Advanced Calculus: MATH 410
Professor David Levermore
21 October 2009

1. Differentiability

1.1. Differentiability at a Point.
Given any function $f : \mathrm{Dom}(f) \to \mathbb{R}$, the slope of the secant line through any two points $(a, f(a))$ and $(b, f(b))$ on its graph is given by

$$(1.1) \qquad \frac{f(b) - f(a)}{b - a}.$$

This quantity is called a *difference quotient*. It is defined over all a, $b \in \mathrm{Dom}(f)$ for which $b \neq a$. It is undefined when $b = a$.

A function f is said to be *differentiable* at a point $a \in \mathrm{Dom}(f)$ whenever

$$(1.2) \qquad \lim_{b \to a} \frac{f(b) - f(a)}{b - a} \quad \text{exists},$$

This will be the case whenever a unique tangent line to the graph at $(a, f(a))$ exists and is not vertical, in which case the *slope of the tangent line* is given by

$$(1.3) \qquad f'(a) = \lim_{b \to a} \frac{f(b) - f(a)}{b - a},$$

whereby the equation of the tangent line is given by

$$(1.4) \qquad y = f(a) + f'(a)(x - a).$$

By replacing b by $a + h$ in (1.3), the slope of this tangent line may be expressed as

$$(1.5) \qquad f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}.$$

It should be evident to you that (1.5) is equivalent to (1.3). 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. A function that is differentiable at every point in $\mathrm{Dom}(f)$ is said to be differentiable. A function that is differentiable at every point in a set $S \subset \mathrm{Dom}(f)$ is said to be differentiable over S.

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 $x \neq a$ one has 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 one sees that

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

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

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 $y = 1$?
(5) g behaves like x^2 for large values of $|x|$?
(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.

1.2. **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,

(1.6) $$\mathrm{Dom}(f') \equiv \left\{ x \in \mathrm{Dom}(f) : f \text{ is differentiable at } x \right\},$$

and by (1.5) the value of $f'(x)$ is given by

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

If f is differentiable then $\mathrm{Dom}(f') = \mathrm{Dom}(f)$. Otherwise $\mathrm{Dom}(f')$ is a strict subset of $\mathrm{Dom}(f)$.

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

$$\mathrm{Dom}(f'') = \left\{ x \in \mathrm{Dom}(f') : f' \text{ is differentiable at } x \right\}.$$

In a similar way the nth 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

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

If f has all its derivatives at a point a, it is said to be *infinitely differentiable* at a. If f is infinitely differentiable at every point in $\mathrm{Dom}(f)$, it is said to be *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

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

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.

1.3. **Differentiable Functions.** The classical differentiatian rules that you recall from calculus can now be derived.

1.3.1. *Linear Combinations of Differentiable Functions.* Given any two differentiable functions u and v, and 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:*

(1.8)
$$\frac{d}{dx}(ku) = k\,\frac{du}{dx}\,, \qquad \frac{d}{dx}(u+v) = \frac{du}{dx} + \frac{dv}{dx}\,.$$

These rules follow from the definition of the derivative (1.7) 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 (1.8) 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 $k_1 u_1 + k_2 u_2 + \cdots + k_n u_n$ 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 (1.8) show that each such linear combination is also differentiable and its derivative is given by the *linear combination rule:*

(1.9)
$$\frac{d}{dx}\big(k_1 u_1 + k_2 u_2 + \cdots + k_n u_n\big) = k_1 \frac{du_1}{dx} + k_2 \frac{du_2}{dx} + \cdots + k_n \frac{du_n}{dx}\,.$$

1.3.2. *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 Leibnitz) rule:*

(1.10)
$$\frac{d}{dx}(uv) = \frac{du}{dx}v + u\frac{dv}{dx}\,.$$

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 one considers the product of three differentiable functions u, v, and w then two applications of (1.10) show that

$$\frac{d}{dx}(uvw) = \frac{du}{dx}vw + u\frac{dv}{dx}w + uv\frac{dw}{dx}\,.$$

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 Leibnitz rule:*

(1.11)
$$\frac{d}{dx}(u_1 u_2 \cdots u_n) = \frac{du_1}{dx}u_2 \cdots u_n + u_1\frac{du_2}{dx}\cdots u_n + \cdots + u_1 u_2 \cdots \frac{du_n}{dx}\,.$$

A consequence of setting $v = 1/u$ in the product rule (1.10) is the *reciprocal rule:*

$$(1.12) \qquad \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{u}\right) = -\frac{1}{u^2}\frac{\mathrm{d}u}{\mathrm{d}x} \qquad \text{wherever } u \neq 0.$$

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

$$(1.13) \qquad \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{u}{v}\right) = \frac{\dfrac{\mathrm{d}u}{\mathrm{d}x}v - u\dfrac{\mathrm{d}v}{\mathrm{d}x}}{v^2} \qquad \text{wherever } v \neq 0.$$

If the general Leibnitz rule (1.11) is specialized to the case where all the functions u_k are the same function u then it reduces to the *monomial power rule:*

$$(1.14) \qquad \frac{\mathrm{d}}{\mathrm{d}x}u^n = nu^{n-1}\frac{\mathrm{d}u}{\mathrm{d}x}.$$

The monomial power rule was derived above for positive integers n. When it is combined with the reciprocal rule (1.12), one sees 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:*

$$(1.15) \qquad \frac{\mathrm{d}}{\mathrm{d}x}u^p = pu^{p-1}\frac{\mathrm{d}u}{\mathrm{d}x}.$$

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 (1.14) is then applied to each side of the identity $(u^p)^n = u^m$, one finds 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.

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

$$(1.16) \qquad \frac{\mathrm{d}}{\mathrm{d}x}v(u) = v'(u)\frac{\mathrm{d}u}{\mathrm{d}x}.$$

The chain rule is the most important general rule for differentiation. It is natural to think that it can be derived by letting y approach y 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 does not work 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

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

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 (1.16) then follows from the compostion limit rule and the definition of the derivative (1.7) by letting y approach x.

If one considers the composition of three differentiable functions, w, v, and u, then two applications of (1.16) 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 one considers n differentiable functions $\{u_1, u_2, \cdots, u_n\}$, then $n-1$ applications of (1.16) 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:*

$$(1.17) \qquad \frac{\mathrm{d}}{\mathrm{d}x}u_1(u_2(u_3(\cdots(u_n)\cdots))) = u_1'(u_2(u_3(\cdots(u_n)\cdots)))\,u_2'(u_3(\cdots(u_n)\cdots))\cdots\frac{\mathrm{d}u_n}{\mathrm{d}x}\,.$$

1.3.4. *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 (1.16) can be used to derive the *inverse function rule:*

$$(1.18) \qquad\qquad \frac{\mathrm{d}}{\mathrm{d}x}f^{-1}(u) = \frac{1}{f'\big(f^{-1}(u)\big)}\frac{\mathrm{d}u}{\mathrm{d}x}\,.$$

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 (1.18).

1.4. **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.

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

Definition 1.1. *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).*

Points that are either a local minimizer or a local maximizer of f over D are called local extremizers *and their corresponding values are called* local extrema. *One similarly defines* global extremizers *and* global extrema.

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

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

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

Proposition 1.1. 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*

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

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

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

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. One 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

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

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

Proof: By the definition of the derivative one has

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

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 $x \in D$

$$0 < |x - p| < \delta \quad \implies \quad \left| \frac{f(x) - f(p)}{x - p} - f'(p) \right| < f'(p) \implies \quad \frac{f(x) - f(p)}{x - p} > 0.$$

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 $x \in D$

$$0 < |x - p| < \delta \quad \implies \quad \left| \frac{f(x) - f(p)}{x - p} - f'(p) \right| < -f'(p) \implies \quad \frac{f(x) - f(p)}{x - p} < 0.$$

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

1.4.3. *One-Sided Limit Point Test.*

Definition 1.2. *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 1.2. 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*

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

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.

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

Proposition 1.3. 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*

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

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

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

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. $\square$

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 $D \cap (-\infty, p)$. For example, this will be 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 1.3. *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 1.4. 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.*

1.5. **Intermediate-Value and Sign Dichotomy Theorems.**

1.5.1. *Derivative Intermediate-Value Theorem.* When the Extreme-Value Theorem and the Fermat Critical Point Theorem are combined with the One-Sided Limit Point Theorem, we obtain a result that lies at the heart of some of the tests for analyzing the monotonicity of a function.

Proposition 1.5. 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 be any value between $f'(a)$ and $f'(b)$, so 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 t:
- 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 a and b are each 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 $g'(p) = f'(p) - m = 0$. Hence, $f'(p) = m$ for some p in (a, b). The case where $f'(a) > f'(b)$ is argued similarly. $\qquad \square$

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.

1.5.2. *Derivative Sign Dichotomy Theorem.* The most useful consequence of the Derivative Intermediate-Value Theorem is the following.

Proposition 1.6. 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*

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

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 I, so our supposition must be false. Hence, the values of f' can only take one sign over I. $\qquad \square$

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).

2. 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-Raphson method, error bounds for the Taylor approximation, and various l'Hospital rules for evaluating limits of indeterminant form.

2.1. **Rolle Theorem and Lagrange Mean-Value Theorem.** The theorem of Rolle is a special case of the Lagrange Mean-Value Theorem, from which the full theorem follows easily. It might be more accurate to call it the Rolle Lemma, because it isolates the key step in the proofs of the mean-value theorems of both Lagrange and Cauchy. We will however stick with its classical moniker. Its proof rests upon a combination of the Extreme-Value Theorem with the Fermat Critical-Point Theorem.

Proposition 2.1. 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 following 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:
- $k < f(\overline{p})$;
- $f(\underline{p}) < k$;
- $f(\underline{p}) = f(\overline{p}) = k$.

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

An immediate consequence of the Rolle Theorem is the following extension.

Proposition 2.2. 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

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

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 $(b, f(b))$.

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. $\qquad\square$

2.2. **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 2.3. 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*

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

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 (2.1) 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. $\qquad\square$

2.3. **Monotonicity.** 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 Rolle Theorem, which follows from the Extreme-Value Theorem.

Proposition 2.4. 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).*
- *if $f' > 0$ over (a,b) then f is increasing over I;*
- *if $f' < 0$ over (a,b) then f is decreasing over I;*
- *if $f' \geq 0$ over (a,b) then f is nondecreasing over I;*
- *if $f' \leq 0$ over (a,b) then f is nonincreasing over I;*
- *if $f' = 0$ over (a,b) then f is constant over I.*

Proof: Suppose $f' > 0$ over (a,b). Consider any two points x and 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), one must have $f'(p) > 0$, whereby $f(y) - f(x) = f'(p)(y - x) > 0$. Hence, f is therefore increasing over I. The other cases are argued similarly. $\square$

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 Theorem.

Proposition 2.5. Monotonicity Tests 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. If f has no critical points in (a,b) then the following are equivalent:*
- (i) *f is increasing over I;*
- (ii) *$f(q) < f(r)$ for some q and r in I with $q < r$;*
- (iii) *$f'(p) > 0$ for some p in (a,b);*
- (iv) *$f' > 0$ over (a,b).*

Similarly, the following are equivalent:
- (v) *f is decreasing over I;*
- (vi) *$f(q) > f(r)$ for some q and r in I with $q < r$;*
- (vii) *$f'(p) < 0$ for some p in (a,b);*
- (viii) *$f' < 0$ over (a,b).*

Proof: We will prove that (i) $\implies$ (ii) $\implies$ (iii) $\implies$ (iv) $\implies$ (i). The proof of the equivalence of (v-viii) is similar.

It is clear from the definition of "increasing over I" that (i) implies (ii). Given (ii), 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, (ii) implies (iii). The fact that (iii) implies (iv) follows from the Derivative Sign Dichotomy Theorem. Finally, (iv) implies (i) is just the first assertion of the Monotonicity Theorem (Proposition 2.4). $\square$

2.4. **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 you 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

$$(2.2) \qquad \lim_{x \to c} \frac{R_c f(x)}{x - c} = \lim_{x \to c} \frac{f(x) - f(c) - f'(c)(x - c)}{x - c} = 0 .$$

This states that $|R_c f(x)|$ vanishes faster than $|x - c|$ as x approaches c. This is the best you can expect to say if all you know is that f is differentiable at c. However, if f has more regularity then you can say how much faster $|R_c f(x)|$ vanishes.

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.

Proposition 2.6. Tangent Line Remainder Theorem: *Let $f : (a, b) \to \mathbb{R}$ be twice differentiable over an interval (a, b). 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*

$$(2.3) \qquad f(x) = f(c) + f'(c)(x - c) + \tfrac{1}{2} f''(p) (x - c)^2 .$$

Remark: For a given c the point p will also depend on x, and this theorem does not give you a clue as to what that dependence might be. However, formula (2.3) does allow you to bound the size of the remainder by bounding the possible values of $f''(p)$. For example, if you can find a number K such that $|f''(z)| < K$ for every $z \in (a, b)$, then you see that for every $x \in (a, b)$ one has

$$(2.4) \qquad |R_c f(x)| = |f(x) - f(c) - f'(c)(x - c)| = \tfrac{1}{2}|f''(p)| (x - c)^2 \leq \tfrac{1}{2} K (x - c)^2 .$$

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 (2.2), which only said the remainder vanishes faster than $x - c$ as x approaches c.

Remark: Formula (2.3) also allows you determine the sign of the remainder when you know the sign of $f''(p)$. For example, if you know that $f''(z) > 0$ for every $z \in (a, b)$, then you know that the tangent line approximation lies below f.

Remark: Finally, when f'' is continuous at c you can refine (2.2) even further by using (2.3) 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 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, one sees 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

$$g'(t) = -f''(t)(x - t) + M (x - t) = \left(M - f''(t)\right) (x - t) .$$

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

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

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

2.5. **Convergence of the Newton-Raphson 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 Newton's 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

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

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 (2.5) 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-Raphson 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 one can see the following.

- If f is increasing and convex near x_*, or is decreasing and concave near x_*, then the sequence $\{x_n\}$ will approach x_* monotonically from above.
- If f is increasing and concave near x_*, or is decreasing and 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 you take your 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 2.7. Newton-Raphson 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

$$(2.6) \qquad |x_n - x_*| \leq \frac{1}{K}\big(K|x_0 - x_*|\big)^{2^n} < \frac{1}{K}\big(K(b - a)\big)^{2^n},$$

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 (2.5) the Newton iterates satisfy

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

On the other hand, the 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), one has

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

If we set $R_n = K|x_n - x_*|$ then the above inequality takes the form $R_{n+1} \leq R_n^2$. You can easily use induction to show that $R_n \leq R_0^{2^n}$. Bound (2.6) then follows because $R_0 = K|x_0 - x_*| < K(b - a)$. $\qquad\square$

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 you 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.

2.6. **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 nth order Taylor approximation to $f(x)$ at c is given by the polynomial

$$T_c^n f(x) \equiv f(c) + f'(c)(x - c) + \frac{1}{2}f''(c)(x - c)^2 + \cdots + \frac{1}{n!}f^{(n)}(c)(x - c)^n$$

(2.7)

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

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 you add to the approximation to recover the exact value of $f(x)$. It is the negative of the *error*.

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

Proposition 2.8. 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 nth 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*

(2.8)

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

Remark: The last term in (2.8) is called the *remainder* or *correction* of the Taylor approximation because it is what you add to the approximation to recover the exact value of $f(x)$. It is the negative of the *error*.

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 (2.7) except that instead of $f^{(n+1)}$ being evaluated at c, it is being evaluated at some unspecified point p that lies between c and x.

Remark: For a given c the point p will also depend on both x and n, and this formula does not give you a clue as to what those dependences might be. However, it does allow you to bound the size of the error by bounding the possible values of $f^{(n+1)}(p)$. For example, if you can find a number K such that $|f^{(n+1)}(z)| < K$ for every $z \in (a, b)$, then you see that

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

It also allows you determine the sign of the error when $n + 1$ is even and you know the sign of $f^{(n+1)}(p)$.

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 the above series is absolutely convergent for every $x \in \mathbb{R}$ is easy to see from, for example, the ratio test. What we are showing here is that it converges to e^x.

Let $f(x) = e^x$. Then

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

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

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

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

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

This bound also holds when $x = 0$, so it holds for every $x \in \mathbb{R}$. Because for every $x \in \mathbb{R}$ one has

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

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

Exercise: Prove that for every $x \in \mathbb{R}$ one has

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

We now give a proof of the Lagrange Remainder Theorem. You should note the similarity with the argument used to establish the Tangent Line Remainder Theorem.

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)

$$\frac{\mathrm{d}}{\mathrm{d}t} T_t^n f(x) = \frac{\mathrm{d}}{\mathrm{d}t} \left(f(t) + \sum_{k=1}^{n} \frac{(x-t)^k}{k!} f^{(k)}(t) \right)$$

(2.9)
$$= f'(t) + \sum_{k=1}^{n} \left(\frac{(x-t)^k}{k!} f^{(k+1)}(t) - \frac{(x-t)^{(k-1)}}{(k-1)!} f^{(k)}(t) \right)$$

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

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

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

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

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

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

2.7. **Cauchy Mean-Value Theorem.** The following useful extension of the Lagrange Mean-Value Theorem is attributed to Cauchy. It also is a consequence of the Rolle Theorem (and hence, of the Extreme-Value Theorem).

Proposition 2.9. 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 $p \in (a, b)$ one has*

(2.10) $$\left(f(b) - f(a)\right)g'(p) = \left(g(b) - g(a)\right)f'(p)\,.$$

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

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

Remark: The geometric interpretation of this theorem is not 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

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

However, the p and q produced by this argument will generally not be equal. It is the fact that f' and g' are evaluated at the same point in (2.10) that gives the Cauchy Mean-Value Theorem its power.

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

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

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

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

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 (2.10) holds for this p.

Now assume that $g'(x) \neq 0$ for every $x \in (a, b)$. Notice that equation (2.11) follows directly from (2.10) 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, $g(b) - g(a) \neq 0$. $\qquad\square$

Here is an alternative proof of the Lagrange Remainder Theorem (Proposition 2.8) that is based on the Cauchy Mean-Value Theorem. Some students find this proof easier to understand than the one based on the observation (2.9) 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. $\qquad \square$

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 more complicated Cauchy Mean-Value Theorem. However, this proof is appealing because it does not require the insight of observation (2.9).

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

Proposition 2.10. 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*

$$(2.12) \qquad \lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0 \,,$$

or that

$$(2.13) \qquad \lim_{x \to a} f(x) = \lim_{x \to a} g(x) = \infty \,.$$

If

$$(2.14) \qquad \lim_{x \to a} \frac{f'(x)}{g'(x)} = L \qquad \text{for some } L \in \mathbb{R}_{\text{ex}} \,,$$

then

$$(2.15) \qquad \lim_{x \to a} \frac{f(x)}{g(x)} = L \,.$$

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

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 (2.12). Let $\epsilon > 0$. By (2.14) there exists $d_\epsilon \in (a, b)$ such that

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

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

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

Upon taking the limit of the last inequality above as y approaches a while using the fact that f and g satisfy (2.12), we see that

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

Hence, the limit (2.15) holds.

Now suppose that f and g satisfy (2.13). Let $\epsilon > 0$. By (2.14) 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 (2.13) 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

(2.16)
$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(p)}{g'(p)}.$$

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 $x \in (a, c_\epsilon)$ one has the bounds

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

whereby for every $x \in (a, c_\epsilon)$ one has the bound

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

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 (2.15) holds. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

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

Proposition 2.11. 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 nth 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 nth 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 at c with $F^{(k)}(c) = G^{(k)}(c) = 0$ for every $k = 0, 1, \ldots, n - 1$. Moreover,

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

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

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

Because that $G^{(k)}(x) \neq 0$ for every $x \neq c$ and every $k = 0, 1, \ldots, n - 1$, by $(n-1)$ applications of the l'Hospital rule we see that

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

But this implies that

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

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

3. Uniform Continuity

Uniform continuity is a very useful concept. Here we introduce it in the context of real-valued functions with domains in $\mathbb{R}$.

Definition 3.1. *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 $x, y \in D$ one has*

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

Remark. This is a stronger 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 $x \in D$ one has

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

Here δ depends on y and ϵ ($\delta = \delta_{y,\epsilon}$), while in Definition 3.1 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.

By the above remark we obviously have the following.

Proposition 3.1. *Let $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$ be uniformly continuous over D. Then f is continuous over D.*

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.

3.1. **Some Uniformly Continuous Functions.** We first 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 $x, y \in D$ one has

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

The following should be pretty clear.

Proposition 3.2. *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 $x, y \in D$

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

$\square$

There many uniformly continuous functions because there are many Lipschitz continuous functions. Recall we have shown 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

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

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 3.2 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

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

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 $x, y \in D$ one has

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

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.

3.2. **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 3.1. *Let $D \subset \mathbb{R}$. Then $f : D \to \mathbb{R}$ is uniformly continuous over D if and only if for every $\{x_n\}_{n\in\mathbb{N}}, \{y_n\}_{n\in\mathbb{N}} \subset D$ one has*

$$\lim_{n\to\infty} (x_n - y_n) = 0 \qquad \implies \qquad \lim_{n\to\infty} \left(f(x_n) - f(y_n) \right) = 0.$$

Remark: This characterization is taken as the definition of uniform continuity in the text.

Remark: You 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, you seek a sequence $\{x_n\}_{n\in\mathbb{N}} \subset D$ such that $\{x_n + z_n\}_{n\in\mathbb{N}} \subset D$ and

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

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

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

Theorem 3.1 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 $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+$ one has $\{x_n + z_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+$ 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

$$f(x_n) - f(x_n + z_n) = \frac{1}{2z_n} \nrightarrow 0 \qquad \text{as } n \to \infty.$$

Hence, f cannot be uniformly continuous over $\mathbb{R}_+$ by Theorem 3.1.

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 $\{x_n\}_{n \in \mathbb{N}} \subset \mathbb{R}_+$ one has $\{x_n + z_n\}_{n \in \mathbb{N}} \subset \mathbb{R}_+$ 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

$$f(x_n) - f(x_n + z_n) = -2 - z_n^2 \nrightarrow 0 \qquad \text{as } n \to \infty.$$

Hence, f cannot be uniformly continuous over $\mathbb{R}_+$ by Theorem 3.1.

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.

Now let us turn to the proof of Theorem 3.1. 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 $x, y \in D$ one has

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

Because $(x_n - y_n) \to 0$ as $n \to \infty$, we know $|x_n - y_n| < \delta$ ultimately as $n \to \infty$. Because $|x_n - y_n| < \delta$ implies $|f(x_n) - f(y_n)| < \epsilon$, it follows that $|f(x_n) - f(y_n)| < \epsilon$ ultimately 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. $\qquad\square$

3.3. **Sequential Compactness and Uniform Continuity.** The following theorem shows that if D is closed and bounded then continuity implies uniform continuity. What lies behind this result is the fact that D is sequentially compact when it is closed and bounded.

Theorem 3.2. *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: We will establish the uniform continuity of f by using the characterization of Theorem 3.1. 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

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

Suppose not. Then there exists $\epsilon_o > 0$ such that

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

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

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

and

(3.1) $$\big|f(x_{n_k}) - f(y_{n_k})\big| \geq \epsilon_o \qquad \text{for every } k \in \mathbb{N}.$$

Because D is sequentially compact, the subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ has a further subsequence $\{x_{n_{k_l}}\}_{l\in\mathbb{N}}$ that converges to some $x_* \in D$. Because

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

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

$$\lim_{l\to\infty} y_{n_{k_l}} = \lim_{l\to\infty} x_{n_{k_l}} + \lim_{l\to\infty} \big(y_{n_{k_l}} - x_{n_{k_l}}\big) = x_* + 0 = x_*.$$

Because f is continuous at $x_* \in D$, we know that

$$\lim_{l\to\infty} \big(f(x_{n_{k_l}}) - f(y_{n_{k_l}})\big) = f(x_*) - f(x_*) = 0.$$

But this contradicts our supposition, which by (3.1) implies that

$$\big|f(x_{n_{k_l}}) - f(y_{n_{k_l}})\big| \geq \epsilon_o \qquad \text{for every } l \in \mathbb{N}.$$

Therefore

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

whereby f is uniformly continuous by Theorem 3.1. $\square$

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. This is easily seen from the definition by taking $\delta < 1$. However, the next proposition shows that the hypothesis D is closed cannot be dropped.

Proposition 3.3. *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 uniformly 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 uniformly continuous over D by showing that for every $\delta > 0$ there exists $x, y \in D$ such that

$$|x - y| < \delta, \qquad \text{and} \qquad |f(x) - f(y)| \geq 1.$$

Because x_* is a limit point of D there exists a sequence $\{x_n\}_{n \in \mathbb{N}} \subset D$ such that $x_n \to x_*$ as $n \to \infty$. Let $\delta > 0$ be arbitrary. Let $m \in \mathbb{N}$ such that

$$n \geq m \qquad \Longrightarrow \qquad |x_n - x_*| < \frac{\min\{\delta, 1\}}{2}.$$

Then for every $k \in \mathbb{N}$ one has

$$|x_{m+k} - x_m| \leq |x_{m+k} - x_*| + |x_m - x_*| < \frac{\delta}{2} + \frac{\delta}{2} = \delta,$$

while

$$|x_{m+k} - x_*| < \frac{1}{2}.$$

This last inequality implies that for every $k \in \mathbb{N}$ one has

$$|f(x_{m+k}) - f(x_m)| = \frac{|x_m - x_{m+k}|}{|x_{m+k} - x_*|\,|x_m - x_*|} \geq 2 \frac{|x_m - x_{m+k}|}{|x_m - x_*|}.$$

Because

$$\lim_{k \to \infty} \frac{|x_m - x_{m+k}|}{|x_m - x_*|} = 1,$$

we may pick a $k \in \mathbb{N}$ large enough so that

$$\frac{|x_m - x_{m+k}|}{|x_m - x_*|} > \frac{1}{2}.$$

Then for this k we have

$$|x_{m+k} - x_m| < \delta, \quad \text{and} \quad |f(x_{m+k}) - f(x_m)| \geq 2 \frac{|x_m - x_{m+k}|}{|x_m - x_*|} > 2\frac{1}{2} = 1.$$

Because $\delta > 0$ was arbitrary, we can conclude that f is not uniformly continuous over D. $\qquad \square$

3.4. **Uniform Continuity and Cauchy Sequences.** The property of uniform continuity is closely related to that of mapping Cauchy sequences into Cauchy sequences.

Proposition 3.4. *Let $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$ be uniformly continuous over D. If $\{x_n\} \subset D$ is a Cauchy sequence then $\{f(x_n)\}$ is a converent sequence.*

Remark. This is generally false for continuous functions. For example, let $f : \mathbb{R}_+ \to \mathbb{R}$ be given by $f(x) = 1/x$. Then $\{1/n\}$ is Cauchy, but $\{f(1/n)\} = \{n\}$ is not convergent.

Proof. Let $\{x_n\} \subset D$ be Cauchy. Let $\epsilon > 0$. Because f is uniformly continuous there exists $\delta > 0$ such that for every $x, y \in D$

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

Because $\{x_n\} \subset D$ is Cauchy, there exists N_δ such that

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

whereby

$$m, n > N_\delta \quad \implies \quad |f(x_m) - f(x_n)| < \epsilon \,.$$

Therefore $\{f(x_n)\}$ is Cauchy, and thereby is convergent. $\qquad\qquad\qquad\square$

The converse of Proposition 3.4 is false. For example, let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$. Because f is continuous, it maps convergent sequences into convergent sequences. Because a sequence is convergent in $\mathbb{R}$ if and only if it is Cauchy, f therefore maps Cauchy sequences into convergent sequences. However, we have already shown that f is not uniformly continuous.

A converse of Proposition 3.4 should not be expected. This is because uniform continuity is a concept related to the entire domain of a function while mapping Cauchy sequences into convergent sequences is a local concept. A local concept that is weaker than uniform continuity over the entire domain is that of being uniformly continuous over bounded subsets of the domain. Because every Cauchy sequence is bounded, this weaker concept still enough to imply that Cauchy sequences map into convergent sequences. Remarkably, the converse is also true. We begin with a lemma.

Lemma 3.1. *Let $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$ map Cauchy sequences into convergent sequences. Then there exists a unique function $\hat{f} : D^c \to \mathbb{R}$ that is continuous and that extends f — i.e. $\hat{f}(x) = f(x)$ for every $x \in D$.*

Proof. Let $x \in D^c$. Let $\{x_n\} \subset D$ converge to x. Because $\{x_n\}$ is a Cauchy, $\{f(x_n)\}$ is convergent. We claim that its limit is the same for every sequence in D that converges to x. Indeed, let $\{x_n\}, \{y_n\} \subset D$ be any two sequences that converge to x. Define a new sequence $\{z_n\}$ by

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

It is clear that $\{z_n\}$ converges to x, whereby $\{f(z_n)\}$ is convergent. Its limit must be the common limit of the subsequences $\{f(z_{2n})\} = \{f(x_{2n})\}$ and $\{f(z_{2n+1})\} = \{f(y_{2n+1})\}$. Hence, $\{f(x_n)\}$ and $\{f(y_n)\}$ converge to the same limit. We therefore define $\hat{f}(x)$ to be this common limit. It is clear that $\hat{f}(x) = f(x)$ for every $x \in D$.

We now have to prove that $\hat{f} : D^c \to \mathbb{R}$ is continuous. Let $x \in D^c$ and $\{x_n\} \subset D^c$ such that $x_n \to x$. Because each $x_n \in D^c$, there exists $\{x_{n,k}\}_k \subset D$ such that $x_{n,k} \to x_n$ as $k \to \infty$.

Theorem 3.3. *Let $D \subset \mathbb{R}$. Then a function $f : D \to \mathbb{R}$ is uniformly continuous over every bounded subset of D if and only if it maps Cauchy sequences in D into Cauchy sequences.*

Proof. ($\implies$) Let $\{x_n\} \subset D$ be Cauchy. Because Cauchy sequences are bounded, there exists $B \subset D$ such that $\{x_n\} \subset B$ and B is bounded. Because f is uniformly continuous over B while $\{x_n\} \subset B$ is Cauchy, Proposition 3.4 implies that $\{f(x_n)\}$ is a Cauchy sequence.

($\impliedby$)