The Bisection Method

A Single Technique Yielding Simple Proofs of the Four "Hard" Theorems of Calculus

Anderson Trimm

Illinois Mathematics and Science Academy

Institute Day
March 3, 2023

The Four Hard Theorems

The following theorems are indispensable in calculus, but are almost never proved in calculus textbooks. These are

The Four Hard Theorems

The following theorems are indispensable in calculus, but are almost never proved in calculus textbooks. These are

1. **Intermediate Value Theorem:** A continuous function f on a closed interval $[a, b]$ takes all values between $f(a)$ and $f(b)$.

The Four Hard Theorems

The following theorems are indispensable in calculus, but are almost never proved in calculus textbooks. These are

1. **Intermediate Value Theorem:** A continuous function f on a closed interval $[a, b]$ takes all values between $f(a)$ and $f(b)$.
2. **Boundedness Theorem**: A continuous function on a closed interval is bounded.

The Four Hard Theorems

The following theorems are indispensable in calculus, but are almost never proved in calculus textbooks. These are

1. **Intermediate Value Theorem:** A continuous function f on a closed interval $[a, b]$ takes all values between $f(a)$ and $f(b)$.
2. **Boundedness Theorem**: A continuous function on a closed interval is bounded.
3. **Extreme Value Theorem**: A continuous function on a closed interval has an absolute maximum and an absolute minimum.

The Four Hard Theorems

The following theorems are indispensable in calculus, but are almost never proved in calculus textbooks. These are

1. **Intermediate Value Theorem:** A continuous function f on a closed interval $[a, b]$ takes all values between $f(a)$ and $f(b)$.
2. **Boundedness Theorem**: A continuous function on a closed interval is bounded.
3. **Extreme Value Theorem**: A continuous function on a closed interval has an absolute maximum and an absolute minimum.
4. **Integrability Theorem**: A continuous function on a closed interval is integrable.

The Four Hard Theorems

The following theorems are indispensable in calculus, but are almost never proved in calculus textbooks. These are

1. **Intermediate Value Theorem:** A continuous function f on a closed interval $[a, b]$ takes all values between $f(a)$ and $f(b)$.
2. **Boundedness Theorem**: A continuous function on a closed interval is bounded.
3. **Extreme Value Theorem**: A continuous function on a closed interval has an absolute maximum and an absolute minimum.
4. **Integrability Theorem**: A continuous function on a closed interval is integrable.

What makes proving these theorems hard is that they all assume continuity not just at a single point, but on a whole interval. Proofs of such "global" properties are generally left to a course in Real Analysis, since they require introducing "topological" ideas.

The Four Hard Theorems

To see why these are left out of most Calculus texts, consider the proofs given in Walter Rudin's *Principles of Mathematical Analysis*.

- To prove the Intermediate Value Theorem (Theorem 4.23), Rudin first proves that $[a, b]$ is connected, and then uses continuity of f to show that $f([a, b])$ is also connected, from which the IVT follows.
- To prove the Boundedness Theorem (Theorem 4.15), Rudin first proves the Heine-Borel Theorem, uses it to show $[a, b]$ is compact, and then uses continuity of f to show that $f([a, b])$ is compact, and therefore bounded by Heine-Borel again.
- The Extreme Value Theorem (Theorem 4.16) follows from the Boundedness Theorem.
- The Integrability Theorem follows from Theorem 4.19, whose rather technical proof uses compactness of $[a, b]$ to show that a continuous function f on $[a, b]$ is uniformly continuous.

Michael Spivak's *Calculus* gives less abstract proofs of these theorems (without introducing the topological notions of connectedness and compactness), but they are quite technical and rather long.

Why prove the IVT in first-year calculus?

Here is a point of view about teaching calculus, which I largely agree with:

> "Every aspect of this book was influenced by the desire to **present calculus not merely as a prelude to but as the first real encounter with mathematics**. Since the foundations of analysis provided the arena in which modern modes of mathematical thinking developed, **calculus ought to be the place in which to expect, rather than avoid, the strengthening of insight with logic**. In addition to developing the students' intuition about the beautiful concepts of analysis, it is surely equally important to persuade them that **precision and rigor are neither deterrents to intuition, nor ends in themselves, but the natural medium in which to formulate and think about mathematical questions**." Michael Spivak, *Calculus* (emphasis mine)

While I wouldn't dare teach BC Calculus from Spivak's book, one important theme I try to get across to students is the need for precise definitions and for making careful arguments. My students, for instance, are expected to at least be able to follow and even produce (with guidance) the proofs that are given in appendices of most calculus books.

However, when I get to the IVT, I find I am more or less forced to revert back to drawing pictures and reasoning from the informal definition of continuity, which I had previously argued was unreliable!

Why prove the IVT in first-year calculus?

Searching the web for at least a decent sketch of a proof that high school students would find accessible, I found that my frustration was articulated quite well in an article in The College Mathematics Journal by Steven M. Walk (Vol. 42, No.4, September 2011, https://www.jstor.org/stable/10.4169/college.math.j.42.4.254).

"In a bizarre one-two punch, we tell students that the IVT is obvious (e.g., you can't get from one side of a river to the other without getting wet!), but then, paradoxically, insist that its proof is far above their heads."

"I have heard it said that the proof of the IVT can be skipped because students have an intuition about the real line that they can "transfer" to the graphs of continuous functions. That would be fine - if only intuition were a reliable source! Intuition, after all, is what tells students that $(x+y)^2 = x^2 + y^2$, $(fg)'(x) = f'(x)g'(x)$, and that $\frac{\ln x}{x}$ reduces to a mysterious object called 'ln'."

"Rather than proving the IVT property, we look at (maybe) three cases where it seems to be true, and then we encourage students to leap to the conclusion that it is true in *all* cases. This is the kind of behavior that will make us cringe when students do it in later courses... and we wonder where they learn it!"

A method of proof appropriate for Calculus I

Walk then points out that there *is* a simple proof of the Intermediate Value Theorem which is appropriate for first year calculus students, which is based on repeatedly bisecting the interval $[a, b]$. I call this technique the *Bisection Method*.

Moreover, the same technique can be used to give proofs of the other theorems mentioned above, which are just as simple. The Bisection Method then becomes a common proof technique throughout the course that students can add to their toolbox, giving them additional opportunities to exercise their reasoning abilities.

The Bisection Method

The Bisection Method is already familiar to Calculus students as a method of approximating roots of continuous functions. For example, consider the polynomial $f(x) = x^3 - x - 2$. Since $f(1) = -2 < 0$ and $f(2) = 4 > 0$, by the Intermediate Value Theorem f must have a zero on the interval $(1, 2)$. To get a better estimate of the root, apply what Computer Scientists call the *Binary Search Algorithm*: Begin by checking the midpoint of $[1, 2]$. If f is zero there, we are done. Otherwise, the function must change sign on either $[1, 1.5]$ or $[1.5, 2]$. By the IVT again, f must be zero somewhere in the subinterval on which the sign changes, so we repeat for this subinterval, and so on. Since the length of the interval is halved at after each iteration, the process converges to the root.

Iteration	a_n	b_n	c_n	$f(c_n)$
1	1	2	1.5	−0.125
2	1.5	2	1.75	1.6093750
3	1.5	1.75	1.625	0.6660156
4	1.5	1.625	1.5625	0.2521973
5	1.5	1.5625	1.5312500	0.0591125
6	1.5	1.5312500	1.5156250	−0.0340538
7	1.5156250	1.5312500	1.5234375	0.0122504
8	1.5156250	1.5234375	1.5195313	−0.0109712
9	1.5195313	1.5234375	1.5214844	0.0006222
10	1.5195313	1.5214844	1.5205078	−0.0051789
11	1.5205078	1.5214844	1.5209961	−0.0022794
12	1.5209961	1.5214844	1.5212402	−0.0008289
13	1.5212402	1.5214844	1.5213623	−0.0001034
14	1.5213623	1.5214844	1.5214233	0.0002594
15	1.5213623	1.5214233	1.5213928	0.0000780

The Bisection Method

What seems to be less well-known is that one can use the Bisection Method to *prove* the Intermediate Value Theorem.

The proof presented in Walk's article is not original. He cites a couple of textbooks from the 1960s. A similar version of the proof outlined in Exercise 15 in Chapter 8 of Spivak's *Calculus*. However, Walk does give an interesting discussion on how he teaches the proof, including how he works his way up to it.

Walk and I part ways quite a bit when it comes to how to teach these. I personally prefer the versions of these proofs sketched in the exercises in Spivak's book, so in the following I will describe the lessons I have created based around these.

I have taught these lessons a couple of times now, and they have been well-received by virtually all of my students. (Maybe I should add here that I use a Mastery Grading system in calculus, which allows students to fail forward without grade penalty, so my students are not afraid to take academic risks on things like this.)

Background Material

I spend the first few days of class on some background material, which is revisited from time to time until we finally get to continuity and the IVT.

Each bisection proof is a proof by contradiction, so students need to be familiar with this technique. In class I go through the standard proof that $\sqrt{2}$ is irrational, and then have the students repeat the proof with 2 replaced by some other prime, and maybe give some other practice for homework. This is easily absorbed by all students after one day.

The hard theorems all depend on completeness of the real numbers, which I give as an axiom, stated as the least upper bound property:

Completeness Axiom: *Every nonempty subset A of $\mathbb{R}$ which is bounded above has a least upper bound.*

We work out $\sup A$ and $\inf A$ for sets like $A = \{\frac{1}{n}\}$, and prove things like $c \geq 0$ implies $c \sup A = \sup(cA)$, and have students work out variations. We also prove that $\mathbb{N}$ is unbounded and that there is a rational between any two reals as consequences of the completeness axiom. Most students have this digested by the end of the first week. I try to keep the exercises easy to build confidence (e.g., I prove something about $\sup A$ and then have students give the dual proof for $\inf A$).

The Nested Intervals Theorem

Here is the technical part (Chapter 8, Exercise 14 in Spivak). After proving this (students can do it in one class period, with help), the rest is simple.

Nested Intervals Theorem

Theorem. For each $n \in \mathbb{N}$, assume we are given a closed interval

$$I_n = [a_n, b_n] = \{x \in \mathbb{R} : a_n \leq x \leq b_n\}$$

such that

(i) $a_n \leq a_{n+1}$ and $b_{n+1} \leq b_n$ for all $n \in \mathbb{N}$, so that $I_1 \supseteq I_2 \supseteq I_3 \supseteq I_4 \supseteq \cdots$ (see the picture below).

Then

$$\boxed{\bigcap_{n=1}^{\infty} I_n \neq \emptyset}.$$

If, in addition,

(ii) $\inf\{b_n - a_n\} = 0$,

then

$$\boxed{\bigcap_{n=1}^{\infty} I_n = \{x\}, \text{ where } x = \sup\{a_n\} = \inf\{b_n\}}.$$

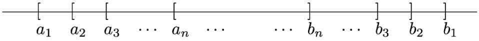

Figure 1: Nested intervals satisfying (i) and (ii).

The Nested Intervals Theorem

Proof: Since the intervals are nested, $a_m \leq b_n$ for all m, n. This shows that every b_n is an upper bound for the set $\{a_m\}$ and every a_m is a lower bound for the set $\{b_n\}$. Let $a = \sup\{a_n\}$ and $b = \inf\{b_n\}$. By definition, $a_n \leq a$ for all n, and since b_n is an upper bound for $\{a_n\}$, $a \leq b_n$, so we have $a_n \leq a \leq b_n$ for all n, which says that $a \in I_n$ for every n and therefore $a \in \bigcap_{n=1}^{\infty} I_n$, so $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$. Similarly, $b \in \bigcap_{n=1}^{\infty} I_n$. Now, if $b < a$, then $b < \frac{a+b}{2} < a$. Since $\frac{a+b}{2} < a$, $\frac{a+b}{2}$ is not an upper bound for $\{a_n\}$ so there exists some $a_k > \frac{a+b}{2}$. Similarly, since $b < \frac{a+b}{2}$, $\frac{a+b}{2}$ is not a lower bound for $\{b_n\}$, so there exists some $b_\ell < \frac{a+b}{2}$, but then $b_\ell < a_k$, which is a contradiction since $a_m \leq b_n$ for all m, n. Thus, $a \leq b$. Assume now that $\inf\{b_n - a_n\} = 0$. Note that $a_n \leq a \leq b \leq b_n$ implies $0 \leq b - a \leq b_n - a_n$ for all n, which says that $b - a$ is a lower bound for $\{b_n - a_n\}$ and therefore $0 \leq b - a \leq \inf\{b_n - a_n\} = 0$, which implies $a = b$. Finally, let $y \in \bigcap_{n=1}^{\infty}$. Then $a_n \leq y \leq b_n$ for every $n \in \mathbb{N}$. The first inequality says that y is an upper bound for $\{a_n\}$, hence $a \leq y$. The second inequality says that y is a lower bound for $\{b_n\}$, hence $y \leq b = a$. Since $a \leq y \leq a$, this implies $y = a$. Hence, $\bigcap_{n=1}^{\infty} I_n \subseteq \{a\}$. Since we also have $\{a\} \subseteq \bigcap_{n=1}^{\infty} I_n$, this implies $\bigcap_{n=1}^{\infty} I_n = \{a\}$, where $a = \sup\{a_n\} = \inf\{b_n\}$. □

Exercise: $\mathbb{R}$ is uncountable

After proving the Nested Intervals Theorem, in the same class period I have students use it to give a quick proof that $\mathbb{R}$ is uncountable.

Theorem. There is no surjection $f : \mathbb{N} \to \mathbb{R}$.

Proof. For each $n \in \mathbb{N}$, write $f(n) = x_n$. Let $I_1 = [a_1, b_1]$ be any interval not containing x_1. Let $I_2 = [a_2, b_2]$ be any closed subinterval of I_1 not containing x_2, $I_3 = [a_3, b_3]$ any closed subinterval of I_2 not containing x_3, and so on. By the Nested Intervals Theorem, $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$. But no x_k can be in the intersection (since $x_k \notin I_k$), so there exists some real number $y \in \bigcap_{n=1}^{\infty} I_n \subseteq \mathbb{R}$ such that $y \neq f(n)$ for any $n \in \mathbb{N}$. Hence, f is not a surjection. $\square$

(Note that we did not need hypothesis (ii) of the Nested Intervals Theorem.)

The Bisection Method

Finally, we need to show that the Bisection Method produces a sequence of intervals satisfying both hypotheses of the Nested Intervals Theorem, which is easy at this point.

Consider the following procedure:

- Begin with a closed interval $I_1 = [a_1, b_1]$.
- Bisect I_1 to obtain two closed subintervals $[a_1, \frac{a_1+b_1}{2}]$ and $[\frac{a_1+b_1}{2}, b_1]$.
- Select one of the two subintervals above, and call it $I_2 = [a_2, b_2]$.
- Keep repeating this process to obtain a sequence of intervals $I_1, I_2, I_3, I_4, I_5, \ldots$.

Show that the sequence of intervals $I_1, I_2, I_3, I_4, I_5, \ldots$ obtained above satisfies both hypotheses of the Nested Intervals Theorem. That is, show that

i. $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$ and $b_{n+1} \leq b_n$ for all $n \in \mathbb{N}$, so that $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$, and
ii. $\inf\{b_n - a_n\} = 0$.

(i): Since for each n we have $a_n < \frac{a_n+b_n}{2} < b_n$, so either $a_{n+1} = a_n$ or moves to the right, and similarly either $b_{n+1} = b_n$ or moves to the left.

(ii): Note that $b_n - a_n = \frac{b_1-a_1}{2^{n-1}}$, so $\inf\{b_n - a_n\} = (b_1 - a_1)\inf\{\frac{1}{2^{n-1}}\} = 0$, by a previous exercise.

Rationals between reals, revisited

As their first bisection proof, I have the students give an alternative proof that there is a rational between any two reals, which I restate as follows.

Theorem. Given any $x \in \mathbb{R}$ and $\epsilon > 0$, show that $(x - \epsilon, x + \epsilon)$ contains a rational number.

We will need the following Lemma. (The name is due to Walk.)

Lemma. (The Capture Theorem) Let A be a nonempty subset of $\mathbb{R}$. If A is bounded above, then any open interval containing $\sup A$ contains an element of A. Similarly, if A is bounded below, then then any open interval containing $\inf A$ contains an element of A.

Proof. Let (x, y) be an open interval such that $x < \sup A < y$. If (x, y) didn't contain an element of A, then x would be an upper bound for A, which is a contradiction since $x < \sup A$. □

Rationals between reals, revisited

Theorem. Given any $x \in \mathbb{R}$ and $\epsilon > 0$, show that $(x-\epsilon, x+\epsilon)$ contains a rational number.

Proof. If x is rational we are done, so assume x is irrational. Let b_1 be the smallest integer greater than x, and let $a_1 = b_1 - 1$. Then $I_1 = [a_1, b_1]$ contains x and has rational endpoints. It follows that x is contained in either $(a_1, \frac{a_1+b_1}{2})$ or $(\frac{a_1+b_1}{2}, b_1)$. Let I_2 be the (closed) subinterval containing x. Continuing in this way, we obtain a sequence of closed intervals $I_1 \supseteq I_2 \supseteq \cdots$ satisfying the hypotheses of the Nested Intervals Theorem, where each I_n contains x and has rational endpoints. By the Nested Intervals Theorem $\bigcap_{n=1}^{\infty} I_n = \{y\}$, where $y = \sup\{a_n\} = \inf\{b_n\}$. Since $x \in I_n$ for all n, $x \in \bigcap_{n=1}^{\infty} I_n$, and therefore $y = x$. Since $x = \sup\{a_n\}$, by the Capture Theorem the open interval $(x-\epsilon, x+\epsilon)$ contains a_m for some $m \in \mathbb{N}$. Since a_m is rational, this completes the proof. □

Intermediate Value Theorem

Time passes, and we eventually come to the IVT. I state the theorem, give the intuition and examples, and students apply it in various ways. Now we are ready to prove it.

We will need the following Lemma. (The name is again due to Walk.)

Lemma. (Aura Theorem) Let f be continuous at a.

- (a) If $f(a) > 0$, then $f(x) > 0$ for all x in some open interval containing a.
- (b) If $f(a) < 0$, then $f(x) < 0$ for all x in some open interval containing a.

Proof. (I prove part (a) and leave part (b) as an exercise.) Assume $f(a) > 0$. Then corresponding to $\frac{f(a)}{2} > 0$ there exists a corresponding $\delta > 0$ such that $|x - a| < \delta$ implies

$$\begin{aligned} |f(x) - (a)| < \frac{f(a)}{2} &\iff -\frac{f(a)}{2} < f(x) - f(a) < \frac{f(a)}{2} \\ &\iff 0 < \frac{f(a)}{2} < f(x) < \frac{3f(a)}{2}, \end{aligned}$$

hence $f(x) > 0$ for all $x \in (a - \delta, a + \delta)$. $\square$

Intermediate Value Theorem

We first prove a special case, from which the general case follows easily.

Theorem. (Bolzano's Theorem) Let f be a continuous function defined on $[a, b]$. If $f(a) < 0$ and $f(b) > 0$, then there exists $x \in [a, b]$ such that $f(x) = 0$.

Proof. Let $I_1 = [a_1, b_1] = [a, b]$. If $f(\frac{a_1+b_1}{2}) = 0$, we are done. Otherwise, f changes sign on either $[a_1, \frac{a_1+b_1}{2}]$ or $[\frac{a_1+b_1}{2}, b_1]$. Let $I_2 = [a_2, b_2]$ be the subinterval on which f changes sign and repeat. By the Nested Intervals Theorem, $\bigcap_{n=1}^{\infty} I_n = \{x\}$. Suppose $f(x) > 0$. By the Aura Theorem, f must be positive on an open interval containing x. Since $x = \sup\{a_n\}$, by the Capture Theorem this open interval must contain some a_m. But $f(a_m) < 0$, which is a contradiction. Similarly, $f(x)$ can't be negative, since if it were then it must be negative on an open interval containing $x = \inf\{b_n\}$, which must contain some b_k, but $f(b_k) > 0$. Hence, $f(x) = 0$. $\square$

If $f(a) > 0$ and $f(b) < 0$, then $g := -f$ satisfies the hypotheses of Bolzano's Theorem, and therefore $g(x) = -f(x) = 0$ for some $x \in (a, b)$, and therefore $f(x) = 0$.

General case: If $f(a) < c < f(b)$, then $g(x) := f(x) - c$ is continuous and satisfies the hypotheses of Bolzano's Theorem, so there exists some $x \in (a, b)$ such that $g(x) = f(x) - c = 0$ and therefore $f(x) = c$.

Boundedness Theorem

The Extreme Value Theorem follows from the Boundedness Theorem.

Theorem. (Boundedness Theorem) If f is continuous on $[a, b]$, then f is bounded on $[a, b]$.

We will need the following lemma.

Lemma. If f is continuous at a, then f is bounded on some open interval containing a.

Proof. Since f is continuous at a, corresponding to $1 > 0$, say, there exists $\delta > 0$ such that $|x - a| < \delta$ implies $|f(x) - f(a)| < 1$. That is, $x \in (a - \delta, a + \delta)$ implies $f(a) - 1 < f(x) < f(a) + 1$, which shows that f is bounded on the open interval $(a - \delta, a + \delta)$. □

Boundedness Theorem

Also, we have seen previously that when the intervals $I_n = [a_n, b_n]$ arise from bisection, any open interval containing x (where $\bigcap_{n=1}^{\infty} I_n = \{x\}$) necessarily contains an a_k and a b_ℓ. Since the intervals I_n are nested, this implies something stronger: namely, that such an open interval contains one of the intervals I_N.

To see this, note that there are three possibilities:

- if $k = \ell$, then the open interval contains I_k.
- if $k < \ell$, then the open interval contains $a_k \le a_{k+1} \le \cdots \le a_\ell \le b_\ell$, so the open interval contains I_ℓ.
- if $k > \ell$, then the open interval contains $a_k \le b_k \le \cdots \le b_{\ell+1} \le b_\ell$, so the open interval contains I_k.

Boundedness Theorem

Theorem. (Boundedness Theorem) If f is continuous on $[a, b]$, then f is bounded on $[a, b]$.

Proof. Let $I_1 = [a_1, b_1] = [a, b]$. Suppose f is continuous on $[a, b]$ but not bounded. Then f is either unbounded on $[a_1, \frac{a_1+b_1}{2}]$ or $[\frac{a_1+b_1}{2}, b_1]$ (since, otherwise, f would be bounded on their union and hence on all of I_1). Let I_2 be the subinterval on which f is unbounded and repeat. By the Nested Intervals Theorem, $\bigcap_{n=1}^{\infty} I_n = \{x\}$, where $x = \sup\{a_n\} = \inf\{b_n\}$. Since f is continuous at x, f is bounded on some open interval containing x. However, as we have seen, such an open interval contains one of the intervals I_N, which is a contradiction since f is unbounded on each I_N. Hence, f is bounded on $[a, b]$. □

Extreme Value Theorem

Theorem. (Extreme Value Theorem) A continuous function on $[a, b]$ attains both an absolute maximum and an absolute minimum on $[a, b]$.

The proof is more or less the standard one.

Proof. We first prove f has a maximum on $[a, b]$. Since f is continuous on $[a, b]$, by the Boundedness Theorem f is bounded on $[a, b]$. Since f is bounded, its image set is a nonempty subset of $\mathbb{R}$ which is bounded above, so by the Completeness Axiom it has a least upper bound. Let $M = \sup f([a, b])$. By definition of M, $f(x) \leq M$ for all $x \in [a, b]$. Suppose, by way of contradiction, that $f(x) < M$ for all $x \in [a, b]$. Then $g(x) := \frac{1}{M - f(x)}$ is continuous on $[a, b]$ and hence bounded on $[a, b]$ by the Boundedness Theorem again, so there exists $K > 0$ such that $\frac{1}{M - f(x)} \leq K$ for all $x \in [a, b]$. It follows that $f(x) \leq M - \frac{1}{K}$ for all $x \in [a, b]$, which says that $M - \frac{1}{K}$ is an upper bound for $f([a, b])$. Since $K > 0$, $M - \frac{1}{K} < M$, so this contradicts the fact that $M = \sup f([a, b])$. Hence, there must exist $c \in [a, b]$ such that $f(c) = M$.

To see that f has a minimum on $[a, b]$, just note that $g := -f$ is continuous on $[a, b]$ and therefore has a maximum on $[a, b]$ by the first part. This means there exists $c \in [a, b]$ such that, for all $x \in [a, b]$,
$g(x) \leq g(c) \iff -f(x) \leq -g(c) \iff f(x) \geq f(c)$, which says that $f(c)$ is a minimum. $\square$

Uniform Continuity Theorem

The Integrability Theorem follows from the Uniform Continuity Theorem.

Recall that a function is *uniformly continuous* on an interval I if for every $\epsilon > 0$ there exists a $\delta > 0$ such that, for all $x, y \in I$, $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$.

Theorem. If f is continuous on $[a, b]$ then f is uniformly continuous on $[a, b]$.

We will need the following lemma.

Lemma. Suppose $a \leq c \leq b$. Then, if f is continuous on $[a, b]$ and uniformly continuous on $[a, c]$ and $[c, b]$, then f is uniformly continuous on $[a, b]$.

Uniform Continuity Theorem

Lemma. Suppose $a \leq c \leq b$. Then, if f is continuous on $[a, b]$ and uniformly continuous on $[a, c]$ and $[c, b]$, then f is uniformly continuous on $[a, b]$.

Proof. Let $\epsilon > 0$. Since f is uniformly continuous on $[a, c]$ there exists $\delta_1 > 0$ such that $x, y \in [a, c]$ and $|x - y| < \delta_1$ implies $|f(x) - f(y)| < \epsilon$. Since f is uniformly continuous on $[c, b]$ there exists $\delta_2 > 0$ such that $x, y \in [b, c]$ and $|x - y| < \delta_2$ implies $|f(x) - f(y)| < \epsilon$. What if $x < c < y$ or $y < c < x$? To take care of this case, we use continuity of f. Since f is continuous at c, there exists $\delta_3 > 0$ such that $|x - c| < \delta_3$ implies $|f(x) - f(c)| < \frac{\epsilon}{2}$. It follows that if $|x - c| < \delta_3$ and $|y - c| < \delta_3$, then $|f(x) - f(y)| = |f(x) - f(c) + f(c) - f(y)| \leq |f(x) - f(c)| + |f(c) - f(y)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. Let $\delta = \min\{\delta_1, \delta_2, \delta_3\}$ and suppose $|x - y| < \delta$. If x and y are both in $[a, c]$ or $[c, b]$, then we have seen that $|f(x) - f(y)| < \epsilon$. Otherwise, either $x < c < y$ or $y < c < x$. In the first case, subtracting x gives $0 < c - x < y - x < |y - x| < \delta$. In each case $|x - y| < \delta$ implies $|x - c| < \delta$ and $|y - c| < \delta$, which implies $|f(x) - f(y)| < \epsilon$. □

Uniform Continuity Theorem

We can now use the Bisection Method to prove the Uniform Continuity Theorem.

Theorem. If f is continuous on $[a, b]$ then f is uniformly continuous on $[a, b]$.

Proof. Suppose f is continuous on $[a, b]$ but not uniformly continuous. Since f is not uniformly continuous on $[a, b]$, there exists $\epsilon > 0$ such that $|f(x) - f(y)| \geq \epsilon$ for all $x, y \in [a, b]$. Let $I_1 = [a_1, b_1] = [a, b]$. By the lemma, it follows that f is either not uniformly continuous on $[a_1, \frac{a_1+b_1}{2}]$ or not uniformly continuous on $[\frac{a_1+b_1}{2}, b_1]$. Let I_2 be the subinterval on which f is not uniformly continuous. Then $|f(x) - f(y)| \geq \epsilon$ for all $x, y \in I_2$. Continuing in this way, we obtain a sequence of intervals $I_1 \supseteq I_2 \supseteq \cdots$ satisfying the hypotheses of the Nested Intervals Theorem and such that, for each $n \in \mathbb{N}$, $|f(x) - f(y)| \geq \epsilon$ for all $x, y \in I_n$. By the Nested Intervals Theorem, $\bigcap_{n=1}^{\infty} I_n = \{x_0\}$, where $x = \sup\{a_n\} = \inf\{b_n\}$. Since f is continuous at x_0, corresponding to the ϵ above there exists $\delta > 0$ such that $|x - x_0| < \delta$ implies $|f(x) - f(x_0)| < \frac{\epsilon}{2}$. Then, for any $x, y \in (x_0 - \delta, x_0 + \delta)$, we have $|x - x_0| < \delta$ and $|y - y_0| < \delta$, so
$|f(x)-f(y)| = |f(x)-f(x_0)+f(x_0)-f(y)| \leq |f(x)-f(x_0)|+|f(x_0)-f(y)| < \frac{\epsilon}{2}+\frac{\epsilon}{2} = \epsilon.$
But, as we have seen, $(x_0 - \delta, x_0 + \delta)$ contains some I_N and $|f(x) - f(y)| \geq \epsilon$ for all $x, y \in I_N$, so we reach a contradiction. Hence, f is uniformly continuous on $[a, b]$. □

Integration

I introduce the integral axiomatically, following the notes of Pete Clark (http://alpha.math.uga.edu/~pete/2400full.pdf):

(I0) Continuous functions are integrable.
(I1) If $f = c$ is constant, then c is integrable and $\int_a^b c = c(b-a)$.
(I2) If f_1 and f_2 are integrable with $f_1(x) \leq f_2(x)$ for all $x \in [a,b]$, then $\int_a^b f_1 \leq \int_a^b f_2$.
(I3) If f is integrable and $c \in (a,b)$, then $\int_a^b f = \int_a^c f + \int_c^b f$.

The proof of Theorem 8.1 in Clark's notes shows that the Fundamental Theorem of Calculus follows from these axioms.

It remains then to construct the function $f \mapsto \int_a^b f$ and to describe the domain.

To prove that a continuous function is integrable, it is easier to use the Darboux integral than the Riemann integral. This is the route taken in Spivak's text. Equivalence of the two integrals is proved in Theorem 8.26 of Clark's notes.

The Darboux Integral

The Darboux integral is defined in terms of upper and lower sums, rather than Riemann sums. Let $P = \{a = x_0 < x_1 < x_2 < \cdots < x_n = b\}$ be a partition of $[a, b]$. For a bounded function on $[a, b]$, let

$$M_i = \sup\{f(x) : x_{i-1} \leq x \leq x_i\},$$
$$m_i = \inf\{f(x) : x_{i-1} \leq x \leq x_i\},$$

and define the upper sum and lower sums for f on $[a, b]$ relative to P by

$$U(f, P) = \sum_{i=1}^{n} M_i(x_i - x_{i-1})$$
$$L(f, P) = \sum_{i=1}^{n} m_i(x_i - x_{i-1}).$$

The upper and lower integrals of f on $[a, b]$ are then defined by

$$\overline{\int_a^b} f := \inf\{U(f, P) : P \text{ a partion of } [a, b]\}$$
$$\underline{\int_a^b} f := \sup\{L(f, P) : P \text{ a partion of } [a, b]\}.$$

A function is *integrable* if $\overline{\int_a^b} f = \underline{\int_a^b} f$, and we denote the common value by $\int_a^b f$.

The Darboux Integral

Clark proves in Theorem 8.8 that the Darboux Integral $\int_a^b f$ satisfies axioms (I1)-(I3), and therefore that the Fundamental Theorem of Calculus holds for the Darboux Integral.

Rather than working with sups and infs, an equivalent condition for integrability is given by Theorem 8.7 of Clark's notes:

Darboux's Integrability condition: A function is Darboux integrable if and only if for every $\epsilon > 0$ there exists a partition P_ϵ of $[a, b]$ such that $U(f, P_\epsilon) - L(f, P_\epsilon) < \epsilon$.

The Integrability Theorem

We are finally ready to prove the Integrability Theorem (Ch 13, Theorem 3, Spivak).

Theorem. (The Integrability Theorem) If f is continuous on $[a, b]$, then f is integrable on $[a, b]$.

Proof. First note that, since f is continuous on $[a, b]$, it is bounded on $[a, b]$, which we have required to be integrable. Let $\epsilon > 0$. Since f is continuous on $[a, b]$, it is uniformly continuous on $[a, b]$, so corresponding to $\frac{\epsilon}{b-a} > 0$ there exists $\delta > 0$ such that for all x and y in $[a, b]$, $|x - y| < \delta$ implies $|f(x) - f(y)| < \frac{\epsilon}{b-a}$. If we choose a partition P such that $|x_i - x_{i-1}| < \delta$ for each i, then for each i we have $|f(x) - f(y)| < \frac{\epsilon}{b-a}$ for all $x, y \in [x_{i-1}, x_i]$. By the Extreme Value Theorem, $M_i = f(x)$ and $m_i = f(y)$ for some $x, y \in [x_{i-1}, x_i]$, so in particular we have $M_i - m_i < \frac{\epsilon}{b-a}$ for each i, and therefore

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1}) \\ &< \frac{\epsilon}{b-a} \sum_{i=1}^{n} (x_i - x_{i-1}) \\ &= \frac{\epsilon}{b-a}(b - a) = \epsilon. \end{aligned}$$

□