Real Analysis Notes

Stan Tuznik

November 1, 2020

Contents

Chapter 1

Continuity

Prop 1.1. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map represented by the matrix $A \in \mathbb{R}^{m \times n}$. Then we have*

$$\|T(h)\| = \|Ah\| \le M \|h\|$$

for some constant M.

> *Proof.* Note that this is just a statement about matrix-vector multiplication. Let $[Ah]_i$ denote the ith entry of the matrix product $Ah \in \mathbb{R}^m$. That is,
>
> $$[Ah]_i = \sum_{j=1}^{n} A_{ij} h_j$$
>
> Then let
>
> $$r = \max_{i,j} |A_{ij}|$$
>
> Also note that since
>
> $$\|h\|^2 = \sum_{j=1}^{n} |h_j|^2$$
>
> we have
>
> $$|h_j| \le \|h\|$$
>
> Taken together, we have
>
> $$[Ah]_i = \sum_{j=1}^{n} A_{ij} h_j \le \sum_{j=1}^{n} |A_{ij}| |h_j| \le rn \|h\|$$
>
> so that
>
> $$\|Ah\|^2 = \sum_{i=1}^{m} [Ah]_i^2 \le \sum_{i=1}^{m} r^2 n^2 \|h\|^2 = m r^2 n^2 \|h\|^2$$
>
> and so we have
>
> $$\|Ah\| \le rn\sqrt{m} \|h\|$$
>
> Letting $M = rn\sqrt{m}$, the conclusion holds. ∎

Lemma 1.2. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map represented by the matrix $A \in \mathbb{R}^{m \times n}$. Then for $a \in \mathbb{R}^n$, T is continuous at a.*

Proof. We must show that

$$\lim_{h \to 0} \|T(a + h) - T(a)\| = 0$$

Note that, by linearity,

$$\|T(a + h) - T(a)\| = \|T(a) + T(h) - T(a)\| = \|T(h)\| \le M \|h\|$$

and the right-hand side clearly goes to 0 as $h \to 0$. $\blacksquare$

The next seems to be an alternative approach. For arbitrary linear map $A \in \mathbb{R}^{m \times n}$, we can define a matrix norm

$$\|A\| := \sum_{i=1,\ldots,m} \sum_{j=1,\ldots,n} A_{i,j}^2$$

This is technically the square of the element-wise matrix p-norm:

$$\|A\|_p = \left(\sum_{\text{rows}} \sum_{\text{cols}} A_{i,j}^p \right)^{1/p}$$

Then we can look at bounds on matrix-vector and matrix-matrix products. In fact, the following proposition is a generalization of the one above.

Prop 1.3. *Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, and $\vec{b} \in \mathbb{R}^n$. Then*

$$\left\| A\vec{b} \right\| \le \|A\| \left\| \vec{b} \right\|$$

$$\|AB\| \le \|A\| \|B\|$$

Proof. For the first inequality, notice that the magnitude on the left is a vector norm. That is,

$$\left\| A\vec{b} \right\|^2 = \sum_{i=1}^m \left(A\vec{b} \right)_i^2$$

and that each entry in this vector is the inner product (usual Euclidean dot product) of $\vec{b}$ with a row of A:

$$\left(A\vec{b} \right)_i = A_{i*}\vec{b}$$

Since this is an inner product, we may apply the Cauchy-Schwarz inequality to it and obtain

$$\left(A\vec{b} \right)_i = A_{i*}\vec{b} \le \left| A_{i*}\vec{b} \right| \le \|A_{i*}\| \left\| \vec{b} \right\|$$

Taking this in the original vector magnitude, we have

$$\left\| A\vec{b} \right\|^2 = \sum_{i=1}^{m} \left(A\vec{b} \right)_i^2$$

$$\leq \sum_{i=1}^{m} \left\| A_{i*} \right\|^2 \left\| \vec{b} \right\|^2$$

$$= \left(\sum_{i=1}^{m} \left\| A_{i*} \right\|^2 \right) \left\| \vec{b} \right\|^2$$

$$= \left\| A \right\|^2 \left\| \vec{b} \right\|^2$$

and so $\left\| A\vec{b} \right\| \leq \left\| A \right\| \left\| \vec{b} \right\|$ as we wanted to prove.

The next inequality is even more general and follows from the matrix-vector product case. Note that every column of the matrix product AB is of the form $A\vec{b}$ where $\vec{b} = B_{*j}$. Thus

$$\left\| AB \right\|^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(AB \right)_{i,j}^2$$

$$= \sum_{j=1}^{n} \left| AB_{*j} \right|^2$$

$$\leq \sum_{j=1}^{n} \left\| A \right\|^2 \left\| B_{*j} \right\|^2$$

$$= \left\| A \right\|^2 \left(\sum_{j=1}^{n} \left\| B_{*j} \right\|^2 \right)$$

$$= \left\| A \right\|^2 \left\| B \right\|^2$$

and so

$$\left\| AB \right\| \leq \left\| A \right\| \left\| B \right\|$$

as desired.

∎

The second inequality defines the matrix norm $\left\| \cdot \right\|$ (and more generally the matrix 2-norm) as a *submultiplicative* norm. With these general properties in hand, we can easily verify continuity of a linear function.

Proof. The linear function $T : \mathbb{R}^n \to \mathbb{R}^m$ will be continuous if we have

$$\lim_{x \to a} T\left(x \right) = T\left(a \right)$$

Let $\epsilon > 0$ be arbitrary. Then

$$\left| \vec{x} - \vec{a} \right| < \epsilon / \left\| A \right\|$$

4

implies that

$$\|T(\vec{x}) - T(\vec{a})\| = \|A\vec{x} - A\vec{a}\|$$
$$= \|A(\vec{x} - \vec{a})\|$$
$$\leq \|A\| \, \|\vec{x} - \vec{a}\|$$
$$< \epsilon$$

Note that linearity of T allowed us to express its action as a matrix multiplication, and the associative (linearity) property of matrix multiplication allowed us to "pull it off" of the $\vec{x}$ and $\vec{a}$ and onto their vector difference. Then, the previous proposition allowed us to bound this matrix-vector product by the norm of the vector and a quantity which measures in some sense the size of the matrix A. ∎

Chapter 2

The Derivative

2.1   Local Properties

Just as with continuity, the derivative is a local notion: its definition at a point depends only on the behavior of f on a small neighborhood of that point. Recall that we were able to use the continuity of f to make certain claims about the local behavior of f. For example, we showed rather easily that if $f(x) > 0$, then $f(t) > 0$ for t in some neighborhood of x. Since a differentiable function is continuous, these claims still hold. We can go further, now, since the derivative gives us a bit more information. Let's compare the definitions to try to understand this.

The continuity of f at x says that for any $\epsilon > 0$, there exists a $\delta > 0$ so that

$$|t - x| < \delta \implies |f(t) - f(x)| < \epsilon$$

That is, continuity at a point x tells us that the function values can be made arbitrarily close to the function value $f(x)$ — i.e., what we would expect — by restricting t to be sufficiently close to x in the domain.

Differentiability of f at x says that there exists a number $f'(x) \in \mathbb{R}$ so that for any $\epsilon > 0$, there is a $\delta > 0$ with

$$|t - x| < \delta \implies \left| \frac{f(t) - f(x)}{t - x} - f'(x) \right| < \epsilon$$

This definition tells us that the slope of the secant line can be made arbitrarily close to the number $f'(x)$ by restricting us to be sufficiently close to x in the domain.

Thus, continuity deals with the function values directly while differentiability deals with rates of change of functions.

The key difference for many applications of differentiability is that knowledge about the derivative $f'(x)$, that is, the number, will inform us about possible function values $f(t)$ at t nearby x; the notion of "locality" again rears its head.

Intuitively, since $f'(x)$ is the limiting slope of the secant line "at x," we expect that $f'(x) > 0$ implies that $f(t) > f(x)$ for $t > x$ in some small region. We can make this official.

Prop 2.1. *Let f be differentiable at $x \in \mathbb{R}$ and assume that $f'(x) > 0$. Then $f(t) > f(x)$ for $t \in (x, x + \delta)$ for some sufficiently small $\delta > 0$. Similarly, $f(t) < f(x)$ for $t \in (x - \delta, x)$.*

Proof. This will be an almost obvious consequence of the definition of derivative. Since $f'(x) > 0$, we have for arbitrary $\epsilon > 0$ some $\delta > 0$ with

$$|t - x| < \delta \implies \left| \frac{f(t) - f(x)}{t - x} - f'(x) \right| < \epsilon$$

Since we want to compare $f(t)$ and $f(x)$, we see them in the numerator and wish to get rid of the absolute value. We can do this by explicitly writing out what the absolute value means:

$$-\epsilon < \frac{f(t) - f(x)}{t - x} - f'(x) < \epsilon$$

or

$$f'(x) - \epsilon < \frac{f(t) - f(x)}{t - x} < f'(x) + \epsilon$$

Since, ϵ can be made arbitrarily small, and since $f'(x) > 0$, we see that we can make the left-most terms positive by suitably restricting t; that is, we appeal to the definition of derivative again and choose $\epsilon = \frac{1}{2}f'(x)$ and choose the appropriate δ. Thus, if $|t - x| < \delta$, then we now have

$$0 < \frac{1}{2}f'(x) < \frac{f(t) - f(x)}{t - x}$$

Lastly, we can restrict $t > x$ so that $0 < t - x < \delta$, or $x < t < x + \delta$, and we have

$$0 < f(t) - f(x)$$

which is what we sought to prove. On the other hand, if we let $t < x$ so that $-\delta < t - x < 0$, or $x - \delta < t < x$, we have

$$0 > f(t) - f(x)$$

which was the other part of what we want to prove. ∎

In this example, we used information about how the function is changing at a point, instantaneously, to inform us about the relative values of the function near that point. With only continuity, we were unable to make such comparisons; we only knew that function values were arbitrarily close by.

Another extremely helpful conclusion is that the derivative of a function is zero at a local extremum. This is particularly helpful in identifying potential local extrema of functions.

Prop 2.2. *Let f have a local maximum at x. If f is differentiable at x, then we have $f'(x) = 0$.*

Proof. We prove this directly. The proof would naturally proceed very similarly to that of the previous proposition; we instead simply use the results.

Assume f has a local maximum at x, then we have $f(t) < f(x)$ for all t sufficiently near x. If $f'(x) > 0$, then we would have $f(t) > f(x)$ for all $t \in (x, x + \delta)$, by the

7

previous proposition. This would contradict the local maximum at x, so we know that $f'(x) \le 0$.

On the other hand, if $f'(x) < 0$, then we would have $f(t) > f(x)$ for all $t \in (x - \delta, x)$, by (an unproven analogue of) the previous proposition. This would again contradict the local maximum of f at x, so we must have $f'(x) \ge 0$.

Taken together, we see that we must have $f'(x) = 0$. ∎

2.2   The Mean-value Theorem

A key result in elementary real analysis is the mean value theorem for derivatives. The idea is that we start with a closed interval $[a, b] \subset \mathbb{R}$ and consider the values of a function f which is continuous on $[a, b]$ and differentiable on (a, b). Given the function values at the interval endpoints, $f(a)$ and $f(b)$, we can draw a secant line between the endpoints $(a, f(a))$ and $(b, f(b))$. This secant line naturally has the slope

$$\frac{f(b) - f(a)}{b - a}$$

and the mean-value theorem asserts that we can always find some point within the interval, say $c \in (a, b)$, where the derivative of the function f equals this slope of the secant line. As it will turn out, this theorem has numerous important applications (including the fundamental theorem of calculus), so it is important to understand.

First, we prove a special case in which the secant line is horizontal (the slope is zero).

Theorem 2.3 (Rolle). *Let f be continuous on $[a, b]$ and differentiable on (a, b), and $f(a) = f(b)$. Then there is some $c \in (a, b)$ with*

$$f'(c) = 0$$

Proof. If f is constant on $[a, b]$, then we are done, since every interior point of $[a, b]$ will have zero derivative.

On the other hand, assume without loss of generality that there is some point $d \in (a, b)$ with $f(d) > f(a)$.

Since f is a continuous function on a compact set, the image $f([a, b])$ is also compact; since compact subsets of $\mathbb{R}$ are exactly those which are closed and bounded, $f([a, b])$ is closed and bounded. Hence, f is bounded above on $[a, b]$. Let $M = \sup_{x \in [a,b]} f(x)$. Since this image is closed, $M \in f([a, b])$, i.e., this supremum is actually a maximum of f which is attained at some point, say c, in $[a, b]$: $f(c) = M$. Also, note that since we have $f(d) > f(a)$, we know that $c \ne a$ and $c \ne b$, so it occurs in the interior of the closed interval. Since f has a maximum at c, we must have $f'(c) = 0$, and we are done.

The logic proceeds similarly if we considered the existence of some $d \in (a, b)$ with $f(d) < f(a)$, in which case we would deal with an infimum and obtain the same

8

conclusion. ∎

We can use this special case to prove a more general version:

Theorem 2.4 (Mean-value Theorem). *Let f be continuous on $[a, b]$ and differentiable on (a, b). Then there is some $c \in (a, b)$ with*

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

Proof. The standard trick is to look at an auxiliary function to which we can apply Rolle's theorem, but which involves the derivative f' of f, so that we might extract the desired result. For example, we might try letting $h(t) = \alpha f(t)$ where α is a constant to be determined. Then we have $h'(t) = f'(t)$. Unfortunately, we do not have $h(a) = h(b)$, so we cannot apply Rolle's theorem to this choice of h.

Instead, we may try a function such as

$$h(t) = f(t) - \alpha t$$

which has derivative

$$h'(t) = f'(t) - \alpha$$

By clever choice of α, we may now ensure that Rolle's theorem applies to h: $h(a) = h(b)$ requires

$$f(a) - \alpha a = f(b) - \alpha b$$

or

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

With this choice of α, Rolle's theorem applies to h and so we have a $c \in (a, b)$ with

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

which proves the theorem. ∎

Theorem 2.5 (Extended Mean-value Theorem). *Let f, g be continuous on $[a, b]$ and differentiable on (a, b). Then there is some $c \in (a, b)$ with*

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

Proof. Note that if we divide both sides of the result by $b - a$, we obtain

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

9

which reduces to the standard mean-value theorem result when $g(t) = t$ is the identity function. We use nearly the same technique to prove this version: come up with a suitable function to which Rolle's theorem applies, and which gives us the final result.

First, assume that $g(a) \neq g(b)$; otherwise, the result follows directly from Rolle's theorem applied to g. Then let

$$h(t) = f(t) - \alpha g(t).$$

To force Rolle's theorem to apply, we need $h(a) = h(b)$, which requires

$$\alpha = \frac{f(b) - f(a)}{g(b) - g(a)}$$

which is where the assumption that $g(b) \neq g(a)$ becomes handy. Then we apply Rolle's theorem to h to conclude that there is a $c \in (a, b)$ such that

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

and the proof is complete.

(Note that we could also have used

$$h(t) = [f(b) - f(a)] g(t) - [g(b) - g(a)] f(t)$$

as is done in [Rud64].) $\blacksquare$

2.3 Taylor's Theorem

Chapter 3

Integration

3.1 Darboux Integral (Green-blue)

One of the most rudimentary types of integrals is defined in terms of Darboux sums. This is the definition of integral most commonly found in introductory calculus courses, but formalized, and is equivalent to the Riemann integral (discussed next). Upper and lower Darboux sums are computed over a partition, and the infimum and supremum of these, respectively, are taken and are the upper and lower Darboux integrals. When these are equal, the function is Darboux integrable, and the Darboux integral is the common value.

3.2 Riemann Integral (Red)

The Riemann integral, while equivalent to the Darboux integral, is usually formulated in a more complicated manner. We start with a function on a closed interval, together with a tagged partition of that interval. A tagged partition is a normal partition which also comes with a set of points, each of which lies in a distinct interval of the partition. The Riemann sum of the function over this interval with respect to the tagged partition is

$$\sum_{i=1}^{n} f\left(t_i\right)\left(x_i - x_{i-1}\right)$$

(Note that the Riemann sum is related to the Darboux sums, which instead of the tagged function values $f\left(t_i\right)$ use the supremum and infimum of the function over the intervals. In cases where f is continuous on the intervals, the intervals (being compact) map to a compact image under the continuous f, and so the image is closed and bounded; f thus attains its maximum and minimum on each interval, and we can take these values as the tags to obtain a Riemann sum equal to the upper and lower Darboux sums, respectively.)

The integral involves, as expected, a limiting process over smaller and smaller partitions. We must make this precise. Similar to the derivative, the Riemann integral of f over an interval is usually defined as some number, I, such that we can make the Riemann sums arbitrarily close to I by restricting the mesh of the partition (the maximum subinterval

length) to be less than some value. This may look something like the following:

$$I = \int_a^b f(x)\, dx := \lim_{n \to \infty} \sum_{k=1}^n f(x_k^*)\, \Delta x_k$$

and where $\max \Delta x_k \to 0$ as $n \to \infty$. More consistent with the $\epsilon - \delta$ definition is the following: f is Riemann integrable with Riemann integral I if for any $\epsilon > 0$ there is some $\delta > 0$ with

$$\text{mesh}\,(P) < \delta \implies |RS\,(f, P) - I| < \epsilon$$

where P indicates a partition and $RS\,(f, P)$ is the Riemann sum of f over the partition P.

Note that this definition makes no mention of a particular partition: it claims that as we shrink any partition, we eventually get close to the value.

Prop 3.1. *The Riemann integral of $f : [a, b] \to \mathbb{R}$, when it exists, is uniquely defined.*

Proof. The proof is extremely similar to that of the uniqueness of functional limits. Assume by way of contradiction that there are two integrals, say I and I'. Then let $\epsilon > 0$. By the definition of the integral, we know that we have $\delta > 0$ so that any tagged partition P with $\text{mesh}\,(P) < \delta$ has $|RS\,(f, P) - I| < \frac{\epsilon}{2}$. Similarly, there is a $\delta' > 0$ so that any tagged partition P with $\text{mesh}\,(P) < \delta'$ has $|RS\,(f, P) - I'| < \frac{\epsilon}{2}$. Not surprisingly, take $\text{mesh}\,(P) < \min\{\delta, \delta'\}$ and obtain

$$|I - I'| \le |RS\,(f, P) - I| + |RS\,(f, P) - I'| < 2\frac{\epsilon}{2} = \epsilon$$

which proves that $I = I'$, since $\epsilon > 0$ can be made arbitrarily small. ∎

Let's try to work through an example, to see this definition in action.

Prop 3.2. $\int_a^b x\, dx = \frac{b^2 - a^2}{2}$ *for* $0 \le a < b$.

Proof. Let $P = \{[x_{i-1}, x_i]\}_{i=1}^n$ be an arbitrary tagged partition of $[a, b]$. Given $\epsilon > 0$, we wish to choose a $\delta > 0$ to force $\left| RS\,(f, P) - \frac{b^2 - a^2}{2} \right| < \epsilon$ for $\text{mesh}\,(P) < \delta$. ∎

Theorem 3.3. *a function $f : [a, b] \to \mathbb{R}$ is Darboux integrable if and only if it is Riemann integrable. If so, the integrals are equal.*

3.3 Riemann-Stieltjes Integral

This is the integral which is introduced in [Rud64].

3.4 Lebesgue Integral

Chapter 4

Fundamental Theorems

4.1 Fundamental Theorem of Calculus

Theorem 4.1 (FTOC I [Rud64]). *Let $f : [a,b] \to \mathbb{R}$. Then*

$$F(x) = \int_a^x f(t)\, dt$$

is continuous on $[a,b]$ (as a function of x). If f continuous at $x_0 \in [a,b]$, then F is differentiable at x_0 with

$$F'(x_0) = f(x_0).$$

Proof. Since we assume f is finite, it is bounded on $[a,b]$. Let M be such a bound. We want to show continuity of F at an arbitrary point, say $x \in [a,b]$. The definition of continuity tells us to look at

$$|F(t) - F(x)| = \left| \int_x^y f(t)\, dt \right| \leq M\,|y - x| < \epsilon$$

where we used a property of integrals for the central inequality. Thus, taking $\delta = \epsilon/M$ gives the desired continuity of F.

Next, we assume that f is continuous at x_0. We want to prove a derivative of F, so let's resort to the definition; let $\epsilon > 0$, and consider

$$\left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| = \left| \frac{\int_{x_0}^x f(u)\, du}{x - x_0} - (x - x_0)\frac{f(x_0)}{x - x_0} \right| \tag{4.1}$$

$$= \left| \frac{\int_{x_0}^x f(u)\, du}{x - x_0} - \frac{\int_{x_0}^x f(x_0)\, du}{x - x_0} \right| \tag{4.2}$$

$$= \left| \frac{\int_{x_0}^x [f(u) - f(x_0)]\, du}{x - x_0} \right| \tag{4.3}$$

where $u \in [x_0, x]$ is the dummy variable of integration. Note that the difference in the integrand is similar to what is usually seen in the definition of continuity; further, we are assuming that f is continuous at x_0. Hence, we can make u close to x_0 by making x, the upper limit of integration, close to x_0. Pick δ so that

$$|x - x_0| < \delta \implies |f(x) - f(x_0)| < \epsilon.$$

Then we have

$$\left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| = \left| \frac{\int_{x_0}^{x} [f(u) - f(x_0)] \, du}{x - x_0} \right| \tag{4.4}$$

$$< \frac{\epsilon |x - x_0|}{|x - x_0|} = \epsilon \tag{4.5}$$

This is the definition of derivative, and we have shown that

$$F'(x_0) = f(x_0).$$

∎

Corollary 4.1.1 ([Spi08]). *Let $f : [a, b] \to \mathbb{R}$ be continuous and assume we have a function F such that $f = F'$. Then*

$$\int_a^b f(t) \, dt = F(b) - F(a).$$

Proof. Write $G(x) = \int_a^x f(t) \, dt$. Then we have $G'(x) = f(x) = F'(x)$ by the FTOC I, so that we have $G = F + c$ for some $c \in \mathbb{R}$. Since $G(a) = 0$, $c = -F(a)$. Hence, we have

$$G(x) = F(x) - F(a) = \int_a^x f(t) \, dt$$

which is what we sought to prove. ∎

Note: this should not be confused with a definition for the integral. In fact, it may be that a function is integrable but not the derivative of another function. Thus, we should think of this as a shortcut for computing an integral: if we know an antiderivative for f, then we can compute the integral easily.

Theorem 4.2 (FTOC II [Rud64]). *Let $f : [a, b] \to \mathbb{R}$ be integrable and assume we have a function F, differentiable on $[a, b]$, such that $F' = f$. Then we have*

$$\int_a^b f(t) \, dt = F(b) - F(a).$$

Note: this looks similar to the previous corollary, but differs in that we do not assume that f is continuous! Thus, this is a stronger result, since we have a weaker hypothesis (since integrability does not necessitate continuity).

Proof. Our goal is to compute the integral on the left in terms of the antiderivative F. We will apply the mean-value theorem to F on each subinterval of a partition. Let $\epsilon > 0$ and $P = \{[x_{i-1}, x_i]\}_{i=1}^{n}$ be a tagged partition of $[a, b]$ such that

$$\left| \sum_{i=1}^{n} f(t_i) \Delta x_i - \int_a^b f(x) \, dx \right| < \epsilon$$

(this is just the definition of Riemann integrability of f).

Next, consider the partition subinterval $[x_{i-1}, x_i]$. Then we have the tags $t_0, t_1, \ldots, t_n$ so that

$$f(t_i) = F'(t_i) = \frac{F(x_i) - F(x_{i-1})}{\Delta x_i}$$

or, upon rearranging, we have

$$f(t_i) \Delta x_i = F(x_i) - F(x_{i-1})$$

and we have the telescoping sum

$$\sum_{i=1}^{n} f(t_i) \Delta x_i = F(b) - F(a).$$

Ultimately, then, we have

$$\left| F(b) - F(a) - \int_a^b f(x) \, dx \right| < \epsilon$$

where $\epsilon > 0$ is arbitrary. Thus, $\int_a^b f(x) \, dx = F(b) - F(a)$, and we are done. $\blacksquare$

Bibliography

[Rud64] Walter Rudin. *Principles of mathematical analysis.* McGraw-Hill New York, 3 edition, 1964.

[Spi08] Michael Spivak. *Calculus.* Publish or Perish, Inc. Houston, Texas, 2008.