Advanced Calculus: MATH 410

Professor David Levermore

13 November 2006

1. Differentiation

1.1. **Extrema, Local Extrema, and Critical Points.**

1.1.1. *Extrema.* In introductory calculus you learned that differential calculus can be used to find a minimum or maximum of a given function.

Definition 1.1. *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*.*

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

A general function f defined over a set D may have neither a minimum nor a maximum. For example, consider

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

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

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

And some may have both. For example, consider

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

There is one general theorem regarding the existence of extrema that we have already established. It will play a central role in the proofs of many subsequent propositions.

Proposition 1.1. Extreme-Value Theorem: *Let $D \subset \mathbb{R}$ be closed and bounded. Let $f : D \to \mathbb{R}$ be continuous. Then f has both a minimum and a maximum over D.*

You should know how to prove this theorem. You should know examples which illustrate that none of its hypotheses can simply be dropped.

1.1.2. *Local Extrema.* The concept of local extrema arises natrually when differential calculus is used to find extrema.

Definition 1.2. *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)*.*

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: Sometimes the terms *relative* and *absolute* are used in place of *local* and *global* respectively.

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

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

1.2. **Transversality Lemma and Critical Points.**

1.2.1. *Transversality Lemma.* Calculus provides tools that can be used to find local extrema, thereby narrowing the hunt for global extrema. The first step in developing those tools is the following.

Proposition 1.2. 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.

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 $x \in D$

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

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 $x \in D$

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

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

□

1.2.2. *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*

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

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

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

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 $D \cap (-\infty, p)$. 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 definitions.

Definition 1.3. *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 $D \cap (-\infty, p)$.*

Definition 1.4. *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 1.4. 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.*

1.2.3. *One-Sided Limit Point Test.* Another consequence of the Transversality Lemma is the following test for when a one-sided limit point is a local minimizer or maximizer.

Proposition 1.5. 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*

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

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.

1.3. **Rolle's Theorem and Lagrange Mean-Value Theorem.**

1.3.1. *Rolle's Theorem.* When the Extreme-Value Theorem is combined with the Critical-Point Theorem, we obtain a result that plays a central role in many subsequent proofs.

Proposition 1.6. Rolle's Theorem: *Let $a < b$ and*

- $f : [a, b] \to \mathbb{R}$ *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 such a 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 p_{min} and p_{max} in $[a, b]$ such that

$$f(p_{min}) \leq f(x) \leq f(p_{max}) \quad \text{for every } x \in [a, b] .$$

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

$$f(p_{min}) \leq k \leq f(p_{max}) .$$

At least one of the following cases must then hold:

- $k < f(p_{max})$;
- $f(p_{min}) < k$;
- $f(p_{min}) = f(p_{max}) = k$.

If $k < f(p_{max})$ then p_{max} must be in (a, b). But then f is differentiable at p_{max} and by the Critical Point Theorem $f'(p_{max}) = 0$. The argument when $f(p_{min}) < k$ goes similarly, yielding $f'(p_{min}) = 0$. Finally, if $f(p_{min}) = f(p_{max}) = k$ 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. □

1.3.2. *Lagrange Mean-Value Theorem.* An immediate consequence of Rolle's Theorem is the following extension.

Proposition 1.7. Lagrange Mean-Value Theorem: *Let $a < b$ and*

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

Then

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

Proof: Define $g : [a, b] \to \mathbb{R}$ by

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

Clearly,

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

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

1.3.3. *Lipschitz Bounds.* Recall that $f : D \to \mathbb{R}$ is Lipschitz continuous over $D \subset \mathbb{R}$ 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. The following is an easy consequence of the Lagrange Mean-Value Theorem.

Proposition 1.8. Lipschitz Bound 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 $f' : (a, b) \to \mathbb{R}$ is bounded then f satifies the Lipschitz bound*

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

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

Proof: Exercise. □

1.3.4. *Monotonicity.* Another consequence of the Lagrange Mean-Value Theorem (and hence, of Rolle's Theorem) is the following.

Proposition 1.9. 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 $f' = 0$ 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 cases where $f' < 0$ over (a, b) and $f' = 0$ over (a, b) are argued similarly. □

1.4. **Intermediate-Value Theorem and Monotonicity.**

1.4.1. *Intermediate-Value Theorem.* When the Extreme-Value Theorem and the 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 1.10. 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 case $f'(a) = f'(b)$ is true because there are no values between $f'(a)$ and $f'(b)$. Now consider the case $f'(a) < f'(b)$. Let m be any value between $f'(a)$ and $f'(b)$, so that

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

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

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

Clearly,

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

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

Proposition 1.11. Derivative Sign Dichotomy 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 differentiable. If f has no critical points over I then either*

$$f' > 0 \text{ over } I \qquad \text{or} \qquad f' < 0 \text{ over } I\,.$$

Proof: Suppose not. Then there are points q and r in I 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 is a critical point of f over I. 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. □

1.4.3. *Monotonicity Tests.* A consequence of the Lagrange Mean-Value Theorem and of the Derivative Sign Dichotomy Theorem (and hence, of the Intermediate-Value Theorem) is the following.

Proposition 1.12. 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) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv) $\Longrightarrow$ (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 1.9). □

1.5. **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 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

$$\text{(1)} \qquad \lim_{x \to c} \frac{R_c f(x)}{x - c} = \lim_{x \to c} \frac{f(x) - f(c) - f'(c)(x - c)}{x - c} = 0\,.$$

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 Rolle's Theorem (and hence, of the Extreme-Value Theorem) is that it yields the following expression for the remainder of the tangent line approximation.

Proposition 1.13. 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*

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

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) 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

$$(3) \qquad |R_c f(x)| = |f(x) - f(c) - f'(c)(x-c)| \le \tfrac{1}{2}K\,(x-c)^2\,.$$

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

Remark: Formula (2) 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 (1) even further by using (2) to show that

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

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 $z \in [c,x]$ define $g(z)$ by

$$g(z) \equiv f(x) - f(z) - f'(z)(x-z) - \frac{1}{2}M\,(x-z)^2\,.$$

Clearly, one sees that as a function of z:

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

$$g'(z) = -f''(z)(x-z) + M\,(x-z) = \big(M - f''(z)\big)\,(x-z)\,.$$

Rolle's 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. □

1.6. **Convergence of Newton's 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

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}\,. \tag{4}$$

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 (4) 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.

Newton's 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 concavity 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 concave up near x_*, or is decreasing and concave down near x_*, then the sequence $\{x_n\}$ will approach x_* from above.
- If f is increasing and concave down near x_*, or is decreasing and convex up near x_*, then the sequence $\{x_n\}$ will approach x_* from below.

These observations can be expressed as follows.

- If $f'(x_*)f''(x_*) > 0$ then the sequence $\{x_n\}$ will approach x_* from above.
- If $f'(x_*)f''(x_*) < 0$ then the sequence $\{x_n\}$ will approach x_* 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 1.14. Newton's 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

$$|x_n - x_*| \leq \frac{1}{K}\big(K\,|x_0 - x_*|\big)^{2^n} < \frac{1}{K}\big(K(b-a)\big)^{2^n}\,, \tag{5}$$

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 (4) 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

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

By letting $R_n = K|x_n - x_*|$, the above inequality takes the form $R_{n+1} \leq R_n^2$. By induction you can easily show that $R_n \leq R_0^{2^n}$. The result 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 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.

1.7. **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

$$\begin{aligned} T_c^{(n)} f(x) &\equiv f(c) + f'(c)(x-c) + \frac{1}{2} f''(c)(x-c)^2 \\ &\qquad + \cdots + \frac{1}{n!} f^{(n)}(c)(x-c)^n \\ &= \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(c)(x-c)^k \,. \end{aligned} \tag{6}$$

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 1.15. 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*

$$f(x) = T_c^{(n)} f(x) + \frac{1}{(n+1)!} f^{(n+1)}(p)(x-c)^{n+1} \,. \tag{7}$$

Remark: The last term in (7) 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 (6) 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 absoluely 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 the proof of the Lagrange Remainder Theorem. You should note the similarity with the argument used to establish the Tangent Line Remainder Theorem.

Proof: Consider the case when $c < x < b$. Fix this x and let M be determined by the relation

$$\begin{aligned} f(x) = f(c) + f'(c)(x-c) + \frac{1}{2}f''(c)(x-c)^2 + \cdots \\ + \frac{1}{n!}f^{(n)}(c)(x-c)^n + \frac{1}{(n+1)!}M\,(x-c)^{n+1}\,. \end{aligned}$$

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

$$\begin{aligned} g(z) \equiv f(x) - f(z) - f'(z)(x-z) - \cdots \\ - \frac{1}{n!}f^{(n)}(z)(x-z)^n - \frac{1}{(n+1)!}M\,(x-z)^{n+1}\,. \end{aligned}$$

Clearly, as a function of z,

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

$$\begin{aligned} g'(z) &= -\frac{1}{n!}f^{(n+1)}(z)(x-z)^n + \frac{1}{n!}M\,(x-z)^n \\ &= \frac{1}{n!}\big(M - f^{(n+1)}(z)\big)\,(x-z)^n\,. \end{aligned}$$

Rolle's Theorem then implies that $g'(p) = 0$ for some p in (a, 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 (a, x). The case $x < a$ is argued similarly. □

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

Proposition 1.16. 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*

$$\big(f(b) - f(a)\big)g'(p) = \big(g(b) - g(a)\big)f'(p)\,. \tag{8}$$

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

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(p)}{g'(p)}\,. \tag{9}$$

Remark: Of course, this theorem reduces to the Lagrange Mean-Value Theorem in the case $g(x) = x$.

Remark: It is worth noting that this theorem does not follow by naively 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, 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 (8) that 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

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

Rolle's 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 (8) holds for this p.

Now assume that $g'(x) \neq 0$ for every $x \in (a,b)$. Notice that equation (9) follows directly from (8) 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, $g(b) - g(a) \neq 0$. □

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

Proposition 1.17. l'Hospital's 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*

$$\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0\,, \tag{10}$$

or that

$$\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = \infty\,. \tag{11}$$

If

$$\lim_{x\to a} \frac{f'(x)}{g'(x)} = L \qquad \textit{for some } L \in \mathbb{R}_{\text{ex}}\,, \tag{12}$$

then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = L\,. \tag{13}$$

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 l'Hospital's 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 (10). Let $\epsilon > 0$. By (12) there exists $d_\epsilon \in (a, b)$ such that

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

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

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

Upon taking the limit of the last inequality above as y approaches a while using the fact that f and g satisfy (10), we see that

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

Hence, the limit (13) holds.

Now suppose that f and g satisfy (11). Let $\epsilon > 0$. By (12) 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 (11) 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

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(p)}{g'(p)}\,. \tag{14}$$

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

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

A short calculation shows that the right-hand side above becomes ϵ if we choose $\eta_\epsilon = \frac{1}{2}\epsilon/(\epsilon + |L|)$. Therefore, the limit (13) holds. $\square$

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

Proposition 1.18. Taylor Polynomial Approximation Theorem: *Let $f : (a,b) \to \mathbb{R}$ be n times differentiable for some $n \in \mathbb{Z}_+$. Let $c \in (a,b)$. 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: The theorem states that the remainder of the n^{th} order Taylor approximation 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 not even assuming that $f^{(n)}$ is continuous, so we cannot take this approach. Rather, we will apply l'Hospital's 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 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 l'Hospital's 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 $n!G(x) = (x-c)^n$. □