x) - f_N(c)}{x - c} \right| \leq \frac{\epsilon}{3}. \tag{3}$$

Finally, the inequalities in (1), (2), and (3), together imply that for x satisfying $0 < |x - c| < \delta$,

$$\begin{aligned} \left| \frac{f(x) - f(c)}{x - c} - g(c) \right| \quad &\leq \quad \left| \frac{f(x) - f(c)}{x - c} - \frac{f_N(x) - f_N(c)}{x - c} \right| \\ &\quad + \left| \frac{f_N(x) - f_N(c)}{x - c} - f'_N(c) \right| + |f'_N(c) - g(c)| \\ &< \quad \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

□

The hypothesis in the Differentiable Limit Theorem is unnecessarily strong. We actually do not need to assume that $f_n(x) \to f(x)$ at each point in the domain because the assumption that the sequence of derivatives (f_n') converges uniformly is nearly strong enough to *prove* that (f_n) converges, uniformly in fact. Two functions with the same derivative may differ by a constant, so we must assume that there is at least one point x_0 where $f_n(x_0) \to f(x_0)$.

Theorem 6.3.2. *Let (f_n) be a sequence of differentiable functions defined on the closed interval $[a, b]$, and assume (f_n') converges uniformly on $[a, b]$. If there exists a point $x_0 \in [a, b]$ where $f_n(x_0)$ is convergent, then (f_n) converges uniformly on $[a, b]$.*

Proof. Exercise 6.3.7. □

Combining the last two results produces a stronger version of Theorem 6.3.1.

Theorem 6.3.3. *Let (f_n) be a sequence of differentiable functions defined on the closed interval $[a, b]$, and assume (f_n') converges uniformly to a function g on $[a, b]$. If there exists a point $x_0 \in [a, b]$ for which $f_n(x_0)$ is convergent, then (f_n) converges uniformly. Moreover, the limit function $f = \lim f_n$ is differentiable and satisfies $f' = g$.*

Exercises

Exercise 6.3.1. Consider the sequence of functions defined by

$$g_n(x) = \frac{x^n}{n}.$$

(a) Show (g_n) converges uniformly on $[0, 1]$ and find $g = \lim g_n$. Show that g is differentiable and compute $g'(x)$ for all $x \in [0, 1]$.

(b) Now, show that (g_n') converges on $[0, 1]$. Is the convergence uniform? Set $h = \lim g_n'$ and compare h and g'. Are they the same?

Exercise 6.3.2. Consider the sequence of functions

$$h_n(x) = \sqrt{x^2 + \frac{1}{n}}.$$

(a) Compute the pointwise limit of (h_n) and then prove that the convergence is uniform on $\mathbf{R}$.

(b) Note that each h_n is differentiable. Show $g(x) = \lim h_n'(x)$ exists for all x, and explain how we can be certain that the convergence is *not* uniform on any neighborhood of zero.

Exercise 6.3.3. Consider the sequence of functions

$$f_n(x) = \frac{x}{1 + nx^2}.$$

(a) Find the points on $\mathbf{R}$ where each $f_n(x)$ attains its maximum and minimum value. Use this to prove (f_n) converges uniformly on $\mathbf{R}$. What is the limit function?

(b) Let $f = \lim f_n$. Compute $f_n'(x)$ and find all the values of x for which $f'(x) = \lim f_n'(x)$.

Exercise 6.3.4. Let

$$h_n(x) = \frac{\sin(nx)}{\sqrt{n}}.$$

Show that $h_n \to 0$ uniformly on $\mathbf{R}$ but that the sequence of derivatives (h_n') diverges for every $x \in \mathbf{R}$.

Exercise 6.3.5. Let

$$g_n(x) = \frac{nx + x^2}{2n},$$

and set $g(x) = \lim g_n(x)$. Show that g is differentiable in two ways:

(a) Compute $g(x)$ by algebraically taking the limit as $n \to \infty$ and then find $g'(x)$.

(b) Compute $g_n'(x)$ for each $n \in \mathbf{N}$ and show that the sequence of derivatives (g_n') converges uniformly on every interval $[-M, M]$. Use Theorem 6.3.3 to conclude $g'(x) = \lim g_n'(x)$.

(c) Repeat parts (a) and (b) for the sequence $f_n(x) = (nx^2 + 1)/(2n + x)$.

Exercise 6.3.6. Provide an example or explain why the request is impossible. Let's take the domain of the functions to be all of $\mathbf{R}$.

(a) A sequence (f_n) of nowhere differentiable functions with $f_n \to f$ uniformly and f everywhere differentiable.

(b) A sequence (f_n) of differentiable functions such that (f_n') converges uniformly but the original sequence (f_n) does not converge for any $x \in \mathbf{R}$.

(c) A sequence (f_n) of differentiable functions such that both (f_n) and (f_n') converge uniformly but $f = \lim f_n$ is not differentiable at some point.

Exercise 6.3.7. Use the Mean Value Theorem to supply a proof for Theorem 6.3.2. To get started, observe that the triangle inequality implies that, for any $x \in [a, b]$ and $m, n \in \mathbf{N}$,

$$|f_n(x) - f_m(x)| \leq |(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))| + |f_n(x_0) - f_m(x_0)|.$$

6.4 Series of Functions

Definition 6.4.1. For each $n \in \mathbf{N}$, let f_n and f be functions defined on a set $A \subseteq \mathbf{R}$. The infinite series

$$\sum_{n=1}^{\infty} f_n(x) = f_1(x) + f_2(x) + f_3(x) + \cdots$$

converges pointwise on A to $f(x)$ if the sequence $s_k(x)$ of partial sums defined by

$$s_k(x) = f_1(x) + f_2(x) + \cdots + f_k(x)$$

converges pointwise to $f(x)$. The series *converges uniformly on* A to f if the sequence $s_k(x)$ converges uniformly on A to $f(x)$.

In either case, we write $f = \sum_{n=1}^{\infty} f_n$ or $f(x) = \sum_{n=1}^{\infty} f_n(x)$, always being explicit about the type of convergence involved.

If we have a series $\sum_{n=1}^{\infty} f_n$ where the functions f_n are continuous, then the Algebraic Continuity Theorem (Theorem 4.3.4) guarantees that the partial sums—because they are finite sums—will be continuous as well. A corresponding observation is true if we are dealing with differentiable functions. As a consequence, we can immediately translate the results for sequences in the previous sections into statements about the behavior of infinite series of functions.

Theorem 6.4.2 (Term-by-term Continuity Theorem). *Let f_n be continuous functions defined on a set $A \subseteq \mathbf{R}$, and assume $\sum_{n=1}^{\infty} f_n$ converges uniformly on A to a function f. Then, f is continuous on A.*

Proof. Apply the Continuous Limit Theorem (Theorem 6.2.6) to the partial sums $s_k = f_1 + f_2 + \cdots + f_k$. □

Theorem 6.4.3 (Term-by-term Differentiability Theorem). *Let f_n be differentiable functions defined on an interval A, and assume $\sum_{n=1}^{\infty} f_n'(x)$ converges uniformly to a limit $g(x)$ on A. If there exists a point $x_0 \in [a, b]$ where $\sum_{n=1}^{\infty} f_n(x_0)$ converges, then the series $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly to a differentiable function $f(x)$ satisfying $f'(x) = g(x)$ on A. In other words,*

$$f(x) = \sum_{n=1}^{\infty} f_n(x) \quad \text{and} \quad f'(x) = \sum_{n=1}^{\infty} f_n'(x).$$

Proof. Apply the stronger form of the Differentiable Limit Theorem (Theorem 6.3.3) to the partial sums $s_k = f_1 + f_2 + \cdots + f_k$. Observe that Theorem 5.2.4 implies that $s_k' = f_1' + f_2' + \cdots + f_k'$. □

In the vocabulary of infinite series, the Cauchy Criterion takes the following form.

Theorem 6.4.4 (Cauchy Criterion for Uniform Convergence of Series). *A series $\sum_{n=1}^{\infty} f_n$ converges uniformly on $A \subseteq \mathbf{R}$ if and only if for every $\epsilon > 0$ there exists an $N \in \mathbf{N}$ such that*

$$|f_{m+1}(x) + f_{m+2}(x) + f_{m+3}(x) + \cdots + f_n(x)| < \epsilon$$

whenever $n > m \geq N$ and $x \in A$.

The benefits of uniform convergence over pointwise convergence suggest the need for some ways of determining when a series converges uniformly. The following corollary to the Cauchy Criterion is the most common such tool. In particular, it will be quite useful in our upcoming investigations of power series.

Corollary 6.4.5 (Weierstrass M-Test). *For each $n \in \mathbf{N}$, let f_n be a function defined on a set $A \subseteq \mathbf{R}$, and let $M_n > 0$ be a real number satisfying*

$$|f_n(x)| \leq M_n$$

for all $x \in A$. If $\sum_{n=1}^{\infty} M_n$ converges, then $\sum_{n=1}^{\infty} f_n$ converges uniformly on A.

Proof. Exercise 6.4.1. □

Exercises

Exercise 6.4.1. Supply the details for the proof of the Weierstrass M-Test (Corollary 6.4.5).

Exercise 6.4.2. Decide whether each proposition is true or false, providing a short justification or counterexample as appropriate.

(a) If $\sum_{n=1}^{\infty} g_n$ converges uniformly, then (g_n) converges uniformly to zero.

(b) If $0 \leq f_n(x) \leq g_n(x)$ and $\sum_{n=1}^{\infty} g_n$ converges uniformly, then $\sum_{n=1}^{\infty} f_n$ converges uniformly.

(c) If $\sum_{n=1}^{\infty} f_n$ converges uniformly on A, then there exist constants M_n such that $|f_n(x)| \leq M_n$ for all $x \in A$ and $\sum_{n=1}^{\infty} M_n$ converges.

Exercise 6.4.3. (a) Show that

$$g(x) = \sum_{n=0}^{\infty} \frac{\cos(2^n x)}{2^n}$$

is continuous on all of $\mathbf{R}$.

(b) The function g was cited in Section 5.4 as an example of a continuous nowhere differentiable function. What happens if we try to use Theorem 6.4.3 to explore whether g is differentiable?

Exercise 6.4.4. Define

$$g(x) = \sum_{n=0}^{\infty} \frac{x^{2n}}{(1 + x^{2n})}.$$

Find the values of x where the series converges and show that we get a continuous function on this set.

Exercise 6.4.5. (a) Prove that

$$h(x) = \sum_{n=1}^{\infty} \frac{x^n}{n^2} = x + \frac{x^2}{4} + \frac{x^3}{9} + \frac{x^4}{16} + \cdots$$

is continuous on $[-1, 1]$.

(b) The series

$$f(x) = \sum_{n=1}^{\infty} \frac{x^n}{n} = x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots$$

converges for every x in the half-open interval $[-1, 1)$ but does not converge when $x = 1$. For a fixed $x_0 \in (-1, 1)$, explain how we can still use the Weierstrass M-Test to prove that f is continuous at x_0.

Exercise 6.4.6. Let

$$f(x) = \frac{1}{x} - \frac{1}{x+1} + \frac{1}{x+2} - \frac{1}{x+3} + \frac{1}{x+4} - \cdots .$$

Show f is defined for all $x > 0$. Is f continuous on $(0, \infty)$? How about differentiable?

Exercise 6.4.7. Let

$$f(x) = \sum_{k=1}^{\infty} \frac{\sin(kx)}{k^3}.$$

(a) Show that $f(x)$ is differentiable and that the derivative $f'(x)$ is continuous.

(b) Can we determine if f is twice-differentiable?

Exercise 6.4.8. Consider the function

$$f(x) = \sum_{k=1}^{\infty} \frac{\sin(x/k)}{k}.$$

Where is f defined? Continuous? Differentiable? Twice-differentiable?

Exercise 6.4.9. Let

$$h(x) = \sum_{n=1}^{\infty} \frac{1}{x^2 + n^2}.$$

(a) Show that h is a continuous function defined on all of $\mathbf{R}$.

(b) Is h differentiable? If so, is the derivative function h' continuous?

Exercise 6.4.10. Let $\{r_1, r_2, r_3, \ldots\}$ be an enumeration of the set of rational numbers. For each $r_n \in \mathbf{Q}$, define

$$u_n(x) = \begin{cases} 1/2^n & \text{for } x > r_n \\ 0 & \text{for } x \leq r_n. \end{cases}$$

Now, let $h(x) = \sum_{n=1}^{\infty} u_n(x)$. Prove that h is a monotone function defined on all of $\mathbf{R}$ that is continuous at every irrational point.

6.5 Power Series

It is time to put some mathematical teeth into our understanding of functions expressed in the form of a power series; that is, functions of the form

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots .$$

The first order of business is to determine the points $x \in \mathbf{R}$ for which the resulting series on the right-hand side converges. This set certainly contains $x = 0$, and, as the next result demonstrates, it takes a very predictable form.

Theorem 6.5.1. *If a power series $\sum_{n=0}^{\infty} a_n x^n$ converges at some point $x_0 \in \mathbf{R}$, then it converges absolutely for any x satisfying $|x| < |x_0|$.*

Proof. If $\sum_{n=0}^{\infty} a_n x_0^n$ converges, then the sequence of terms $(a_n x_0^n)$ is bounded. (In fact, it converges to 0.) Let $M > 0$ satisfy $|a_n x_0^n| \leq M$ for all $n \in \mathbf{N}$. If $x \in \mathbf{R}$ satisfies $|x| < |x_0|$, then

$$|a_n x^n| = |a_n x_0^n| \left| \frac{x}{x_0} \right|^n \leq M \left| \frac{x}{x_0} \right|^n .$$

But notice that

$$\sum_{n=0}^{\infty} M \left| \frac{x}{x_0} \right|^n$$

is a geometric series with ratio $|x/x_0| < 1$ and so converges. By the Comparison Test, $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely. □

The main implication of Theorem 6.5.1 is that the set of points for which a given power series converges must necessarily be $\{0\}$, $\mathbf{R}$, or a bounded interval centered around $x = 0$. Because of the strict inequality in Theorem 6.5.1, there is some ambiguity about the endpoints of the interval, and it is possible that the set of convergent points may be of the form $(-R, R)$, $[-R, R)$, $(-R, R]$, or $[-R, R]$.

The value of R is referred to as the *radius of convergence* of a power series, and it is customary to assign R the value 0 or ∞ to represent the set $\{0\}$ or $\mathbf{R}$, respectively. Some of the standard devices for computing the radius of convergence for a power series are explored in the exercises. Of more interest to us here is the investigation of the properties of functions defined in this way. Are they continuous? Are they differentiable? If so, can we differentiate the series term-by-term? What happens at the endpoints?

Establishing Uniform Convergence

The positive answers to the preceding questions, and the usefulness of power series in general, are largely due to the fact that they converge uniformly on compact sets contained in their domain of convergent points. As we are about to see, a complete proof of this fact requires a fairly delicate argument attributed to the Norwegian mathematician Niels Henrik Abel. A significant amount of progress, however, can be made with the Weierstrass M-Test (Corollary 6.4.5).

Theorem 6.5.2. *If a power series $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely at a point x_0, then it converges uniformly on the closed interval $[-c, c]$, where $c = |x_0|$.*

Proof. This proof requires a straightforward application of the Weierstrass M-Test. The details are requested in Exercise 6.5.3. □

For many applications, Theorem 6.5.2 is good enough. For instance, because any $x \in (-R, R)$ is contained in the interior of a closed interval $[-c, c] \subseteq (-R, R)$, it now follows that a power series that converges on an open interval is necessarily continuous on this interval.

But what happens if we know that a series converges at an endpoint of its interval of convergence? Does the good behavior of the series on $(-R, R)$ necessarily extend to the endpoint $x = R$? If the convergence of the series at $x = R$ is absolute convergence, then we can again rely on Theorem 6.5.2 to conclude that the series converges uniformly on the set $[-R, R]$. The remaining interesting open question is what happens if a series converges *conditionally* at a point $x = R$. We may still use Theorem 6.5.1 to conclude that we have pointwise convergence on the interval $(-R, R]$, but more work is needed to establish uniform convergence on compact sets containing $x = R$.

Abel's Theorem

We should remark that if the power series $g(x) = \sum_{n=0}^{\infty} a_n x^n$ converges conditionally at $x = R$, then it is possible for it to diverge when $x = -R$. The series

$$\sum_{n=1}^{\infty} \frac{(-1)^n x^n}{n}$$

with $R = 1$ is an example. To keep our attention fixed on the convergent endpoint, we will prove uniform convergence on the set $[0, R]$.

The first step in the argument is an estimate that should be compared to Abel's Test for convergence of series, developed back in Chapter 2 (Exercise 2.7.13).

Lemma 6.5.3 (Abel's Lemma). *Let b_n satisfy $b_1 \geq b_2 \geq b_3 \geq \cdots \geq 0$, and let $\sum_{n=1}^{\infty} a_n$ be a series for which the partial sums are bounded. In other words, assume there exists $A > 0$ such that*

$$|a_1 + a_2 + \cdots + a_n| \leq A$$

for all $n \in N$. Then, for all $n \in \mathbf{N}$,

$$|a_1 b_1 + a_2 b_2 + a_3 b_3 + \cdots + a_n b_n| \leq A b_1.$$

Proof. Let $s_n = a_1 + a_2 + \cdots + a_n$. Using the summation-by-parts formula derived in Exercise 2.7.12, we can write

$$\begin{aligned} \left| \sum_{k=1}^{n} a_k b_k \right| &= \left| s_n b_{n+1} + \sum_{k=1}^{n} s_k (b_k - b_{k+1}) \right| \\ &\leq A b_{n+1} + \sum_{k=1}^{n} A (b_k - b_{k+1}) \\ &= A b_{n+1} + (A b_1 - A b_{n+1}) = A b_1. \end{aligned}$$ □

It is worth observing that if A were an upper bound on the partial sums of $\sum |a_n|$ (note the absolute value bars), then the proof of Lemma 6.5.3 would be a simple exercise in the triangle inequality. The point of the matter is that because we are only assuming conditional convergence, the triangle inequality is not going to be of any use in proving Abel's Theorem, but we are now in possession of an inequality that we can use in its place.

Theorem 6.5.4 (Abel's Theorem). *Let $g(x) = \sum_{n=0}^{\infty} a_n x^n$ be a power series that converges at the point $x = R > 0$. Then the series converges uniformly on the interval $[0, R]$. A similar result holds if the series converges at $x = -R$.*

Proof. To set the stage for an application of Lemma 6.5.3, we first write

$$g(x) = \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} (a_n R^n) \left(\frac{x}{R} \right)^n.$$

Let $\epsilon > 0$. By the Cauchy Criterion for Uniform Convergence of Series (Theorem 6.4.4), we will be done if we can produce an N such that $n > m \geq N$ implies

$$(1) \qquad \left| (a_{m+1} R^{m+1}) \left(\frac{x}{R} \right)^{m+1} + (a_{m+2} R^{m+2}) \left(\frac{x}{R} \right)^{m+2} + \cdots + (a_n R^n) \left(\frac{x}{R} \right)^n \right| < \epsilon.$$

Because we are assuming that $\sum_{n=0}^{\infty} a_n R^n$ converges, the Cauchy Criterion for convergent series of real numbers guarantees that there exists an N such that

$$|a_{m+1}R^{m+1} + a_{m+2}R^{m+2} + \cdots + a_n R^n| < \frac{\epsilon}{2}$$

whenever $n > m \geq N$. But now, for any fixed $m \in \mathbf{N}$, we can apply Abel's Lemma (Lemma 6.5.3) to the sequences obtained by omitting the first m terms. Using $\epsilon/2$ as a bound on the partial sums of $\sum_{j=1}^{\infty} a_{m+j}R^{m+j}$ and observing that $(x/R)^{m+j}$ is monotone decreasing, an application of Abel's Lemma to equation (1) yields

$$\left|(a_{m+1}R^{m+1})\left(\frac{x}{R}\right)^{m+1} + (a_{m+2}R^{m+2})\left(\frac{x}{R}\right)^{m+2} + \cdots + (a_n R^n)\left(\frac{x}{R}\right)^{n}\right| \leq \frac{\epsilon}{2}\left(\frac{x}{R}\right)^{m+1} < \epsilon.$$

□

The Success of Power Series

An economical way to summarize the conclusions of Theorem 6.5.2 and Abel's Theorem is with the following statement.

Theorem 6.5.5. *If a power series converges pointwise on the set $A \subseteq \mathbf{R}$, then it converges uniformly on any compact set $K \subseteq A$.*

Proof. A compact set contains both a maximum x_1 and a minimum x_0, which by hypothesis must be in A. Abel's Theorem implies the series converges uniformly on the interval $[x_0, x_1]$ and thus also on K. □

This fact leads to the desirable conclusion that a power series is continuous at every point at which it converges. To make an argument for differentiability, we would like to appeal to Theorem 6.4.3; however, this result has a slightly more involved set of hypotheses. In order to conclude that a power series $\sum_{n=0}^{\infty} a_n x^n$ is differentiable, and that term-by-term differentiation is allowed, we need to know beforehand that the differentiated series $\sum_{n=1}^{\infty} n a_n x^{n-1}$ converges uniformly.

Theorem 6.5.6. *If $\sum_{n=0}^{\infty} a_n x^n$ converges for all $x \in (-R, R)$, then the differentiated series $\sum_{n=1}^{\infty} n a_n x^{n-1}$ converges at each $x \in (-R, R)$ as well. Consequently, the convergence is uniform on compact sets contained in $(-R, R)$.*

Proof. Exercise 6.5.5. □

We should point out that it is possible for a series to converge at an endpoint $x = R$ but for the differentiated series to diverge at this point. The series $\sum_{n=1}^{\infty} x^n/n$ has this property when $x = -1$. On the other hand, if the differentiated series does converge at the point $x = R$, then Abel's Theorem applies and the convergence of the differentiated series is uniform on compact sets that contain R.

With all the pieces in place, we summarize the impressive conclusions of this section.

Theorem 6.5.7. *Assume*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

converges on an interval $A \subseteq \mathbf{R}$. *The function* f *is continuous on* A *and differentiable on any open interval* $(-R, R) \subseteq A$. *The derivative is given by*

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

Moreover, f *is infinitely differentiable on* $(-R, R)$*, and the successive derivatives can be obtained via term-by-term differentiation of the appropriate series.*

Proof. The details for why f is continuous have been discussed. Theorem 6.5.6 justifies the application of the Term-by-term Differentiability Theorem (Theorem 6.4.3), which verifies the formula for f'.

A differentiated power series is a power series in its own right, and Theorem 6.5.6 implies that, although the series may no longer converge at a particular endpoint, the radius of convergence does not change. By induction, then, power series are differentiable an infinite number of times. □

Exercises

Exercise 6.5.1. Consider the function g defined by the power series

$$g(x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \cdots .$$

(a) Is g defined on $(-1, 1)$? Is it continuous on this set? Is g defined on $(-1, 1]$? Is it continuous on this set? What happens on $[-1, 1]$? Can the power series for $g(x)$ possibly converge for any other points $|x| > 1$? Explain.

(b) For what values of x is $g'(x)$ defined? Find a formula for g'.

Exercise 6.5.2. Find suitable coefficients (a_n) so that the resulting power series $\sum a_n x^n$ has the given properties, or explain why such a request is impossible.

(a) Converges for every value of $x \in \mathbf{R}$.

(b) Diverges for every value of $x \in \mathbf{R}$.

(c) Converges absolutely for all $x \in [-1, 1]$ and diverges off of this set.

(d) Converges conditionally at $x = -1$ and converges absolutely at $x = 1$.

(e) Converges conditionally at both $x = -1$ and $x = 1$.

Exercise 6.5.3. Use the Weierstrass M-Test to prove Theorem 6.5.2.

Exercise 6.5.4 (Term-by-term Antidifferentiation). Assume $f(x) = \sum_{n=0}^{\infty} a_n x^n$ converges on $(-R, R)$.

(a) Show

$$F(x) = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}$$

is defined on $(-R, R)$ and satisfies $F'(x) = f(x)$.

(b) Antiderivatives are not unique. If g is an arbitrary function satisfying $g'(x) = f(x)$ on $(-R, R)$, find a power series representation for g.

Exercise 6.5.5. (a) If s satisfies $0 < s < 1$, show ns^{n-1} is bounded for all $n \geq 1$.

(b) Given an arbitrary $x \in (-R, R)$, pick t to satisfy $|x| < t < R$. Use this start to construct a proof for Theorem 6.5.6.

Exercise 6.5.6. Previous work on geometric series (Example 2.7.5) justifies the formula

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \cdots, \qquad \text{for all } |x| < 1.$$

Use the results about power series proved in this section to find values for $\sum_{n=1}^{\infty} n/2^n$ and $\sum_{n=1}^{\infty} n^2/2^n$. The discussion in Section 6.1 may be helpful.

Exercise 6.5.7. Let $\sum a_n x^n$ be a power series with $a_n \neq 0$, and assume

$$L = \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right|$$

exists.

(a) Show that if $L \neq 0$, then the series converges for all x in $(-1/L, 1/L)$. (The advice in Exercise 2.7.9 may be helpful.)

(b) Show that if $L = 0$, then the series converges for all $x \in \mathbf{R}$.

(c) Show that (a) and (b) continue to hold if L is replaced by the limit

$$L' = \lim_{n \to \infty} s_n \quad \text{where} \quad s_n = \sup \left\{ \left| \frac{a_{k+1}}{a_k} \right| : k \geq n \right\}.$$

(General properties of the *limit superior* are discussed in Exercise 2.4.7.)

Exercise 6.5.8. (a) Show that power series representations are unique. If we have

$$\sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} b_n x^n$$

for all x in an interval $(-R, R)$, prove that $a_n = b_n$ for all $n = 0, 1, 2, \ldots$.

(b) Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ converge on $(-R, R)$, and assume $f'(x) = f(x)$ for all $x \in (-R, R)$ and $f(0) = 1$. Deduce the values of a_n.

Exercise 6.5.9. Review the definitions and results from Section 2.8 concerning products of series and Cauchy products in particular. At the end of Section 2.9, we mentioned the following result: If both $\sum a_n$ and $\sum b_n$ converge conditionally to A and B respectively, then it is possible for the Cauchy product,

$$\sum d_n \quad \text{where} \quad d_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0,$$

to diverge. However, if $\sum d_n$ does converge, then it must converge to AB. To prove this, set

$$f(x) = \sum a_n x^n, \quad g(x) = \sum b_n x^n, \quad \text{and} \quad h(x) = \sum d_n x^n.$$

Use Abel's Theorem and the result in Exercise 2.8.7 to establish this result.

Exercise 6.5.10. Let $g(x) = \sum_{n=0}^{\infty} b_n x^n$ converge on $(-R, R)$, and assume $(x_n) \to 0$ with $x_n \neq 0$. If $g(x_n) = 0$ for all $n \in \mathbf{N}$, show that $g(x)$ must be identically zero on all of $(-R, R)$.

Exercise 6.5.11. A series $\sum_{n=0}^{\infty} a_n$ is said to be *Abel-summable to* L if the power series

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

converges for all $x \in [0, 1)$ and $L = \lim_{x \to 1^-} f(x)$.

(a) Show that any series that converges to a limit L is also Abel-summable to L.

(b) Show that $\sum_{n=0}^{\infty} (-1)^n$ is Abel-summable and find the sum.

6.6 Taylor Series

Our study of power series has led to some enthusiastic conclusions about the nature of functions of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \cdots .$$

Despite their infinite character, power series can be manipulated more or less as though they are polynomials. On its interval of convergence, a power series is continuous and infinitely differentiable, and successive derivatives or antiderivatives can be computed by performing the desired operation on each individual term in the series—just as it is done for polynomials.

In Section 6.1 we informally encountered the powerful idea that familiar functions such as $\arctan(x)$ and $\sqrt{1+x}$ can be represented as power series. This is a game changing revelation. If a function can be represented as a power series, and a power series can be treated like a polynomial, then vast new possibilities are suddenly available for the kinds of calculations that can be undertaken. Given this state of affairs, it is natural to wonder whether *all* of the well-behaved—i.e., infinitely differentiable—functions of calculus might have representations as power series.

In the examples and exercises in this section, we will assume the familiar properties of the trigonometric, inverse trigonometric, exponential, and logarithmic functions. Rigorously defining these functions is an important exercise in analysis. In fact, one of the most common methods for providing proper definitions is through power series, a point of view that is explored in Section 8.4. The point of this discussion, however, is to come at this question from the other direction. Assuming we are in possession of an infinitely differentiable function such as $\sin(x)$, can we find suitable coefficients a_n so that

$$\sin(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \cdots$$

for at least some nonzero values of x?

Manipulating Series

In Section 6.1 we generated several new series representations starting from the formula

$$(1) \qquad \frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \cdots, \quad \text{for all } |x| < 1$$

proved in Example 2.7.5. At the time, we were not concerned with supplying rigorous proofs, but we have since done the bulk of the work necessary to confidently assert that the manipulations in Section 6.1 are perfectly valid.

Example 6.6.1. Theorem 6.5.7 applied to equation (1) gives

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + 5x^4 + \cdots, \quad \text{for all } |x| < 1.$$

What about the series we generated for $\arctan(x)$? The substitution of $-x^2$ for x in (1) doesn't cause any problem:

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + x^8 - \cdots, \quad \text{for all } |x| < 1.$$

The content of Exercise 6.5.4 is that we can take the term-by-term antiderivative of this series and arrive at an antiderivative for $1/(1+x^2)$. Noting that $\arctan(0) = 0$, it follows that

$$\arctan(x) = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \cdots, \tag{2}$$

for all $x \in (-1, 1)$. In fact, this formula is also valid for $x = \pm 1$. (Exercise 6.6.1.) Similar methods can be used to find series representations for functions such as $\log(1+x)$ and $x/(1+x^2)^2$.

Taylor's Formula for the Coefficients

Manipulating old series to produce new ones was a well-honed craft in the 17th and 18th centuries, but there also emerged a formula for producing the coefficients from "scratch"—a recipe for generating a power series representation using only the function in question and its derivatives. The technique is named after the mathematician Brook Taylor (1685–1731) who published it in 1715, although it was certainly known previous to this date.

Given an infinitely differentiable function f defined on some interval centered at zero, the idea is to assume that f has a power series expansion and deduce what the coefficients must be.

Theorem 6.6.2 (Taylor's Formula). *Let*

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 + \cdots \tag{3}$$

be defined on some nontrivial interval centered at zero. Then,

$$a_n = \frac{f^{(n)}(0)}{n!}.$$

Proof. Exercise 6.6.3 □

Let's use Taylor's formula to produce the so-called *Taylor series* for $\sin(x)$. For the constant term we get $a_0 = \sin(0) = 0$. Then, $a_1 = \cos(0) = 1$, $a_2 = -\sin(0)/2! = 0$, and $a_3 = -\cos(0)/3! = -1/3!$. Continuing on, we are led to the series

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots.$$

So can we say that this series *equals* $\sin(x)$? Well, we need to be very clear about what we have proved to this point. To derive Taylor's formula, *we assumed that f actually had a power series representation.* The conclusion is that if f can be expressed in the form

$$f(x) = \sum_{n=0}^{\infty} a_n x^n,$$

then it must be that

$$a_n = \frac{f^{(n)}(0)}{n!}.$$

But what about the converse question? Assume f is infinitely differentiable in a neighborhood of zero. If we let

$$a_n = \frac{f^{(n)}(0)}{n!},$$

does the resulting series

$$\sum_{n=0}^{\infty} a_n x^n$$

converge to $f(x)$ on some nontrivial set of points? Does it converge at all? If it does converge, we know that the limit function is an infinitely differentiable function whose derivatives at zero are exactly the same as the derivatives of f. Is it possible for this limit to be different from f? In other words, might the Taylor series of a function converge to the wrong thing?

Let

$$S_N(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_N x^N.$$

The polynomial $S_N(x)$ is a partial sum of the Taylor series expansion for the function $f(x)$. Thus, we are interested in whether or not

$$\lim_{N\to\infty} S_N(x) = f(x)$$

for some values of x besides zero.

Lagrange's Remainder Theorem

A powerful tool for analyzing this question was provided by Joseph Louis Lagrange (1736–1813). The idea is to consider the difference

$$E_N(x) = f(x) - S_N(x),$$

which represents the error between f and the partial sum S_N.

Theorem 6.6.3 (Lagrange's Remainder Theorem). *Let f be differentiable $N+1$ times on $(-R, R)$, define $a_n = f^{(n)}(0)/n!$ for $n = 0, 1, \ldots, N$, and let*

$$S_N(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_N x^N.$$

Given $x \neq 0$ in $(-R, R)$, there exists a point c satisfying $|c| < |x|$ where the error function $E_N(x) = f(x) - S_N(x)$ satisfies

$$E_N(x) = \frac{f^{(N+1)}(c)}{(N+1)!} x^{N+1}.$$

Before embarking on a proof, let's examine the significance of this result. Proving $S_N(x) \to f(x)$ is equivalent to showing $E_N(x) \to 0$. There are three components to the expression for $E_N(x)$. In the denominator, we have $(N+1)!$, which helps to make E_N small as N tends to infinity. In the numerator, we have x^{N+1}, which potentially grows depending on the size of x. Thus, we should expect that a Taylor series is less likely to converge the farther x is chosen from the origin. Finally, we have $f^{(N+1)}(c)$, which is a bit of a mystery. For functions with straightforward derivatives, this term can often be handled using a suitable upper bound.

Example 6.6.4. Consider the Taylor series for $\sin(x)$ generated earlier. How well does

$$S_5(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5$$

approximate $\sin(x)$ on the interval $[-2, 2]$? Lagrange's Remainder Theorem asserts that the difference between these two functions is

$$E_5(x) = \sin(x) - S_5(x) = \frac{-\sin(c)}{6!}x^6$$

for some c in the interval $(-|x|, |x|)$. Not knowing the value of c, we can still be quite certain that $|\sin(c)| \leq 1$. Because $x \in [-2, 2]$, we have

$$|E_5(x)| \leq \frac{2^6}{6!} \approx .089.$$

To prove that $S_N(x)$ converges uniformly to $\sin(x)$ on $[-2, 2]$, we observe that the $f^{(N+1)}(c)$ term in the Lagrange formula will never exceed 1 in absolute value. Thus,

$$|E_N(x)| = \left| \frac{f^{(N+1)}(c)}{(N+1)!} x^{N+1} \right| \leq \frac{1}{(N+1)!} 2^{N+1}$$

for $x \in [-2, 2]$. Because factorials grow significantly faster than exponentials, it follows that $E_N(x) \to 0$ uniformly on $[-2, 2]$.

Replacing the constant 2 with an arbitrary constant R has no effect on the validity of the argument, and so the Taylor series converges uniformly to $\sin(x)$ on every interval of the form $[-R, R]$.

Proof of Lagrange's Remainder Theorem: The Taylor coefficients are chosen so that the function f and the polynomial S_N have the same derivatives at zero, at least up through the Nth derivative, after which S_N becomes the zero function. In other words, $f^{(n)}(0) = S_N^{(n)}(0)$ for all $0 \leq n \leq N$, which implies the error function $E_N(x) = f(x) - S_N(x)$ satisfies

$$E_N^{(n)}(0) = 0 \quad \text{for all } n = 0, 1, 2, \ldots, N.$$

The key ingredient in this argument is the Generalized Mean Value Theorem (Theorem 5.3.5) from Chapter 5. To simplify notation, let's assume $x > 0$ and apply the Generalized Mean Value Theorem to the functions $E_N(x)$ and x^{N+1} on the interval $[0, x]$. Thus, there exists a point $x_1 \in (0, x)$ such that

$$\frac{E_N(x)}{x^{N+1}} = \frac{E'_N(x_1)}{(N+1)x_1^N}.$$

Now apply the Generalized Mean Value Theorem to the functions $E'_N(x)$ and $(N+1)x^N$ on the interval $[0, x_1]$ to get that there exists a point $x_2 \in (0, x_1)$ where

$$\frac{E_N(x)}{x^{N+1}} = \frac{E'_N(x_1)}{(N+1)x_1^N} = \frac{E''_N(x_2)}{(N+1)Nx_2^{N-1}}.$$

Continuing in this manner we find

$$\frac{E_N(x)}{x^{N+1}} = \frac{E_N^{(N+1)}(x_{N+1})}{(N+1)!}$$

where $x_{N+1} \in (0, x_N) \subseteq \cdots \subseteq (0, x)$. Now set $c = x_{N+1}$. Because $S_N^{(N+1)}(x) = 0$, we have $E_N^{(N+1)}(x) = f^{(N+1)}(x)$ and it follows that

$$E_N(x) = \frac{f^{(N+1)}(c)}{(N+1)!} x^{N+1}$$

as desired. □

Taylor Series Centered at $a \neq 0$.

Throughout this chapter we have focused our attention on series expansions centered at zero, but there is nothing special about zero other than notational simplicity. If f is defined in some neighborhood of $a \in \mathbf{R}$ and infinitely differentiable at a, then the Taylor series expansion around a takes the form

$$\sum_{n=0}^{\infty} c_n (x-a)^n \quad \text{where} \quad c_n = \frac{f^{(n)}(a)}{n!}.$$

Setting $E_N(x) = f(x) - S_N(x)$ as usual, Lagrange's Remainder Theorem in this case says that there exists a value c between a and x where

$$E_N(x) = \frac{f^{(N+1)}(c)}{(N+1)!}(x-a)^{N+1}.$$

In Exercise 6.6.9, we derive an alternate remainder formula due to Cauchy that requires these more general expansions for its derivation.

A Counterexample

Lagrange's Remainder Theorem is extremely useful for determining how well the partial sums of the Taylor series approximate the original function, but it leaves unresolved the central question of whether or not the Taylor series necessarily converges to the function that generated it. The appearance of $f^{(N+1)}(c)$ in the error formula makes any general statement impossible. The Cauchy form of the remainder just mentioned provides another way to represent the error between the partial sum $S_N(x)$ and the function $f(x)$, and there are others still, but none lend themselves to a proof that $S_N \to f$. This is because no such proof exists! Let

$$g(x) = \begin{cases} e^{-1/x^2} & \text{for } x \neq 0, \\ 0 & \text{for } x = 0. \end{cases}$$

Computing the Taylor coefficients for this function, it's clear that $a_0 = g(0) = 0$. To compute a_1 we write

$$a_1 = g'(0) = \lim_{x\to 0} \frac{g(x) - g(0)}{x - 0} = \lim_{x\to 0} \frac{e^{-1/x^2}}{x} = \lim_{x\to 0} \frac{1/x}{e^{1/x^2}}$$

where both numerator and denominator tend to ∞ as x approaches zero. Applying the ∞/∞ version of L'Hospital's Rule (Theorem 5.3.8) we see

$$a_1 = \lim_{x\to 0} \frac{-1/x^2}{e^{1/x^2}(-2/x^3)} = \lim_{x\to 0} \frac{x}{2e^{1/x^2}} = 0.$$

This tells us that g is flat at the origin. In Exercise 6.6.6, we outline the rest of the proof showing that $g^{(n)}(0) = 0$ for all $n \in \mathbf{N}$; in other words, g is *extremely* flat at the origin.

The implications of this example are highly significant. The function g is infinitely differentiable, and every one of its Taylor coefficients is equal to zero. By default, then, its Taylor series converges uniformly on all of $\mathbf{R}$ to the zero function. But other than at $x = 0$, $g(x)$ is never equal to zero. *The Taylor series for $g(x)$ converges, but it does not converge to $g(x)$* except at the center point $x = 0$. The unmistakable conclusion is that not every infinitely differentiable function can be represented by its Taylor series.

Exercises

Exercise 6.6.1. The derivation in Example 6.6.1 shows the Taylor series for $\arctan(x)$ is valid for all $x \in (-1, 1)$. Notice, however, that the series also converges when $x = 1$. Assuming that $\arctan(x)$ is continuous, explain why the value of the series at $x = 1$ must necessarily be $\arctan(1)$. What interesting identity do we get in this case?

Exercise 6.6.2. Starting from one of the previously generated series in this section, use manipulations similar to those in Example 6.6.1 to find Taylor series representations for each of the following functions. For precisely what values of x is each series representation valid?

(a) $x\cos(x^2)$

(b) $x/(1+4x^2)^2$

(c) $\log(1+x^2)$

Exercise 6.6.3. Derive the formula for the Taylor coefficients given in Theorem 6.6.2.

Exercise 6.6.4. Explain how Lagrange's Remainder Theorem can be modified to prove

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots = \log(2).$$

Exercise 6.6.5. (a) Generate the Taylor coefficients for the exponential function $f(x) = e^x$, and then prove that the corresponding Taylor series converges uniformly to e^x on any interval of the form $[-R, R]$.

(b) Verify the formula $f'(x) = e^x$.

(c) Use a substitution to generate the series for e^{-x}, and then informally calculate $e^x \cdot e^{-x}$ by multiplying together the two series and collecting common powers of x.

Exercise 6.6.6. Review the proof that $g'(0) = 0$ for the function

$$g(x) = \begin{cases} e^{-1/x^2} & \text{for } x \neq 0, \\ 0 & \text{for } x = 0. \end{cases}$$

introduced at the end of this section.

(a) Compute $g'(x)$ for $x \neq 0$. Then use the definition of the derivative to find $g''(0)$.

(b) Compute $g''(x)$ and $g'''(x)$ for $x \neq 0$. Use these observations and invent whatever notation is needed to give a general description for the nth derivative $g^{(n)}(x)$ at points different from zero.

(c) Construct a general argument for why $g^{(n)}(0) = 0$ for all $n \in \mathbf{N}$.

Exercise 6.6.7. Find an example of each of the following or explain why no such function exists.

(a) An infinitely differentiable function $g(x)$ on all of $\mathbf{R}$ with a Taylor series that converges to $g(x)$ only for $x \in (-1, 1)$.

(b) An infinitely differentiable function $h(x)$ with the same Taylor series as $\sin(x)$ but such that $h(x) \neq \sin(x)$ for all $x \neq 0$.

(c) An infinitely differentiable function $f(x)$ on all of $\mathbf{R}$ with a Taylor series that converges to $f(x)$ if and only if $x \leq 0$.

Exercise 6.6.8. Here is a weaker form of Lagrange's Remainder Theorem whose proof is arguably more illuminating than the one for the stronger result.

(a) First establish a lemma: If g and h are differentiable on $[0, x]$ with $g(0) = h(0)$ and $g'(t) \leq h'(t)$ for all $t \in [0, x]$, then $g(t) \leq h(t)$ for all $t \in [0, x]$.

(b) Let f, S_N, and E_N be as Theorem 6.6.3, and take $0 < x < R$. If $|f^{(N+1)}(t)| \leq M$ for all $t \in [0, x]$, show

$$|E_N(x)| \leq \frac{Mx^{N+1}}{(N+1)!}.$$

Exercise 6.6.9 (Cauchy's Remainder Theorem). Let f be differentiable $N+1$ times on $(-R, R)$. For each $a \in (-R, R)$, let $S_N(x, a)$ be the partial sum of the Taylor series for f centered at a; in other words, define

$$S_N(x, a) = \sum_{n=0}^{N} c_n (x-a)^n \quad \text{where} \quad c_n = \frac{f^{(n)}(a)}{n!}.$$

Let $E_N(x, a) = f(x) - S_N(x, a)$. Now fix $x \neq 0$ in $(-R, R)$ and consider $E_N(x, a)$ as a function of a.

(a) Find $E_N(x, x)$.

(b) Explain why $E_N(x, a)$ is differentiable with respect to a, and show

$$E_N'(x, a) = \frac{-f^{(N+1)}(a)}{N!}(x-a)^N.$$

(c) Show

$$E_N(x) = E_N(x, 0) = \frac{f^{(N+1)}(c)}{N!}(x-c)^N x$$

for some c between 0 and x. This is Cauchy's form of the remainder for Taylor series centered at the origin.

Exercise 6.6.10. Consider $f(x) = 1/\sqrt{1-x}$.

(a) Generate the Taylor series for f centered at zero, and use Lagrange's Remainder Theorem to show the series converges to f on $[0, 1/2]$. (The case $x < 1/2$ is more straightforward while $x = 1/2$ requires some extra care.) What happens when we attempt this with $x > 1/2$?

(b) Use Cauchy's Remainder Theorem proved in Exercise 6.6.9 to show the series representation for f holds on $[0, 1)$.

6.7 The Weierstrass Approximation Theorem

Karl Weierstrass's name is attached to a number of significant results discussed already. The Bolzano-Weierstrass Theorem was fundamental to understanding the relationship between convergence, completeness, and compactness worked out in the early chapters. In this chapter, the Weierstrass M-Test emerged as the primary tool for demonstrating uniform convergence of infinite series.

As discussed in Section 5.4, Weierstrass was also responsible for one of the earliest examples of a continuous, nowhere differentiable function, making this discovery in 1872.

In 1885, Weierstrass proved a result that served as an interesting counterpoint to his nowhere differentiable function. This theorem, which also bears his name, would become the catalyst for a new branch of analysis called approximation theory.

Theorem 6.7.1 (Weierstrass Approximation Theorem). *Let $f : [a, b] \to \mathbf{R}$ be continuous. Given $\epsilon > 0$, there exists a polynomial $p(x)$ satisfying*

$$|f(x) - p(x)| < \epsilon$$

for all $x \in [a, b]$.

A restatement of the Weierstrass Approximation Theorem (WAT) without all the symbols is that every continuous function on a closed interval can be uniformly approximated by a polynomial.

Exercise 6.7.1. Assuming WAT, show that if f is continuous on $[a, b]$, then there exists a sequence (p_n) of polynomials such that $p_n \to f$ uniformly on $[a, b]$.

Our work in the previous section provides a nice starting point for understanding what WAT is saying. Given a function such as $\sin(x)$, we saw in Example 6.6.4 that the resulting Taylor series converges uniformly on compact sets back to $\sin(x)$. Because the partial sums of a Taylor series are polynomials, this example constitutes a proof of WAT in the very special case of $f(x) = \sin(x)$. It should be clear, however, that Taylor series won't work in general. To construct a Taylor series, we need f to be an infinitely differentiable function (and even then the Taylor series might fail to approximate f), while WAT requires only that f be continuous.

So should we be surprised that such a theorem is true? This is hard to say. On a purely intuitive level, if we consider a smooth curve like $f(x) = \sqrt{1-x}$ on $[-1, 1]$, then it doesn't take too much imagination to believe that a polynomial might exist that tracks closely with $\sqrt{1-x}$ as x moves over the domain. But one of the lessons of Section 5.4 is that a continuous function does not have to be smooth. Although it is not Weierstrass's original example, a careful look at the nowhere differentiable function shown in Figure 5.7 makes the point just as well. Despite the unimaginably jagged nature of the graph, according to WAT, it is still possible to find a polynomial that uniformly approximates this unruly function to any prescribed degree of accuracy.

Interpolation

Weierstrass's theorem deals with approximating polynomials, but a good way to get a feel for the content of this result is to temporarily replace the polynomials in WAT with the collection of all continuous, piecewise-linear functions.

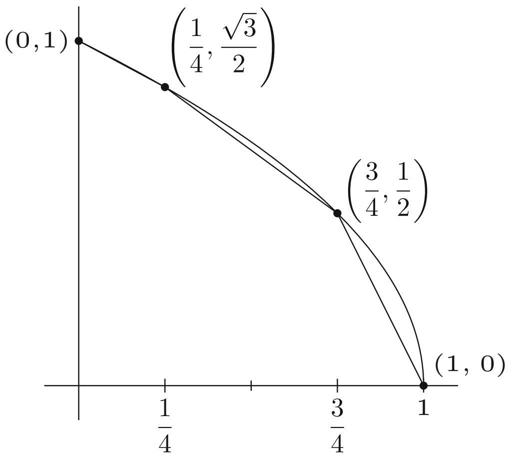

Figure 6.6: Polygonal approximation of $f(x) = \sqrt{1-x}$.

Definition 6.7.2. A continuous function $\phi : [a, b] \to \mathbf{R}$ is *polygonal* if there is a partition

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

of $[a, b]$ such that ϕ is linear on each subinterval $[x_{i-1}, x_i]$, where $i = 1, \ldots n$.

The term "interpolation" refers to the process of finding a function whose graph passes through a given set of points. If, for example, we take the points

$$(0, 1), \left(\frac{1}{4}, \frac{\sqrt{3}}{2}\right), \left(\frac{3}{4}, \frac{1}{2}\right), (1, 0)$$

then there is an obvious polygonal function that interpolates these points: it is just the function we get by connecting the points with line segments. Now these four points all lie on the graph of $f(x) = \sqrt{1-x}$, and notice that the resulting polygonal interpolation does a reasonable job of imitating the graph of f (Fig. 6.6). This is not an accident.

Theorem 6.7.3. *Let $f : [a, b] \to \mathbf{R}$ be continuous. Given $\epsilon > 0$, there exists a polygonal function ϕ satisfying*

$$|f(x) - \phi(x)| < \epsilon$$

for all $x \in [a, b]$.

Exercise 6.7.2. Prove Theorem 6.7.3.

Notice how similar Theorem 6.7.3 is to WAT, the only difference being that we have substituted a polygonal function in place of the polynomial.

The strategy for the proof of Theorem 6.7.3 is to first choose an appropriate numbers of points on the graph of f, and then show that the resulting polygonal interpolation of these points does the trick. It's not unreasonable to suspect that a similar strategy might lead to a proof of the Weierstrass Approximation Theorem. Can we prove WAT by constructing a polynomial interpolation of points on the graph of f? Well, no as it turns out, but this is not so easy to see.

Exercise 6.7.3. (a) Find the second degree polynomial $p(x) = q_0 + q_1 x + q_2 x^2$ that interpolates the three points $(-1, 1), (0, 0)$, and $(1, 1)$ on the graph of $g(x) = |x|$. Sketch $g(x)$ and $p(x)$ over $[-1, 1]$ on the same set of axes.

(b) Find the fourth degree polynomial that interpolates $g(x) = |x|$ at the points $x = -1, -1/2, 0, 1/2$, and 1. Add a sketch of this polynomial to the graph from (a).

The previous exercise may still give the impression that a polynomial interpolation approach is going to lead to a proof of WAT, but that isn't the case. Continuing on with larger and larger numbers of equally spaced points yields high degree polynomials that oscillate very rapidly and actually do a poor job of approximating g *between* the interpolating points. In fact, it turns out that the resulting sequence of polynomials only converges to $g(x)$ when $x = -1, 0,$ or 1.

Approximating the Absolute Value Function

Having reached a temporary dead end, we need to back up a bit and take a different turn. Let's return to Theorem 6.7.3 which asserts that every continuous function can be uniformly approximated by a polygonal function. This should feel like a promising first step toward a proof of WAT and indeed it is. If we can find a way to approximate an arbitrary polygonal function with polynomials, then a triangle inequality argument would finish the proof.

Before we get too excited about this line of attack, keep in mind that the absolute value function from Exercise 6.7.3 is an example of a polygonal function and we are currently unsure how to produce polynomials to approximate it. What has changed, however, is our motivation for doing so. A moment's thought reveals that handling the absolute value function might be the key to solving the whole problem. Why is this? Every polygonal function is made up of line segments that meet at corners. If we can find polynomials that uniformly approximate $g(x) = |x|$ with its right angled corner at the origin, then with a little cleverness we ought to be able to handle more general polygonal functions and prove WAT using Theorem 6.7.3.

Cauchy's Remainder Formula for Taylor Series

One elegant way to show $g(x) = |x|$ is the uniform limit of polynomials is via Taylor series, which is a bit surprising given that $|x|$ is not differentiable. The trick, as we will see, is to start by computing the Taylor series for the infinitely differentiable function $\sqrt{1-x}$.

Exercise 6.7.4. Show that $f(x) = \sqrt{1-x}$ has Taylor series coefficients a_n where $a_0 = 1$ and

$$a_n = \frac{-1 \cdot 3 \cdot 5 \cdots (2n-3)}{2 \cdot 4 \cdot 6 \cdots 2n}$$

for $n \geq 1$.

Our goal is to show

$$\sqrt{1-x} = \sum_{n=0}^{\infty} a_n x^n \tag{1}$$

for all $x \in [-1, 1]$ by showing that the error function

$$E_N(x) = \sqrt{1-x} - \sum_{n=0}^{N} a_n x^n$$

tends to 0 as $N \to \infty$. To this point, Lagrange's Remainder Theorem has been the featured tool for jobs like this, but it comes up short in this case. To see exactly why, fix $x \in (0, 1]$. Then Theorem 6.6.3 asserts that there exists a $c \in (0, x)$ (dependent on N) such that

$$\begin{aligned} E_N(x) &= \frac{f^{(N+1)}(c)}{(N+1)!} x^{N+1} \\ &= \frac{1}{(N+1)!} \left(\frac{-1 \cdot 3 \cdot 5 \cdots (2N-1)}{2^{N+1}(1-c)^{N+1/2}} \right) x^{N+1} \\ &= \left(\frac{-1 \cdot 3 \cdot 5 \cdots (2N-1)}{2 \cdot 4 \cdot 6 \cdots (2N+2)} \right) \left(\frac{x}{1-c} \right)^{N+1/2} x^{1/2} . \end{aligned}$$

The problem is that $x/(1-c)$ is largest when $c = x$, and $(x/(1-x))^{N+1/2}$ goes exponentially to infinity when x is bigger than $1/2$. This doesn't mean our Taylor series is only valid on $[0, 1/2]$; it just means we are using the wrong remainder formula.

Exercise 6.7.5. (a) Follow the advice in Exercise 6.6.9 to prove the Cauchy form of the remainder:

$$E_N(x) = \frac{f^{(N+1)}(c)}{N!} (x-c)^N x$$

for some c between 0 and x.

(b) Use this result to prove equation (1) is valid for all $x \in (-1, 1)$.

Although Cauchy's Remainder Theorem doesn't tell us so, equation (1) is also valid at $x = \pm 1$.

Exercise 6.7.6. (a) Let

$$c_n = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots 2n}$$

for $n \geq 1$. Show $c_n < \frac{2}{\sqrt{2n+1}}$.

(b) Use (a) to show that $\sum_{n=0}^{\infty} a_n$ converges (absolutely, in fact) where a_n is the sequence of Taylor coefficients generated in Exercise 6.7.4.

(c) Carefully explain how this verifies that equation (1) holds for all $x \in [-1, 1]$.

Recall that our goal is to find polynomials that uniformly approximate the absolute value function on an interval containing the non-differentiable point at the origin. Our Taylor series for $\sqrt{1-x}$ provides a clever shortcut for handling this task.

Exercise 6.7.7. (a) Use the fact that $|a| = \sqrt{a^2}$ to prove that, given $\epsilon > 0$, there exists a polynomial $q(x)$ satisfying

$$||x| - q(x)| < \epsilon$$

for all $x \in [-1, 1]$.

(b) Generalize this conclusion to an arbitrary interval $[a, b]$.

Proving WAT

Earlier we suggested that proving WAT for the special case of the absolute value function was the key to the whole proof. Now it is time to fill in the details.

Exercise 6.7.8. (a) Fix $a \in [-1, 1]$ and sketch

$$h_a(x) = \frac{1}{2}(|x-a| + (x-a))$$

over $[-1, 1]$. Note that h_a is polygonal and satisfies $h_a(x) = 0$ for all $x \in [-1, a]$.

(b) Explain why we know $h_a(x)$ can be uniformly approximated with a polynomial on $[-1, 1]$.

(c) Let ϕ be a polygonal function that is linear on each subinterval of the partition

$$-1 = a_0 < a_1 < a_2 < \cdots < a_n = 1 .$$

Show there exist constants $b_0, b_1, \ldots, b_{n-1}$ so that

$$\phi(x) = \phi(-1) + b_0 h_{a_0}(x) + b_1 h_{a_1}(x) + \cdots + b_{n-1} h_{a_{n-1}}(x)$$

for all $x \in [-1, 1]$.

(d) Complete the proof of WAT for the interval $[-1, 1]$, and then generalize to an arbitrary interval $[a, b]$.

Exercise 6.7.9. (a) Find a counterexample which shows that WAT is not true if we replace the closed interval $[a, b]$ with the open interval (a, b).

(b) What happens if we replace $[a, b]$ with the closed set $[a, \infty)$. Does the theorem still hold?

Exercise 6.7.10. Is there a countable subset of polynomials $\mathcal{C}$ with the property that every continuous function on $[a, b]$ can be uniformly approximated by polynomials from $\mathcal{C}$?

Exercise 6.7.11. Assume that f has a continuous derivative on $[a, b]$. Show that there exists a polynomial $p(x)$ such that

$$|f(x) - p(x)| < \epsilon \quad \text{and} \quad |f'(x) - p'(x)| < \epsilon$$

for all $x \in [a, b]$.

6.8 Epilogue

The argument sketched out here for the Weierstrass Approximation Theorem is due to Henri Lebesque, who published his proof in 1898. Its greatest virtue is its relative simplicity. Starting from a single special case—the absolute value function—we managed to bootstrap our way up to an arbitrary continuous function. A downside of this approach is that by the time we reach the case of a general continuous function, there is no practical way to explicitly write down a formula for the polynomial that approximates it.

There are a number of other proofs for WAT that don't have this drawback. A particularly popular one was provided by Sergei Bernstein. Bernstein employs a family of polynomials—now called Bernstein polynomials—that have become important in their own right. Weierstrass's original approach was also quite elegant. His proof has much in common with the proof of Fejér's Theorem in Section 8.5 on Fourier series. Not coincidentally, it is possible to derive yet another proof of WAT as a corollary to Fejér's Theorem. (See Exercise 8.5.11.)

The Weierstrass Approximation Theorem is set on a closed interval $[a, b]$. Exercise 6.7.9 is included to emphasize the importance of the closed and bounded nature of the domain, but it should not be too surprising that the theorem will remain true if we replace $[a, b]$ with an arbitrary compact set. What about replacing the set of polynomials? Are there other collections of relatively simple continuous functions that can be used to approximate an arbitrary continuous function? Sure there are. In Theorem 6.7.3 we saw that polygonal functions have this property, and there are other examples as well. In the late 1930s, Marshall Stone proved a far-reaching generalization of the Weierstrass Approximation Theorem. Stone's version of WAT starts with an arbitrary compact set K and a collection $\mathcal{C}$ of continuous functions on K with the following three properties:

(i) the constant function $k(x) = 1$ is in $\mathcal{C}$,

(ii) if $p, q \in \mathcal{C}$ and $c \in \mathbf{R}$ then $p + q, pq$, and cp are all in $\mathcal{C}$,

(iii) if $x \neq y$ in K, then there exists $p \in \mathcal{C}$ with $p(x) \neq p(y)$.

Under these conditions, Stone showed that any continuous function on K could be uniformly approximated by functions in $\mathcal{C}$. This result, referred to as the Stone–Weierstrass Theorem, has a slightly more involved proof that tracks very closely with Lebesgue's proof of WAT outlined in the previous section. In particular, both arguments depend fundamentally on being able to approximate the absolute value function with polynomials.

A collection of functions that possesses property (ii) of the Stone–Weierstrass Theorem is called an *algebra*. An algebra that possesses property (iii) is said to *separate points*. Having the constant function $k(x) = 1$ in the algebra ensures we don't have some $x_0 \in K$ where $p(x_0) = 0$ for all functions in our algebra. (Why would this be problematic?) It is straightforward to check that the set of polynomials as well as the set of polygonal functions form algebras that separate points, and so both WAT and Theorem 6.7.3 become special cases of Stone's general result. For a new example, consider the collection of polynomials with only even powers on the interval $[0, 1]$. The Stone–Weierstrass Theorem tells us that this subset of polynomials can still uniformly approximate an arbitrary continuous function, although if we were to switch our domain to $[-1, 1]$ then this algebra would no longer separate points. As a final example, consider the set

$$\mathcal{C} = \{a_0 + a_1 \cos(x) + \cdots + a_n \cos(nx) : a_0, a_1, \ldots, a_n \in \mathbf{R}\}.$$

In Section 8.5 we take up the theory of Fourier series which explores when a function has a representation as an infinite series of trigonometric functions. As a precursor to that conversation, notice that the Stone–Weierstrass Theorem tells us at the outset that at least every continuous function on $[0, \pi]$ is the uniform limit of functions from $\mathcal{C}$.

The story from Section 6.6 surrounding Taylor series expansions also deserves a final word. The ingenuity with which Euler and others found and exploited power series representations for the cast of familiar functions from calculus understandably led to speculation that every function could be represented in such a fashion. (The term "function" at this time implicitly referred to functions that were infinitely differentiable.) This point of view effectively ended with Cauchy's discovery in 1821 of the counterexample presented at the end of Section 6.6. So under what conditions does the Taylor series necessarily converge to the generating function? Lagrange's Remainder Theorem states that the difference between the Taylor polynomial $S_N(x)$ and the function $f(x)$ is given by

$$E_N(x) = \frac{f^{(N+1)}(c)}{(N+1)!} x^{N+1}.$$

The $(N+1)!$ term in the denominator grows more rapidly than the x^{N+1} term in the numerator. Thus, if we knew for instance that

$$|f^{(N+1)}(c)| \leq M$$

for all $c \in (-R, R)$ and $N \in \mathbf{N}$, we could be sure that $E_N(x) \to 0$ and hence that $S_N(x) \to f(x)$. This is the case for $\sin(x)$, $\cos(x)$, and e^x, whose derivatives do not grow at all as $N \to \infty$. It is also possible to formulate weaker conditions on the rate of growth of $f^{(N+1)}$ that guarantee convergence.

It is not altogether clear whether Cauchy's counterexample should come as a surprise. The fact that every previous search for a Taylor series ended in success certainly gives the impression that a power series representation is an intrinsic property of infinitely differentiable functions. But notice what we are saying here. A Taylor series for a function f is constructed from the values of f and its derivatives at the origin. If the Taylor series converges to f on some interval $(-R, R)$, then the behavior of f near zero completely determines its behavior at every point in $(-R, R)$. One implication of this would be that if two functions with Taylor series agree on some small neighborhood $(-\epsilon, \epsilon)$, then these two functions would have to be the same everywhere. When it is put this way, we probably should not expect a Taylor series to always converge back to the function from which it was derived. As we have seen, this is not the case for real-valued functions. What is fascinating, however, is that results of this nature *do* hold for functions of a complex variable. The definition of the derivative looks symbolically the same when the real numbers are replaced by complex numbers, but the implications are profoundly different. In this setting, a function that is differentiable at every point in some open disc must necessarily be infinitely differentiable on this set. This supplies the ingredients to construct the Taylor series that in every instance converges uniformly on compact sets to the function that generated it.