

Chapter 5

The Derivative

5.1 Discussion: Are Derivatives Continuous?

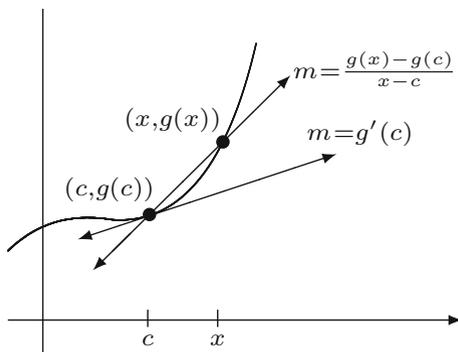
The geometric motivation for the derivative is most likely familiar territory. Given a function $g(x)$, the derivative $g'(x)$ is understood to be the slope of the graph of g at each point x in the domain. A graphical picture (Fig. 5.1) reveals the impetus behind the mathematical definition

$$g'(c) = \lim_{x \rightarrow c} \frac{g(x) - g(c)}{x - c}.$$

The difference quotient $(g(x) - g(c))/(x - c)$ represents the slope of the line through the two points $(x, g(x))$ and $(c, g(c))$. By taking the limit as x approaches c , we arrive at a well-defined mathematical meaning for the slope of the tangent line at $x = c$.

The myriad applications of the derivative function are the topic of much of the calculus sequence, as well as several other upper-level courses in mathematics. None of these applied questions are pursued here in any length, but it should be pointed out that the rigorous underpinnings for differentiation worked out in this chapter are an essential foundation for any applied study. Eventually, as the derivative is subjected to more and more complex manipulations, it becomes crucial to know precisely how differentiation is defined and how it interacts with other mathematical operations.

Although physical applications are not explicitly discussed, we will encounter several questions of a more abstract quality as we develop the theory. Many of these are concerned with the relationship between differentiation and continuity. Are continuous functions always differentiable? If not, how nondifferentiable can a continuous function be? Are differentiable functions continuous? Given that

Figure 5.1: DEFINITION OF $g'(c)$.

a function f has a derivative at every point in its domain, what can we say about the function f' ? Is f' continuous? How accurately can we describe the set of all possible derivatives, or are there no restrictions? Put another way, if we are given an arbitrary function g , is it always possible to find a differentiable function f such that $f' = g$, or are there some properties that g must possess for this to occur? In our study of continuity, we saw that restricting our attention to monotone functions had a significant impact on the answers to questions about sets of discontinuity. What effect, if any, does this same restriction have on our questions about potential sets of nondifferentiable points? Some of these issues are harder to resolve than others, and some remain unanswered in any satisfactory way.

A particularly useful class of examples for this discussion are functions of the form

$$g_n(x) = \begin{cases} x^n \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

When $n = 0$, we have seen (Example 4.2.6) that the oscillations of $\sin(1/x)$ prevent $g_0(x)$ from being continuous at $x = 0$. When $n = 1$, these oscillations are squeezed between $|x|$ and $-|x|$, the result being that g_1 is continuous at $x = 0$ (Example 4.3.6). Is $g'_1(0)$ defined? Using the preceding definition, we get

$$g'_1(0) = \lim_{x \rightarrow 0} \frac{g_1(x)}{x} = \lim_{x \rightarrow 0} \sin(1/x),$$

which, as we now know, does not exist. Thus, g_1 is not differentiable at $x = 0$. On the other hand, the same calculation shows that g_2 is differentiable at zero. In fact, we have

$$g'_2(0) = \lim_{x \rightarrow 0} x \sin(1/x) = 0.$$

At points different from zero, we can use the familiar rules of differentiation (soon to be justified) to conclude that g_2 is differentiable everywhere in \mathbf{R} with

$$g'_2(x) = \begin{cases} -\cos(1/x) + 2x \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

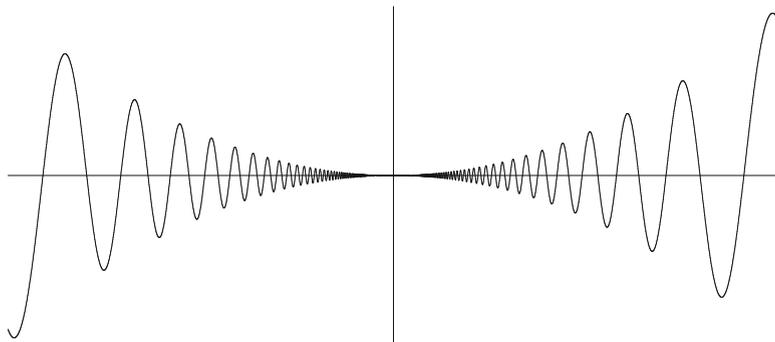


Figure 5.2: THE FUNCTION $g_2(x) = x^2 \sin(1/x)$ NEAR ZERO.

But now consider

$$\lim_{x \rightarrow 0} g_2'(x).$$

Because the $\cos(1/x)$ term is not preceded by a factor of x , we must conclude that this limit does not exist and that, consequently, the derivative function is not continuous. To summarize, the function $g_2(x)$ is continuous and differentiable everywhere on \mathbf{R} (Fig. 5.2), the derivative function g_2' is thus defined everywhere on \mathbf{R} , but g_2' has a discontinuity at zero. The conclusion is that derivatives need not, in general, be continuous!

The discontinuity in g_2' is *essential*, meaning $\lim_{x \rightarrow 0} g_2'(x)$ does not exist as a one-sided limit. But, what about a function with a simple jump discontinuity? For example, does there exist a function h such that

$$h'(x) = \begin{cases} -1 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0. \end{cases}$$

A first impression may bring to mind the absolute value function, which has slopes of -1 at points to the left of zero and slopes of 1 to the right. However, the absolute value function is *not* differentiable at zero. We are seeking a function that is differentiable everywhere, including the point zero, where we are insisting that the slope of the graph be -1 . The degree of difficulty of this request should start to become apparent. Without sacrificing differentiability at any point, we are demanding that the slopes jump from -1 to 1 and not attain any value in between.

Although we have seen that continuity is not a required property of derivatives, the intermediate value property will prove a more stubborn quality to ignore.

5.2 Derivatives and the Intermediate Value Property

Although the definition would technically make sense for more complicated domains, all of the interesting results about the relationship between a function and its derivative require that the domain of the given function be an interval. Thinking geometrically of the derivative as a rate of change, it should not be too surprising that we would want to confine the independent variable to move about a connected domain.

The theory of functional limits from Section 4.2 is all that is needed to supply a rigorous definition for the derivative.

Definition 5.2.1 (Differentiability). Let $g : A \rightarrow \mathbf{R}$ be a function defined on an interval A . Given $c \in A$, the *derivative of g at c* is defined by

$$g'(c) = \lim_{x \rightarrow c} \frac{g(x) - g(c)}{x - c},$$

provided this limit exists. In this case we say g is *differentiable at c* . If g' exists for all points $c \in A$, we say that g is *differentiable on A* .

Example 5.2.2. (i) Consider $f(x) = x^n$, where $n \in \mathbf{N}$, and let c be any arbitrary point in \mathbf{R} . Using the algebraic identity

$$x^n - c^n = (x - c)(x^{n-1} + cx^{n-2} + c^2x^{n-3} + \dots + c^{n-1}),$$

we can calculate the familiar formula

$$\begin{aligned} f'(c) &= \lim_{x \rightarrow c} \frac{x^n - c^n}{x - c} = \lim_{x \rightarrow c} (x^{n-1} + cx^{n-2} + c^2x^{n-3} + \dots + c^{n-1}) \\ &= c^{n-1} + c^{n-1} + \dots + c^{n-1} = nc^{n-1}. \end{aligned}$$

(ii) If $g(x) = |x|$, then attempting to compute the derivative at $c = 0$ produces the limit

$$g'(0) = \lim_{x \rightarrow 0} \frac{|x|}{x},$$

which is $+1$ or -1 depending on whether x approaches zero from the right or left. Consequently, this limit does not exist, and we conclude that g is not differentiable at zero.

Example 5.2.2 (ii) is a reminder that continuity of g does not imply that g is necessarily differentiable. On the other hand, if g is differentiable at a point, then it is true that g must be continuous at this point.

Theorem 5.2.3. *If $g : A \rightarrow \mathbf{R}$ is differentiable at a point $c \in A$, then g is continuous at c as well.*

Proof. We are assuming that

$$g'(c) = \lim_{x \rightarrow c} \frac{g(x) - g(c)}{x - c}$$

exists, and we want to prove that $\lim_{x \rightarrow c} g(x) = g(c)$. But notice that the Algebraic Limit Theorem for functional limits allows us to write

$$\lim_{x \rightarrow c} (g(x) - g(c)) = \lim_{x \rightarrow c} \left(\frac{g(x) - g(c)}{x - c} \right) (x - c) = g'(c) \cdot 0 = 0.$$

It follows that $\lim_{x \rightarrow c} g(x) = g(c)$. \square

Combinations of Differentiable Functions

The Algebraic Limit Theorem (Theorem 2.3.3) led easily to the conclusion that algebraic combinations of continuous functions are continuous. With only slightly more work, we arrive at a similar conclusion for sums, products, and quotients of differentiable functions.

Theorem 5.2.4 (Algebraic Differentiability Theorem). *Let f and g be functions defined on an interval A , and assume both are differentiable at some point $c \in A$. Then,*

- (i) $(f + g)'(c) = f'(c) + g'(c)$,
- (ii) $(kf)'(c) = kf'(c)$, for all $k \in \mathbf{R}$,
- (iii) $(fg)'(c) = f'(c)g(c) + f(c)g'(c)$, and
- (iv) $(f/g)'(c) = \frac{g(c)f'(c) - f(c)g'(c)}{[g(c)]^2}$, provided that $g(c) \neq 0$.

Proof. Statements (i) and (ii) are left as exercises. To prove (iii), we rewrite the difference quotient as

$$\begin{aligned} \frac{(fg)(x) - (fg)(c)}{x - c} &= \frac{f(x)g(x) - f(x)g(c) + f(x)g(c) - f(c)g(c)}{x - c} \\ &= f(x) \left[\frac{g(x) - g(c)}{x - c} \right] + g(c) \left[\frac{f(x) - f(c)}{x - c} \right]. \end{aligned}$$

Because f is differentiable at c , it is continuous there and thus $\lim_{x \rightarrow c} f(x) = f(c)$. This fact, together with the functional-limit version of the Algebraic Limit Theorem (Theorem 4.2.4), justifies the conclusion

$$\lim_{x \rightarrow c} \frac{(fg)(x) - (fg)(c)}{x - c} = f(c)g'(c) + f'(c)g(c).$$

A similar proof of (iv) is possible, or we can use an argument based on the next result. Each of these options is discussed in Exercise 5.2.3. \square

The composition of two differentiable functions also fortunately results in another differentiable function. This fact is referred to as the *Chain Rule*. To discover the proper formula for the derivative of the composition $g \circ f$, we can write

$$\begin{aligned}(g \circ f)'(c) &= \lim_{x \rightarrow c} \frac{g(f(x)) - g(f(c))}{x - c} = \lim_{x \rightarrow c} \frac{g(f(x)) - g(f(c))}{f(x) - f(c)} \cdot \frac{f(x) - f(c)}{x - c} \\ &= g'(f(c)) \cdot f'(c).\end{aligned}$$

With a little polish, this string of equations could qualify as a proof except for the pesky fact that the $f(x) - f(c)$ expression causes problems in the denominator if $f(x) = f(c)$ for x values in arbitrarily small neighborhoods of c . (The function $g_2(x)$ discussed in Section 5.1 exhibits this behavior near $c = 0$.) The upcoming proof of the Chain Rule manages to finesse this problem but in content is essentially the argument just given. Another approach is sketched in Exercise 5.2.4.

Theorem 5.2.5 (Chain Rule). *Let $f : A \rightarrow \mathbf{R}$ and $g : B \rightarrow \mathbf{R}$ satisfy $f(A) \subseteq B$ so that the composition $g \circ f$ is defined. If f is differentiable at $c \in A$ and if g is differentiable at $f(c) \in B$, then $g \circ f$ is differentiable at c with $(g \circ f)'(c) = g'(f(c)) \cdot f'(c)$.*

Proof. Because g is differentiable at $f(c)$, we know that

$$g'(f(c)) = \lim_{y \rightarrow f(c)} \frac{g(y) - g(f(c))}{y - f(c)}.$$

Another way to assert this same fact is to let $d(y)$ be the difference quotient

$$(1) \quad d(y) = \frac{g(y) - g(f(c))}{y - f(c)},$$

and observe that $\lim_{y \rightarrow f(c)} d(y) = g'(f(c))$. At the moment, $d(y)$ is not defined when $y = f(c)$, but it should seem natural to declare that $d(f(c)) = g'(f(c))$, so that d is continuous at $f(c)$.

Now, we come to the finesse. Equation (1) can be rewritten as

$$(2) \quad g(y) - g(f(c)) = d(y)(y - f(c)).$$

Observe that this equation holds for all $y \in B$ including $y = f(c)$. Thus, we are free to substitute $y = f(t)$ for any arbitrary $t \in A$. If $t \neq c$, we can divide equation (2) by $(t - c)$ to get

$$\frac{g(f(t)) - g(f(c))}{t - c} = d(f(t)) \frac{(f(t) - f(c))}{t - c}$$

for all $t \neq c$. Finally, taking the limit as $t \rightarrow c$ and applying the Algebraic Limit Theorem together with Theorem 4.3.9 yields the desired formula. \square

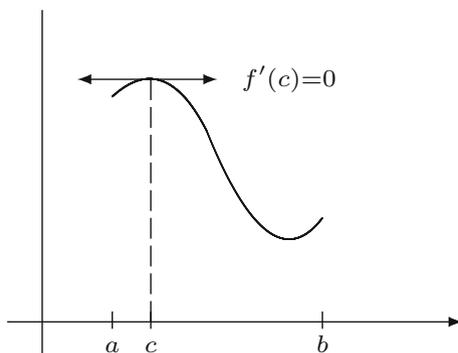


Figure 5.3: THE INTERIOR EXTREMUM THEOREM.

Darboux's Theorem

One conclusion from this chapter's introduction is that although continuity is necessary for the derivative to exist, it is not the case that the derivative function itself will always be continuous. Our specific example was $g_2(x) = x^2 \sin(1/x)$, where we set $g_2(0) = 0$. By tinkering with the exponent of the leading x^2 factor, it is possible to construct examples of differentiable functions with derivatives that are unbounded, or twice-differentiable functions that have discontinuous second derivatives (Exercise 5.2.7). The underlying principle in all of these examples is that by controlling the size of the oscillations of the original function, we can make the corresponding oscillations of the slopes volatile enough to prevent the existence of the relevant limits.

It is significant that for this class of examples, the discontinuities that arise are never simple jump discontinuities. (A precise definition of "jump discontinuity" is presented in Section 4.6.) We are now ready to confirm our earlier suspicions that although derivatives do not in general have to be continuous, they do possess the intermediate value property. (See Definition 4.5.3.) This surprising observation is a fairly straightforward corollary to the more obvious observation that differentiable functions attain maximums and minimums only at points where the derivative is equal to zero (Fig. 5.3).

Theorem 5.2.6 (Interior Extremum Theorem). *Let f be differentiable on an open interval (a, b) . If f attains a maximum value at some point $c \in (a, b)$ (i.e., $f(c) \geq f(x)$ for all $x \in (a, b)$), then $f'(c) = 0$. The same is true if $f(c)$ is a minimum value.*

Proof. Because c is in the open interval (a, b) , we can construct two sequences (x_n) and (y_n) , which converge to c and satisfy $x_n < c < y_n$ for all $n \in \mathbf{N}$. The fact that $f(c)$ is a maximum implies that $f(y_n) - f(c) \leq 0$ for all n , and thus

$$f'(c) = \lim_{n \rightarrow \infty} \frac{f(y_n) - f(c)}{y_n - c} \leq 0$$

by the Order Limit Theorem (Theorem 2.3.4). In a similar way,

$$\frac{f(x_n) - f(c)}{x_n - c} \geq 0$$

for each x_n because both numerator and denominator are negative. This implies that

$$f'(c) = \lim_{n \rightarrow \infty} \frac{f(x_n) - f(c)}{x_n - c} \geq 0,$$

and therefore $f'(c) = 0$, as desired. \square

The Interior Extremum Theorem is the fundamental fact behind the use of the derivative as a tool for solving applied optimization problems. This idea, discovered and exploited by Pierre de Fermat, is as old as the derivative itself. In a sense, finding maximums and minimums is arguably why Fermat invented his method of finding slopes of tangent lines. It was 200 years later that the French mathematician Gaston Darboux (1842–1917) pointed out that Fermat's method of finding maximums and minimums carries with it the implication that if a derivative function attains two distinct values $f'(a)$ and $f'(b)$, then it must also attain every value in between.

Theorem 5.2.7 (Darboux's Theorem). *If f is differentiable on an interval $[a, b]$, and if α satisfies $f'(a) < \alpha < f'(b)$ (or $f'(a) > \alpha > f'(b)$), then there exists a point $c \in (a, b)$ where $f'(c) = \alpha$.*

Proof. We first simplify matters by defining a new function $g(x) = f(x) - \alpha x$ on $[a, b]$. Notice that g is differentiable on $[a, b]$ with $g'(x) = f'(x) - \alpha$. In terms of g , our hypothesis states that $g'(a) < 0 < g'(b)$, and we hope to show that $g'(c) = 0$ for some $c \in (a, b)$.

The remainder of the argument is outlined in Exercise 5.2.11. \square

Exercises

Exercise 5.2.1. Supply proofs for parts (i) and (ii) of Theorem 5.2.4.

Exercise 5.2.2. Exactly one of the following requests is impossible. Decide which it is, and provide examples for the other three. In each case, let's assume the functions are defined on all of \mathbf{R} .

- Functions f and g not differentiable at zero but where fg is differentiable at zero.
- A function f not differentiable at zero and a function g differentiable at zero where fg is differentiable at zero.
- A function f not differentiable at zero and a function g differentiable at zero where $f + g$ is differentiable at zero.
- A function f differentiable at zero but not differentiable at any other point.

Exercise 5.2.3. (a) Use Definition 5.2.1 to produce the proper formula for the derivative of $h(x) = 1/x$.

(b) Combine the result in part (a) with the Chain Rule (Theorem 5.2.5) to supply a proof for part (iv) of Theorem 5.2.4.

(c) Supply a direct proof of Theorem 5.2.4 (iv) by algebraically manipulating the difference quotient for (f/g) in a style similar to the proof of Theorem 5.2.4 (iii).

Exercise 5.2.4. Follow these steps to provide a slightly modified proof of the Chain Rule.

(a) Show that a function $h : A \rightarrow \mathbf{R}$ is differentiable at $a \in A$ if and only if there exists a function $l : A \rightarrow \mathbf{R}$ which is continuous at a and satisfies

$$h(x) - h(a) = l(x)(x - a) \quad \text{for all } x \in A.$$

(b) Use this criterion for differentiability (in both directions) to prove Theorem 5.2.5.

Exercise 5.2.5. Let $f_a(x) = \begin{cases} x^a & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases}$

(a) For which values of a is f continuous at zero?

(b) For which values of a is f differentiable at zero? In this case, is the derivative function continuous?

(c) For which values of a is f twice-differentiable?

Exercise 5.2.6. Let g be defined on an interval A , and let $c \in A$.

(a) Explain why $g'(c)$ in Definition 5.2.1 could have been given by

$$g'(c) = \lim_{h \rightarrow 0} \frac{g(c+h) - g(c)}{h}.$$

(b) Assume A is open. If g is differentiable at $c \in A$, show

$$g'(c) = \lim_{h \rightarrow 0} \frac{g(c+h) - g(c-h)}{2h}.$$

Exercise 5.2.7. Let

$$g_a(x) = \begin{cases} x^a \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Find a particular (potentially noninteger) value for a so that

(a) g_a is differentiable on \mathbf{R} but such that g'_a is unbounded on $[0, 1]$.

- (b) g_a is differentiable on \mathbf{R} with g'_a continuous but not differentiable at zero.
- (c) g_a is differentiable on \mathbf{R} and g'_a is differentiable on \mathbf{R} , but such that g''_a is not continuous at zero.

Exercise 5.2.8. Review the definition of uniform continuity (Definition 4.4.4). Given a differentiable function $f : A \rightarrow \mathbf{R}$, let's say that f is *uniformly differentiable* on A if, given $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\left| \frac{f(x) - f(y)}{x - y} - f'(y) \right| < \epsilon \quad \text{whenever } 0 < |x - y| < \delta.$$

- (a) Is $f(x) = x^2$ uniformly differentiable on \mathbf{R} ? How about $g(x) = x^3$?
- (b) Show that if a function is uniformly differentiable on an interval A , then the derivative must be continuous on A .
- (c) Is there a theorem analogous to Theorem 4.4.7 for differentiation? Are functions that are differentiable on a closed interval $[a, b]$ necessarily uniformly differentiable?

Exercise 5.2.9. Decide whether each conjecture is true or false. Provide an argument for those that are true and a counterexample for each one that is false.

- (a) If f' exists on an interval and is not constant, then f' must take on some irrational values.
- (b) If f' exists on an open interval and there is some point c where $f'(c) > 0$, then there exists a δ -neighborhood $V_\delta(c)$ around c in which $f'(x) > 0$ for all $x \in V_\delta(c)$.
- (c) If f is differentiable on an interval containing zero and if $\lim_{x \rightarrow 0} f'(x) = L$, then it must be that $L = f'(0)$.

Exercise 5.2.10. Recall that a function $f : (a, b) \rightarrow \mathbf{R}$ is *increasing* on (a, b) if $f(x) \leq f(y)$ whenever $x < y$ in (a, b) . A familiar mantra from calculus is that a differentiable function is increasing if its derivative is positive, but this statement requires some sharpening in order to be completely accurate.

Show that the function

$$g(x) = \begin{cases} x/2 + x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is differentiable on \mathbf{R} and satisfies $g'(0) > 0$. Now, prove that g is *not* increasing over any open interval containing 0.

In the next section we will see that f is indeed increasing on (a, b) if and only if $f'(x) \geq 0$ for all $x \in (a, b)$.

Exercise 5.2.11. Assume that g is differentiable on $[a, b]$ and satisfies $g'(a) < 0 < g'(b)$.

- (a) Show that there exists a point $x \in (a, b)$ where $g(a) > g(x)$, and a point $y \in (a, b)$ where $g(y) < g(b)$.

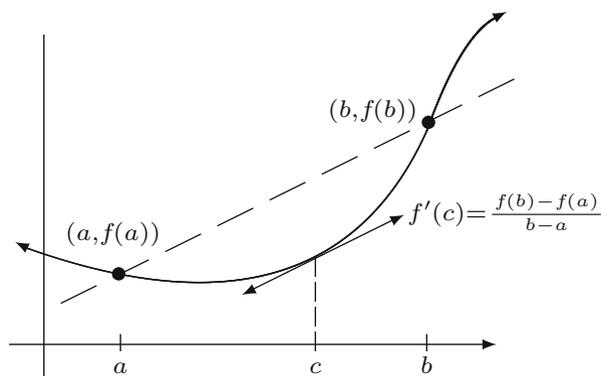


Figure 5.4: THE MEAN VALUE THEOREM.

(b) Now complete the proof of Darboux's Theorem started earlier.

Exercise 5.2.12 (Inverse functions). If $f : [a, b] \rightarrow \mathbf{R}$ is one-to-one, then there exists an inverse function f^{-1} defined on the range of f given by $f^{-1}(y) = x$ where $y = f(x)$. In Exercise 4.5.8 we saw that if f is continuous on $[a, b]$, then f^{-1} is continuous on its domain. Let's add the assumption that f is differentiable on $[a, b]$ with $f'(x) \neq 0$ for all $x \in [a, b]$. Show f^{-1} is differentiable with

$$(f^{-1})'(y) = \frac{1}{f'(x)} \quad \text{where } y = f(x).$$

5.3 The Mean Value Theorems

The Mean Value Theorem (Fig. 5.4) makes the geometrically plausible assertion that a differentiable function f on an interval $[a, b]$ will, at some point, attain a slope equal to the slope of the line through the endpoints $(a, f(a))$ and $(b, f(b))$. More tersely put,

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

for at least one point $c \in (a, b)$.

On the surface, there does not appear to be anything especially remarkable about this observation. Its validity appears undeniable—much like the Intermediate Value Theorem for continuous functions—and its proof is rather short. The ease of the proof, however, is misleading, as it is built on top of some hard-fought accomplishments from the study of limits and continuity. In this regard, the Mean Value Theorem is a kind of reward for a job well done. As we will see, it is a prize of exceptional value. Although the result itself is geometrically unsurprising, the Mean Value Theorem is the cornerstone of the proof for almost every major theorem pertaining to differentiation. We will use it to prove L'Hospital's rules regarding limits of quotients of differentiable functions.

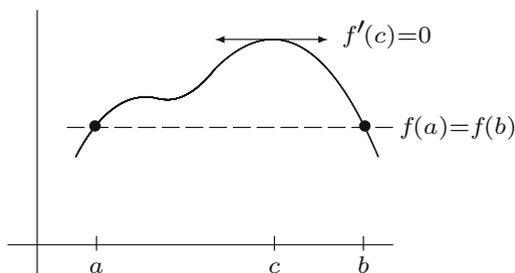


Figure 5.5: ROLLE'S THEOREM.

A rigorous analysis of how infinite series of functions behave when differentiated requires the Mean Value Theorem (Theorem 6.4.3), and it is the crucial step in the proof of the Fundamental Theorem of Calculus (Theorem 7.5.1). It is also the fundamental concept underlying Lagrange's Remainder Theorem (Theorem 6.6.3) which approximates the error between a Taylor polynomial and the function that generates it.

The Mean Value Theorem can be stated in various degrees of generality, each one important enough to be given its own special designation. Recall that the Extreme Value Theorem (Theorem 4.4.2) states that continuous functions on compact sets always attain maximum and minimum values. Combining this observation with the Interior Extremum Theorem for differentiable functions (Theorem 5.2.6) yields a special case of the Mean Value Theorem first noted by the mathematician Michel Rolle (1652–1719) (Fig. 5.5).

Theorem 5.3.1 (Rolle's Theorem). *Let $f : [a, b] \rightarrow \mathbf{R}$ be continuous on $[a, b]$ and differentiable on (a, b) . If $f(a) = f(b)$, then there exists a point $c \in (a, b)$ where $f'(c) = 0$.*

Proof. Because f is continuous on a compact set, f attains a maximum and a minimum. If both the maximum and minimum occur at the endpoints, then f is necessarily a constant function and $f'(x) = 0$ on all of (a, b) . In this case, we can choose c to be any point we like. On the other hand, if either the maximum or minimum occurs at some point c in the interior (a, b) , then it follows from the Interior Extremum Theorem (Theorem 5.2.6) that $f'(c) = 0$. \square

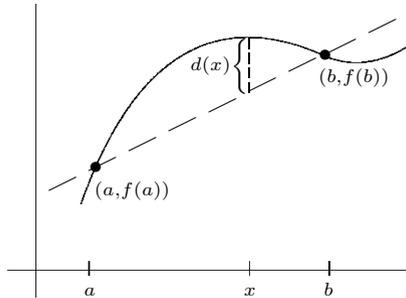
Theorem 5.3.2 (Mean Value Theorem). *If $f : [a, b] \rightarrow \mathbf{R}$ is continuous on $[a, b]$ and differentiable on (a, b) , then there exists a point $c \in (a, b)$ where*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Proof. Notice that the Mean Value Theorem reduces to Rolle's Theorem in the case where $f(a) = f(b)$. The strategy of the proof is to reduce the more general statement to this special case.

The equation of the line through $(a, f(a))$ and $(b, f(b))$ is

$$y = \left(\frac{f(b) - f(a)}{b - a} \right) (x - a) + f(a).$$



We want to consider the difference between this line and the function $f(x)$. To this end, let

$$d(x) = f(x) - \left[\left(\frac{f(b) - f(a)}{b - a} \right) (x - a) + f(a) \right],$$

and observe that d is continuous on $[a, b]$, differentiable on (a, b) , and satisfies $d(a) = 0 = d(b)$. Thus, by Rolle's Theorem, there exists a point $c \in (a, b)$ where $d'(c) = 0$. Because

$$d'(x) = f'(x) - \frac{f(b) - f(a)}{b - a},$$

we get

$$0 = f'(c) - \frac{f(b) - f(a)}{b - a},$$

which completes the proof. \square

The point has been made that the Mean Value Theorem manages to find its way into nearly every proof of any statement related to the geometrical nature of the derivative. As a simple example, if f is a constant function $f(x) = k$ on some interval A , then a straightforward calculation of f' using Definition 5.2.1 shows that $f'(x) = 0$ for all $x \in A$. But how do we prove the converse statement? If we know that a differentiable function g satisfies $g'(x) = 0$ everywhere on A , our intuition suggests that we should be able to prove $g(x)$ is constant. It is the Mean Value Theorem that provides us with a way to articulate rigorously what seems geometrically valid.

Corollary 5.3.3. *If $g : A \rightarrow \mathbf{R}$ is differentiable on an interval A and satisfies $g'(x) = 0$ for all $x \in A$, then $g(x) = k$ for some constant $k \in \mathbf{R}$.*

Proof. Take $x, y \in A$ and assume $x < y$. Applying the Mean Value Theorem to g on the interval $[x, y]$, we see that

$$g'(c) = \frac{g(y) - g(x)}{y - x}$$

for some $c \in A$. Now, $g'(c) = 0$, so we conclude that $g(y) = g(x)$. Set k equal to this common value. Because x and y are arbitrary, it follows that $g(x) = k$ for all $x \in A$. \square

Corollary 5.3.4. *If f and g are differentiable functions on an interval A and satisfy $f'(x) = g'(x)$ for all $x \in A$, then $f(x) = g(x) + k$ for some constant $k \in \mathbf{R}$.*

Proof. Let $h(x) = f(x) - g(x)$ and apply Corollary 5.3.3 to the differentiable function h . \square

The Mean Value Theorem has a more general form due to Cauchy. It is this generalized version of the theorem that is needed to analyze L'Hospital's rules and Lagrange's Remainder Theorem.

Theorem 5.3.5 (Generalized Mean Value Theorem). *If f and g are continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b) , then there exists a point $c \in (a, b)$ where*

$$[f(b) - f(a)]g'(c) = [g(b) - g(a)]f'(c).$$

If g' is never zero on (a, b) , then the conclusion can be stated as

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Proof. This result follows by applying the Mean Value Theorem to the function $h(x) = [f(b) - f(a)]g(x) - [g(b) - g(a)]f(x)$. The details are requested in Exercise 5.3.5. \square

L'Hospital's Rules

The Algebraic Limit Theorem asserts that when taking a limit of a quotient of functions we can write

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow c} f(x)}{\lim_{x \rightarrow c} g(x)},$$

provided that each individual limit exists and $\lim_{x \rightarrow c} g(x)$ is not zero. If the denominator does converge to zero and the numerator has a nonzero limit, then it is not difficult to argue that the quotient $f(x)/g(x)$ grows in absolute value without bound as x approaches c . L'Hospital's Rules are named for the Marquis de L'Hospital (1661–1704), who learned the results from his tutor, Johann Bernoulli (1667–1748), and published them in 1696 in what is regarded as the first calculus text. Stated in different levels of generality, they are an effective tool for handling the indeterminate cases when either numerator and denominator both tend to zero or both tend simultaneously to infinity.

Theorem 5.3.6 (L'Hospital's Rule: 0/0 case). *Let f and g be continuous on an interval containing a , and assume f and g are differentiable on this interval with the possible exception of the point a . If $f(a) = g(a) = 0$ and $g'(x) \neq 0$ for all $x \neq a$, then*

$$\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)} = L \quad \text{implies} \quad \lim_{x \rightarrow a} \frac{f(x)}{g(x)} = L.$$

Proof. This argument follows from a straightforward application of the Generalized Mean Value Theorem. It is requested as Exercise 5.3.11. \square

L'Hospital's Rule remains true if we replace the assumption that $f(a) = g(a) = 0$ with the hypothesis that $\lim_{x \rightarrow a} g(x) = \infty$. To this point we have not been explicit about what it means to say that a limit equals ∞ . The logical structure of such a definition is precisely the same as it is for finite functional limits. The difference is that rather than trying to force the function to take on values in some small ϵ -neighborhood around a proposed limit, we must show that $g(x)$ eventually *exceeds* any proposed upper bound. The arbitrarily small $\epsilon > 0$ is replaced by an arbitrarily large $M > 0$.

Definition 5.3.7. Given $g : A \rightarrow \mathbf{R}$ and a limit point c of A , we say that $\lim_{x \rightarrow c} g(x) = \infty$ if, for every $M > 0$, there exists a $\delta > 0$ such that whenever $0 < |x - c| < \delta$ it follows that $g(x) \geq M$.

We can define $\lim_{x \rightarrow c} g(x) = -\infty$ in a similar way.

The following version of L'Hospital's Rule is typically referred to as the ∞/∞ case even though the hypothesis only requires that the function in the denominator tend to infinity. To simplify the notation of the proof, we state the result using a one-sided limit.

Theorem 5.3.8 (L'Hospital's Rule: ∞/∞ case). *Assume f and g are differentiable on (a, b) and that $g'(x) \neq 0$ for all $x \in (a, b)$. If $\lim_{x \rightarrow a} g(x) = \infty$ (or $-\infty$), then*

$$\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)} = L \quad \text{implies} \quad \lim_{x \rightarrow a} \frac{f(x)}{g(x)} = L.$$

Proof. Let $\epsilon > 0$. Because $\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)} = L$, there exists a $\delta_1 > 0$ such that

$$\left| \frac{f'(x)}{g'(x)} - L \right| < \frac{\epsilon}{2}$$

for all $a < x < a + \delta_1$. For convenience of notation, let $t = a + \delta_1$ and note that t is fixed for the remainder of the argument.

Our functions are not defined at a , but for any $x \in (a, t)$ we can apply the Generalized Mean Value Theorem on the interval $[x, t]$ to get

$$\frac{f(x) - f(t)}{g(x) - g(t)} = \frac{f'(c)}{g'(c)}$$

for some $c \in (x, t)$. Our choice of t then implies

$$(1) \quad L - \frac{\epsilon}{2} < \frac{f(x) - f(t)}{g(x) - g(t)} < L + \frac{\epsilon}{2}$$

for all x in (a, t) .

In an effort to isolate the fraction $\frac{f(x)}{g(x)}$, the strategy is to multiply inequality (1) by $(g(x) - g(t))/g(x)$. We need to be sure, however, that this quantity is positive, which amounts to insisting that $1 \geq g(t)/g(x)$. Because t is fixed and $\lim_{x \rightarrow a} g(x) = \infty$, we can choose $\delta_2 > 0$ so that $g(x) \geq g(t)$ for all $a < x < a + \delta_2$. Carrying out the desired multiplication results in

$$\left(L - \frac{\epsilon}{2}\right) \left(1 - \frac{g(t)}{g(x)}\right) < \frac{f(x) - f(t)}{g(x)} < \left(L + \frac{\epsilon}{2}\right) \left(1 - \frac{g(t)}{g(x)}\right),$$

which after some algebraic manipulations yields

$$L - \frac{\epsilon}{2} + \frac{-Lg(t) + \frac{\epsilon}{2}g(t) + f(t)}{g(x)} < \frac{f(x)}{g(x)} < L + \frac{\epsilon}{2} + \frac{-Lg(t) - \frac{\epsilon}{2}g(t) + f(t)}{g(x)}.$$

Again, let's remind ourselves that t is fixed and that $\lim_{x \rightarrow a} g(x) = \infty$. Thus, we can choose a δ_3 such that $a < x < a + \delta_3$ implies that $g(x)$ is large enough to ensure that both

$$\frac{-Lg(t) + \frac{\epsilon}{2}g(t) + f(t)}{g(x)} \quad \text{and} \quad \frac{-Lg(t) - \frac{\epsilon}{2}g(t) + f(t)}{g(x)}$$

are less than $\epsilon/2$ in absolute value. Putting this all together and choosing $\delta = \min\{\delta_1, \delta_2, \delta_3\}$ guarantees that

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

for all $a < x < a + \delta$. □

Exercises

Exercise 5.3.1. Recall from Exercise 4.4.9 that a function $f : A \rightarrow \mathbf{R}$ is Lipschitz on A if there exists an $M > 0$ such that

$$\left| \frac{f(x) - f(y)}{x - y} \right| \leq M$$

for all $x \neq y$ in A .

- Show that if f is differentiable on a closed interval $[a, b]$ and if f' is continuous on $[a, b]$, then f is Lipschitz on $[a, b]$.
- Review the definition of a contractive function in Exercise 4.3.11. If we add the assumption that $|f'(x)| < 1$ on $[a, b]$, does it follow that f is contractive on this set?

Exercise 5.3.2. Let f be differentiable on an interval A . If $f'(x) \neq 0$ on A , show that f is one-to-one on A . Provide an example to show that the converse statement need not be true.

Exercise 5.3.3. Let h be a differentiable function defined on the interval $[0, 3]$, and assume that $h(0) = 1$, $h(1) = 2$, and $h(3) = 2$.

- (a) Argue that there exists a point $d \in [0, 3]$ where $h(d) = d$.
- (b) Argue that at some point c we have $h'(c) = 1/3$.
- (c) Argue that $h'(x) = 1/4$ at some point in the domain.

Exercise 5.3.4. Let f be differentiable on an interval A containing zero, and assume (x_n) is a sequence in A with $(x_n) \rightarrow 0$ and $x_n \neq 0$.

- (a) If $f(x_n) = 0$ for all $n \in N$, show $f(0) = 0$ and $f'(0) = 0$.
- (b) Add the assumption that f is twice-differentiable at zero and show that $f''(0) = 0$ as well.

Exercise 5.3.5. (a) Supply the details for the proof of Cauchy's Generalized Mean Value Theorem (Theorem 5.3.5).

- (b) Give a graphical interpretation of the Generalized Mean Value Theorem analogous to the one given for the Mean Value Theorem at the beginning of Section 5.3. (Consider f and g as parametric equations for a curve.)

Exercise 5.3.6. (a) Let $g : [0, a] \rightarrow \mathbf{R}$ be differentiable, $g(0) = 0$, and $|g'(x)| \leq M$ for all $x \in [0, a]$. Show $|g(x)| \leq Mx$ for all $x \in [0, a]$.

- (b) Let $h : [0, a] \rightarrow \mathbf{R}$ be twice differentiable, $h'(0) = h(0) = 0$ and $|h''(x)| \leq M$ for all $x \in [0, a]$. Show $|h(x)| \leq Mx^2/2$ for all $x \in [0, a]$.
- (c) Conjecture and prove an analogous result for a function that is differentiable three times on $[0, a]$.

Exercise 5.3.7. A *fixed point* of a function f is a value x where $f(x) = x$. Show that if f is differentiable on an interval with $f'(x) \neq 1$, then f can have at most one fixed point.

Exercise 5.3.8. Assume f is continuous on an interval containing zero and differentiable for all $x \neq 0$. If $\lim_{x \rightarrow 0} f'(x) = L$, show $f'(0)$ exists and equals L .

Exercise 5.3.9. Assume f and g are as described in Theorem 5.3.6, but now add the assumption that f and g are differentiable at a , and f' and g' are continuous at a with $g'(a) \neq 0$. Find a short proof for the $0/0$ case of L'Hospital's Rule under this stronger hypothesis.

Exercise 5.3.10. Let $f(x) = x \sin(1/x^4)e^{-1/x^2}$ and $g(x) = e^{-1/x^2}$. Using the familiar properties of these functions, compute the limit as x approaches zero of $f(x)$, $g(x)$, $f(x)/g(x)$, and $f'(x)/g'(x)$. Explain why the results are surprising but not in conflict with the content of Theorem 5.3.6.¹

Exercise 5.3.11. (a) Use the Generalized Mean Value Theorem to furnish a proof of the 0/0 case of L'Hospital's Rule (Theorem 5.3.6).

(b) If we keep the first part of the hypothesis of Theorem 5.3.6 the same but we assume that

$$\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)} = \infty,$$

does it necessarily follow that

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \infty?$$

Exercise 5.3.12. If f is twice differentiable on an open interval containing a and f'' is continuous at a , show

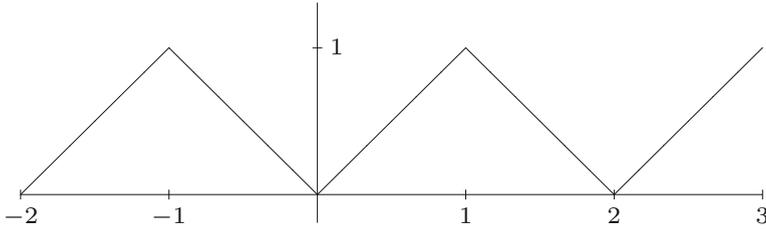
$$\lim_{h \rightarrow 0} \frac{f(a+h) - 2f(a) + f(a-h)}{h^2} = f''(a).$$

(Compare this to Exercise 5.2.6(b).)

5.4 A Continuous Nowhere-Differentiable Function

Exploring the relationship between continuity and differentiability has led to both fruitful results and pathological counterexamples. The bulk of discussion to this point has focused on the continuity of derivatives, but historically a significant amount of debate revolved around the question of whether continuous functions were necessarily differentiable. Early in the chapter, we saw that continuity was a requirement for differentiability, but, as the absolute value function demonstrates, the converse of this proposition is not true. A function can be continuous but not differentiable at some point. But just how nondifferentiable can a continuous function be? Given a *finite* set of points, it is not difficult to imagine how to construct a graph with corners at each of these points, so that the corresponding function fails to be differentiable on this finite set. The trick gets more difficult, however, when the set becomes infinite. For instance, is it possible to construct a function that is continuous on all of \mathbf{R} but fails to be differentiable at every rational point? Not only is this possible, but the situation is even more disconcerting. In 1872, Karl Weierstrass presented an example of a continuous function that was not differentiable at *any* point. (It seems to be

¹A large class of "counterexamples" of this sort to L'Hospital's Rule are explored in [4].

Figure 5.6: THE FUNCTION $h(x)$.

the case that Bernhard Bolzano had his own example of such a beast as early as 1830, but it was not published until much later.)

Weierstrass actually discovered a class of nowhere-differentiable functions of the form

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(b^n x)$$

where the values of a and b are carefully chosen. Such functions are specific examples of Fourier series discussed in Section 8.5. The details of Weierstrass' argument are simplified if we replace the cosine function with a piecewise linear function that has oscillations qualitatively like $\cos(x)$.

Define

$$h(x) = |x|$$

on the interval $[-1, 1]$ and extend the definition of h to all of \mathbf{R} by requiring that $h(x+2) = h(x)$. The result is a periodic "sawtooth" function (Fig. 5.6).

Exercise 5.4.1. Sketch a graph of $(1/2)h(2x)$ on $[-2, 3]$. Give a qualitative description of the functions

$$h_n(x) = \frac{1}{2^n} h(2^n x)$$

as n gets larger.

Now, define

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

The claim is that $g(x)$ is continuous on all of \mathbf{R} but fails to be differentiable at any point.

Infinite Series of Functions and Continuity

The definition of $g(x)$ is a significant departure from the way we usually define functions. For each $x \in \mathbf{R}$, $g(x)$ is defined to be the value of an infinite series.

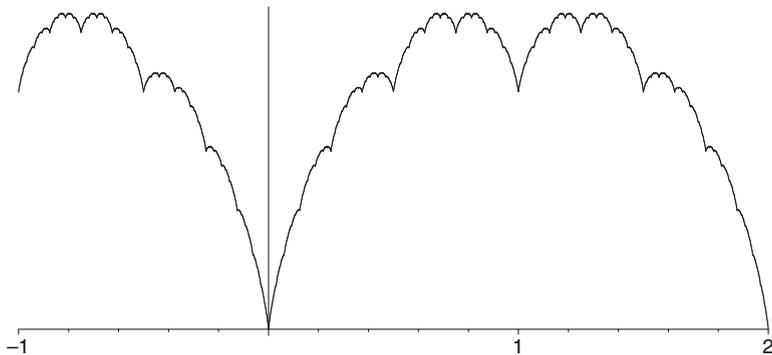


Figure 5.7: A SKETCH OF $g(x) = \sum_{n=0}^{\infty} (1/2^n)h(2^n x)$.

Exercise 5.4.2. Fix $x \in \mathbf{R}$. Argue that the series

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

converges and thus $g(x)$ is properly defined.

Exercise 5.4.3. Taking the continuity of $h(x)$ as given, reference the proper theorems from Chapter 4 that imply that the *finite* sum

$$g_m(x) = \sum_{n=0}^m \frac{1}{2^n} h(2^n x)$$

is continuous on \mathbf{R} .

This brings us to an archetypical question in analysis: When do conclusions that are valid in finite settings extend to infinite ones? A finite sum of continuous functions is certainly continuous, but does this necessarily hold for an infinite sum of continuous functions? In general, we will see that this is *not* always the case. For this particular sum, however, the continuity of the limit function $g(x)$ can be proved. Deciphering when results about finite sums of functions extend to infinite sums is one of the fundamental themes of Chapter 6. Although a self-contained argument for the continuity of g is not beyond our means at this point, we will nevertheless postpone the proof (see, for example, Exercise 6.4.3), leaving it as an enticement for the upcoming study of uniform convergence.

Exercise 5.4.4. As the graph in Figure 5.7 suggests, the structure of $g(x)$ is quite intricate. Answer the following questions, assuming that $g(x)$ is indeed continuous.

- (a) How do we know g attains a maximum value M on $[0, 2]$? What is this value?

- (b) Let D be the set of points in $[0, 2]$ where g attains its maximum. That is $D = \{x \in [0, 2] : g(x) = M\}$. Find one point in D .
- (c) Is D finite, countable, or uncountable?

Nondifferentiability

When the proper tools are in place, the proof that g is continuous is quite straightforward. The more difficult task is to show that g is not differentiable at any point in \mathbf{R} .

Let's first look at the point $x = 0$. Our function g does not appear to be differentiable here, and a rigorous proof is not too difficult. Consider the sequence $x_m = 1/2^m$, where $m = 0, 1, 2, \dots$

Exercise 5.4.5. Show that

$$\frac{g(x_m) - g(0)}{x_m - 0} = m + 1,$$

and use this to prove that $g'(0)$ does not exist.

Any temptation to say something like $g'(0) = \infty$ should be resisted. Setting $x_m = -(1/2^m)$ in the previous argument produces difference quotients heading toward $-\infty$. The geometric manifestation of this is the “cusp” that appears at $x = 0$ in the graph of g .

Exercise 5.4.6. (a) Modify the previous argument to show that $g'(1)$ does not exist. Show that $g'(1/2)$ does not exist.

- (b) Show that $g'(x)$ does not exist for any rational number of the form $x = p/2^k$ where $p \in \mathbf{Z}$ and $k \in \mathbf{N} \cup \{0\}$.

The points described in Exercise 5.4.6 (b) are called *dyadic* points. If $x = p/2^k$ is a dyadic rational number, then the function h_n has a corner at x as long as $n \geq k$. Thus, it should not be too surprising that g fails to be differentiable at points of this form. The argument is more delicate at points between the dyadic points.

Assume x is *not* a dyadic number. For a fixed value of $m \in \mathbf{N} \cup \{0\}$, x falls between two adjacent dyadic points,

$$\frac{p_m}{2^m} < x < \frac{p_m + 1}{2^m}.$$

Set $x_m = p_m/2^m$ and $y_m = (p_m + 1)/2^m$. Repeating this for each m yields two sequences (x_m) and (y_m) satisfying

$$\lim x_m = \lim y_m = x \quad \text{and} \quad x_m < x < y_m.$$

Exercise 5.4.7. (a) First prove the following general lemma: Let f be defined on an open interval J and assume f is differentiable at $a \in J$. If (a_n) and (b_n) are sequences satisfying $a_n < a < b_n$ and $\lim a_n = \lim b_n = a$, show

$$f'(a) = \lim_{n \rightarrow \infty} \frac{f(b_n) - f(a_n)}{b_n - a_n}.$$

(b) Now use this lemma to show that $g'(x)$ does not exist.

Weierstrass's original 1872 paper contained a demonstration that the infinite sum

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(b^n x)$$

defined a continuous nowhere-differentiable function provided $0 < a < 1$ and b was an odd integer satisfying $ab > 1 + 3\pi/2$. The condition on a is easy to understand. If $0 < a < 1$, then $\sum_{n=0}^{\infty} a^n$ is a convergent geometric series, and the forthcoming Weierstrass M-Test (Theorem 6.4.5) can be used to conclude that f is continuous. The restriction on b is more mysterious. In 1916, G.H. Hardy extended Weierstrass' result to include any value of b for which $ab \geq 1$. Without looking at the details of either of these arguments, we nevertheless get a sense that the lack of a derivative is intricately tied to the relationship between the compression factor (the parameter a) and the rate at which the frequency of the oscillations increases (the parameter b).

Exercise 5.4.8. Review the argument for the nondifferentiability of $g(x)$ at nondyadic points. Does the argument still work if we replace $g(x)$ with the summation $\sum_{n=0}^{\infty} (1/2^n)h(3^n x)$? Does the argument work for the function $\sum_{n=0}^{\infty} (1/3^n)h(2^n x)$?

5.5 Epilogue

Far from being an anomaly to be relegated to the margins of our understanding of continuous functions, Weierstrass's example and those like it should actually serve as a guide to our intuition. The image of continuity as a smooth curve in our mind's eye severely misrepresents the situation and is the result of a bias stemming from an overexposure to the much smaller class of differentiable functions. The lesson here is that continuity is a strictly weaker notion than differentiability. In Section 3.6, we alluded to a corollary of the Baire Category Theorem, which asserts that Weierstrass's construction is actually typical of continuous functions. We will see that most continuous functions are nowhere-differentiable, so that it is really the differentiable functions that are the exceptions rather than the rule. The details of how to phrase this observation more rigorously are spelled out in Section 8.2.

To say that the nowhere-differentiable function g constructed in the previous section has "corners" at every point of its domain misses the mark. Weierstrass's original class of nowhere-differentiable functions was constructed from infinite

sums of *smooth* trigonometric functions. It is the densely nested oscillating structure that makes the definition of a tangent line impossible. So what happens when we restrict our attention to monotone functions? How nondifferentiable can an increasing function be? Given a finite set of points, it is not difficult to piece together a monotone function which has actual corners—and thus is not differentiable—at each point in the given set. A natural question is whether there exists a continuous, monotone function that is nowhere-differentiable. Weierstrass suspected that such a function existed but only managed to produce an example of a continuous, increasing function which failed to be differentiable on a countable dense set (Exercise 7.5.11). In 1903, the French mathematician Henri Lebesgue (1875–1941) demonstrated that Weierstrass’s intuition had failed on this account. Lebesgue proved that a continuous, monotone function would have to be differentiable at “almost” every point in its domain. To be specific, Lebesgue showed that, for every $\epsilon > 0$, the set of points where such a function fails to be differentiable can be covered by a countable union of intervals whose lengths sum to less than ϵ . This notion of “zero length,” or “measure zero” as it is called, was encountered in our discussion of the Cantor set and is explored more fully in Section 7.6, where Lebesgue’s substantial contribution to the theory of integration is discussed.

With the relationship between the continuity of f and the existence of f' somewhat in hand, we once more return to the question of characterizing the set of all derivatives. Not every function is a derivative. Darboux’s Theorem forces us to conclude that there are some functions—those with jump discontinuities in particular—that cannot appear as the derivative of some other function. Another way to phrase Darboux’s Theorem is to say that all derivatives must satisfy the intermediate value property. Continuous functions do possess the intermediate value property, and it is natural to ask whether every continuous function is necessarily a derivative. For this smaller class of functions, the answer is yes. The Fundamental Theorem of Calculus, treated in Chapter 7, states that, given a continuous function f , the function $F(x) = \int_a^x f$ satisfies $F' = f$. This does the trick. The collection of derivatives at least contains the continuous functions. The search for a concise characterization of *all* possible derivatives, however, remains largely unsuccessful.

As a final remark, we will see that by cleverly choosing f , this technique of defining F via $F(x) = \int_a^x f$ can be used to produce examples of continuous functions which fail to be differentiable on interesting sets, *provided we can show that $\int_a^x f$ is defined*. The question of just how to define integration became a central theme in analysis in the latter half of the 19th century and has continued on to the present. Much of this story is discussed in detail in Chapter 7 and Section 8.1.