Advanced Calculus: MATH 410
Functions and Regularity

Professor David Levermore

13 October 2018

5. Functions, Continuity, and Limits

5.1. **Functions.** We now turn our attention to the study of real-valued functions that are defined over arbitrary nonempty subsets of $\mathbb{R}$. The subset of $\mathbb{R}$ over which such a function f is defined is called the *domain* of f, and is denoted $\mathrm{Dom}(f)$. We will write $f : \mathrm{Dom}(f) \to \mathbb{R}$ to indicate that f maps elements of $\mathrm{Dom}(f)$ into $\mathbb{R}$. For every $x \in \mathrm{Dom}(f)$ the function f associates the value $f(x) \in \mathbb{R}$. The *range* of f is the subset of $\mathbb{R}$ defined by

$$\mathrm{Rng}(f) = \big\{ f(x) \,:\, x \in \mathrm{Dom}(f) \big\} . \tag{5.1}$$

Sequences correspond to the special case where $\mathrm{Dom}(f) = \mathbb{N}$.

When a function f is given by an expression then, unless it is specified otherwise, $\mathrm{Dom}(f)$ will be understood to be all $x \in \mathbb{R}$ for which the expression makes sense. For example, if functions f and g are given by $f(x) = \sqrt{1 - x^2}$ and $g(x) = 1/(x^2 - 1)$, and no domains are specified explicitly, then it will be understood that

$$\mathrm{Dom}(f) = [-1, 1] , \qquad \mathrm{Dom}(g) = \big\{ x \in \mathbb{R} \,:\, x \neq \pm 1 \big\} .$$

These are *natural domains* for these functions. Of course, if these functions arise in the context of a problem for which x has other natural restrictions then these domains might be smaller. For example, if x represents the population of a species or the amount of a product being manufactured then we must further restrict x to $[0, \infty)$. If $f(x) = \sqrt{1 - x^2}$, $g(x) = 1/(x^2 - 1)$, and no domains are specified explicitly in such a context then it will be understood that

$$\mathrm{Dom}(f) = [0, 1] , \qquad \mathrm{Dom}(g) = \big\{ x \in [0, \infty) \,:\, x \neq 1 \big\} .$$

These are *natural domains* for these functions when x is naturally restricted to $[0, \infty)$.

Given any two functions, $f : \mathrm{Dom}(f) \to \mathbb{R}$ and $g : \mathrm{Dom}(g) \to \mathbb{R}$ with $\mathrm{Dom}(f) \subset \mathbb{R}$ and $\mathrm{Dom}(g) \subset \mathbb{R}$, we define their sum $f + g$, product fg, quotient f/g, and composition $g(f)$ to be the functions given by

$$\begin{aligned}
(f + g)(x) &= f(x) + g(x) && \forall x \in \mathrm{Dom}(f + g) , \\
(fg)(x) &= f(x)g(x) && \forall x \in \mathrm{Dom}(fg) , \\
(f/g)(x) &= f(x)/g(x) && \forall x \in \mathrm{Dom}(f/g) , \\
g(f)(x) &= g\big(f(x)\big) && \forall x \in \mathrm{Dom}\big(g(f)\big) .
\end{aligned} \tag{5.2}$$

where the natural domains appearing above are defined by

$$\begin{aligned}
\mathrm{Dom}(f + g) &= \mathrm{Dom}(f) \cap \mathrm{Dom}(g) , \\
\mathrm{Dom}(fg) &= \mathrm{Dom}(f) \cap \mathrm{Dom}(g) , \\
\mathrm{Dom}(f/g) &= \big\{ x \in \mathrm{Dom}(f) \cap \mathrm{Dom}(g) \,:\, g(x) \neq 0 \big\} , \\
\mathrm{Dom}\big(g(f)\big) &= \big\{ x \in \mathrm{Dom}(f) \,:\, f(x) \in \mathrm{Dom}(g) \big\} .
\end{aligned} \tag{5.3}$$

Notice that these domains are exactly the largest sets for which the respective expressions in (5.2) make sense.

Remark. A common notation for composition is $g \circ f$. We prefer the notation $g(f)$ because it makes the noncommutative aspect of the operation explicit.

Example. Polynomials are a class of functions. A polynomial is classified by a natrual number called its *degree*. A polynomial of degree 0 is a constant. A polynomial p of degree $n > 0$ has the form

$$p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n \,, \qquad \text{where } a_0 \neq 0 \,. \tag{5.4}$$

The natural domain of a polynomial function is $\mathbb{R}$. The class of polynomial functions is closed under addition, multiplication, and composition, but not under division.

Exercise. Show that the class of polynomials is closed under addition, multiplication, and composition, but not under division.

Example. A function r is said to be *rational* if it has the form

$$r(x) = \frac{p(x)}{q(x)} \,, \qquad \text{where } p \text{ and } q \text{ are polynomial functions} \,. \tag{5.5}$$

The natural domain of such a rational function is all $x \in \mathbb{R}$ where $q(x) \neq 0$. The class of rational functions is closed under addition, multiplication, division, and composition.

Exercise. Show that the class of rational functions is closed under addition, multiplication, division, and composition.

Example. A function f is said to be *algebraic* if for some $m > 0$ there exist polynomials $\{p_k(x)\}_{k=0}^m$ with $p_0(x)$ nonzero at some point in $\mathrm{Dom}(f)$ such that $y = f(x)$ solves

$$p_0(x)y^m + p_1(x)y^{m-1} + \cdots + p_{m-1}(x)y + p_m(x) = 0 \quad \text{for every } x \in \mathrm{Dom}(f) \,. \tag{5.6}$$

It is beyond the scope of this course to show that the class of algebraic functions is closed under addition, multiplication, division, and composition.

5.2. **Continuity.** Continuity is one of the most important concepts in mathematics. Here we introduce it in the context of real-valued functions with domains in $\mathbb{R}$.

Definition 5.1. *A function* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *with* $\mathrm{Dom}(f) \subset \mathbb{R}$ *is said to be* continuous at *a point* $x \in \mathrm{Dom}(f)$ *if for every* $\epsilon > 0$ *there exists* $\delta > 0$ *such that for every* $y \in \mathrm{Dom}(f)$ *we have*

$$|y - x| < \delta \qquad \Longrightarrow \qquad |f(y) - f(x)| < \epsilon \,. \tag{5.7}$$

Otherwise f *is said to be* discontinuous at x *or to have a* discontinuity at x. *A function* f *that is continuous at every point in a set* $S \subset \mathrm{Dom}(f)$ *is said to be* continuous over S. *A function* f *that is continuous over* $\mathrm{Dom}(f)$ *is said to be* continuous.

This definition states that f is continuous at x when we can insure that $f(y)$ is arbitarily close to $f(x)$ (within any ϵ of $f(x)$) by requiring that y is sufficiently close to x (within some δ of x). It is important to understand that the δ whose existence is asserted in this definition generally depends on both x and ϵ. Sometimes we will emphasize this dependence by explicitly writing $\delta_{x,\epsilon}$ or δ_ϵ, but more often this dependence will not be shown explicitly.

Being continuous at a point can be characterized in terms of sequences.

Proposition 5.1. *Let* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *with* $\mathrm{Dom}(f) \subset \mathbb{R}$. *If* $x \in \mathrm{Dom}(f)$ *then* f *is continuous at* x *if and only if for every sequence* $\{x_n\} \subset \mathrm{Dom}(f)$ *that converges to* x, *the sequence* $\{f(x_n)\}$ *converges to* $f(x)$ *— i.e. if and only if*

$$\forall \{x_n\} \subset \mathrm{Dom}(f) \qquad \lim_{n\to\infty} x_n = x \qquad \Longrightarrow \qquad \lim_{n\to\infty} f(x_n) = f(x) \,. \tag{5.8}$$

Proof. ($\Longrightarrow$) Let f be continuous at $x \in \mathrm{Dom}(f)$. Let $\{x_n\} \subset \mathrm{Dom}(f)$ be a sequence such that $x_n \to x$ as $n \to \infty$. We must show that $f(x_n) \to f(x)$ as $n \to \infty$.

Let $\epsilon > 0$. Because f is continuous at x there exists $\delta > 0$ such that (5.7) holds. Because $x_n \to x$ as $n \to \infty$ there exist $n_\delta \in \mathbb{N}$ such that $n > n_\delta$ implies $|x_n - x| < \delta$. It thereby follows that

$$n > n_\delta \quad \Longrightarrow \quad |x_n - x| < \delta \quad \Longrightarrow \quad |f(x_n) - f(x)| < \epsilon\,.$$

Therefore $f(x_n) \to f(x)$ as $n \to \infty$.

($\Longleftarrow$) Let (5.8) hold at $x \in \mathrm{Dom}(f)$. We will argue that f is continuous at x by contradiction. Suppose that f is not continuous at x. Upon negating Definition 5.1 we see there exists $\epsilon > 0$ such that for every $\delta > 0$ there exists $y \in \mathrm{Dom}(f)$ such that

$$|y - x| < \delta \quad \text{and} \quad |f(y) - f(x)| \geq \epsilon\,.$$

In particular, for every $n \in \mathbb{N}$ there exists $x_n \in \mathrm{Dom}(f)$ such that

$$|x_n - x| < \frac{1}{2^n} \quad \text{and} \quad |f(x_n) - f(x)| \geq \epsilon\,.$$

It follows that $x_n \to x$ as $n \to \infty$ while, because $|f(x_n) - f(x)| \geq \epsilon$ for every $n \in \mathbb{N}$, the sequence $\{f(x_n)\}$ does not converge to $f(x)$. But this contradicts the fact (5.8) holds at $x \in \mathrm{Dom}(f)$. Therefore f must be continuous at x. ☐

Remark. We can equally well have defined continuity by the sequence characterization given by Proposition 5.1. This is what Fitzpatrick does.

Remark. Roughly speaking, when drawing the graph of a function f that is continuous over an interval, we need not lift the pen or pencil from the paper. This is because (5.1) states that as the pen moves along the graph $(x, f(x))$ it will approach the point $(a, f(a))$ as x tends to a. The graph of f will consequently have no breaks, jumps, or holes over each interval over which it is defined. You should be able to tell by looking at the graph of a function where it is continuous.

The following proposition shows how continuity behaves with respect to combinations of functions.

Proposition 5.2. *Let* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *and* $g : \mathrm{Dom}(g) \to \mathbb{R}$ *where* $\mathrm{Dom}(f)$ *and* $\mathrm{Dom}(g)$ *are subsets of* $\mathbb{R}$.

If f *and* g *are continuous at* $x \in \mathrm{Dom}(f) \cap \mathrm{Dom}(g)$ *then the functions* $f + g$ *and* $f\,g$ *will be continuous at* x*, as will be the function* f/g *provided* $g(x) \neq 0$.

If f *is continuous at* $x \in \mathrm{Dom}\big(g(f)\big)$ *while* g *is continuous at* $f(x)$ *then the function* $g(f)$ *will be continuous at* x.

In particular, if f *and* g *are continuous then so are the combinations* $f + g$, $f\,g$, f/g, *and* $g(f)$ *considered over their natural domains.*

Proof. Exercise. (Do this both using the δ-ϵ definition and the sequence characterization.)

Examples. Every elementary function is continuous. This includes all rational functions, which are built up from combinations of the function x with constant functions. For example, the function $f(x) = 1/x$ is continuous because it is undefined at $x = 0$. This also includes all trigonometric functions that are built up from combinations of the functions $\cos(x)$ and $\sin(x)$ with constant functions. For example, $\tan(x)$, $\cot(x)$, $\sec(x)$, and $\csc(x)$ are continuous because they are undefined at points near which they behave badly.

5.3. **Extreme-Value Theorem.** We now consider the question of when a function whose range is bounded below (above) might take on a smallest (largest) value.

Definition 5.2. *Let* $D \subset \mathbb{R}$ *and* $f : D \to \mathbb{R}$*. We say that* f *has a* minimum (maximum) *over* D *if the set* $f(D) = \{f(x) : x \in D\}$ *has a minimum (maximum). In this case* $\min\{f(D)\}$ *(*$\max\{f(D)\}$*) is called the* minimum (maximum) *of* f *over* D*, and any* $p \in D$ *for which* $f(p) = \min\{f(D)\}$ *(*$f(p) = \max\{f(D)\}$*) is called a* minimizer (maximizer) *of* f *over* D*.*

A point that is either a minimizer or a maximizer of f *over* D *is called an* extremizer *of* f *over* D *and its corresponding value is called an* extremum *of* f *over* D*.*

It should be clear from this definition that a function can have at most one minimum and one maximum, but might have many minimizers or maximizers. Some functions f defined over a set D may have neither a minimum nor a maximum. For example, consider

$$\begin{aligned} f(x) &= \tanh(x) && \text{over } (-\infty, \infty)\,, \\ f(x) &= \tan(x) && \text{over } (-\tfrac{\pi}{2}, \tfrac{\pi}{2})\,, \\ f(x) &= x^3 && \text{over } (-\infty, \infty)\,. \end{aligned}$$

Some may have one but not the other. For example, consider

$$\begin{aligned} f(x) &= \operatorname{sech}(x) && \text{over } (-\infty, \infty)\,, \\ f(x) &= \sec(x) && \text{over } (-\tfrac{\pi}{2}, \tfrac{\pi}{2})\,, \\ f(x) &= (x^2-1)^2 && \text{over } (-\infty, \infty)\,. \end{aligned}$$

And some may have both. For example, consider

$$\begin{aligned} f(x) &= \sin(x) && \text{over } (-\infty, \infty)\,, \\ f(x) &= \frac{x}{1+x^2} && \text{over } (-\infty, \infty)\,, \\ f(x) &= xe^{-x} && \text{over } [0, \infty)\,. \end{aligned}$$

In particular, $f(x) = \sin(x)$ has infinitely many minimizers and maximizers over $(-\infty, \infty)$.

We now establish a theorem that asserts the existence of extrema in settings where the function is continuous and the domain is closed and bounded (hence, sequentially compact). This theorem will play a central role in the proofs of many subsequent propositions.

Proposition 5.3. Extreme-Value Theorem. *Let* $D \subset \mathbb{R}$ *be nonempty, closed and bounded. Let* $f : D \to \mathbb{R}$ *be continuous. Then* f *has both a minimum and a maximum over* D*. (In particular,* $\mathrm{Rng}(f)$ *is bounded.)*

Proof. We first prove that f has a minimum over D. Let $\underline{m} = \inf\{\mathrm{Rng}(f)\}$. There are two possibilities: either $\underline{m} > -\infty$ or $\underline{m} = -\infty$. We claim that in either case we can find a sequence $\{x_k\} \subset D$ such that $f(x_k) \to \underline{m}$ as $k \to \infty$. Indeed, if $\underline{m} > -\infty$ then for every $k \in \mathbb{N}$ there exist $x_k \in \mathrm{Dom}(f)$ such that $f(x_k) \in [\underline{m}, \underline{m} + \frac{1}{2^k})$, whereby $\{f(x_k)\} \to \underline{m}$ as $k \to \infty$. On the other hand, if $\underline{m} = -\infty$ then for every $k \in \mathbb{N}$ there exist $x_k \in \mathrm{Dom}(f)$ such that $f(x_k) < -k$, whereby $\{f(x_k)\} \to -\infty (= \underline{m})$ as $k \to \infty$. In either case $f(x_k) \to \underline{m}$ as $k \to \infty$.

Because D is closed and bounded, it is sequentially compact. Because $\{x_k\} \subset D$ and D is sequentially compact, there exists a subsequence $\{x_{n_k}\}$ of $\{x_k\}$ and a point $x \in D$ such that $x_{n_k} \to x$ as $k \to \infty$. The fact f is continuous over D then implies that $f(x_{n_k}) \to f(x)$ as $k \to \infty$. But we also know that $f(x_{n_k}) \to \underline{m}$ as $k \to \infty$. It follows that $\underline{m} = f(x) > -\infty$, whereby $\underline{m}$ is a minimum and x is a minimizer of f over D.

The proof that f has a maximum over D goes similarly, and is left as an exercise. □

Exercise. Give examples that illustrate that none of the hypotheses in Proposition 5.3 can simply be dropped. Specifically, give examples of (a) a continuous function over a closed domain that has no extremum, (b) a continuous function over a bounded domain that has no extremum, and (c) a discontinuous function over a closed and bounded domain that has no extremum.

5.4. **Intermediate-Value Theorem.** Another important property of continuous functions is established by the following theorem.

Proposition 5.4. Intermediate-Value Theorem. *Let $a < b$ and let $f : [a, b] \to \mathbb{R}$ be continuous. Then f takes all values that lie between $f(a)$ and $f(b)$.*

Proof. There is nothing to prove if $f(a) = f(b)$. We will give the proof for the case $f(a) < f(b)$. The case $f(a) > f(b)$ then follows by applying the first case to $-f$.

Let $q \in \big(f(a), f(b)\big)$. We want to show there exists an $c \in (a, b)$ such that $f(c) = q$. We do this by constructing a nested sequence of closed intervals whose endpoints converge to c. The construction is by the so-called *bisection method*. Set $[a_0, b_0] = [a, b]$. Given $[a_k, b_k]$ for some $k \in \mathbb{N}$ let $m_k = \frac{1}{2}(a_k + b_k)$ denote the midpoint and define

$$[a_{k+1}, b_{k+1}] = \begin{cases} [a_k, m_k] & \text{if } f(m_k) > q\,, \\ [m_k, b_k] & \text{if } f(m_k) \leq q\,. \end{cases}$$

Because $[a_{k+1}, b_{k+1}] \subset [a_k, b_k]$ for every $k \in \mathbb{N}$, $\{[a_k, b_k]\}_{k\in\mathbb{N}}$ is a nested sequence of closed intervals such that $b_k - a_k = (b - a)/2^k$ and $f(a_k) \leq q < f(b_k)$. By the Nested-Interval Theorem there exists $c \in (a, b)$ such that

$$\bigcap_{k=0}^{\infty} [a_k, b_k] = \{c\}\,, \quad \text{where } c = \lim_{k\to\infty} a_k = \lim_{k\to\infty} b_k\,.$$

By the continuity of f and the fact $f(a_k) \leq q < f(b_k)$ we then see that

$$f(c) = \lim_{k\to\infty} f(a_k) \leq q \leq \lim_{k\to\infty} f(b_k) = f(c)\,.$$

Hence, $f(c) = q$. □

An consequence of the Intermediate-Value Theorem is that continuous functions map intervals into intervals.

Proposition 5.5. *Let $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$ be continuous. If $I \subset D$ is an interval then $f(I)$ is an interval.*

Proof. We will employ the Interval Characterization Theorem. Specifically, we will show that if $p, q \in f(I)$ then all points that lie between p and q are also in $f(I)$. The Interval Characterization Theorem then implies that $f(I)$ is an interval.

Let $p, q \in f(I)$ be distinct points. Without loss of generality we may assume that $p < q$. We must show that $(p, q) \subset f(I)$. Let $r \in (p, q)$. Because $p, q \in f(I)$, we know that $p = f(a)$ and $q = f(b)$ for some $a, b \in I$. Either $a < b$ or $b < a$. If $a < b$ then $[a, b] \subset I$ and f is continuous over $[a, b]$. The Intermediate-Value Theorem then implies there exists $c \in (a, b)$ such that $f(c) = r$. On the other hand, if $b < a$ then $[b, a] \subset I$ and f is continuous over $[b, a]$. The Intermediate-Value Theorem then implies there exists $c \in (b, a)$ such that $f(c) = r$. In both cases we conclude that $r \in f(I)$. Therefore $(p, q) \subset f(I)$. □

Remark. While continuous functions map intervals into intervals, a function that maps intervals to intervals need not be continuous. An example defined over $\mathbb{R}$ is

$$f(x) = \begin{cases} \cos(1/x) & \text{if } x \neq 0\,, \\ 0 & \text{if } x = 0\,. \end{cases}$$

Exercise. Show that the function f given in the above remark maps intervals to intervals, yet is not continuous.

5.5. **Limits of a Function.** In this section we introduce three notions of limits of a function: limits at a point, one-sided limits at a point, and limits at infinity.

5.5.1. *Limits at a Point.* Limits of a function at a point are defined as follows.

Definition 5.3. *Given*

- *a function* $f : \text{Dom}(f) \to \mathbb{R}$ *with* $\text{Dom}(f) \subset \mathbb{R}$*,*
- *a limit point* $a \in \mathbb{R}$ *of* $\text{Dom}(f)$*,*
- *a number* $b \in \mathbb{R}$*,*

we say the limit of $f(x)$ as x approaches a is b *when for every* $\epsilon > 0$ *there exists* $\delta > 0$

$$\forall x \in \text{Dom}(f) \qquad 0 < |x - a| < \delta \quad \Longrightarrow \quad |f(x) - b| < \epsilon\,. \tag{5.9}$$

We denote this as

$$\lim_{x \to a} f(x) = b\,,$$

or as

$$f(x) \to b \quad \text{as } x \to a\,.$$

If $\lim_{x\to a} f(x) = b$ *for some* $b \in \mathbb{R}$ *then we say that "*$\lim_{x\to a} f(x)$ *exists." Otherwise we say that "*$\lim_{x\to a} f(x)$ *does not exist."*

These limits can be characterized in terms of convergent sequences.

Proposition 5.6. *Let* $f : \text{Dom}(f) \to \mathbb{R}$ *with* $\text{Dom}(f) \subset \mathbb{R}$*,* $a \in \mathbb{R}$ *be a limit point of* $\text{Dom}(f)$*, and* $b \in \mathbb{R}$*. Then* $\lim_{x\to a} f(x) = b$ *if and only if*

$$\forall \{x_n\} \subset \text{Dom}(f) - \{a\} \qquad \lim_{n\to\infty} x_n = a \quad \Longrightarrow \quad \lim_{n\to\infty} f(x_n) = b\,. \tag{5.10}$$

Proof. Exercise.

The existence of the limit of a function at a point in its domain is related to the continuity of the function at that point by the following.

Proposition 5.7. *A function* $f : \text{Dom}(f) \to \mathbb{R}$ *is continuous at a point* $a \in \text{Dom}(f)$ *if and only if*

$$\lim_{x\to a} f(x) = f(a)\,. \tag{5.11}$$

Remark. Here (5.11) is asserting two things:

- the limit on the left side of (5.11) exists;
- the limit equals $f(a)$.

A function can fail to be continuous at a point in its domain when the limit on the left of (5.11) fails to exist or when the limit exists but does not equal $f(a)$.

Proof. Exercise.

5.5.2. *One-Sided Limits at a Point.* One-Sided limits of a function are defined as follows.

Definition 5.4. *Given*

- *a function* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *with* $\mathrm{Dom}(f) \subset \mathbb{R}$,
- *a limit point* $a \in \mathbb{R}$ *of* $\mathrm{Dom}(f) \cap [a, \infty)$ *(of* $\mathrm{Dom}(f) \cap (-\infty, a]$*)*,
- *a number* $b \in \mathbb{R}$,

we say the limit of $f(x)$ as x approaches a from the right (left) is b *when for every* $\epsilon > 0$ *there exists* $\delta > 0$ *such that*

$$\begin{aligned} &\forall x \in \mathrm{Dom}(f) \qquad 0 < x - a < \delta \quad \Longrightarrow \quad |f(x) - b| < \epsilon \\ &\Big(\forall x \in \mathrm{Dom}(f) \qquad 0 < a - x < \delta \quad \Longrightarrow \quad |f(x) - b| < \epsilon\Big) . \end{aligned} \tag{5.12}$$

We denote this as

$$\lim_{x \to a^+} f(x) = b \qquad \Big(\lim_{x \to a^-} f(x) = b\Big) ,$$

or as

$$f(x) \to b \quad \text{as } x \to a^+ \qquad \Big(f(x) \to b \quad \text{as } x \to a^-\Big) .$$

If $\lim_{x \to a^\pm} f(x) = b$ *for some* $b \in \mathbb{R}$ *then we say that "*$\lim_{x \to a^\pm} f(x)$ *exists." Otherwise we say that "*$\lim_{x \to a^\pm} f(x)$ *does not exist."*

These limits can be characterized in terms of convergent sequences.

Proposition 5.8. *Let* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *with* $\mathrm{Dom}(f) \subset \mathbb{R}$. *Let* $b \in \mathbb{R}$. *If* $a \in \mathbb{R}$ *is a limit point of* $\mathrm{Dom}(f) \cap [a, \infty)$ *then* $\lim_{x \to a^+} f(x) = b$ *if and only if*

$$\forall \{x_n\} \subset \mathrm{Dom}(f) \cap (a, \infty) \qquad \lim_{n \to \infty} x_n = a \quad \Longrightarrow \quad \lim_{n \to \infty} f(x_n) = b . \tag{5.13}$$

If $a \in \mathbb{R}$ *is a limit point of* $\mathrm{Dom}(f) \cap (-\infty, a]$ *then* $\lim_{x \to a^-} f(x) = b$ *if and only if*

$$\forall \{x_n\} \subset \mathrm{Dom}(f) \cap (-\infty, a) \qquad \lim_{n \to \infty} x_n = a \quad \Longrightarrow \quad \lim_{n \to \infty} f(x_n) = b . \tag{5.14}$$

Proof. Exercise.

Theorem 5.7 motivates the following definition of so-called one-sided continuity.

Definition 5.5. *A function* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *is said to be* right (left) continuous *at a point* $a \in \mathrm{Dom}(f)$ *if and only if*

$$\lim_{x \to a^+} f(x) = f(a) \qquad \left(\lim_{x \to a^-} f(x) = f(a)\right) . \tag{5.15}$$

Otherwise f *is said to be* right (left) discontinuous at x *or to have a* right (left) discontinuity at x. *A function* f *that is right (left) continuous at every point in a set* $S \subset \mathrm{Dom}(f)$ *is said to be* right (left) continuous over S. *A function* f *that is right (left) continuous over* $\mathrm{Dom}(f)$ *is said to be* right (left) continuous.

The following is an easy consequence of this definition.

Proposition 5.9. *A function* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *is continuous at a point* $a \in \mathrm{Dom}(f)$ *if and only if it is both right and left continuous at* a *— i.e. if and only if*

$$\lim_{x \to a^-} f(x) = f(a) = \lim_{x \to a^+} f(x) . \tag{5.16}$$

Proof. Exercise.

5.5.3. *Limits at Infinity.* Limits at infinity of a function are defined as follows.

Definition 5.6. *Given*

- *a function* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *with* $\mathrm{Dom}(f) \subset \mathbb{R}$,
- $\mathrm{Dom}(f)$ *in not bounded above (below),*
- *a number* $b \in \mathbb{R}$,

we say the limit of $f(x)$ as x approaches $+\infty$ $(-\infty)$ is b *when for every* $\epsilon > 0$ *there exists* $m \in \mathbb{R}$ *such that*

$$\begin{aligned} &\forall x \in \mathrm{Dom}(f) \qquad x > m \quad \Longrightarrow \quad |f(x) - b| < \epsilon \\ &\Big(\forall x \in \mathrm{Dom}(f) \qquad x < m \quad \Longrightarrow \quad |f(x) - b| < \epsilon\Big)\,. \end{aligned} \tag{5.17}$$

We denote this as

$$\lim_{x\to+\infty} f(x) = b \qquad \Big(\lim_{x\to-\infty} f(x) = b\Big)\,,$$

or as

$$f(x) \to b \quad \text{as } x \to +\infty \qquad \Big(f(x) \to b \quad \text{as } x \to -\infty\Big)\,.$$

If $\lim_{x\to\pm\infty} f(x) = b$ *for some* $b \in \mathbb{R}$ *then we say that "*$\lim_{x\to\pm\infty} f(x)$ *exists." Otherwise we say that "*$\lim_{x\to\pm\infty} f(x)$ *does not exist."*

Remark. It is common to write ∞ in place of $+\infty$. We will often do so too.

These limits can be characterized in terms of sequences.

Proposition 5.10. *Let* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *with* $\mathrm{Dom}(f) \subset \mathbb{R}$ *and* $b \in \mathbb{R}$. *When* $\mathrm{Dom}(f)$ *is not bounded above (below) then* $\lim_{x\to\pm\infty} f(x) = b$ *if and only if*

$$\forall \{x_n\} \subset \mathrm{Dom}(f) \qquad \lim_{n\to\infty} x_n = \pm\infty \quad \Longrightarrow \quad \lim_{n\to\infty} f(x_n) = b\,. \tag{5.18}$$

Proof. Exercise.

5.5.4. *Limits of Combinations of Functions.* The basic theorem regarding limits of an algebraic combination of functions is the following.

Proposition 5.11. *Let* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *and* $g : \mathrm{Dom}(g) \to \mathbb{R}$ *where* $\mathrm{Dom}(f) \subset \mathbb{R}$ *and* $\mathrm{Dom}(g) \subset \mathbb{R}$. *Let* b, $c \in \mathbb{R}$ *such that*

$$\lim f(x) = b\,, \qquad \text{and} \qquad \lim g(x) = c\,,$$

*where "*lim*" stands either for one of*

$$\lim_{x\to a}\,, \qquad \lim_{x\to a^+}\,, \qquad \lim_{x\to a^-}\,, \qquad \text{for some } a \in \mathbb{R}\,,$$

or for one of

$$\lim_{x\to+\infty}\,, \qquad \lim_{x\to-\infty}\,.$$

Then

$$\begin{aligned} \lim \big(f(x) + g(x)\big) &= b + c\,, \\ \lim \big(f(x)\,g(x)\big) &= b\,c\,, \\ \lim \frac{f(x)}{g(x)} &= \frac{b}{c} \qquad \text{provided } c \neq 0\,. \end{aligned}$$

Proof. Exercise.

The story regarding the limits of a composition of functions is more complicated. The simplest result is the following.

Proposition 5.12. *Let* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *and* $g : \mathrm{Dom}(g) \to \mathbb{R}$ *where* $\mathrm{Dom}(f) \subset \mathbb{R}$ *and* $\mathrm{Dom}(g) \subset \mathbb{R}$*. Let* $b \in \mathbb{R}$ *such that*

$$\lim f(x) = b\,,$$

*where "*lim*" stands either for one of*

$$\lim_{x\to a}\,, \qquad \lim_{x\to a^+}\,, \qquad \lim_{x\to a^-}\,, \qquad \text{for some } a \in \mathbb{R}\,,$$

or for one of

$$\lim_{x\to +\infty}\,, \qquad \lim_{x\to -\infty}\,.$$

If $b \in \mathrm{Dom}(g)$ *and* g *is continuous at* b *then*

$$\lim g(f(x)) = g(b)\,.$$

Proof. Exercise.

5.6. **Monotonic Functions.** We now extend the notions associated with monotonic sequences to more general functions.

Definition 5.7. *Given a function* $f : Dom(f) \to \mathbb{R}$ *with* $\mathrm{Dom}(f) \subset \mathbb{R}$*, we say that* f *is*

$$\begin{aligned}
\text{increasing } &\textit{whenever } f(x) < f(y) \textit{ for every } x, y \in \mathrm{Dom}(f) \textit{ with } x < y\,,\\
\text{nondecreasing } &\textit{whenever } f(x) \le f(y) \textit{ for every } x, y \in \mathrm{Dom}(f) \textit{ with } x < y\,,\\
\text{decreasing } &\textit{whenever } f(y) < f(x) \textit{ for every } x, y \in \mathrm{Dom}(f) \textit{ with } x < y\,,\\
\text{nonincreasing } &\textit{whenever } f(y) \le f(x) \textit{ for every } x, y \in \mathrm{Dom}(f) \textit{ with } x < y\,.
\end{aligned}$$

We say that f *is* monotonic *if it is either nondecreasing or nonincreasing. We say that* f *is* strictly monotonic *if it is either increasing or decreasing.*

Remark. Sequences are functions whose domain is $\mathbb{N}$. The definitions give above are consistent with our earlier usage of the same terms in the context of sequences.

An important fact about monotonic functions defined over an interval is that their one-sided limits exists at every point. This restricts both the kind and number of discontinuities such functions can have.

Proposition 5.13. *Let* $f : (a, b) \to \mathbb{R}$ *be monotonic. Then the one-sided limits of* f *exist at every* $x \in (a, b)$*. When* f *is nondecreasing we have*

$$\begin{aligned}
\lim_{y\to x^-} f(y) &= \sup\big\{f(y) \,:\, a < y < x\big\} \ge f(x)\,,\\
\lim_{y\to x^+} f(y) &= \inf\big\{f(y) \,:\, x < y < b\big\} \le f(x)\,.
\end{aligned}$$

When f *is nonincreasing we have*

$$\begin{aligned}
\lim_{y\to x^-} f(y) &= \inf\big\{f(y) \,:\, a < y < x\big\} \le f(x)\,,\\
\lim_{y\to x^+} f(y) &= \sup\big\{f(y) \,:\, x < y < b\big\} \ge f(x)\,.
\end{aligned}$$

Proof. Exercise.

Proposition 5.14. *Let $f:(a,b)\to\mathbb{R}$ be nondecreasing (nonincreasing). Define $\underline{f}:(a,b)\to\mathbb{R}$ and $\overline{f}:(a,b)\to\mathbb{R}$ for every $x\in(a,b)$ by*

$$\underline{f}(x)=\lim_{y\to x^-} f(y)\,, \qquad \overline{f}(x)=\lim_{y\to x^+} f(y)\,.$$

$$\Big(\underline{f}(x)=\lim_{y\to x^+} f(y)\,, \qquad \overline{f}(x)=\lim_{y\to x^-} f(y)\,.\Big)$$

Then $\underline{f}$ and $\overline{f}$ are nondecreasing (nonincreasing) with $\underline{f}(x)\leq f(x)\leq\overline{f}(x)$ for every $x\in(a,b)$. Moreover, $\underline{f}$ is left (right) continuous while $\overline{f}$ is right (left) continuous over (a,b).

Proof. Exercise.

Proposition 5.15. *Let $f:(a,b)\to\mathbb{R}$ be monotonic. Let $\underline{f}:(a,b)\to\mathbb{R}$ and $\overline{f}:(a,b)\to\mathbb{R}$ be defined as in Proposition 5.14. Then f is continuous at $x\in(a,b)$ if and only if $\underline{f}(x)=\overline{f}(x)$. Moreover, there can be at most a countable subset of (a,b) where f is discontinuous.*

Remark. Recall that a set S is countable if and only if there is a mapping from $\mathbb{N}$ onto S. Recall too that a countable union of countable sets is still countable.

Proof. The fact that f is continuous at $x\in(a,b)$ if and only if $\underline{f}(x)=\overline{f}(x)$ follows from Propostions 5.9, 5.13, and 5.14. So all that remains is to prove that there can be at most a countable set of points in (a,b) at which f is discontinuous.

Without loss of generality we can assume that f is nondecreasing. (Otherwise consider $-f$.) For every $[c,d]\subset(a,b)$ and every $\epsilon>0$ define the set $S_\epsilon([c,d])\subset(a,b)$ by

$$S_\epsilon([c,d])=\Big\{x\in[c,d]\,:\,\overline{f}(x)-\underline{f}(x)\geq\epsilon\Big\}\,.$$

Let $\{x_1,x_2,\cdots,x_n\}\subset S_\epsilon([c,d])$ and $x_1<x_2<\cdots<x_n$. Because f is nondecreasing over $[c,d]$

$$\underline{f}(c)\leq\underline{f}(x_1)<\overline{f}(x_1)\leq\underline{f}(x_2)<\overline{f}(x_2)\leq\cdots\leq\underline{f}(x_n)<\overline{f}(x_n)\leq\overline{f}(d)\,.$$

But this implies that

$$\overline{f}(d)-\underline{f}(c)\geq\sum_{k=1}^{n}\big(\overline{f}(x_k)-\underline{f}(x_k)\big)\geq n\epsilon\,.$$

Hence, each set $S_\epsilon([c,d])$ has at most $\big(\overline{f}(d)-\underline{f}(c)\big)/\epsilon$ points in it, and is thereby a finite set.

Now let $\{\epsilon_n\}_{n\in\mathbb{N}}$ be any decreasing sequence that converges to 0 and let $\{[c_n,d_n]\}_{n\in\mathbb{N}}$ be any sequence of intervals such that $[c_n,d_n]\subset[c_{n+1},d_{n+1}]\subset(a,b)$ and

$$\bigcup_{n\in\mathbb{N}}[c_n,d_n]=(a,b)\,.$$

Set $S_n=S_{\epsilon_n}([c_n,d_n])$. We claim that every point in (a,b) at which f is discontinuous must be in some S_n. Therefore the set of all points in (a,b) at which f is discontinuous is

$$\bigcup_{n\in\mathbb{N}}S_n\,.$$

But this is a countable union of finite sets, which is thereby countable. □

Exercise. Prove the claim that every point in (a,b) at which f is discontinuous is in some S_n.

Exercise. Let $\{x_n\}_{n\in\mathbb{N}}$ be any sequence contained in $[0,1]$. Construct an increasing function f over $[0,1]$ that is discontinuous at the points $\{x_n\}_{n\in\mathbb{N}}$ and is continuous elsewhere in $[0,1]$.

6. Differentiability and Derivatives

6.1. **Differentiability.** Given any function $f : \mathrm{Dom}(f) \to \mathbb{R}$ with $\mathrm{Dom}(f) \subset \mathbb{R}$, the equation of the secant line through any two points $(a, f(a))$ and $(b, f(b))$ on its graph is

$$y = f(a) + \frac{f(b) - f(a)}{b - a}\,(x - a)\,. \tag{6.1}$$

The slope of this secant line is given by the *difference quotient*

$$\frac{f(b) - f(a)}{b - a}\,. \tag{6.2}$$

This quantity is defined for every $a, b \in \mathrm{Dom}(f)$ such that $b \neq a$. It is undefined when $b = a$.

Definition 6.1. *A function $f : \mathrm{Dom}(f) \to \mathbb{R}$ is said to be* differentiable at *a point $a \in \mathrm{Dom}(f)$ whenever*

$$\lim_{b \to a} \frac{f(b) - f(a)}{b - a} \quad \textit{exists}\,. \tag{6.3}$$

A function f that is differentiable at every point in a set $S \subset \mathrm{Dom}(f)$ is said to be differentiable over *S. If f is differentiable at every point in $\mathrm{Dom}(f)$ then it is said to be* differentiable.

This definition should be viewed geometrically as follows. When f is differentiable at a we see from (6.1) and (6.3) that the equation of the tangent line is given by

$$y = f(a) + f'(a)(x - a)\,, \tag{6.4}$$

where *slope of the tangent line* is given by

$$f'(a) = \lim_{b \to a} \frac{f(b) - f(a)}{b - a}\,. \tag{6.5}$$

By replacing b by $a + h$ in (6.5), the slope of this tangent line may be expressed as

$$f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}\,. \tag{6.6}$$

It should be evident that (6.6) is equivalent to (6.5). Visually, if the graph of a function f at $(a, f(a))$ either has no unique tangent line or has a vertical tangent line then f is not differentiable at the point a.

It is easy to see that if f is differentiable at the point a then it is continuous at a. Indeed, for every $x \in \mathrm{Dom}(f)$ such that $x \neq a$ we have the identity

$$f(x) = f(a) + \frac{f(x) - f(a)}{x - a}\,(x - a)\,.$$

If we let x approach a in this identity then because f is differentiable at a we see that

$$\begin{aligned} \lim_{x \to a} f(x) &= f(a) + \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \cdot \lim_{x \to a} (x - a) \\ &= f(a) + f'(a) \cdot 0 = f(a)\,, \end{aligned}$$

whereby f is continuous at a. The converse is not true. For example, the functions $|x|$ and $x^{1/3}$ are continuous over $\mathbb{R}$ but are not differentiable at 0 for different reasons. At this stage you should be able to give such examples of functions that are continuous but not differentiable at some point. Later in the course we will construct functions that are continuous everywhere yet are differentiable nowhere. Indeed, most continuous functions are differentiable nowhere.

Examples. Consider the functions f and g given by

$$f(x) = \begin{cases} 0 & \text{for } x = 0 \\ x\sin(1/x) & \text{otherwise}, \end{cases} \qquad g(x) = \begin{cases} 0 & \text{for } x = 0 \\ x^2\cos(1/x) & \text{otherwise}. \end{cases}$$

Can you see that

(1) f and g are even?
(2) f oscillates between the lines $y = x$ and $y = -x$ near zero?
(3) g oscillates between the parabolas $y = x^2$ and $y = -x^2$ near zero?
(4) f has an horizontal asymptote of $y = 1$ as $|x| \to \infty$?
(5) g behaves like x^2 as $|x| \to \infty$?
(6) f and g are continuous at $x = 0$?
(7) f is not differentiable at $x = 0$?
(8) g is differentiable at $x = 0$ with $g'(0) = 0$?

Computers often have difficulty rendering accurate graphs of such functions near zero, so they must be understood analytically.

6.2. **Derivatives.** The derivative of a function f, which is defined at every point x where f is differentiable, is the function f' whose value at x is the slope of the tangent line to the graph of f at x. Hence,

$$\text{Dom}(f') \equiv \big\{x \in \text{Dom}(f) \,:\, f \text{ is differentiable at } x\big\}, \tag{6.7}$$

and by (6.6) the value of $f'(x)$ is given by

$$f'(x) = \frac{\mathrm{d}}{\mathrm{d}x} f(x) \equiv \lim_{y \to x} \frac{f(y) - f(x)}{y - x}. \tag{6.8}$$

Then f is differentiable if and only if $\text{Dom}(f') = \text{Dom}(f)$. Otherwise $\text{Dom}(f')$ is a strict subset of $\text{Dom}(f)$. If f is differentiable and f' is continuous then f is said to be *continuously differentiable.*

The second derivative of f is the derivative of its derivative. It is defined by

$$f''(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2} f(x) \equiv \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}}{\mathrm{d}x} f(x)\right),$$

with

$$\text{Dom}(f'') = \big\{x \in \text{Dom}(f') \,:\, f' \text{ is differentiable at } x\big\}.$$

If $\text{Dom}(f'') = \text{Dom}(f)$ then f is said to be *twice differentiable.* If f is twice differentiable and f'' is continuous then f is said to be *twice continuously differentiable.*

In a similar way the n^{th} derivative of f is defined by

$$f^{(n)}(x) = \frac{d^n}{dx^n} f(x) \equiv \frac{d}{dx}\left(\frac{d^{n-1}}{dx^{n-1}} f(x)\right).$$

with

$$\text{Dom}(f^{(n)}) = \big\{x \in \text{Dom}(f^{(n-1)}) \,:\, f^{(n-1)} \text{ is differentiable at } x\big\}.$$

If $\text{Dom}(f^{(n)}) = \text{Dom}(f)$ then f is said to be n*-times differentiable.* If f is n-times differentiable and $f^{(n)}$ is continuous over $\text{Dom}(f)$ then f is said to be n*-times continuously differentiable.* If f is n-times differentiable for every n then it is said to be *infinitely differentiable* or *smooth.*

If the variable z is a function of the variable x then we will sometimes denote the first, second, and n^{th} derivatives of this function by

$$\frac{\mathrm{d}z}{\mathrm{d}x}, \qquad \frac{\mathrm{d}^2z}{\mathrm{d}x^2}, \qquad \text{and} \qquad \frac{\mathrm{d}^nz}{\mathrm{d}x^n}.$$

There are many other commonly used notations for derivatives. By now you have likely seen a few others. Such a variety is not too surprising when you realize that derivatives are among the most useful objects in all of mathematics.

6.3. **Differentiation.** Differentiation is the processs by which we compute derivatives. The classical differentiation rules that you recall from calculus can now be derived.

6.3.1. *Linear Combinations of Differentiable Functions.* Given any two differentiable functions u and v, and any constant k, the functions ku and $u+v$ are also differentiable and their derivatives are given by the so-called *multiplication rule* and *sum rule:*

$$\frac{\mathrm{d}}{\mathrm{d}x}(ku) = k\,\frac{\mathrm{d}u}{\mathrm{d}x}, \qquad \frac{\mathrm{d}}{\mathrm{d}x}(u+v) = \frac{\mathrm{d}u}{\mathrm{d}x} + \frac{\mathrm{d}v}{\mathrm{d}x}. \tag{6.9}$$

These rules follow from the definition of the derivative (6.8) and the algebraic identities

$$\frac{ku(y)-ku(x)}{y-x} = k\,\frac{u(y)-u(x)}{y-x},$$
$$\frac{u(y)+v(y)-u(x)-v(x)}{y-x} = \frac{u(y)-u(x)}{y-x} + \frac{v(y)-v(x)}{y-x}.$$

The multiplication and sum rules (6.9) express the fact that differentiation is a linear operation. The linear combinations of n given functions $\{u_1, u_2, \cdots, u_n\}$ are all those functions of the form $k_1u_1 + k_2u_2 + \cdots + k_nu_n$ for some choice of n constants $\{k_1, k_2, \cdots, k_n\}$. In other words, the linear combinations are all those function that can be built up from the given functions $\{u_1, u_2, \cdots, u_n\}$ by repeated multiplication by constants and addition. If each of the given functions $\{u_1, u_2, \cdots, u_n\}$ is differentiable then repeated applications of the multiplication and sum rules (6.9) show that each such linear combination is also differentiable and its derivative is given by the *linear combination rule:*

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(k_1u_1 + k_2u_2 + \cdots + k_nu_n\big) = k_1\frac{\mathrm{d}u_1}{\mathrm{d}x} + k_2\frac{\mathrm{d}u_2}{\mathrm{d}x} + \cdots + k_n\frac{\mathrm{d}u_n}{\mathrm{d}x}. \tag{6.10}$$

6.3.2. *Algebraic Combinations of Differentiable Functions.* Given any two differentiable functions u and v, the function uv is also differentiable and its derivative is given by the so-called *product (or Leibniz) rule:*

$$\frac{\mathrm{d}}{\mathrm{d}x}(uv) = \frac{\mathrm{d}u}{\mathrm{d}x}v + u\frac{\mathrm{d}v}{\mathrm{d}x}. \tag{6.11}$$

This is not as simple to express in words as say the sum rule, but may be rendered as "the derivative of a product is the derivative of the first times the second plus the first times the derivative of the second". This rule follows directly from the definition and the algebraic identity

$$\frac{u(y)v(y)-u(x)v(x)}{y-x} = \frac{u(y)-u(x)}{y-x}\,v(y) + u(x)\,\frac{v(y)-v(x)}{y-x}.$$

The product rule is a very important general rule for differentiation. In fact, most other rules in this section will essentially follow from the product rule.

If we consider the product of three differentiable functions u, v, and w then two applications of (6.11) show that

$$\frac{\mathrm{d}}{\mathrm{d}x}(uvw) = \frac{\mathrm{d}u}{\mathrm{d}x}vw + u\frac{\mathrm{d}v}{\mathrm{d}x}w + uv\frac{\mathrm{d}w}{\mathrm{d}x}\,.$$

More generally, given n differentiable functions $\{u_1, u_2, \cdots, u_n\}$, their product $u_1 u_2 \cdots u_n$ is differentiable and its derivative is given by the *general Leibniz rule:*

$$\frac{\mathrm{d}}{\mathrm{d}x}(u_1 u_2 \cdots u_n) = \frac{\mathrm{d}u_1}{\mathrm{d}x}u_2 \cdots u_n + u_1\frac{\mathrm{d}u_2}{\mathrm{d}x} \cdots u_n + \cdots + u_1 u_2 \cdots \frac{\mathrm{d}u_n}{\mathrm{d}x}\,. \tag{6.12}$$

A consequence of setting $v = 1/u$ in the product rule (6.11) is the *reciprocal rule:*

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{u}\right) = -\frac{1}{u^2}\frac{\mathrm{d}u}{\mathrm{d}x} \qquad \text{wherever } u \neq 0\,. \tag{6.13}$$

If the reciprocal rule is combined with the product rule then we obtain the *quotient rule:*

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{u}{v}\right) = \frac{\dfrac{\mathrm{d}u}{\mathrm{d}x}v - u\dfrac{\mathrm{d}v}{\mathrm{d}x}}{v^2} \qquad \text{wherever } v \neq 0\,. \tag{6.14}$$

If the general Leibniz rule (6.12) is specialized to the case where all the functions u_k are the same function u then it reduces to the *monomial power rule:*

$$\frac{\mathrm{d}}{\mathrm{d}x}u^n = nu^{n-1}\frac{\mathrm{d}u}{\mathrm{d}x}\,. \tag{6.15}$$

The monomial power rule was derived above for positive integers n. When it is combined with the reciprocal rule (6.13), we see that it extends to negative integers n. This rule can be extended further. Namely, given any differentiable function u and any rational number p for which u^p is defined, the function u^p is differentiable wherever u^{p-1} is defined and its derivative is given by the *rational power rule:*

$$\frac{\mathrm{d}}{\mathrm{d}x}u^p = pu^{p-1}\frac{\mathrm{d}u}{\mathrm{d}x}\,. \tag{6.16}$$

Wherever $u \neq 0$ this rule can be derived as follows. Because p is rational it can be expressed as $p = m/n$ where m and n are integers and $n > 0$. If the monomial power rule (6.15) is then applied to each side of the identity $(u^p)^n = u^m$, we find that

$$n(u^p)^{n-1}\frac{\mathrm{d}}{\mathrm{d}x}u^p = mu^{m-1}\frac{\mathrm{d}u}{\mathrm{d}x}\,,$$

which is equivalent to the rational power rule wherever $u \neq 0$. Points where $u = 0$ and $p \geq 1$ can be treated directly from the definition of the derivative.

6.3.3. *Compositions of Differentiable Functions.* Given two differentiable functions v and u, the derivative of their composition $v(u)$ is given by the *chain rule:*

$$\frac{\mathrm{d}}{\mathrm{d}x}v(u) = v'(u)\,\frac{\mathrm{d}u}{\mathrm{d}x}\,. \tag{6.17}$$

The chain rule is the most important general rule for differentiation.

It is natural to think that the chain rule can be derived by letting y approach x in the algebraic identity

$$\frac{v(u(y)) - v(u(x))}{y - x} = \frac{v(u(y)) - v(u(x))}{u(y) - u(x)} \, \frac{u(y) - u(x)}{y - x} \,.$$

However, this argument fails because the identity breaks down wherever the $u(y) - u(x)$ that appears in the denominator becomes zero. This difficulty is overcome by observing that if v is differentiable at a point b then a *continuous difference quotient* may be defined for every $z \in \mathrm{Dom}(v)$ by

$$Q_b v(z) \equiv \begin{cases} \dfrac{v(z) - v(b)}{z - b} & \text{for } z \neq b \,, \\ v'(b) & \text{for } z = b \,. \end{cases} \tag{6.18}$$

This is a continuous function of z at b and satisfies

$$v(z) - v(b) = Q_b v(z)\,(z - b) \,.$$

Now set $b = u(x)$ and $z = u(y)$ in this relation and divide by $y - x$ to obtain

$$\frac{v(u(y)) - v(u(x))}{y - x} = Q_{u(x)} v\big(u(y)\big) \, \frac{u(y) - u(x)}{y - x} \,.$$

The chain rule (6.17) then follows from the composition limit rule of Proposition 5.12 and the definition of the derivative (6.8) by letting y approach x.

If we consider the composition of three differentiable functions, w, v, and u, then two applications of (6.17) show that

$$\frac{\mathrm{d}}{\mathrm{d}x} w(v(u)) = w'(v(u))\, v'(u) \, \frac{\mathrm{d}u}{\mathrm{d}x} \,.$$

More generally, if we consider n differentiable functions $\{u_1, u_2, \cdots, u_n\}$, then $n-1$ applications of (6.17) show their composition $u_1(u_2(u_3(\cdots(u_n)\cdots)))$ is differentiable and its derivative is given by the *linked chain rule:*

$$\frac{\mathrm{d}}{\mathrm{d}x} u_1(u_2(u_3(\cdots(u_n)\cdots))) = u_1'(u_2(u_3(\cdots(u_n)\cdots)))\, u_2'(u_3(\cdots(u_n)\cdots)) \cdots \frac{\mathrm{d}u_n}{\mathrm{d}x} \,. \tag{6.19}$$

6.3.4. *Inverses of Differentiable Functions.* Because a function f is "undone" when composed with its inverse function f^{-1} in the sense that $u = f(f^{-1}(u))$, the chain rule (6.17) can be used to derive the *inverse function rule:*

$$\frac{\mathrm{d}}{\mathrm{d}x} f^{-1}(u) = \frac{1}{f'\big(f^{-1}(u)\big)} \, \frac{\mathrm{d}u}{\mathrm{d}x} \,. \tag{6.20}$$

To find the derivative formula for $v = f^{-1}(u)$, we derive the identity $f(v) = u$ to obtain

$$f'(v) \, \frac{\mathrm{d}v}{\mathrm{d}x} = \frac{\mathrm{d}u}{\mathrm{d}x} \,.$$

Then solve for dv/dx and use $v = f^{-1}(u)$ to eliminate the v in $f'(v)$. This gives (6.20). This derivation assumes that $v = f^{-1}(u)$ is differentiable at x.

However, if we use the continuous difference quotient (6.18) that was used to prove the chain rule rather than the chain rule itself then we can prove the following.

Proposition 6.1. *Let* $u : \mathrm{Dom}(u) \to \mathbb{R}$ *with* $\mathrm{Dom}(u) \subset \mathbb{R}$. *Let* $f : \mathrm{Dom}(f) \to \mathbb{R}$ *be one-to-one over* $\mathrm{Dom}(f) \subset \mathbb{R}$. *Let* $\mathrm{Rng}(u) \subset \mathrm{Rng}(f)$, *so that* $v = f^{-1}(u)$ *has natural domain* $\mathrm{Dom}(v) = \mathrm{Dom}(u)$. *Let* $x \in \mathrm{Dom}(u)$ *such that*

- *u is differentiable at x and*
- *f is differentiable at $v(x)$ with $f'(v(x)) \neq 0$.*

Then $v = f^{-1}(u)$ is differentiable at x with

$$v'(x) = \frac{1}{f'(v(x))}\, u'(x)\,. \tag{6.21}$$

Proof. Because $u = f(v)$, we have the identity

$$Q_{v(x)}f(v(y))\,\frac{v(y)-v(x)}{y-x} = \frac{u(y)-u(x)}{y-x} \qquad \text{for every } x, y \in \mathrm{Dom}(u)\,.$$

Because our assumptions imply that

$$\lim_{y\to x}\frac{u(y)-u(x)}{y-x} = u'(x)\,, \qquad \lim_{y\to x} Q_{v(x)}f(v(y)) = f'(v(x)) \neq 0\,,$$

we conclude that v is differentiable at x with

$$v'(x) = \lim_{y\to x}\frac{v(y)-v(x)}{y-x} = \lim_{y\to x}\frac{1}{Q_{v(x)}f(v(y))}\,\lim_{y\to x}\frac{u(y)-u(x)}{y-x} = \frac{1}{f'(v(x))}\,u'(x)\,.$$

This proves (6.21). □

6.4. **Local Extrema and Critical Points.** In introductory calculus you learned how to use derivatives to find a minimum or maximum of a given function. Here we put those methods on a firm theoretical foundation.

6.4.1. *Local Extrema.* We begin with the concept of local extrema, which arises natrually when calculus is used to find extrema.

Definition 6.2. *Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$. We say that $p \in D$ is a* local minimizer (maximizer) *of f over D if p is a minimizer (maximizer) of f restricted to $D \cap (p-\delta, p+\delta)$ for some $\delta > 0$. The value $f(p)$ is then called a* local minimum (maximum) *of f over D. In this context, a minimizer (maximizer) of f over D is referred to as a* global minimizer (maximizer) *while a minimum (maximum) of f over D is referred to as a* global minimum (maximum).

A point that is either a local minimizer or local maximizer of f over D is called a local extremizer *and its corresponding value is called a* local extremum. *We similarly defines* global extremizer *and* global extremum.

Remark. The terms *relative* and *absolute* are sometimes used rather than *local* and *global.*

Remark. It is clear that every global extremum of a function is also a local extremum. However, a function can have many local extrema without having any global extremum. For example, consider

$$f(x) = x + 2\sin(x) \qquad \text{over } (-\infty, \infty)\,.$$

6.4.2. *Transversality Lemma.* A key step in developing calculus tools for finding local extrema is the following lemma.

Proposition 6.2. Transversality Lemma. *Let $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$ be differentiable at $p \in D$. If $f'(p) > 0$ then there exists a $\delta > 0$ such that*

$$\begin{aligned} x \in D \cap (p-\delta, p) \qquad &\Longrightarrow \qquad f(x) < f(p)\,, \\ x \in D \cap (p, p+\delta) \qquad &\Longrightarrow \qquad f(x) > f(p)\,, \end{aligned}$$

while if $f'(p) < 0$ then there exists a $\delta > 0$ such that

$$\begin{aligned} x \in D \cap (p-\delta, p) \qquad &\Longrightarrow \qquad f(x) > f(p)\,, \\ x \in D \cap (p, p+\delta) \qquad &\Longrightarrow \qquad f(x) < f(p)\,. \end{aligned}$$

Remark. The lemma states that if $f'(p) \neq 0$ the graph of f will lie below the line $y = f(p)$ on one side of p, and above it on the other. In other words, it says the graph of f is transversal to the line $y = f(p)$. Hence, it is called the Transversality Lemma. We cannot expect much more. For example, it is not generally true that if $f : D \to \mathbb{R}$ is differentiable at $p \in D$ and $f'(p) > 0$ ($f'(p) < 0$) that then f is increasing (decreasing) near p. This is seen from the example

$$f(x) = \begin{cases} 0 & \text{for } x = 0 \\ mx + x^2 \cos(1/x) & \text{otherwise}\,, \end{cases}$$

where $m \in (0, 1)$. Because

$$f'(x) = \begin{cases} m & \text{for } x = 0 \\ m + \sin(1/x) + 2x\cos(1/x) & \text{otherwise}\,, \end{cases}$$

we see that $f'(0) = m > 0$, yet f is not an increasing function over any interval containing 0.

Proof. By the definition of the derivative we have

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

When $f'(p) > 0$ we use the ϵ-δ characterization of this limit with $\epsilon = f'(p)$ to conclude that there exists a $\delta > 0$ such that for every $x \in D$

$$0 < |x - p| < \delta \qquad \Longrightarrow \qquad \left| \frac{f(x) - f(p)}{x - p} - f'(p) \right| < f'(p) \;\Longrightarrow\; \qquad \frac{f(x) - f(p)}{x - p} > 0\,.$$

This implication is equivalent to the first assertion of the Lemma.

Similarly, when $f'(p) < 0$ we use the ϵ-δ characterization of the limit with $\epsilon = -f'(p)$ to conclude that there exists a $\delta > 0$ such that for every $x \in D$

$$0 < |x - p| < \delta \qquad \Longrightarrow \qquad \left| \frac{f(x) - f(p)}{x - p} - f'(p) \right| < -f'(p) \;\Longrightarrow\; \qquad \frac{f(x) - f(p)}{x - p} < 0\,.$$

This implication is equivalent to the second assertion of the Lemma. $\square$

6.4.3. *One-Sided Limit Point Test.*

Definition 6.3. *Let $D \subset \mathbb{R}$ and p be a limit point of D. Then p is called a* one-sided limit point *of D whenever p is not a limit point of both $D \cap (p, \infty)$ and $D \cap (-\infty, p)$.*

One consequence of the Transversality Lemma is the following test for when a one-sided limit point is a local minimizer or maximizer.

Proposition 6.3. One-Sided Limit Point Test. *Let $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$ be differentiable at $p \in D$. If p is not a limit point of $D \cap (p, \infty)$ $\big(D \cap (-\infty, p)\big)$ then*

$$\text{if} \quad f'(p) > 0 \quad \text{then } p \text{ is a local maximizer (minimizer) of } f \text{ over } D\,,$$
$$\text{if} \quad f'(p) < 0 \quad \text{then } p \text{ is a local minimizer (maximizer) of } f \text{ over } D\,,$$
$$\text{if} \quad f'(p) = 0 \quad \text{then there is no information}\,.$$

Proof. Exercise.

Remark. When D is either $[a, b]$, $[a, b)$, or $(a, b]$ then this test applies to a or b when it is a closed endpoint of D.

6.4.4. *Critical Points.* The following corollary of the Transversality Lemma states that certain points cannot be local extremizers.

Proposition 6.4. Transversality Corollary. *Let $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$ be differentiable at $p \in D$. If p is a limit point of $D \cap (p, \infty)$ $\big(D \cap (-\infty, p)\big)$ then*

$$f'(p) > 0 \quad \Longrightarrow \quad p \text{ is not a local maximizer (minimizer) of } f \text{ over } D\,,$$
$$f'(p) < 0 \quad \Longrightarrow \quad p \text{ is not a local minimizer (maximizer) of } f \text{ over } D\,.$$

In particular, if p is a limit point of both $D \cap (p, \infty)$ and $D \cap (-\infty, p)$ then

$$f'(p) \neq 0 \quad \Longrightarrow \quad p \text{ is not a local extremizer of } f \text{ over } D\,.$$

Proof. Observe that if p is a limit point of $D \cap (p, \infty)$ then for every $\delta > 0$ the set $D \cap (p, p+\delta)$ is nonempty. Similarly, if p is a limit point of $D \cap (-\infty, p)$ then for every $\delta > 0$ the set $D \cap (p - \delta, p)$ is nonempty. Given these observations, the result follows from the Transversality Lemma. The details are left as an exercise. □

Remark. When $f : D \to \mathbb{R}$ is differentiable at $p \in D$, the definition requires p to be a limit point of D. It follows that p must be a limit point of at least one of $D \cap (p, \infty)$ or $D \cap (-\infty, p)$. However, p does not generally have to be a limit point of both $D \cap (p, \infty)$ and $D \cap (-\infty, p)$. This is the case when D is either $[a, b]$, $[a, b)$, or $(a, b]$ and p is a closed endpoint of D.

The above corollary motivates the following definition.

Definition 6.4. *Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$. Then $p \in D$ is called a* critical point *of f over D if either*

- *f is not differentiable at p,*
- *$f'(p) = 0$,*
- *or p is a one-sided limit point of D.*

The last assertion of the Transversality Corollary can then be recast as follows.

Proposition 6.5. Fermat Critical Point Theorem. *Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$. Then every local extremizer of f over D is a critical point of f over D.*

6.5. **Intermediate-Value and Sign Dichotomy Theorems.** The Extreme-Value Theorem and the Fermat Critical Point Theorem combine with the One-Sided Limit Point Theorem to give a result that lies at the heart of some tests for analyzing the monotonicity of a function.

Proposition 6.6. Derivative Intermediate-Value Theorem. *Let $a < b$ and $f : [a, b] \to \mathbb{R}$ be differentiable. Then f' takes all values that lie between $f'(a)$ and $f'(b)$.*

Proof. The theorem holds when $f'(a) = f'(b)$ because in that case there are no values between $f'(a)$ and $f'(b)$. Now consider the case when $f'(a) < f'(b)$. Let $m \in \mathbb{R}$ such that

$$f'(a) < m < f'(b)\,.$$

Define a function $g : [a, b] \to \mathbb{R}$ for every $x \in [a, b]$ by

$$g(x) \equiv f(x) - m\,x\,.$$

Clearly, as a function of x:

- g is continuous over $[a, b]$;
- g is differentiable over $[a, b]$ with $g'(x) = f'(x) - m$;
- $g'(a) = f'(a) - m < 0$ while $g'(b) = f'(b) - m > 0$.

The One-Sided Limit Point Theorem then implies that both a and b are local maxima and not local minima of g over $[a, b]$. But by the Extreme-Value Theorem g must therefore have a global minimum at some p in (a, b). Because g is differentiable over (a, b), the Fermat Critical Point Theorem implies that $g'(p) = f'(p) - m = 0$. Therefore $f'(p) = m$ for some p in (a, b). The case where $f'(b) < f'(a)$ is argued similarly. □

Remark. The Derivative Intermediate-Value Theorem is stronger than the Intermediate-Value Theorem for continuous functions that we studied earlier. We know that derivatives are not generally continuous, so this theorem does not follow from the earlier one. It will be a consequence of the Second Fundamental Theorem of Calculus that every function that is continuous over an interval $[a, b]$ is the derivative of some other function over that interval. The class of functions considered by the Derivative Intermediate-Value Theorem is therefore strictly larger than that considered by the earlier theorem.

We will employ the following consequence of the Derivative Intermediate-Value Theorem to obtain tests for analyzing the monotonicity of a function.

Proposition 6.7. Derivative Sign Dichotomy Theorem. *Let $a < b$ and $f : (a, b) \to \mathbb{R}$ be differentiable. If f has no critical points in (a, b) then either*

$$f' > 0 \text{ over } (a, b) \qquad \text{or} \qquad f' < 0 \text{ over } (a, b)\,.$$

Proof. Suppose not. Then there are points q, $r \in (a, b)$ such that $f'(q) < 0 < f'(r)$. In the case $q < r$, the Derivative Intermediate-Value Theorem applied to f over $[q, r]$ implies that there exists a $p \in (q, r)$ such that $f'(p) = 0$. This would imply that $p \in (a, b)$ is a critical point of f. The case $q > r$ leads to the same conclusion. However f has no critical points over (a, b), so our supposition must be false. Hence, the values of f' can only take one sign over (a, b). □

Remark. The converse of this theorem is trivially true because if f' is either always positive over (a, b) or always negative over (a, b) then it is never zero over (a, b), whereby f has no critical points in (a, b).

6.6. **Convex and Concave Functions.** We begin by defining what it means for a function to be either convex or concave over an interval.

Definition 6.5. *Let $f : D \to \mathbb{R}$ for some $D \subset \mathbb{R}$. Let $I \subset D$ be an interval. The function f is said to be* convex (strictly convex) over I *if for every $x, z \in I$ with $x < z$*

$$\begin{aligned} &f\big(tx + (1-t)z\big) \leq tf(x) + (1-t)f(z) \qquad \forall t \in (0,1)\,. \\ &\Big(f\big(tx + (1-t)z\big) < tf(x) + (1-t)f(z) \qquad \forall t \in (0,1)\,.\Big) \end{aligned} \tag{6.22}$$

The function f is said to be concave (strictly concave) over I *if for every $x, z \in I$ with $x < z$*

$$\begin{aligned} &f\big(tx + (1-t)z\big) \geq tf(x) + (1-t)f(z) \qquad \forall t \in (0,1)\,. \\ &\Big(f\big(tx + (1-t)z\big) > tf(x) + (1-t)f(z) \qquad \forall t \in (0,1)\,.\Big) \end{aligned} \tag{6.23}$$

Remark. The geometric picture is that chords drawn between two points on the graph of a function over an interval will lie above the graph if the function is strictly convex over the interval, and lie below the graph if the function is strictly concave over the interval.

Remark. A function f is convex (strictly convex) over an interval I if and only if the function $-f$ is concave (strictly concave) over I. We will therefore often only state results or give proofs for the convex or strictly convex cases.

A useful characterization of these concepts in terms of the monotonicity of difference quotients is provided by the following proposition.

Proposition 6.8. *Let $f : D \to \mathbb{R}$ for some $D \subset \mathbb{R}$. Let $I \subset D$ be an interval. Then*

- *f is convex over I if and only if for every $x, y, z \in I$*
$$x < y < z \quad \Longrightarrow \quad \frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(y)}{z - y}\,.$$
- *f is strictly convex over I if and only if for every $x, y, z \in I$*
$$x < y < z \quad \Longrightarrow \quad \frac{f(y) - f(x)}{y - x} < \frac{f(z) - f(y)}{z - y}\,.$$
- *f is concave over I if and only if for every $x, y, z \in I$*
$$x < y < z \quad \Longrightarrow \quad \frac{f(y) - f(x)}{y - x} \geq \frac{f(z) - f(y)}{z - y}\,.$$
- *f is strictly concave over I if and only if for every $x, y, z \in I$*
$$x < y < z \quad \Longrightarrow \quad \frac{f(y) - f(x)}{y - x} > \frac{f(z) - f(y)}{z - y}\,.$$

Proof. We will only prove the first characterization. The other characterizations have similar proofs, which are left as exercises.

($\Longrightarrow$) Suppose that f is convex over I. Let $x, y, z \in I$ such that $x < y < z$. By setting $t = \frac{z-y}{z-x}$ into (6.22) we obtain

$$f(y) \leq \frac{z - y}{z - x}\, f(x) + \frac{y - x}{z - x}\, f(z)\,.$$

But this inequality is equivalent to the inequality given in the first characterization.

($\Longleftarrow$) Suppose that f has the first property given in the proposition. Let $x, z \in I$ with $x < z$. Let $t \in (0, 1)$ be arbitrary. Set $y = tx + (1-t)z$. Then $x < y < z$, whereby the inequality given in the first characterization is satisfied. But this inequality is equivalent to

$$f(y) \leq \frac{z-y}{z-x}\, f(x) + \frac{y-x}{z-x}\, f(z)\,,$$

which when expressed in terms of t becomes

$$f\big(tx + (1-t)z\big) \leq t f(x) + (1-t) f(z)\,,$$

where $t \in (0, 1)$ was arbitrary. Therefore f satifies (6.22) and is thereby convex. □

Remark. Many elementary calculus textbooks now use the terms *concave up* for *convex* and *concave down* for *concave*. This terminology was introduced because students had trouble remembering which picture went with convex and which picture went with concave. It became widely used by elementary textbooks in the second half of the twentith century, but advanced textbooks and reseach papers have largely stuck with the traditional terminology used here. Indeed, the term *convex* is used more often in advanced literature than *concave*. You will find many works about *convex sets, convex analysis, convex optimization, convex programming,* and *locally convex topological vector spaces*, but far fewer that use *concave* in a similar way.

Remark. Other elementary calculus textbooks use the terms *convex up* for *convex* and *concave down* for *concave*. Unfortunately, some students still get confused and will say that $f(x) = -x^2$ is convex down. It would be smarter for elementary calculus textbooks to use the terms *convex up* for *convex* and *convex down* for *concave*. There are two reasons for this. First, *convex* is the dominant term they will see in more advanced textbooks, so seeing it in their first course is better than not seeing it. Second, for a convex function the region above its graph (up) is convex while for a concave function the region below its graph (down) is convex.

We now use Proposition 6.8 to show that a function that is either convex or concave over an open interval is continuous over that interval.

Proposition 6.9. *Let $f : D \to \mathbb{R}$ for some $D \subset \mathbb{R}$. If f is convex or concave over $(a, b) \subset D$ then f is continuous over (a, b).*

Proof. We will give the proof only for the case when f is convex over (a, b). Let $x \in (a, b)$. Let $q, r, s, t \in (a, b)$ such that $q < r < x < s < t$. Then by Proposition 6.8 we can show that

$$\frac{f(r) - f(q)}{r - q} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(t) - f(s)}{t - s} \qquad \text{for every } y \in (r, s) \text{ such that } y \neq x\,.$$

From these inequalities we can derive the bound

$$|f(y) - f(x)| \leq L\,|y - x| \qquad \text{for every } y \in (r, s)\,,$$

where

$$L = \max\left\{ \left| \frac{f(r) - f(q)}{r - q} \right|, \left| \frac{f(t) - f(s)}{t - s} \right| \right\}.$$

This bound implies that f is continuous at x. Because $x \in (a, b)$ was arbitrary, we conclude that f is continuous over (a, b). □

Exercise. Show that the assertion of this proposition is false if we replace (a, b) with either $(a, b]$, $[a, b)$, or $[a, b]$.

Proposition 6.8 can also be used to show that a function that is either convex or concave over an open interval is more than continuous over that interval. It is almost differentiable over that interval.

Proposition 6.10. *Let $f : (a, b) \to \mathbb{R}$ be convex or concave over (a, b). Then the one-sided derivatives of f exist and are monotonic over (a, b).*

If f is convex then for every $y \in (a, b)$ the left and right derivatives are given by

$$\begin{aligned} \mathrm{D}_- f(y) &\equiv \lim_{x \to y^-} \frac{f(x) - f(y)}{x - y} = \sup\left\{ \frac{f(x) - f(y)}{x - y} \,:\, x \in (a, y) \right\} , \\ \mathrm{D}_+ f(y) &\equiv \lim_{z \to y^+} \frac{f(z) - f(y)}{z - y} = \inf\left\{ \frac{f(z) - f(y)}{z - y} \,:\, z \in (y, b) \right\} . \end{aligned} \tag{6.24}$$

The functions $\mathrm{D}_- f$ and $\mathrm{D}_+ f$ are nondecreasing over (a, b). For every $x, y, z \in (a, b)$ with $x < y < z$ they satisfy

$$\mathrm{D}_+ f(x) \le \frac{f(x) - f(y)}{x - y} \le \mathrm{D}_- f(y) \le \mathrm{D}_+ f(y) \le \frac{f(z) - f(y)}{z - y} \le \mathrm{D}_- f(z) . \tag{6.25}$$

Moreover, $\mathrm{D}_- f$ is left continuous over (a, b) and $\mathrm{D}_+ f$ is right continuous over (a, b) with

$$\mathrm{D}_- f(y) = \underline{\mathrm{D}_- f}(y) = \underline{\mathrm{D}_+ f}(y) , \qquad \overline{\mathrm{D}_- f}(y) = \overline{\mathrm{D}_+ f}(y) = \mathrm{D}_+ f(y) , \tag{6.26}$$

where $\underline{\mathrm{D}_\pm f}$ and $\overline{\mathrm{D}_\pm f}$ are defined as in Proposition 5.14 for nondecreasing functions.

If f is concave then for every $y \in (a, b)$ the left and right derivatives are given by

$$\begin{aligned} \mathrm{D}_- f(y) &\equiv \lim_{x \to y^-} \frac{f(x) - f(y)}{x - y} = \inf\left\{ \frac{f(x) - f(y)}{x - y} \,:\, x \in (a, y) \right\} , \\ \mathrm{D}_+ f(y) &\equiv \lim_{z \to y^+} \frac{f(z) - f(y)}{z - y} = \sup\left\{ \frac{f(z) - f(y)}{z - y} \,:\, z \in (y, b) \right\} . \end{aligned} \tag{6.27}$$

The functions $\mathrm{D}_- f$ and $\mathrm{D}_+ f$ are nonincreasing over (a, b). For every $x, y, z \in (a, b)$ with $x < y < z$ they satisfy

$$\mathrm{D}_+ f(x) \ge \frac{f(x) - f(y)}{x - y} \ge \mathrm{D}_- f(y) \ge \mathrm{D}_+ f(y) \ge \frac{f(z) - f(y)}{z - y} \ge \mathrm{D}_- f(z) . \tag{6.28}$$

Moreover, $\mathrm{D}_- f$ is left continuous over (a, b) while $\mathrm{D}_+ f$ is right continuous over (a, b) with

$$\mathrm{D}_- f(y) = \overline{\mathrm{D}_- f}(y) = \overline{\mathrm{D}_+ f}(y) , \qquad \underline{\mathrm{D}_- f}(y) = \underline{\mathrm{D}_+ f}(y) = \mathrm{D}_+ f(y) , \tag{6.29}$$

where $\underline{\mathrm{D}_\pm f}$ and $\overline{\mathrm{D}_\pm f}$ are defined as in Proposition 5.14 for nonincreasing functions.

Proof. Exercise.

Remark. This shows that if a function is convex or concave over an open interval then, in addition to being continuous over that interval, it is differentiable where its one-sided derivatives are equal. These one-sided derivatives are monotonic over the interval, so by Proposition 5.15 they are continuous and equal at all but at most a countable number of points in that interval. Therefore the derivative of the function is monotonic over the subset of the interval over which it is defined.

7. Mean-Value Theorems and Their Applications

In this section we study the mean-value theorems of Lagrange and Cauchy. Their proofs rest upon the Extreme-Value Theorem and the Fermat Critical-Point Theorem. Their usefulness will be illustrated by using them to establish the monotonicity tests that you used in calculus, an error bound for the tangent line approximation, a convergence estimate for the Newton method, error bounds for the Taylor approximation, and various l'Hospital rules for evaluating limits of indeterminant form.

7.1. **Lagrange Mean-Value Theorem.** We begin by proving a special case of the Lagrange Mean-Value Theorem, from which the full theorem follows easily. This special case is called the Rolle Theorem. Because it isolates the key step in the proofs of both the Lagrange and Cauchy Mean-Value Theorems, it might be more accurate to call it the Rolle Lemma. However, we will stick with its classical moniker. Its proof simply specializes Lagrange's proof to the special case considered. It rests upon a combination of the Extreme-Value Theorem with the Fermat Critical-Point Theorem.

Proposition 7.1. Rolle Theorem. *Let $a, b \in \mathbb{R}$ such that $a < b$. Let*

- *$f : [a, b] \to \mathbb{R}$ be continuous;*
- *$f(a) = f(b)$;*
- *f be differentiable over (a, b).*

Then $f'(p) = 0$ for some $p \in (a, b)$.

Remark. This result can be motivated by simply graphing any such function and noticing that f' will vanish at points in (a, b) where f takes extreme values. Indeed, this intuition is all that lies behind the proof.

Proof. The Extreme-Value Theorem asserts that there exist points $\underline{p}$ and $\overline{p}$ in $[a, b]$ such that

$$f(\underline{p}) \leq f(x) \leq f(\overline{p}) \quad \text{for every } x \in [a, b] \,.$$

Let $k = f(a) = f(b)$. By setting $x = a$ or $x = b$ above, we see that

$$f(\underline{p}) \leq k \leq f(\overline{p}) \,.$$

At least one of the following three cases must then hold:

- $f(\underline{p}) = k = f(\overline{p})$;
- $f(\underline{p}) < k$;
- $k < f(\overline{p})$.

If $f(\underline{p}) = k = f(\overline{p})$ then $f(x) = k$ over $[a, b]$ and $f'(p) = 0$ for every p in (a, b). If $f(\overline{p}) < k$ then $\underline{p}$ must be in (a, b). But because f is thereby differentiable at $\underline{p}$, the Fermat Critical-Point Theorem then implies that $f'(\underline{p}) = 0$. Finally, the argument when $k < f(\overline{p})$ goes similarly, yielding $f'(\overline{p}) = 0$. At least one such p can therefore be found in each case. □

We are now ready for the full Lagrange Mean-Value Theorem.

Proposition 7.2. Lagrange Mean-Value Theorem. *Let $a, b \in \mathbb{R}$ such that $a < b$. Let*

- *$f : [a, b] \to \mathbb{R}$ be continuous;*
- *f be differentiable over (a, b).*

Then

$$f'(p) = \frac{f(b) - f(a)}{b - a} \quad \textit{for some } p \textit{ in } (a, b) \,.$$

Remark. The geometric interpretation of this theorem is that $(p, f(p))$ is a point on the graph of f where the slope of the tangent line equals the slope of the secant line through the points $(a, f(a))$ and $(b, f(b))$. Notice that this theorem reduces to the Rolle Theorem in the case when $f(a) = f(b)$.

Proof. Define $g : [a, b] \to \mathbb{R}$ for every $x \in [a, b]$ by

$$g(x) \equiv f(x) - f(a) - m\,(x - a)\,, \quad \text{where} \quad m = \frac{f(b) - f(a)}{b - a}\,.$$

Clearly, as a function of x:

- g is continuous over $[a, b]$;
- $g(a) = g(b) = 0$;
- g is differentiable over (a, b) with $g'(x) = f'(x) - m$.

The Rolle Theorem then implies that there exists $p \in (a, b)$ such that $g'(p) = f'(p) - m = 0$. Hence, $f'(p) = m$ for this p. □

7.2. **Lipschitz Bounds.** An easy consequence of the Lagrange Mean-Value Theorem is the existence of so-called Lipschitz bounds for functions with a bounded derivative.

Definition. If $D \subset \mathbb{R}$ then $f : D \to \mathbb{R}$ is said to be Lipschitz continuous over D if there exists a constant L such that

$$|f(x) - f(y)| \leq L|x - y| \qquad \text{for every } x, y \in D\,.$$

Such a bound is called a Lipschitz bound or Lipschitz condition, while L is called a Lipschitz constant.

Proposition 7.3. Lipschitz Bound Theorem. *Let $I \subset \mathbb{R}$ be either (a, b), $[a, b)$, $(a, b]$ or $[a, b]$ for some $a < b$. Let $f : I \to \mathbb{R}$ be continuous over I and differentiable over (a, b). If $f' : (a, b) \to \mathbb{R}$ is bounded then f satifies the Lipschitz bound*

$$|f(x) - f(y)| \leq L|x - y| \qquad \text{for every } x, y \in I\,, \tag{7.1}$$

where $L = \sup\{|f'(z)| \,:\, z \in (a, b)\}$. Moreover, this is the smallest possible Lipschitz constant for f over I.

Proof. Let $x, y \in I$. If $x = y$ then bound (7.1) holds for every $L \geq 0$. If $x < y$ then by the Lagrange Mean-Value Theorem there exists $p \in (x, y)$ such that

$$f'(p) = \frac{f(y) - f(x)}{y - x}\,.$$

It then follows that

$$|f(x) - f(y)| = |f'(p)||x - y| \leq L|x - y|\,.$$

The case when $y < x$ goes similarly. The proof that L is the smallest possible Lipschitz constant for f over I is left as an exercise. □

7.3. **Monotonicity.** We now recall the notions of monotonicity of a function over a set.

Definition 7.1. *Given a function $f : Dom(f) \to \mathbb{R}$ with* $\mathrm{Dom}(f) \subset \mathbb{R}$ *and a set* $S \subset \mathrm{Dom}(f)$*, we say that* f *is*

$$\begin{aligned} \text{increasing over } S \quad & \textit{whenever } f(x) < f(y) \textit{ for every } x, y \in S \textit{ with } x < y\,, \\ \text{nondecreasing over } S \quad & \textit{whenever } f(x) \leq f(y) \textit{ for every } x, y \in S \textit{ with } x < y\,, \\ \text{decreasing over } S \quad & \textit{whenever } f(y) < f(x) \textit{ for every } x, y \in S \textit{ with } x < y\,, \\ \text{nonincreasing over } S \quad & \textit{whenever } f(y) \leq f(x) \textit{ for every } x, y \in S \textit{ with } x < y\,. \end{aligned}$$

We say that f *is* monotonic over S *if it is either nondecreasing or nonincreasing over* S*. We say that* f *is* strictly monotonic over S *if it is either increasing or decreasing over* S*.*

In calculus you learned how to determine the monotonicity of a function through a sign analysis of its first derivative. You probably used the following theorem, which is a consequence of the Lagrange Mean-Value Theorem. Of course, that theorem is a consequence of the Extreme-Value and Fermat Critical-point Theorems.

Proposition 7.4. Monotonicity Theorem. *Let* I *be either* (a, b)*,* $[a, b)$*,* $(a, b]$ *or* $[a, b]$ *for some* $a < b$*. Let* $f : I \to \mathbb{R}$ *be continuous over* I *and differentiable over* (a, b)*. Then*

(i) $f' > 0$ *over* (a, b) $\Longrightarrow$ f *is increasing over* I*;*
(ii) $f' < 0$ *over* (a, b) $\Longrightarrow$ f *is decreasing over* I*;*
(iii) $f' \geq 0$ *over* (a, b) $\Longleftrightarrow$ f *is nondecreasing over* I*;*
(iv) $f' \leq 0$ *over* (a, b) $\Longleftrightarrow$ f *is nonincreasing over* I*;*
(v) $f' = 0$ *over* (a, b) $\Longleftrightarrow$ f *is constant over* I*.*

Proof. ($\Longrightarrow$) We will prove only case (i). The other cases are argued similarly. Suppose that $f' > 0$ over (a, b). Let $x, y \in I$ with $x < y$. The Lagrange Mean-Value Theorem states that there exists a p such that $x < p < y$ and $f(y) - f(x) = f'(p)(y - x)$. Because any such p must lie in (a, b), we must have $f'(p) > 0$, whereby $f(y) - f(x) = f'(p)(y - x) > 0$. By the arbitrariness of x and y, we conclude that f is increasing over I.

($\Longleftarrow$) We will prove only case (iii). Cases (iv) and (v) are argued similarly. Suppose that f is nondecreasing over I. Let $x \in (a, b)$. Because f is nondecreasing over I we know that

$$\frac{f(y) - f(x)}{y - x} \geq 0 \qquad \text{for every } y \in I \text{ with } y \neq x\,.$$

Because f is differentiable at x, the above inequality implies that

$$f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x} \geq 0\,.$$

By the arbitrariness of x, we conclude that $f' \geq 0$ over (a, b). □

Remark. The converses of (i) and (ii) are false because the derivative of a strictly monotonic function can vanish at isolated points. The next proposition will add a hypothesis under which the converses of (i) and (ii) hold.

In practice, you may have also used the following theorem, which is a consequence of both the Lagrange Mean-Value Theorem and the Derivative Sign Dichotomy Theorem. That latter theorem is a consequence of the Derivative Intermediate-Value Theorem, which also follows from the Extreme-Value and Fermat Critical-Point Theorems.

Proposition 7.5. Strict Monotonicity Theorem. *Let I be either (a,b), $[a,b)$, $(a,b]$ or $[a,b]$ for some $a<b$. Let $f : I \to \mathbb{R}$ be continuous over I and differentiable over (a,b). If f has no critical points in (a,b) then the following are equivalent:*

(i) *$f' > 0$ over (a,b);*
(ii) *f is increasing over I;*
(iii) *$f(q) < f(r)$ for some q and r in I with $q<r$;*
(iv) *$f'(p) > 0$ for some p in (a,b).*

Similarly, the following are equivalent:

(i′) *$f' < 0$ over (a,b);*
(ii′) *f is decreasing over I;*
(iii′) *$f(q) > f(r)$ for some q and r in I with $q<r$;*
(iv′) *$f'(p) < 0$ for some p in (a,b).*

Remark. This proposition is usually used by applying criterion (iv) or (iv′) to infer the monotonicity of f over I. In other words, it allows us to read off the monotonicity from by checking the sign of f' at a single point in (a,b).

Proof. We will prove that (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv) $\Longrightarrow$ (i). The proof of the equivalence of (i′-iv′) is similar.

The fact (i) implies (ii) is just the first assertion of the Monotonicity Theorem (Proposition 7.4). It is clear from the definition of "increasing over I" that (ii) implies (iii). Given (iii), the Lagrange Mean-Value Theorem implies there exists $p \in (q,r) \subset (a,b)$ such that

$$f'(p) = \frac{f(r) - f(q)}{r - q} > 0\,.$$

Hence, (iii) implies (iv). Finally, the fact that (iv) implies (i) is a consequence of the Derivative Sign Dichotomy Theorem. $\square$

7.4. **Convexity.** Finding the intervals over which a given function is either convex or concave is called analyzing the *convexity* of the function. In calculus you learned that the convexity of a function is related to the monotonicity of its first derivative. The following characterizations are a consequence of the Lagrange Mean-Value Theorem (Proposition 7.2) and the characterizations of convexity by difference quotients given in Proposition 6.8.

Proposition 7.6. Convexity Characterization Theorem. *Let I be either (a,b), $(a,b]$, $[a,b)$, or $[a,b]$ for some $a<b$. Let $f : I \to \mathbb{R}$ be continuous over I and differentiable over (a,b). Then*

(i) *f' is increasing over (a,b)* $\Longleftrightarrow$ *f is strictly convex over I;*
(ii) *f' is decreasing over (a,b)* $\Longleftrightarrow$ *f is strictly concave over I;*
(iii) *f' is nondecreasing over (a,b)* $\Longleftrightarrow$ *f is convex over I;*
(iv) *f' is nonincreasing over (a,b)* $\Longleftrightarrow$ *f is concave over I;*
(v) *f' is constant over (a,b)* $\Longleftrightarrow$ *f is affine over I.*

Proof. We will prove only characterization (iii). The others are left as an exercise. The proofs of characterizations (iv) and (v) are very similar. The proofs of characterizations (i) and (ii) will require care because strict inequalities must be established.

($\Longrightarrow$) Suppose f' is nondecreasing over (a,b). Let $x,y,z \in I$ with $x < y < z$. By the Lagrange Mean-Value Theorem there exists $p \in (x,y)$ and $q \in (y,z)$ such that

$$\frac{f(y)-f(x)}{y-x} = f'(p)\,, \qquad \frac{f(z)-f(y)}{z-y} = f'(q)\,.$$

Because f' is nondecreasing over (a,b) while $p,q \in (a,b)$ with $p < q$, we see that $f'(p) \le f'(q)$. This implies that

$$\frac{f(y)-f(x)}{y-x} \le \frac{f(z)-f(y)}{z-y}\,.$$

Because this inequality holds for every $x,y,z \in I$ with $x < y < z$, the difference quotient criterion of Proposition 6.8 implies that f is convex over I.

($\Longleftarrow$) Suppose f is convex over I. Let $x,z \in (a,b)$ with $x < z$. By Proposition 6.8

$$\frac{f(y)-f(x)}{y-x} \le \frac{f(z)-f(y)}{z-y} \qquad \text{for every } y \in (x,z)\,.$$

By letting $y \to x^+$ and $y \to z^-$ in this inequality we find that

$$f'(x) = \lim_{y\to x^+} \frac{f(y)-f(x)}{y-x} \le \lim_{y\to x^+} \frac{f(z)-f(y)}{z-y} = \frac{f(z)-f(x)}{z-x}\,,$$

$$\frac{f(z)-f(x)}{z-x} = \lim_{y\to z^-} \frac{f(y)-f(x)}{y-x} \le \lim_{y\to z^-} \frac{f(z)-f(y)}{z-y} = f'(z)\,,$$

whereby $f'(x) \le f'(z)$. Because this holds for every $x,z \in (a,b)$ with $x < z$, we conclude that f' is nondecreasing over (a,b). ☐

In calculus you learned how to determine the convexity of a function through a sign analysis of its second derivative. You probably used the following theorem, which is a consequence of the Monotonicity Theorem (Proposition 7.4) and the Convexity Characterization Theorem (Proposition 7.6).

Proposition 7.7. Convexity Theorem. *Let I be either (a,b), $[a,b)$, $(a,b]$ or $[a,b]$ for some $a < b$. Let $f : I \to \mathbb{R}$ be continuous over I and twice differentiable over (a,b). Then*

(i) $f'' > 0$ *over* (a,b) $\Longrightarrow$ f *is strictly convex over* I;
(ii) $f'' < 0$ *over* (a,b) $\Longrightarrow$ f *is strictly concave over* I;
(iii) $f'' \ge 0$ *over* (a,b) $\Longleftrightarrow$ f *is convex over* I;
(iv) $f'' \le 0$ *over* (a,b) $\Longleftrightarrow$ f *is concave over* I;
(v) $f'' = 0$ *over* (a,b) $\Longleftrightarrow$ f *is affine over* I.

Proof. Because $f' : (a,b) \to \mathbb{R}$ is differentiable over (a,b), it follows from Proposition 7.4 that

(i) $f'' > 0$ over (a,b) $\Longrightarrow$ f' is increasing over (a,b);
(ii) $f'' < 0$ over (a,b) $\Longrightarrow$ f' is decreasing over (a,b);
(iii) $f'' \ge 0$ over (a,b) $\Longleftrightarrow$ f' is nondecreasing over (a,b);
(iv) $f'' \le 0$ over (a,b) $\Longleftrightarrow$ f' is nonincreasing over (a,b);
(v) $f'' = 0$ over (a,b) $\Longleftrightarrow$ f' is constant over (a,b).

The assertions of Proposition 7.7 then follow from Proposition 7.6. ☐

Remark. The converses of (i) and (ii) are false because the second derivative of a function that is strictly convex or strictly concave can vanish at isolated points. The next proposition will add a hypothesis under which the converses of (i) and (ii) hold.

In practice, you may have also used the following theorem, which is a consequence of the Strict Monotonicity Theorem (Proposition 7.5) and the Convexity Characterization Theorem (Proposition 7.6).

Proposition 7.8. Strict Convexity Theorem. *Let I be either (a,b), $[a,b)$, $(a,b]$ or $[a,b]$ for some $a<b$. Let $f : I \to \mathbb{R}$ be continuous over I and twice differentiable over (a,b). If f' has no critical points in (a,b) then the following are equivalent:*

(i) *$f''>0$ over (a,b);*
(ii) *f is strictly convex over I;*
(iii) *f' is increasing over (a,b);*
(iv) *$f'(q)<f'(r)$ for some q and r in (a,b) with $q<r$;*
(v) *$f''(p)>0$ for some p in (a,b).*

Similarly, the following are equivalent:

(i′) *$f''<0$ over (a,b);*
(ii′) *f is strictly concave over I;*
(iii′) *f' is decreasing over (a,b);*
(iv′) *$f'(q)>f'(r)$ for some q and r in (a,b) with $q<r$;*
(v′) *$f''(p)<0$ for some p in (a,b).*

Remark. This proposition is usually used by applying criterion (v) or (v′) to infer the convexity of f over I. In other words, it allows us to read off the convexity from by checking the sign of f'' at a single point in (a,b).

Remark. A critical point of f' is sometimes called a *degenerate point* of f.

Proof. Equivalences (ii) $\iff$ (iii) and (ii′) $\iff$ (iii′) were established by Proposition 7.6. Equivalences (i) $\iff$ (iii) $\iff$ (iv) $\iff$ (v) and (i′) $\iff$ (iii′) $\iff$ (iv′) $\iff$ (v′) follow by applying Proposition 7.5 to f' considered over (a,b). □

7.5. **Error of the Tangent Line Approximation.** Recall that if $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$ is differentiable at $c \in D$ then the tangent line approximation to f and c is given by

$$f(x) \approx f(c) + f'(c)(x-c)\,.$$

For every $x \in D$ we define $R_c f(x)$ by the relation

$$f(x) = f(c) + f'(c)(x-c) + R_c f(x)\,.$$

The function $R_c f : D \to \mathbb{R}$ is called is called the *remainder* or *correction* of the tangent line approximation at c because it is what we add to the approximation to recover the exact value of $f(x)$. It is the negative of the *error*.

It follows from the definition of differentiability that

$$\lim_{x\to c} \frac{R_c f(x)}{x-c} = \lim_{x\to c} \frac{f(x)-f(c)-f'(c)(x-c)}{x-c} = 0\,. \tag{7.2}$$

This states that $|R_c f(x)|$ vanishes faster than $|x-c|$ as x approaches c. This is the best we can expect to say if all we know is that f is differentiable at c. However, if f has more regularity then we can say how much faster $|R_c f(x)|$ vanishes. For example, if f is differentiable over (a,b) and f' satisfies a continuity bound then we have the following result.

Proposition 7.9. Tangent Line Remainder Bound. *Let $f : (a,b) \to \mathbb{R}$ be differentiable. Let $c \in (a,b)$, $\alpha \in (0,\infty)$ and $K \in (0,\infty)$ such that f' satisfies the continuity bound*

$$|f'(z) - f'(c)| \leq K|z-c|^{\alpha} \quad \textit{for every } z \in (a,b)\,. \tag{7.3}$$

Then the remainder of the tangent line approximation satisfies the bound

$$|R_c f(x)| \leq K|x-c|^{1+\alpha} \quad \textit{for every } x \in (a,b)\,. \tag{7.4}$$

Remark. Bound (7.4) implies that $|R_c f(x)|$ vanishes at least as fast as $|x-c|^{1+\alpha}$ as x approaches c, which is certainly faster than $|x-c|$.

Remark. This proposition is usually applied when f'' does not exist at c, in which case $\alpha \in (0,1]$. Later we will find better bounds when $f''(c)$ exists.

Proof. Because $R_c f(c) = 0$, bound (7.4) clearly holds when $x = c$. Let $x \in (a,b)$ with $x \neq c$. By the Lagrange Mean-Value Theorem there exists p between c and x such that $f(x) - f(c) = f'(p)(x-c)$. We can thereby express $R_c f(x)$ as

$$\begin{aligned} R_c f(x) &= f(x) - f(c) - f'(c)(x-c) \\ &= f'(p)(x-c) - f'(c)(x-c) = \big(f'(p) - f'(c)\big)(x-c)\,. \end{aligned}$$

Then by bound (7.3) and the fact that $|p-c| < |x-c|$ we have

$$|R_c f(x)| = |f'(p) - f'(c)|\,|x-c| \leq K|p-c|^{\alpha}|x-c| \leq K|x-c|^{1+\alpha}\,.$$

Therefore bound (7.4) holds for every $x \in (a,b)$. □

Remark. If f is twice differentiable over (a,b) and f'' is bounded over (a,b) then the Lipschitz Bound Theorem (Proposition 7.3) implies that f' satisfies bound (7.3) for every $c \in (a,b)$ with $\alpha = 1$ and

$$K = \sup\{|f''(z)| \,:\, z \in (a,b)\}\,.$$

Proposition 7.9 then implies that $|R_c f(x)|$ is bounded above by $K|x-c|^2$. The following result allows us to obtain a better bound for this case.

Another consequence of the Rolle Theorem (and hence, of the Extreme-Value Theorem) is the following expression for the remainder of the tangent line approximation due to Lagrange.

Proposition 7.10. Lagrange Tangent Line Remainder Theorem. *Let $f : (a,b) \to \mathbb{R}$ be twice differentiable. Let $c \in (a,b)$. Then for every $x \in (a,b)$ such that $x \neq c$ there exists a point p between c and x such that*

$$f(x) = f(c) + f'(c)(x-c) + \tfrac{1}{2} f''(p)\,(x-c)^2\,. \tag{7.5}$$

Remark. For a given c the point p will also depend on x, and this theorem does not give us a clue as to what that dependence might be. However, formula (7.5) does allow us to bound the size of the remainder by bounding the possible values of $f''(p)$. For example, if we can find a number K such that $|f''(z)| \leq K$ for every $z \in (a,b)$, then we see that for every $x \in (a,b)$ we have

$$|R_c f(x)| = |f(x) - f(c) - f'(c)(x-c)| = \tfrac{1}{2}|f''(p)|\,(x-c)^2 \leq \tfrac{1}{2}K\,(x-c)^2\,. \tag{7.6}$$

This bound shows that the remainder vanishes at least as fast as $(x-c)^2$ as x approaches c. This is a stronger statement than (7.2), which only said the remainder vanishes faster than $x-c$ as x approaches c. Moreover, this bound is better than the one obtained from (7.4) with $\alpha = 1$ by the factor $\frac{1}{2}$.

Remark. Formula (7.5) also allows us to determine the sign of the remainder when we know the sign of $f''(p)$. For example, if we know that $f''(z) > 0$ for every $z \in (a,b)$, then we know that the tangent line approximation lies below f.

Remark. Finally, when f'' is continuous at c we can refine (7.2) even further by using (7.5) to show that

$$\lim_{x\to c} \frac{R_c f(x)}{(x-c)^2} = \lim_{x\to c} \frac{f(x) - f(c) - f'(c)(x-c)}{(x-c)^2} = \lim_{x\to c} \tfrac{1}{2} f''(p) = \tfrac{1}{2} f''(c)\,.$$

This limit follows because f'' is continuous at c and because p is trapped between c and x as x approaches c. It shows that when $f''(c) \neq 0$ the remainder vanishes exactly as fast as $(x-c)^2$ as x approaches c, and that when $f''(c) = 0$ it vanishes faster than $(x-c)^2$ as x approaches c.

We now prove the Lagrange Tangent Line Remainder Theorem.

Proof. First consider the case when $c < x < b$. Fix this x and let M be determined by the equation

$$f(x) = f(c) + f'(c)(x-c) + \frac{1}{2} M\,(x-c)^2\,.$$

For each $t \in [c,x]$ define $g(t)$ by

$$g(t) \equiv f(x) - f(t) - f'(t)(x-t) - \frac{1}{2} M\,(x-t)^2\,.$$

Clearly, we see that as a function of t:

- g is continuous over the interval $[c,x]$;
- $g(c) = g(x) = 0$;
- g is differentiable over (c,x) with

$$g'(t) = -f''(t)(x-t) + M\,(x-t) = \big(M - f''(t)\big)\,(x-t)\,.$$

The Rolle Theorem then implies there exists $p \in (c,x)$ such that $g'(p) = 0$. Hence,

$$0 = g'(p) = \big(M - f''(p)\big)\,(x-p)\,,$$

whereby $M = f''(p)$ for some $p \in (c,x)$. The case $a < x < c$ is argued similarly. □

7.6. **Convergence of the Newton Method.** The zeros of a function f are the solutions of the equation $f(x) = 0$. One of the fastest ways to compute the zeros of a differentiable function is the Newton method. It iteratively constructs a sequence $\{x_n\}_{n\in\mathbb{N}}$ of approximate zeros as follows. Given the guess x_n, we let our next guess x_{n+1} be the x-intercept of the tangent line approximation to f at x_n. In other words, we let x_{n+1} be the solution of

$$f(x_n) + f'(x_n)(x - x_n) = 0\,.$$

Provided $f'(x_n) \neq 0$ this can be solved to obtain

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}\,. \tag{7.7}$$

The points so-obtained are called Newton iterates. Of course, they depend on the initial guess x_0. The process will terminate at some n either if $f'(x_n) = 0$ or if x_{n+1} given by (7.7) lies outside the domain of f. Otherwise it produces a sequence of iterates $\{x_n\}_{n\in\mathbb{N}}$ which may or may not converge.

The Newton method works best if a single root has been isolated in an interval without critical points. Some bounds on the error made by the iterates can then be obtained by analyzing the convexity of f near the root. For example, if we denoted the root by x_* then we can see the following.

- If f is increasing and strictly convex near x_*, or is decreasing and strictly concave near x_*, then the sequence $\{x_n\}$ will approach x_* monotonically from above.
- If f is increasing and strictly concave near x_*, or is decreasing and strictly convex near x_*, then the sequence $\{x_n\}$ will approach x_* monotonically from below.

These observations can be expressed as follows.

- If $f'(x_*)f''(x_*) > 0$ then the sequence $\{x_n\}$ will approach x_* monotonically from above.
- If $f'(x_*)f''(x_*) < 0$ then the sequence $\{x_n\}$ will approach x_* monotonically from below.

Hence, the sequence $\{x_n\}$ will always approach x_* from the side on which $f(x)f''(x) > 0$. If we take our initial guess x_0 on this side the sequence $\{x_n\}$ will be strictly monotonic. It will converge very quickly, eventually doubling the number of correct digits with each new iterate. This fast rate of convergence is governed by the following theorem.

Proposition 7.11. Newton Method Convergence Theorem. *Let $f : [a, b] \to \mathbb{R}$ be twice differentiable over $[a, b]$. Let $f(a)f(b) < 0$. Let L and M be positive constants such that*

- *$L \leq |f'(z)|$ for every $z \in (a, b)$;*
- *$|f''(z)| \leq M < \infty$ for every $z \in (a, b)$;*
- *$b - a < 2L/M$.*

Let $\{x_n\}_{n\in\mathbb{N}}$ be any sequence of Newton iterates that lies within $[a, b]$. Then f has a unique zero $x_* \in (a, b)$ and the Newton iterates satisfy

$$|x_n - x_*| \leq \frac{1}{K}\big(K\,|x_0 - x_*|\big)^{2^n} < \frac{1}{K}\big(K(b-a)\big)^{2^n}, \tag{7.8}$$

where $K = M/(2L)$, so that $K(b - a) < 1$.

Proof. Because $f(a)f(b) < 0$ and f is continuous over $[a, b]$, f must have a zero in (a, b) by the Intermediate-Value Theorem. Because $L \leq |f'(z)|$ for every $z \in (a, b)$, f has no critical points in (a, b), and is thereby strictly montonic over $[a, b]$. It must therefore have a unique zero in (a, b). Let x_* denote this zero.

By (7.7) the Newton iterates satisfy

$$0 = f(x_n) + f'(x_n)(x_{n+1} - x_n)\,.$$

On the other hand, the Lagrange Tangent Line Remainder Theorem states that

$$0 = f(x_*) = f(x_n) + f'(x_n)(x_* - x_n) + \frac{1}{2}f''(p_n)(x_* - x_n)^2\,,$$

for some p_n between x_* and x_n. Subtracting this from the previous equation yields

$$f'(x_n)(x_{n+1} - x_*) = \frac{1}{2}f''(p_n)(x_* - x_n)^2\,.$$

Hence, because x_n and p_n are in (a, b), we have

$$|x_{n+1} - x_*| = \frac{|f''(p_n)|}{2|f'(x_n)|}\,(x_* - x_n)^2 \leq \frac{M}{2L}\,|x_n - x_*|^2 = K|x_n - x_*|^2\,.$$

If we set $R_n = K|x_n - x_*|$ then the above inequality takes the form $R_{n+1} \leq R_n^2$. We can easily use induction to show that $R_n \leq R_0^{2^n}$. Bound (7.8) then follows because $R_0 = K|x_0 - x_*| < K(b - a)$. ∎

Remark. The proof actually shows that once $K|x_n - x_*| < .1$ for some n then $K|x_{n+2} - x_*| < .0001$, $K|x_{n+3} - x_*| < .00000001$, and $K|x_{n+4} - x_*| < .0000000000000001$. This means that once we have an iterate for which Kx_n is correct to within one decimal point, it will be correct to within machine round-off in three or four iterations.

7.7. **Error of the Taylor Polynomial Approximation.** Recall that if $f : (a,b) \to \mathbb{R}$ is n times differentiable at a point $c \in (a,b)$ then the n^{th} order Taylor approximation to $f(x)$ at c is given by the polynomial

$$\begin{aligned} T_c^n f(x) &\equiv f(c) + f'(c)(x-c) + \frac{1}{2}f''(c)(x-c)^2 + \cdots + \frac{1}{n!}f^{(n)}(c)(x-c)^n \\ &= \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(c)(x-c)^k \,. \end{aligned} \tag{7.9}$$

For every $x \in (a,b)$ we define $R_c^n f(x)$ by the relation

$$f(x) = T_c^n f(x) + R_c^n f(x)\,.$$

The function $R_c^n f : (a,b) \to \mathbb{R}$ is called is called the *remainder* or *correction* of the Taylor approximation at c because it is what we add to the approximation to recover the exact value of $f(x)$. It is the negative of the *error*.

The method used to prove the Lagrange Tangent Line Remainder Theorem can be extended to yield an expression for the remainder of the Taylor polynomial approximation.

Proposition 7.12. Lagrange Remainder Theorem. *Let $f : (a,b) \to \mathbb{R}$ be $(n+1)$ times differentiable. Let $c \in (a,b)$. Let $T_c^n f(x)$ denote the n^{th} order Taylor approximation to f at c. Then for every $x \in (a,b)$ such that $x \neq c$ there exists a point p between c and x such that*

$$f(x) = T_c^n f(x) + \frac{1}{(n+1)!} f^{(n+1)}(p)(x-c)^{n+1}\,. \tag{7.10}$$

Remark. This formula is easy to remember because it has the same form as the new term that would appear in the $(n+1)^{st}$ order Taylor polynomial (7.9) except that instead of $f^{(n+1)}$ being evaluated at c, it is being evaluated at an unspecified point p that lies between c and x.

Proof. This proof is built upon the observation is that $T_t^n f(x)$ is a differentiable function of t over (a,b) with (notice the telescoping sum)

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} T_t^n f(x) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(f(t) + \sum_{k=1}^{n} \frac{(x-t)^k}{k!} f^{(k)}(t) \right) \\ &= f'(t) + \sum_{k=1}^{n} \left(\frac{(x-t)^k}{k!} f^{(k+1)}(t) - \frac{(x-t)^{(k-1)}}{(k-1)!} f^{(k)}(t) \right) \\ &= \frac{(x-t)^n}{n!} f^{(n+1)}(t)\,. \end{aligned} \tag{7.11}$$

First consider the case when $c < x < b$. Fix this x and let M be determined by the relation

$$f(x) = T_c^n f(x) + \frac{1}{(n+1)!} M\,(x-c)^{n+1}\,.$$

Define $g(t)$ for every $t \in [c, x]$ by

$$g(t) \equiv f(x) - T_t^n f(x) - \frac{1}{(n+1)!} M\,(x-t)^{n+1}\,.$$

Clearly, as a function of t,

- g is continuous over $[c, x]$;
- $g(c) = g(x) = 0$;
- g is differentiable over (c, x) with

$$g'(t) = -\frac{1}{n!} f^{(n+1)}(t)(x-t)^n + \frac{1}{n!} M\,(x-t)^n = \frac{1}{n!}\big(M - f^{(n+1)}(t)\big)\,(x-t)^n\,.$$

The Rolle Theorem then implies that $g'(p) = 0$ for some p in (c, x). Hence,

$$g'(p) = \frac{1}{n!}\big(M - f^{(n+1)}(p)\big)\,(x-p)^n = 0\,,$$

whereby $M = f^{(n+1)}(p)$ for some p in (c, x). The case $a < x < c$ is argued similarly. □

Remark. Notice the similarity of the foregoing proof with the argument used to establish the Lagrange Tangent Line Remainder Theorem. Indeed, the earlier proof is contained within this one.

There are two common approaches towards using the Lagrange Remainder Theorem to obtain bounds on the Taylor remainder $R_c^n f(x)$. The most common is the first, which is used whenever f is $(n+1)$ times differentiable over (a, b) and $f^{(n+1)}$ can be bounded over (a, b). In this case we express $R_c^n f(x)$ by the formula

$$R_c^n f(x) = \frac{1}{(n+1)!} f^{(n+1)}(p)(x-c)^{n+1} \quad \text{for some } p \text{ between } c \text{ and } x\,. \tag{7.12}$$

For a given c the point p will also depend on both x and n, and this formula does not give us a clue as to what those dependences might be. However, it does allow us to bound the size of the error by bounding the possible values of $f^{(n+1)}(p)$. For example, if we can find $K \in (0, \infty)$ such that $f^{(n+1)}$ satisfies the bound

$$|f^{(n+1)}(z)| < K \quad \text{for every } z \in (a, b)\,,$$

then we see from (7.12) that

$$\big|f(x) - T_c^n f(x)\big| \leq \frac{1}{(n+1)!}\, K\,|x-c|^{n+1}\,. \tag{7.13}$$

Given f and (a, b), the K that appears above depends only on n. Formula (7.12) also allows us to determine the sign of the remainder when $n+1$ is even and we know the sign of $f^{(n+1)}(p)$.

The second common approach towards using the Lagrange Remainder Theorem to obtain bounds on the Taylor remainder $R_c^n f(x)$ is used whenever f is n times differentiable over (a, b) and $f^{(n+1)}$ does not exist at c. In this case we express $R_c^n f(x)$ by the formula

$$R_c^n f(x) = \frac{1}{n!}\big(f^{(n)}(p) - f^{(n)}(c)\big)(x-c)^n \quad \text{for some } p \text{ between } c \text{ and } x\,. \tag{7.14}$$

For a given c the point p will also depend on both x and n, and this formula does not give us a clue as to what those dependences might be. However, it does allow us to bound the size of the error by bounding the possible values of $f^{(n)}(p) - f^{(n)}(c)$. For example, if we can find $\alpha \in (0,1]$ and $K \in (0,\infty)$ such that $f^{(n)}$ satisfies the continuity bound

$$\left|f^{(n)}(z) - f^{(n)}(c)\right| \leq K\,|z-c|^{\alpha} \quad \text{for every } z \in (a,b)\,,$$

then because $|p-c| \leq |x-c|$ we see from (7.14) that

$$\left|f(x) - T_c^n f(x)\right| \leq \frac{1}{n!}\,K\,|x-c|^{n+\alpha}\,. \tag{7.15}$$

Given f, (a,b), and c, the K that appears above depends only on n.

Exercise. Prove the foregoing assertion.

For some functions the first approach can be used to prove that the Taylor series of the function at a point c converges to the function in an interval about c.

Example. We can use the Lagrange Remainder Theorem to prove that

$$e^x = \sum_{k=0}^{\infty} \frac{1}{k!}\,x^k \qquad \text{for every } x \in \mathbb{R}\,.$$

The fact that this series is absoluely convergent for every $x \in \mathbb{R}$ can be seen from any one of the tests for absolute convergence of a series, for example, from the ratio test. However, none of these tests can show that the series converges to e^x!

Let $f(x) = e^x$. Because $f^{(k)}(x) = e^x$ for every $k \in \mathbb{N}$, we see that $f^{(k)}(0) = 1$ for every $k \in \mathbb{N}$. The n^{th} order Taylor approximation to e^x at 0 is thereby given by the polynomial

$$T_0^n f(x) = \sum_{k=0}^{n} \frac{1}{k!}\,x^k\,.$$

These are the partial sums of the infinite sum we are considering. They will converge to $f(x) = e^x$ provided the remainder vanishes as $n \to \infty$.

The Lagrange Remainder Theorem implies that for every $x \neq 0$ there exists a p between 0 and x such that

$$\left|f(x) - T_0^n f(x)\right| = \frac{1}{(n+1)!}\,e^p |x|^{n+1}\,.$$

Because $p \in (-|x|, |x|)$ and because $x \mapsto e^x$ is increasing, we know that $e^p < e^{|x|}$, whereby

$$\left|f(x) - T_0^n f(x)\right| \leq \frac{1}{(n+1)!}\,e^{|x|} |x|^{n+1}\,.$$

This bound also holds when $x = 0$, so it holds for every $x \in \mathbb{R}$. For every $x \in \mathbb{R}$ we can show

$$\lim_{n\to\infty} \frac{1}{(n+1)!}\,e^{|x|} |x|^{n+1} = 0\,.$$

We thereby conclude the series converges to $f(x) = e^x$ for every $x \in \mathbb{R}$.

Exercise. Prove that for every $x \in \mathbb{R}$ we have

$$\cos(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}\,.$$

7.8. **Cauchy Mean-Value Theorem.** We now present a useful extension of the Lagrange Mean-Value Theorem that is attributed to Cauchy. It also is a consequence of the Rolle Theorem (and hence, of the Extreme-Value Theorem).

Proposition 7.13. Cauchy Mean-Value Theorem. *Let* $f : [a, b] \to \mathbb{R}$ *and* $g : [a, b] \to \mathbb{R}$ *be continuous over* $[a, b]$ *and differentiable over* (a, b)*. Then for some* $p \in (a, b)$ *we have*

$$\big(f(b) - f(a)\big)g'(p) = \big(g(b) - g(a)\big)f'(p)\,. \tag{7.16}$$

If moreover $g'(x) \neq 0$ *for every* $x \in (a, b)$ *then*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(p)}{g'(p)}\,. \tag{7.17}$$

Remark. This theorem does not have a geometric interpretation as simple as the tangent line interpretation of the Lagrange Mean-Value Theorem. Of course, it reduces to that theorem when $g(x) = x$.

Remark. This theorem does not follow by simply applying the Lagrange Mean-Value Theorem separately to f and g. That would yeild a $p \in (a, b)$ such that $f(b) - f(a) = f'(p)(b - a)$ and a $q \in (a, b)$ such that $g(b) - g(a) = g'(q)(b - a)$, which leads to

$$\big(f(b) - f(a)\big)g'(q) = \big(g(b) - g(a)\big)f'(p)\,.$$

However, p and q produced by this argument will not generally be equal. The fact that f' and g' are evaluated at the same point in (7.16) gives the Cauchy Mean-Value Theorem its power.

Proof. For every $x \in [a, b]$ define $h(x)$ by

$$h(x) = \big(f(b) - f(a)\big)g(x) - \big(g(b) - g(a)\big)f(x)\,.$$

Clearly,

- h is continuous over $[a, b]$;
- $h(a) = h(b) = f(b)g(a) - g(b)f(a)$;
- h is differentiable over (a, b) with

$$h'(x) = \big(f(b) - f(a)\big)g'(x) - \big(g(b) - g(a)\big)f'(x)\,.$$

The Rolle Theorem then implies that there exists $p \in (a, b)$ such that $h'(p) = 0$. Upon using the above expression for $h'(x)$, we see that equation (7.16) holds for this p.

Now assume that $g'(x) \neq 0$ for every $x \in (a, b)$. Notice that equation (7.17) follows directly from (7.16) provided there is no division by zero. By the Derivative Sign Dichotomy Theorem, either $g' > 0$ or $g' < 0$ over (a, b). By the Monotonicity Theorem g is strictly monotonic over (a, b). Hence, $g(b) - g(a) \neq 0$. ∎

Here is an alternative proof of the Lagrange Remainder Theorem (Proposition 7.12) that is based on the Cauchy Mean-Value Theorem. Some students find this proof easier to understand than the one based on the observation (7.11) that we gave earlier.

Proof. Define $F : (a, b) \to \mathbb{R}$ and $G : (a, b) \to \mathbb{R}$ for every $x \in (a, b)$ by

$$F(x) = f(x) - T_c^n f(x) \,, \qquad G(x) = \frac{1}{(n+1)!}(x-c)^{n+1} \,.$$

Clearly F and G are $(n+1)$ times differentiable over (a, b) with

$$F^{(k)}(c) = 0 \quad \text{and} \quad G^{(k)}(c) = 0 \quad \text{for every } k = 0, 1, \cdots, n \,,$$

and with

$$F^{(n+1)}(x) = f^{(n+1)}(x) \,, \qquad G^{(n+1)}(x) = 1 \,.$$

It is also clear that $G^{(k)}(x) \neq 0$ for every $x \neq c$ and every $k = 0, 1, \cdots, n+1$.

First consider the case $c < x < b$. By the Cauchy Mean-Value Theorem there exists $p_1 \in (c, x)$ such that

$$\frac{F(x)}{G(x)} = \frac{F(x) - F(c)}{G(x) - G(c)} = \frac{F'(p_1)}{G'(p_1)} \,.$$

By the Cauchy Mean-Value Theorem there exists $p_2 \in (c, p_1)$ such that

$$\frac{F(x)}{G(x)} = \frac{F'(p_1)}{G'(p_1)} = \frac{F'(p_1) - F'(c)}{G'(p_1) - G'(c)} = \frac{F''(p_2)}{G'(p_2)} \,.$$

After repeating this argument $n+1$ times, we obtain a set of points $\{p_k\}_{k=1}^{n+1}$ such that

$$c < p_{n+1} < p_n < \cdots < p_2 < p_1 < x \,,$$

and

$$\frac{F(x)}{G(x)} = \frac{F'(p_1)}{G'(p_1)} = \frac{F''(p_2)}{G''(p_2)} = \quad \cdots \quad = \frac{F^{(n)}(p_n)}{G^{(n)}(p_n)} = \frac{F^{(n+1)}(p_{n+1})}{G^{(n+1)}(p_{n+1})} = f^{(n+1)}(p_{n+1}) \,.$$

Upon setting $p = p_{n+1}$, we obtain $F(x) = f^{(n+1)}(p)G(x)$ for some $p \in (c, x)$, which is the desired result. The case $a < x < c$ is argued similarly. $\square$

Remark. Our earlier proof is appealing because it requires only one application of the Lagrange Mean-Value Theorem rather than $n+1$ applications of the Cauchy Mean-Value Theorem. However, this proof is appealing because it does not require the insight of observation (7.11).

7.9. **l'Hospital Rule.** The most important application of the Cauchy Mean-Value Theorem is to the proof of the l'Hospital rule.

Proposition 7.14. l'Hospital Rule Theorem. *Let $f : (a, b) \to \mathbb{R}$ and $g : (a, b) \to \mathbb{R}$ be differentiable with $g'(x) \neq 0$ for every $x \in (a, b)$. Suppose either that*

$$\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0 \,, \tag{7.18}$$

or that

$$\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = \infty \,. \tag{7.19}$$

If

$$\lim_{x \to a} \frac{f'(x)}{g'(x)} = L \qquad \textit{for some } L \in \mathbb{R}_{\text{ex}} \,, \tag{7.20}$$

then

$$\lim_{x \to a} \frac{f(x)}{g(x)} = L \,. \tag{7.21}$$

Remark. The theorem is stated for the right-sided limit $\lim_{x \to a}$. Of course, the theorem also holds for the left-sided limit $\lim_{x \to b}$. In particular, the theorem statement includes the limit $\lim_{x \to -\infty}$ when $a = -\infty$, and the theorem also holds for the limit $\lim_{x \to \infty}$. We can apply the l'Hospital rule to any two-sided limit by thinking of it as two one-sided limits.

Proof. We will give the proof for the case $L \in \mathbb{R}$. The cases $L = \pm\infty$ are left as an exercise. The proof will be given so that it covers the cases $a \in \mathbb{R}$ and $a = -\infty$ at the same time.

First suppose that f and g satisfy (7.18). Let $\epsilon > 0$. By (7.20) there exists $d_\epsilon \in (a, b)$ such that

$$a < x < d_\epsilon \quad \Longrightarrow \quad \left| \frac{f'(x)}{g'(x)} - L \right| < \frac{\epsilon}{2} \,.$$

For every $x, y \in (a, d_\epsilon)$ with $y < x$ the Cauchy Mean-Value Theorem implies there exists $p \in (y, x)$ such that.

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(p)}{g'(p)} \,.$$

Because $p \in (y, x) \subset (a, d_\epsilon)$, it follows that

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

Hence, we have shown that

$$a < y < x < d_\epsilon \quad \Longrightarrow \quad \left| \frac{f(x) - f(y)}{g(x) - g(y)} - L \right| < \frac{\epsilon}{2} \,.$$

Upon taking the limit of the last inequality above as y approaches a while using the fact that f and g satisfy (7.18), we see that

$$a < x < d_\epsilon \quad \Longrightarrow \quad \left| \frac{f(x)}{g(x)} - L \right| \leq \frac{\epsilon}{2} < \epsilon \,.$$

Hence, the limit (7.21) holds.

Now suppose that f and g satisfy (7.19). Let $\epsilon > 0$. By (7.20) there exists $d_\epsilon \in (a, b)$ such that

$$a < x < d_\epsilon \quad \Longrightarrow \quad \left| \frac{f'(x)}{g'(x)} - L \right| < \frac{\epsilon}{2} \,.$$

Because f and g satisfy (7.19) we may assume that

$$a < x < d_\epsilon \quad \Longrightarrow \quad f(x) > 0 \,, \quad g(x) > 0 \,.$$

Here we fix $y \in (a, d_\epsilon)$. For every $x \in (a, y)$ the Cauchy Mean-Value Theorem implies there exists $p \in (x, y)$ such that

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(p)}{g'(p)} \,. \tag{7.22}$$

The idea is now to rewrite the above relation as

$$\frac{f(x)}{g(x)} = \frac{f'(p)}{g'(p)} \frac{1 - \frac{g(y)}{g(x)}}{1 - \frac{f(y)}{f(x)}} \,,$$

and to argue that the first factor on the right-hand side is near L while the second can be made near enough to 1 as x approaches a.

Let $r(x)$ denote this second factor — specifically, let

$$r(x) = \frac{1 - \frac{g(y)}{g(x)}}{1 - \frac{f(y)}{f(x)}} \,.$$

Because

$$\lim_{x \to a} \frac{f(y)}{f(x)} = \lim_{x \to a} \frac{g(y)}{g(x)} = 0 \,,$$

for any $\eta_\epsilon > 0$ (to be chosen) there exists $c_\epsilon \in (a, y)$ such that

$$a < x < c_\epsilon \quad \Longrightarrow \quad 0 < \frac{f(y)}{f(x)} < \eta_\epsilon \,, \quad 0 < \frac{g(y)}{g(x)} < \eta_\epsilon \,.$$

Provided $\eta_\epsilon < 1$, for every $x \in (a, c_\epsilon)$ we have the bounds

$$r(x) < \frac{1}{1 - \eta_\epsilon} \,, \qquad |1 - r(x)| < \frac{\eta_\epsilon}{1 - \eta_\epsilon} \,,$$

whereby for every $x \in (a, c_\epsilon)$ we have the bound

$$\left| \frac{f(x)}{g(x)} - L \right| = \left| \frac{f'(p)}{g'(p)} r(x) - L \right| \leq \left| \frac{f'(p)}{g'(p)} - L \right| r(x) + |L| |1 - r(x)| < \frac{\epsilon}{2} \frac{1}{1 - \eta_\epsilon} + \frac{|L| \, \eta_\epsilon}{1 - \eta_\epsilon} \,.$$

A short calculation shows that the right-hand side above becomes ϵ if we choose $\eta_\epsilon = \frac{1}{2}\epsilon/(\epsilon + |L|)$. We thereby see that

$$a < x < c_\epsilon \quad \Longrightarrow \quad \left| \frac{f(x)}{g(x)} - L \right| < \epsilon \,.$$

Hence, the limit (7.21) holds. □

Exercise. Prove the l'Hospital Theorem for the case $L = \infty$ when f and g satisfy (7.18).

Exercise. Prove the l'Hospital Theorem for the case $L = \infty$ when f and g satisfy (7.19).

An nice application of the l'Hospital rule is the following.

Proposition 7.15. Taylor Polynomial Approximation Theorem. *Let $f : (a, b) \to \mathbb{R}$ be $(n-1)$ times differentiable over (a, b) for some $n \in \mathbb{Z}_+$. Let $c \in (a, b)$ and let $f^{(n-1)}$ be differentiable at c. Let $T_c^n f(x)$ denote the n^{th} order Taylor approximation to f at c. Then*

$$\lim_{x \to c} \frac{f(x) - T_c^n f(x)}{(x-c)^n} = 0 \,.$$

Remark. This proposition states that the n^{th} order Taylor remainder vanishes faster than $(x-c)^n$ as x approaches c. Of course, if f was $(n+1)$ times differentiable then the Lagrange Remainder Theorem would imply that this remainder vanishes at least as fast as $(x-c)^{n+1}$ as x approaches c. However, here we are assuming that $f^{(n)}$ exists only at c and nowhere else, so we cannot take this approach. Rather, we will apply the l'Hospital rule $(n-1)$ times.

Proof. Define $F : (a, b) \to \mathbb{R}$ and $G : (a, b) \to \mathbb{R}$ by

$$F(x) = f(x) - T_c^{(n-1)} f(x) \,, \qquad G(x) = \frac{1}{n!}(x-c)^n \,.$$

Clearly these functions are $(n-1)$ times differentiable over (a, b) with $F^{(k)}(c) = G^{(k)}(c) = 0$ for every $k = 0, 1, \ldots, n-1$. Because $G^{(k)}(x) \neq 0$ for every $x \neq c$ and every $k = 0, 1, \ldots, n-1$, we can apply the l'Hospital rule $(n-1)$ times to obtain

$$\lim_{x \to c} \frac{F(x)}{G(x)} = \lim_{x \to c} \frac{F'(x)}{G'(x)} = \cdots = \lim_{x \to c} \frac{F^{(n-1)}(x)}{G^{(n-1)}(x)} \,.$$

Because

$$F^{(n-1)}(x) = f^{(n-1)}(x) - f^{(n-1)}(c) \,, \qquad G^{(n-1)}(x) = x - c \,,$$

and because $f^{(n-1)}$ is differentiable at c we know that

$$\lim_{x \to c} \frac{F^{(n-1)}(x)}{G^{(n-1)}(x)} = \lim_{x \to c} \frac{f^{(n-1)}(x) - f^{(n-1)}(c)}{x - c} = f^{(n)}(c) \,.$$

It follows that

$$\lim_{x \to c} \frac{F(x)}{G(x)} = \lim_{x \to c} \frac{F^{(n-1)}(x)}{G^{(n-1)}(x)} = f^{(n)}(c) \,.$$

But this implies that

$$\lim_{x \to c} \frac{F(x) - f^{(n)}(c) G(x)}{G(x)} = 0 \,.$$

The result follows because $f(x) - T_c^n f(x) = F(x) - f^{(n)}(c) G(x)$ while $(x-c)^n = n! \, G(x)$. □

Remark. The l'Hospital rule given by Proposition 7.14 is not the rule given by l'Hospital in the first calculus text. His rule was stated for functions f and g that have formal Taylor expansions centered at some point $c \in \mathbb{R}$. If the first nonzero terms in these expansions appear at orders m and n respectively, then his rule was more like the recipe

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \begin{cases} 0 & \text{when } m > n \,, \\ \dfrac{f^{(m)}(c)}{g^{(m)}(c)} & \text{when } m = n \,, \\ \text{undefined} & \text{when } m < n \,. \end{cases}$$

8. Cauchy Continuity and Uniform Continuity

Here we introduce two regularity notions that are stronger than continuity but weaker than differentability — namely, Cauchy continuity and uniform continuity. These notions are closely related and play a central role in analysis.

8.1. Cauchy Continuity.

This regularity notion was not introduced by Cauchy, but is given that name because it relates to Cauchy sequences.

Definition 8.1. *Let $D \subset \mathbb{R}$. A function $f : D \to \mathbb{R}$ is said to be* Cauchy continuous *over D when every Cauchy sequence $\{x_n\}_{n \in \mathbb{N}}$ that lies in D has an image $\{f(x_n)\}_{n \in \mathbb{N}}$ that is also a Cauchy sequence.*

Remark. This says that a Cauchy continuous function maps Cauchy sequences in its domain to Cauchy sequences. By the Cauchy Criterion Theorem, a sequence in $\mathbb{R}$ is Cauchy if and only if it is convergent. In the setting of $\mathbb{R}$ we could have replaced the word "Cauchy" in the above definition with the word "convergent" without changing the meaning of being Cauchy convergent. However, in more general settings not every Cauchy sequence is convergent, so such a replacement would change the meaning. The wording of the definition used above will carry over into these more general settings.

Remark. There is a very important difference between continuity and Cauchy continuity. Continuity is defined to be a property of a function at a point. A function is then said to be continuous over a set if it is continuous at each point in the set. Cauchy continuity is defined to be a property of a function over a set. It makes no sense to talk about a function being Cauchy continuous at a single point.

We now give three propositions that relate Cauchy continuity to continuity. The first states that Cauchy continuity implies continuity — i.e. that it is stronger than continuity.

Proposition 8.1. *Let $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$ be Cauchy continuous over D. Then f is continuous over D.*

Proof. Let $x \in D$ be arbitrary. We will use the sequence characterization of continuity to show that f is continuous at x. Let $\{x_n\} \subset D$ be arbitrary. We must show that

$$\lim_{n\to\infty} x_n = x \qquad \Longrightarrow \qquad \lim_{n\to\infty} f(x_n) = f(x)\,.$$

Let $\{x_n\}$ converge to x. Then $\{x_n\}$ is Cauchy. Because f is Cauchy continuous $\{f(x_n)\}$ is also Cauchy, and therefore convergent. Let

$$\lim_{n\to\infty} f(x_n) = L\,.$$

We must show that $L = f(x)$. To do this we construct a new sequence $\{y_n\}$ by setting

$$y_n = \begin{cases} x_n & \text{for } n \text{ even}\,, \\ x & \text{for } n \text{ odd}\,. \end{cases}$$

It is easy to show that $\{y_n\}$ converges to x, and is thereby Cauchy. Because f is Cauchy continuous $\{f(y_n)\}$ is also Cauchy, and therefore convergent. Because every subsequence of a convergent sequence will converge to the same limit, it follows that

$$L = \lim_{k\to\infty} f(x_{2k}) = \lim_{k\to\infty} f(y_{2k}) = \lim_{k\to\infty} f(y_{2k+1}) = f(x)\,.$$

Therefore f is continuous at x. But $x \in D$ was arbitrary, so f is continuous. □

Exercise. Show that the sequence $\{y_n\}$ defined in the above proof converges to x.

Our second proposition states that continuity and Cauchy continuity are equivalent over closed domains — i.e. over closed domains continuous functions are Cauchy continuous.

Proposition 8.2. *Let $D \subset \mathbb{R}$ be closed. Let $f : D \to \mathbb{R}$ be continuous over D. Then f is Cauchy continuous over D.*

Proof. Let $\{x_n\} \subset D$ be Cauchy. We must show that $\{f(x_n)\}$ is Cauchy. Because $\{x_n\}$ is Cauchy, it is convegent. Let x be its limit. Because $\{x_n\} \subset D$ and D is closed, we see that $x \in D$. Because $\{x_n\} \subset D$ converges to x while f is continuous at x, it follows that $\{f(x_n)\}$ converges to $f(x)$. Therefore $\{f(x_n)\}$ is Cauchy. □

Our third proposition states that over every domain that is not closed there exists a continuous function that is not Cauchy continuous. In other words, over domains that are not closed there are more continuous functions than Cauchy continuous functions. In particular, it shows that the hypothesis that D is closed in Proposition 8.2 was necessary.

Proposition 8.3. *Let $D \subset \mathbb{R}$. If D is not closed then there exists a function $f : D \to \mathbb{R}$ that is continuous over D, but that is not Cauchy continuous over D.*

Proof. Because D is not closed there exists a limit point x_* of D that is not in D. Consider the function $f : D \to \mathbb{R}$ defined for every $x \in D$ by $f(x) = 1/(x - x_*)$. It should be clear to you that this function is continuous over D. We will show that it is not Cauchy continuous over D. Because $x_* \in D^c$ there exists a sequence $\{x_n\} \subset D$ such that $\{x_n\}$ converges to x_*. Because $\{x_n\}$ converges to x_* while $f(x_n) = 1/(x_n - x_*)$, it follows that

$$\{x_n\} \text{ is convergent} \quad \text{while} \quad \{f(x_n)\} \text{ is divergent}\,.$$

Hence,

$$\{x_n\} \text{ is Cauchy} \quad \text{while} \quad \{f(x_n)\} \text{ is not Cauchy}\,.$$

Therefore f is not Cauchy continuous over D. □

8.2. **Uniform Continuity.** Here we introduce uniform continuity in the context of real-valued functions with domains in $\mathbb{R}$.

Definition 8.2. *Let $D \subset \mathbb{R}$. A function $f : D \to \mathbb{R}$ is said to be* uniformly continuous *over D when for every $\epsilon > 0$ there exists $\delta > 0$ such that for every $x, y \in D$ we have*

$$|x - y| < \delta \qquad \Longrightarrow \qquad |f(x) - f(y)| < \epsilon\,.$$

Remark. This is a stronger regularity concept than that of continuity over D. Indeed, a function $f : D \to \mathbb{R}$ is continuous over D when for every $y \in D$ and every $\epsilon > 0$ there exists $\delta > 0$ such that for every $x \in D$ we have

$$|x - y| < \delta \qquad \Longrightarrow \qquad |f(x) - f(y)| < \epsilon\,.$$

Here δ depends on y and ϵ ($\delta = \delta_{y,\epsilon}$), while in Definition 8.2 of uniform continuity δ depends only on ϵ ($\delta = \delta_\epsilon$). In other words, when f is uniformly continuous over D a δ_ϵ can be found that works uniformly for every $y \in D$ — hence, the terminology.

Remark. There is a very important difference between continuity and uniform continuity. Continuity is defined to be a property of a function at a point. A function is then said to be continuous over a set if it is continuous at each point in the set. Uniform continuity is defined to be a property of a function over a set. It makes no sense to talk about a function being uniformly continuous at a single point. In this sense unifrom continuity and Cauchy continuity are similar. We will soon see that they are closely related.

By the first remark above we know that if $f : D \to \mathbb{R}$ is uniformly continuous over D then it is continuous over D. The following shows that more is true.

Proposition 8.4. *Let $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$ be uniformly continuous over D. Then f is Cauchy continuous over D.*

Proof. Let $\{x_n\}$ be any Cauchy sequence contained in D. We must show that $\{f(x_n)\}$ is a Cauchy sequence. Let $\epsilon > 0$. Because f is uniformly continuous over D there exists $\delta > 0$ such that for every x, $y \in D$

$$|x - y| < \delta \qquad \Longrightarrow \qquad |f(x) - f(y)| < \epsilon \, .$$

Because $\{x_n\}$ is a Cauchy sequence there exists $N \in \mathbb{N}$ such that for every m, $n \in \mathbb{N}$

$$m, n > N \qquad \Longrightarrow \qquad |x_m - x_n| < \delta \, .$$

Hence, because $\{x_n\}$ is contained in D, for every m, $n \in \mathbb{N}$

$$m, n > N \qquad \Longrightarrow \qquad |f(x_m) - f(x_n)| < \epsilon \, .$$

Therefore the sequence $\{f(x_n)\}$ is Cauchy. □

Remark. The converse of Proposition 8.4 is false. For example, let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$. Because f is continuous and $\mathbb{R}$ is closed, Proposition 8.2 implies f is Cauchy continuous. However, in the next section we will show that f is not uniformly continuous.

We now show that there are many uniformly continuous functions. Recall that a function $f : D \to \mathbb{R}$ is Lipschitz continuous over D provided there exists an $L \geq 0$ such that for every $x, y \in D$ we have

$$|f(x) - f(y)| \leq L|x - y| \, .$$

The following should be pretty clear.

Proposition 8.5. *Let $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$ be Lipschitz continuous over D. Then f is uniformly continuous over D.*

Proof. Let $\epsilon > 0$. Pick $\delta > 0$ so that $L\delta < \epsilon$. Then for every $x, y \in D$

$$|x - y| < \delta \qquad \Longrightarrow \qquad |f(x) - f(y)| \leq L|x - y| \leq L\delta < \epsilon \, .$$

□

There many uniformly continuous functions because there are many Lipschitz continuous functions. Recall that we showed in Proposition 7.3 that if D is either either (a, b), $[a, b)$, $(a, b]$ or $[a, b]$ for some $a < b$ while $f : D \to \mathbb{R}$ is continuous over D and differentiable over (a, b) with f' bounded then f is Lipschitz continuous over D with

$$L = \sup\{|f'(x)| \, : \, x \in (a, b)\} \, .$$

By Proposition 8.5, every such function is uniformly continuous.

While there are many uniformly continuous functions, there are also many functions that are not uniformly continuous.

Examples. The functions $f : \mathbb{R}_+ \to \mathbb{R}$ given by

$$f(x) = \frac{1}{x}\,, \qquad f(x) = x^2\,, \qquad f(x) = \sin\left(\frac{1}{x}\right),$$

are not uniformly continuous. We will give one approach to showing this in the next section.

Notice that the derivatives in the above examples are all unbounded over $\mathbb{R}_+$:

$$f'(x) = -\frac{1}{x^2}\,, \qquad f'(x) = 2x\,, \qquad f'(x) = -\frac{1}{x^2}\cos\left(\frac{1}{x}\right).$$

Propostion 8.5 shows that every differentiable function that is not uniformly continuous over an open interval must have an unbounded derivative. However, as the following exercise shows, the converse does not hold.

Exercise. Show the function $f : \mathbb{R}_+ \to \mathbb{R}$ given by $f(x) = x^{\frac{1}{2}}$ is uniformly continous over $\mathbb{R}_+$. Hint: First establish the inequality

$$\left|y^{\frac{1}{2}} - x^{\frac{1}{2}}\right| \leq |y - x|^{\frac{1}{2}} \qquad \text{for every } x, y \in \mathbb{R}_+\,.$$

The next exercise introduces an important extension of Lipschitz continuity called *Hölder continuity*, and shows that it also implies uniform continuity.

Exercise. Let $D \subset \mathbb{R}$. A function $f : D \to \mathbb{R}$ is said to be Hölder continuous of order $\alpha \in (0, 1]$ if there exists a $C \in \mathbb{R}_+$ such that for every $x, y \in D$ we have

$$|f(x) - f(y)| \leq C\,|x - y|^\alpha\,.$$

Show that if $f : D \to \mathbb{R}$ is Hölder continuous of order α for some $\alpha \in (0, 1]$ then it is uniformly continuous over D.

8.3. **Sequence Characterization of Uniform Continuity.** The following theorem gives a characterization of uniform continuity in terms of sequences that is handy for showing that certain functions are not uniformly continuous.

Theorem 8.1. *Let $D \subset \mathbb{R}$. Then $f : D \to \mathbb{R}$ is uniformly continuous over D if and only if for every $\{x_n\}_{n\in\mathbb{N}}, \{y_n\}_{n\in\mathbb{N}} \subset D$ we have*

$$\lim_{n\to\infty} (x_n - y_n) = 0 \qquad \Longrightarrow \qquad \lim_{n\to\infty} \big(f(x_n) - f(y_n)\big) = 0\,.$$

Remark. This characterization is used as the definition of uniform continuity in some textbooks.

Remark. We can use this characterization to show that a given function $f : D \to \mathbb{R}$ is not uniformly continuous by starting with a sequence $\{z_n\}_{n\in\mathbb{N}}$ such that $z_n \to 0$ as $n \to \infty$. Next, we seek a sequence $\{x_n\}_{n\in\mathbb{N}} \subset D$ such that $\{x_n + z_n\}_{n\in\mathbb{N}} \subset D$ and

$$\lim_{n\to\infty} \big(f(x_n) - f(x_n + z_n)\big) \neq 0\,.$$

Upon setting $y_n = x_n + z_n$, we will have then found sequences $\{x_n\}_{n\in\mathbb{N}}$, $\{y_n\}_{n\in\mathbb{N}} \subset D$ such that

$$\lim_{n\to\infty} (x_n - y_n) = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \big(f(x_n) - f(y_n)\big) \neq 0\,.$$

Theorem 8.1 then implies the function f is not uniformly continuous over D.

Example. The function $f : \mathbb{R}_+ \to \mathbb{R}$ given by $f(x) = 1/x$ is not uniformly continuous. Let $\{z_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+$ such that $z_n \to 0$ as $n \to \infty$. Then for every $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+$ we have $\{x_n + z_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+$ and

$$f(x_n) - f(x_n + z_n) = \frac{1}{x_n} - \frac{1}{x_n + z_n} = \frac{z_n}{x_n(x_n + z_n)}\,.$$

If we choose $x_n = z_n$ for every $n \in \mathbb{N}$ then

$$f(x_n) - f(x_n + z_n) = \frac{1}{2z_n} \not\to 0 \qquad \text{as } n \to \infty\,.$$

Hence, f cannot be uniformly continuous over $\mathbb{R}_+$ by Theorem 8.1.

Example. The function $f : \mathbb{R}_+ \to \mathbb{R}$ given by $f(x) = x^2$ is not uniformly continuous. Let $\{z_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+$ such that $z_n \to 0$ as $n \to \infty$. Then for every $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+$ we have $\{x_n + z_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+$ and

$$f(x_n) - f(x_n + z_n) = x_n^2 - (x_n + z_n)^2 = -2x_n z_n - z_n^2\,.$$

If we choose $x_n = 1/z_n$ for every $n \in \mathbb{N}$ then

$$f(x_n) - f(x_n + z_n) = -2 - z_n^2 \not\to 0 \qquad \text{as } n \to \infty\,.$$

Hence, f cannot be uniformly continuous over $\mathbb{R}_+$ by Theorem 8.1.

Exercise. Show the function $f : \mathbb{R}_+ \to \mathbb{R}$ given by $f(x) = \sin(1/x)$ is not uniformly continuous. Hint: Proceed as in the first example above, but choose a particular $\{z_n\}_{n\in\mathbb{N}}$ to simplify things.

Exercise. Show the function $f : \mathbb{R}_+ \to \mathbb{R}$ given by $f(x) = x^p$ for some $p > 1$ is not uniformly continuous.

Exercise. Show the function $f : \mathbb{R}_+ \to \mathbb{R}$ given by $f(x) = x^{-p}$ for some $p > 0$ is not uniformly continuous.

Now let us turn to the proof of Theorem 8.1. The proof is similar to the proof of the characterization of continuity at a point in terms of convergent sequences.

Proof. ($\Longrightarrow$) Let $\{x_n\}_{n\in\mathbb{N}}, \{y_n\}_{n\in\mathbb{N}} \subset D$ such that

$$\lim_{n\to\infty} (x_n - y_n) = 0\,.$$

We need to show that

$$\lim_{n\to\infty} \big(f(x_n) - f(y_n)\big) = 0\,.$$

Let $\epsilon > 0$. Because f is uniformly continuous over D there exists $\delta > 0$ such that for every $x, y \in D$ we have

$$|x - y| < \delta \qquad \Longrightarrow \qquad |f(x) - f(y)| < \epsilon\,.$$

Because $(x_n - y_n) \to 0$ as $n \to \infty$, we know $|x_n - y_n| < \delta$ eventually as $n \to \infty$. Because $|x_n - y_n| < \delta$ implies $|f(x_n) - f(y_n)| < \epsilon$, it follows that $|f(x_n) - f(y_n)| < \epsilon$ eventually as $n \to \infty$. Because $\epsilon > 0$ was arbitrary, we have shown that $\big(f(x_n) - f(y_n)\big) \to 0$ as $n \to \infty$.

($\Longleftarrow$) Suppose f is not uniformly continuous over D. Then there exist $\epsilon_o > 0$ such that for every $\delta > 0$ there exists $x, y \in D$ such that

$$|x - y| < \delta \qquad \text{and} \qquad |f(x) - f(y)| \geq \epsilon_o\,.$$

Hence, for every $n \in \mathbb{N}$ there exists $x_n, y_n \in D$ such that

$$|x_n - y_n| < \frac{1}{2^n} \qquad \text{and} \qquad |f(x_n) - f(y_n)| \geq \epsilon_o\,.$$

Clearly, $\{x_n\}_{n\in\mathbb{N}}, \{y_n\}_{n\in\mathbb{N}} \subset D$ such that

$$\lim_{n\to\infty}(x_n - y_n) = 0 \qquad \text{and} \qquad \lim_{n\to\infty}\big(f(x_n) - f(y_n)\big) \neq 0\,.$$

But this contradicts the part of our hypothesis that requires that $\big(f(x_n) - f(y_n)\big) \to 0$ as $n \to \infty$. Therefore f must be uniformly continuous over D. □

8.4. **Bounded Domains and Uniform Continuity.** The following theorem shows that if D is bounded then Cauchy continuity implies uniform continuity. What lies behind this result is the fact that bounded sequences must have converging subsequences.

Theorem 8.2. *Let $D \subset \mathbb{R}$ be bounded. Let $f : D \to \mathbb{R}$ be Cauchy continuous. Then f is uniformly continuous over D.*

Proof. We will show that if f is not uniformly continuous over D then it is not Cauchy continuous over D. If f is not uniformly continuous over D then the characterization of uniform continuity given by Theorem 8.1 implies that there exists sequences $\{x_n\}_{n\in\mathbb{N}}, \{y_n\}_{n\in\mathbb{N}} \subset D$ such that

$$\lim_{n\to\infty}(x_n - y_n) = 0\,.$$

but that

$$\lim_{n\to\infty}\big(f(x_n) - f(y_n)\big) \neq 0\,.$$

Then there exists $\epsilon_o > 0$ such that

$$|f(x_n) - f(y_n)| \geq \epsilon_o \qquad \text{frequently}\,.$$

Hence, there exists subsequences $\{x_{n_m}\}_{m\in\mathbb{N}}, \{y_{n_m}\}_{m\in\mathbb{N}} \subset D$ such that

$$\lim_{m\to\infty}(x_{n_m} - y_{n_m}) = 0\,,$$

and

$$\big|f(x_{n_m}) - f(y_{n_m})\big| \geq \epsilon_o \qquad \text{for every } m \in \mathbb{N}\,. \tag{8.1}$$

Because D is bounded, the subsequence $\{x_{n_m}\}_{m\in\mathbb{N}}$ has a further subsequence $\{x_{n_{m_k}}\}_{k\in\mathbb{N}} \subset D$ that converges to some $x_* \in \mathbb{R}$. Because

$$\lim_{k\to\infty}\big(y_{n_{m_k}} - x_{n_{m_k}}\big) = 0\,,$$

we see that $\{y_{n_{m_k}}\}_{k\in\mathbb{N}}$ also converges with

$$\lim_{k\to\infty} y_{n_{m_k}} = \lim_{k\to\infty} x_{n_{m_k}} + \lim_{k\to\infty}\big(y_{n_{m_k}} - x_{n_{m_k}}\big) = x_* + 0 = x_*\,.$$

Now define a sequence $\{z_l\}_{l\in\mathbb{N}} \subset D$ by

$$z_l = \begin{cases} x_{n_{m_k}} & \text{if } l = 2k\,, \\ y_{n_{m_k}} & \text{if } l = 2k+1\,. \end{cases}$$

The sequence $\{z_l\}_{l\in\mathbb{N}}$ also converges to x_*, and thereby is Cauchy. However the sequence $\{f(z_l)\}_{l\in\mathbb{N}}$ is not Cauchy because we see from (8.1) that

$$\big|f(z_{2k+1}) - f(z_{2k})\big| = \big|f(y_{n_{m_k}}) - f(x_{n_{m_k}})\big| \geq \epsilon_o \qquad \text{for every } k \in \mathbb{N}\,.$$

Therefore f is not Cauchy continuous over D. □

Remark. The conclusion of the above theorem can still hold for some cases where D is unbounded. For example, if $D = \mathbb{Z}$ then every function is uniformly continuous. This is easily seen from the definition of uniform continuity by taking $\delta < 1$.

An immediate consequence of the foregoing theorem is the following important theorem stating that continuity implies uniform continuity when the domain is closed and bounded.

Theorem 8.3. *Let $D \subset \mathbb{R}$ be closed and bounded. Let $f : D \to \mathbb{R}$ be continuous. Then f is uniformly continuous over D.*

Proof. Because f is continuous over D and D is closed, Proposition 8.2 implies that f is Cauchy continuous over D. Because f is Cauchy continuous over D and D is bounded, Theorem 8.2 implies that f is uniformly continuous over D. □

Remark. The conclusion of the above theorem can still hold for some cases where D is closed but unbounded. For example, if $D = \mathbb{Z}$ then every function is uniformly continuous. However, Propositions 8.3 and 8.4 combine to show that the hypothesis D is closed cannot be dropped.

8.5. **Continuity and Restrictions.** The relationship between Cauchy continuity and uniform continuity can be characterized with the aid of the notion of a restriction of a function. The notions of restriction and extension for functions are defined as follows.

Definition 8.3. *Let $f : D \to \mathbb{R}$ and $g : E \to \mathbb{R}$ with $D, E \subset \mathbb{R}$. If*

$$D \subset E \quad \text{and} \quad f(x) = g(x) \text{ for every } x \in D\,,$$

then we say that f is the restriction *of g to D and that g is* an extension *of f to E.*

It should be clear that the restrictions of a given function are uniquely determined by their domains. Moreover, restrictions will inherit certain regularity properties.

Proposition 8.6. *Let $E \subset \mathbb{R}$ and $g : E \to \mathbb{R}$. For every $D \subset E$ there is a unique restriction of g to D, which we will denote $g|_D$.*

If g is continuous over E then $g|_D$ is continuous over D.

If g is Cauchy continuous over E then $g|_D$ is Cauchy continuous over D.

If g is uniformly continuous over E then $g|_D$ is uniformly continuous over D.

Proof. Exercise.

We now give the following characterization of a Cauchy continuous function.

Theorem 8.4. *Let $E \subset \mathbb{R}$ and $g : E \to \mathbb{R}$. Then g is Cauchy continuous over E if and only if $g|_D$ is uniformly continuous over D for every bounded $D \subset E$.*

Proof. ($\Longrightarrow$) Let $g : E \to \mathbb{R}$ be Cauchy continuous over E. Let $B \subset E$ be bounded. Proposition 8.6 implies that $g|_D$ is Cauchy continuous over D. Because D is bounded and $g|_D$ is Cauchy continuous over D, Theorem 8.2 implies that $g|_D$ is uniformly continuous over D.

($\Longleftarrow$) Let $g : E \to \mathbb{R}$ such that $g|_D$ is uniformly continuous over D for every bounded $D \subset E$. Let $\{x_n\}_{n\in\mathbb{N}} \subset E$ be a Cauchy sequence. Because every Cauchy sequence is bounded, let $D \subset E$ be a bounded set such that $\{x_n\}_{n\in\mathbb{N}} \subset D$. Then $g|_D$ is uniformly continuous over D, which implies that $g|_D$ is Cauchy continuous over D. Hence, $\{g(x_n)\}_{n\in\mathbb{N}}$ is a Cauchy sequence. But the Cauchy sequence $\{x_n\}_{n\in\mathbb{N}} \subset E$ was arbitrary. Therefore g is Cauchy continuous over E. □

8.6. **Continuous Extensions.** It should be clear that the extensions of a given function are not uniquely determined by their domains. However there can be a unique extension g of a given function f if sufficent regularity is imposed on both f and g and if E is not too big. Cauchy continuity plays a central role in characterizing when a function can be extended to a larger set as a continuous function.

In this section we consider extensions g that are continuous over their domain E, so-called *continuous extensions.* It is clear from Proposition 8.6 that f must be continuous in order to have any continuous extension. The following theorem characterizes when a function f has a continuous extension g to the closure of its domain.

Theorem 8.5. (Continuous Extension Theorem) *Let $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$. Then there exists an extension of f to D^c that is continuous if and only if f is Cauchy continuous. Moreover, in that case there is a unique extension g of f to D^c that is continuous.*

Remark. Because D^c is closed, Proposition 8.2 showed that the concepts of continuity and Cauchy continuity coincide. Therefore the extension g of f to D^c is Cauchy continuous.

Proof. ($\Longrightarrow$) Let $g : D^c \to \mathbb{R}$ be an extension of f that is continuous. Let B be any bounded subset of D. Then B^c is also bounded. Because g is continuous over D^c and $B^c \subset D^c$, Proposition 8.6 implies that $g|_{B^c}$ is continuous. Because B^c is closed and bounded while $g|_{B^c}$ is continuous, Theorem 8.3 implies that g_{B^c} is uniformly continuous. But then Proposition 8.6 implies that $g|_B$ is uniformly continuous. Because $B \subset D$, we see that $f|_B = g_B$, whereby $f|_B$ is uniformly continuous. But B was an arbitrary bounded subset of D. Therefore the restriction of f to any bounded subset of D is uniformly continuous.

($\Longleftarrow$) The proof of this direction is more difficult because the desired extension g must be constructed from f. Let $x \in D^c$. If $x \in D$ then set $g(x) = f(x)$. If $x \notin D$ then there exists a sequence $\{x_n\} \subset D$ such that $\{x_n\}$ converges to x. The idea will be to set

$$g(x) = \lim_{n\to\infty} f(x_n)\,. \tag{8.2}$$

However, for the value of $g(x)$ to be well-defined we must show that it does not depend on the choice of the sequence $\{x_n\}$.

Let $\{x_n\}$ and $\{y_n\}$ be any sequences contained in D that converge to x. Because convergent sequences are bounded, the set $B = \{x_n\} \cup \{y_n\}$ is a bounded subset of D. Because $f|_B$ is uniformly continuous and B is bounded, Proposition 8.4 implies that $f|_B$ is Cauchy continuous. Because the sequences $\{x_n\}$ and $\{y_n\}$ are Cauchy while $f|_B$ is Cauchy continuous, the sequences $\{f(x_n)\}$ and $\{f(y_n)\}$ are also Cauchy, and therefore convegent. We must show that they have the same limit. To do this we construct a new sequence $\{z_n\} \subset B$ by setting

$$z_n = \begin{cases} x_n & \text{for } n \text{ even}\,, \\ y_n & \text{for } n \text{ odd}\,. \end{cases}$$

It is easy to show that $\{z_n\}$ converges to x, and is thereby Cauchy. Because $f|_B$ is Cauchy continuous, $\{f(z_n)\}$ is also Cauchy, and thereby convergent. Because every subsequence of a convergent sequence will converge to the same limit, it follows that

$$\begin{aligned} \lim_{n\to\infty} f(x_n) &= \lim_{k\to\infty} f(x_{2k}) = \lim_{k\to\infty} f(z_{2k}) = \lim_{n\to\infty} f(z_n) \\ &= \lim_{k\to\infty} f(z_{2k+1}) = \lim_{k\to\infty} f(y_{2k+1}) = \lim_{n\to\infty} f(y_n)\,. \end{aligned}$$

Therefore all sequences in D that converge to x will produce the same value in formula (8.2). Therefore the function $g : D^c \to \mathbb{R}$ defined by formula (8.2) is well-defined.

Next, it is clear that if is to be a continuous extension of f then its value at any $x \in D^c$ must be given by formula (8.2). This extension is therefore unique.

Finally, we have to prove that $g : D^c \to \mathbb{R}$ is continuous. We leave this last step as an exercise for the interested student; it is not easy. □