Real Analysis

Supplementary course notes for MA338

Evan Randles

Draft date May 8, 2024

Cover Image by Meredith Green

Contents

Preface

These are the course notes for Real Analysis (MA 338). Many of the ideas presented in these notes are not original and, at least in part, have been influenced by a number of excellent texts on the subject, including *Principles of Mathematical Analysis* by W. Rudin [**?**], *The Way of Analysis* by R. S. Strichartz [**?**], *Mathematical Analysis* by T. M. Apostol [**?**], and *An Introduction to Analysis* by W. R. Wade [**?**]. As these notes represent an active working draft, please update/download them frequently as I will often make corrections and changes without explicit warning. Any block of text or word in red is just a note for me (one I'm leaving to myself while editing); these should be disregarded. Words in blue are often hyperlinks to an interesting reference, usually a video, and you should click on them if you're reading along digitally. Also, if you find or suspect an error or typo – no matter how trivial – please email me to let me know!

Acknowledgment:

These course notes are a growing document and are, at present, pretty rough. I owe a great deal of gratitude to my students, who have served as test subjects for these notes as well as editors and proof readers. In particular, I would like to thank Cole Hoffman, Simge Ildes, Lora LaRochelle, Lucas Lefebvre, Jasper Loverude, Alexander Lyon, and Grace Moberg. Of course, I am responsible for the errors that inevitably remain.

Chapter 1

The Real Numbers

Chapter 2

Point-set Topology

Chapter 3

Numerical Sequence and Series

Chapter 4

Continuity

Chapter 5

Differentiation

5.1 Real-valued functions of a real variable

5.2 Differentiability of functions from $\mathbb{R}^d$ to $\mathbb{R}^{d'}$

5.3 Complex functions of a real variable

We will soon integrate complex-valued functions of a real variable, e.g., functions $f : I \to \mathbb{C}$ where $I = [a, b]$. As we discussed previously in the course, $\mathbb{C}$ is simply $\mathbb{R}^2$ with an additional multiplication structure. Its metric is given by the norm/modulus

$$|z| = |a + ib| = |(a, b)| = \sqrt{a^2 + b^2}$$

for $z = a + ib \in \mathbb{C}$. The following proposition simply translates our general notion of continuity (for functions between metric spaces) into the context of the complex modulus and the real and imaginary parts of a complex-valued function.

Proposition 5.1. *Let $I \subseteq \mathbb{R}$ be an interval*[1] *and let $f : I \to \mathbb{C}$. We write $f = u + iv$ where $u = \mathrm{Re}(f)$ and $v = \mathrm{Im}(f)$ are the real and imaginary parts of f, respectively, both of which are necessarily real-valued functions on I.*

1. *For a point $x_0 \in I$, f is continuous at x_0 if, for all $\epsilon > 0$, there is a $\delta = \delta(\epsilon, x)$ for which*

$$|f(x) - f(x_0)| = \sqrt{(u(x) - u(x_0))^2 + (v(x) - v(x_0))^2} < \epsilon$$

   whenever

$$|x - x_0| < \delta.$$

2. *For a point $x_0 \in I$, f is continuous at x_0 if and only if its real and imaginary parts are continuous at x_0. In this case,*

$$f(x_0) = \lim_{x \to x_0} f(x) = \left(\lim_{x \to x_0} u(x)\right) + i \left(\lim_{x \to x_0} v(x)\right) = u(x_0) + iv(x_0).$$

3. *f is continuous on I if and only if its real and imaginary parts are continuous on I.*

4. *f is uniformly continuous on I if and only if its real and imaginary parts are uniformly continuous on I.*

[1] That is, $I = (a, b)$, $(a, b]$, $[a, b)$, or $[a, b]$) where $-\infty \le a < b \le \infty$.

As an exercise, you should prove (or convince yourself that you could prove) the proposition above. Let's now talk about differentiability. Viewing $\mathbb{C}$ as $\mathbb{R}^2$, we can recognize the real and imaginary parts of $f : I \to \mathbb{C}$ as the components of f, i.e., $f = (\mathrm{Re}(f), \mathrm{Im}(f))^\top$. In this sense, f is differentiable at $x_0 \in I$ if

$$f(x_0 + h) = f(x_0) + Df(x_0)h + \mathcal{E}(h)|h|$$

where $\mathcal{E}(h) \to 0$ as $h \to 0$ where Df is a 2×1 column vector consisting of the "partial" derivatives of the components of f. The following proposition connects our vector-valued notion of differentiability to a (new) complex-valued one. While it might appear obvious, the proposition is stronger than that which guarantees the existence of partial derivatives (Theorem 9.17 of Rudin) we discussed in class.

Proposition 5.2. *Let $f : I \to \mathbb{C}$ where I is an interval. Given $x_0 \in I$, f is differentiable at x_0 if and only if $u = \mathrm{Re}(f)$ and $v = \mathrm{Im}(f)$ are differentiable (as real-valued functions) at x_0. In this case,*

$$\lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \left(\lim_{h\to 0} \frac{u(x_0 + h) - u(x_0)}{h}\right) + i \left(\lim_{h\to 0} \frac{v(x_0 + h) - v(x_0)}{h}\right) = u'(x_0) + iv'(x_0).$$

We shall recognize the above limit as the derivative of f at x_0 (instead of the (equivalent) 2×1 derivative matrix) and denote it by $f'(x_0)$ or $\frac{df}{dx}(x_0)$.

Exercise 5.1:

Let $f : I \to \mathbb{C}$ where I is an interval[a].

1. Prove the proposition above.
2. Assume that f and g are complex-valued functions on I, both of which are differentiable at x_0. Use the proposition (and your knowledge of the algebra of derivatives of real-valued functions of a real variable) to prove the following statements:
   (a) For $z = a + ib \in \mathbb{C}$, the function $x \mapsto zf(x)$ is differentiable at x_0 with derivative $(zf)'(x_0) = zf'(x_0)$.
   (b) $f + g$ is differentiable at x_0 with $(f + g)'(x_0) = f'(x_0) + g'(x_0)$.
   (c) fg is differentiable at x_0 with $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$.

[a] You may assume I is open for simplicity.

Exercise 5.2:

In this exercise, you can assume that the sine and cosine functions are everywhere differentiable on $\mathbb{R}$, have the derivatives you know them to have, and satisfy the trigonometric identities $\cos(A+B) = \cos(A)\cos(B) - \sin(A)\sin(B)$ and $\sin(A + B) = \sin(A)\cos(B) + \sin(B)\cos(A)$. With this, define $\mathrm{Cis} : \mathbb{R} \to \mathbb{C}$ by

$$\mathrm{Cis}(\theta) = \cos(\theta) + i\sin(\theta)$$

for $\theta \in \mathbb{R}$.

1. Show that $|\mathrm{Cis}(\theta)| = 1$ for all $\theta \in \mathbb{R}$.
2. Show that $\mathrm{Cis}(\theta_1 + \theta_2) = \mathrm{Cis}(\theta_1)\,\mathrm{Cis}(\theta_2)$.
3. Using the previous proposition, show that Cis is differentiable at $\theta_0 = 0$ and $\mathrm{Cis}'(0) = i$.
4. Use the above to show that Cis is everywhere differentiable and $\mathrm{Cis}'(\theta) = i\,\mathrm{Cis}(\theta)$ for all $\theta \in \mathbb{R}$.

5. It is customary to write $\mathrm{Cis}(x) = e^{i\theta}$ (a fact which will be later justified by series) and, henceforth, we shall adopt this notation completely. In this new notation, write out all conclusions to the above four items.

6. Show that every complex number $z \in \mathbb{C}$ can be written as
$$z = |z|\, e^{i\theta}$$
form some $\theta \in (-\pi, \pi]$, called the *phase*[a] of z. Note here

[a]When $z \neq 0$, θ can be shown to be unique in this range.

5.3.1 Some Notation

We have recently been talking about continuous and differentiable functions. It's helpful to give some notation to collections of such functions; we shall later come back and discuss metrics and norms on them.

Definition 5.3. *Let X and Y be non-empty sets.*

1. *We say that a function a real or complex-valued function f on X is bounded proved that*
$$\|f\|_\infty := \sup_{x\in X} |f(x)| < \infty.$$
We shall denote the collection of bounded real and complex-valued function on X by $B(X;\mathbb{R})$ and $B(X;\mathbb{C})$ respectively. When the context of $\mathbb{R}$ or $\mathbb{C}$ is made clear, we may simply write $B(X)$ to denote the relevant choice of these sets.

2. *In the case that X and Y are metric spaces (with metrics d_X and d_Y), we denote by $C^0(X;Y)$ the set of continuous functions $f : X \to Y$, i.e.,*
$$C^0(X;Y) = \{f : X \to Y \mid f \text{ is continuous on } X\}.$$
We shall pay special attention to the cases in which $Y = \mathbb{R}$ or $\mathbb{C}$.

3. *In the case that $X = I = [a,b]$, we shall denote by $C^n(I;\mathbb{R})$ the set of functions f on X which are n-times differentiable and*
$$f^{(n)} = \frac{d^n f}{dx^n} \in C^0(I;\mathbb{R}).$$
Similarly, $C^n(I;\mathbb{C})$ is the set of complex-valued functions f on I with $f^{(n)} = \frac{d^n f}{dx^n} \in C^0(I;\mathbb{C})$. When the context is clear, we may drop the second entry and simply write $C^n(I)$ to mean $C^n(I;\mathbb{R})$ or $C^n(I;\mathbb{C})$.

4. *Finally, the set of **smooth** real-valued functions on I is*
$$C^\infty(I;\mathbb{R}) = \bigcap_{n=1}^{\infty} C^n(I;\mathbb{R})$$
*and, similarly, the set of **smooth** complex-valued functions on I is*
$$C^\infty(I;\mathbb{C}) = \bigcap_{n=1}^{\infty} C^n(I;\mathbb{C}).$$

Chapter 6

The Riemann-Darboux integral

In this chapter, we cover the basic properties of the Riemann-Darboux integral, whose name gives homage to Bernhard Riemann and Jean Gaston Darboux. As stated in lecture, it turns out that even this integral –the integral you've known and studied since your first brush with calculus – is insufficient for a comprehensive theory of analysis. To treat the comprehensive theory, in earnest, one needs the Lebesgue theory of integration. Though we will try to explore the necessity of Lebesgue integration later (while illustrating the shortcomings of the Riemann-Darboux integral), we first need to lay the groundwork for the Riemann-Darboux integral. This is the subject to which we now turn.

6.1 The Riemann-Darboux Integral for Real-Valued Functions

Definition 6.1. *Consider an interval $I = [a, b]$ where $-\infty < a < b < \infty$.*

1. *A **partition** P of I is a finite subset $P = \{x_0, x_1, x_2, \ldots, x_K\}$ of I such that*

$$a = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_K = b.$$

2. *Given such a partition P, we shall write*

$$\Delta x_k = x_k - x_{k-1}$$

*for $k = 1, 2, \ldots, K$. The **norm** or **size** of the partition is, by definition,*

$$\|P\| = \max_{k=1,2,\ldots,K} \Delta x_k.$$

3. *If P and Q are partitions of I, we say that Q is a **refinement of** P if $P \subseteq Q$.*

Though a partition P is simply a finite subset of I (which is enumerated, increasing, and includes both endpoints), you should picture P as dividing up the interval I into the subintervals $[x_{k-1}, x_k]$ of length Δx_k for $k = 1, 2, \ldots, K$.

Definition 6.2. *Given a bounded real-valued function $f \in B(I)$ and a partition P of I, define*

$$m_k = \inf_{x_{k-1} \le x \le x_k} f(x) \qquad \textit{and} \qquad M_k = \sup_{x_{k-1} \le x \le x_k} f(x)$$

*for each $k = 1, 2, \ldots, k$. With these, we define the **upper and lower Darboux sums** of f with respect to the partition P respectively by*

$$U(f, P) = \sum_{k=1}^{K} M_k \Delta x_k \qquad \textit{and} \qquad L(f, P) = \sum_{k=1}^{K} m_k \Delta x_k.$$

Because f is bounded on I, its infimum and supremum exists on ever subinterval of I and therefore $U(f,P)$ and $L(f,P)$ will always exists (as finite numbers) for any bounded function f and any partition P of I. The numbers $U(f,P)$ and $L(f,P)$ are respectively overestimates and underestimates for the (signed) area under the graph of f on the interval I, when this area is a sensible notion. These estimates are produced by forming rectangles above and below the graph of f where the width of the rectangles are determined by the subdivisions of I produced by the partition P. Note here

By properties of the supremum and infimum, observe that

$$L(f,P) \leq U(f,P), \tag{6.1}$$

an inequality which holds for every partition P and every bounded function $f : I \to \mathbb{R}$.

It is helpful to think about a refinement Q of a partition P as one which produces, generally, finer subdivisions than those given by P – hence the name "refinement". With the aim of comparing upper and lower sums, we need the following lemma. The lemma says essentially that finer divisions of I yield "better" estimates for the area under the graph of f.

Lemma 6.3. *Let P and Q be partitions of I and suppose that Q is a refinement of P. For any $f \in B(I)$,*

$$L(f,P) \leq L(f,Q) \qquad \text{and} \qquad U(f,Q) \leq U(f,P).$$

Proof. Let $f \in B(I)$ and P be a partition of I. For any $y \in I \setminus P$, observe that $P \cup \{y\}$ it a refinement of P (with one extra element) and, for some $j = 1, 2, \dots, K$, it must be that

$$x_{j-1} < y < x_j,$$

i.e., y falls in the jth subinterval of the original partition P. In this case, we have

$$L(f,P) = \sum_{k=1}^{K} m_n \Delta x_k = m_j(x_j - x_{j-1}) + \sum_{k=1, k\neq j}^{K} m_k \Delta x_k.$$

Observe that, for $m_j = \inf_{x_{j-1} \leq x \leq x_j} f(x)$,

$$m_j \leq \inf_{x_{j-1} \leq x \leq y} f(x) := m(x_{j-1}, y) \qquad \text{and} \qquad m_j \leq \inf_{y \leq x \leq x_j} f(x) := m(y, x_j)$$

since both infima above are taken over smaller sets. Consequently,

$$\begin{aligned} L(f,P) &= m_j(x_j - x_{j-1}) + \sum_{k=1,k\neq j}^{K} m_k \Delta x_k \\ &= m_j(x_j - y) + m_j(y - x_{j-1}) + \sum_{k=1,k\neq j}^{K} m_k \Delta x_k \\ &\leq m(y, x_j)(x_j - y) + m(x_{j-1}, y)(y - x_{j-1}) + \sum_{k=1,k\neq j}^{K} m_k \Delta x_k. \end{aligned}$$

Since the partition $P \cup \{y\}$ gives all the same subintervals of I as P except that it splits the subinterval $[x_{j-1}, x_j]$ into two subintervals, $[x_{j-1}, y]$ and $[y, x_j]$, we recognize that the final summation above is simply the lower sum, $L(f, P \cup \{y\})$. Hence

$$L(f,P) \leq L(f, P \cup \{y\}). \tag{6.2}$$

For the upper sum, we see that

$$\sup_{x_{j-1} \leq x \leq y} f(x) \leq M_j \qquad \text{and} \qquad \sup_{y \leq x \leq x_j} f(x) \leq M_j$$

and with this, an analogous argument to that made for lower sums yields

$$U(f, P \cup \{y\}) \leq U(f, P). \tag{6.3}$$

With these two inequalities, we let Q be any refinement of P so that we may write

$$Q = P \cup \{y_1, y_2, \dots, y_S\}$$

where $y_s \in I \setminus P$ for $s = 1, 2, \dots, s$. By repeated application of the inequality (6.2), we find

$$L(f, P) \leq L(f, P \cup \{y_1\}) \leq L(f, P \cup \{y_1\} \cup \{y_2\}) \leq \cdots \leq L(f, P \cup \{y_1\} \cup \{y_2\} \cup \cdots \cup \{y_S\}) = L(f, Q).$$

By an analogous argument, making use of (6.3), we find

$$U(f, P) \geq U(f, P \cup \{y_1\}) \geq U(f, P \cup \{y_1\} \cup \{y_2\}) \geq \cdots \geq U(f, P \cup \{y_1\} \cup \{y_2\} \cup \cdots \cup \{y_S\}) = U(f, Q)$$

and so the proof is complete. $\square$

Thinking back to our picture of the area under the graph, which we will soon interpret as the integral, we expect the lower sums to be underestimates for this area and the upper sums to be overestimates. Equivalently, we can start to think of the integral as a number which sits below all of the upper sums and above all of the lower sums. To think about how to approximate this number, we need to invoke the notion of supremum and infimum. To this end, we'll need another lemma which will help us to make sure the infimum and supremum exist.

Lemma 6.4. *Let $f \in B(I)$ and let P and Q be partitions of I. Then*

$$(b-a)\left(\inf_{x \in I} f(x)\right) \leq L(f, P) \leq U(f, Q) \leq (b-a)\left(\sup_{x \in I} f(x)\right)$$

Proof. We first note that the trivial partition $T = \{a, b\} = \{x_0, x_1\}$ has

$$L(f, T) = \sum_{k=1}^{1} m_k(x_k - x_{k-1}) = m_1(x_1 - x_0) = \left(\inf_{x_0 \leq x \leq x_1} f(x)\right)(x_1 - x_0) = (b-a)\left(\inf_{x \in I} f(x)\right)$$

and

$$U(f, T) = \sum_{k=1}^{1} M_k(x_k - x_{k-1}) = M_1(x_1 - x_0) = \left(\sup_{x_0 \leq x \leq x_1} f(x)\right)(x_1 - x_0) = (b-a)\left(\sup_{x \in I} f(x)\right).$$

Thus, for any partitions P and Q, Lemma 6.3 guarantees that

$$(b-a)\left(\inf_{x \in I} f(x)\right) = L(f, T) \leq L(f, P) \qquad \text{and} \qquad U(f, Q) \leq U(f, T) = (b-a)\left(\sup_{x \in I} f(x)\right)$$

because P and Q are necessarily refinements of T. It remains to establish the inner inequality.

To this end, observe that the union $R = P \cup Q$ is also a partition of I for it is necessarily a finite subset of I which contains a and b. Further, R is a refinement of both partitions P and Q. Thus, by another appeal to Lemma 6.3 and in view of (6.1), we have

$$L(f, P) \leq L(f, R) \leq U(f, R) \leq U(f, Q)$$

which guarantees that $L(f, P) \leq U(f, Q)$ as was asserted. $\square$

Let's isolate some conclusions of the preceding lemma. First, it says that, for any partition P of I,

$$L(f, P) \leq (b-a)\left(\sup_{x \in I} f(x)\right).$$

Hence, the set

$$\{L(f,P) : P \text{ is a partition of } I\}$$

is a set of real numbers which is bounded above (it is, in fact, bounded above by every upper sum) and hence its supremum exists (and is finite). Thus, we define

$$L(f) = \sup_P L(f,P)$$

where this supremum is taken over all partitions P of I. This is called the **lower Darboux sum of f on I**; it is also sometimes referred to as the **lower Darboux integral**. Analogously, Lemma 6.4 guarantees that the infimum of all upper sums exists and so we define **the upper Darboux sum of f on I** as

$$U(f) = \inf_P U(f,P);$$

we may also refer to this as the **upper Darboux integral**. As we've established quite a few inequalities involving upper and lower sums pertaining to the same and different partitions of I, it's helpful to have some sense of how $U(f)$ and $L(f)$ compare for a given bounded function $f : I \to \mathbb{R}$. To this end, lets fix a partition Q of I and note that, in view of Lemma 6.4,

$$L(f,P) \leq U(f,Q)$$

for all partitions P of I. Thus, $U(f,Q)$ is an upper bound of the set of real numbers $\{L(f,P) : P \text{ is a parition of } I\}$. By the defining property of the supremum, we have

$$L(f) = \sup_P L(f,P) \leq U(f,Q).$$

Noting however that Q was arbitrary, we see that $L(f)$ is a lower bound for $U(f,Q)$ for all partitions Q of I. By the defining property of the infimum, we have

$$L(f) \leq \inf_Q U(f,Q) = U(f).$$

Let's summarize this information.

Proposition 6.5. *Let $f : I \to \mathbb{R}$ be a bounded function, i.e., $f \in B(I;\mathbb{R})$. Then the upper and lower Darboux sums,*

$$U(f) = \inf_P U(f,P) \qquad \text{and} \qquad L(f) = \sup_P L(f,P),$$

both exist. Furthermore,

$$L(f) \leq U(f).$$

Exercise 6.1:

This exercise will give you an idea of what's going on in the above construction. In what follows, we will focus on the interval $I = [0,1]$. For each $N = 1, 2, \dots,$, we shall consider the (regular) partition

$$P_N = \{x_0 < x_1 < \cdots < x_N = 1\} = \left\{x_n = \frac{n}{N} : n = 0, 1, 2, \dots, N\right\}$$

of the interval I.

1. For the function $f(x) = 1$ for $0 \leq x \leq 1$, compute $U(f,P_N)$ and $L(f,P_N)$.
   (a) Is it true that $L(f,P_N) \leq U(f,P_N)$?
   (b) Show that $\lim_{N\to\infty}(U(f,P_N) - L(f,P_N)) = 0$.
2. For the function $f(x) = x$ for $0 \leq x \leq 1$, compute $U(f,P_N)$ and $L(f,P_N)$.

(a) Is it true that $L(f, P_N) \leq U(f, P_N)$?

(b) Show that $\lim_{N\to\infty}(U(f, P_N) - L(f, P_N)) = 0$.

3. For the Dirichlet function f defined by

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$$

for $0 \leq x \leq 1$, compute $U(f, P_N)$ and $L(f, P_N)$.

(a) Is it true that $L(f, P_N) \leq U(f, P_N)$?

(b) Does $\lim_{N\to\infty}(U(f, P_N) - L(f, P_N)) = 0$?

4. For the first two examples above, you've seen a sequence (an enumerated collection) of partitions $\{P_N\}$ for which

$$\lim_{N\to\infty}(U(f, P_N) - L(f, P_N)) = 0.$$

In view of Proposition 6.5 and the above fact, does it suffice to conclude that

$$L(f) = U(f)?$$

Prove your assertion (or find a counter example).

5. Is it true that if there is a sequence of partitions $\{P_N\}$ for which

$$\lim_{N\to\infty}(U(f, P_N) - L(f, P_N)) \neq 0.,$$

then

$$L(f) \neq U(f)?$$

Prove your assertion (or find a counter example).

Finding motivation in the preceding example and returning again to our intuition of areas, we would hope that a sensible notion of area under the graph could be gotten by approximating the area from above by upper sums and from below by lower sums. Thus, if such an area does exist, we would hope that the supremum of all the lower sums coincides with the supremum of all the lower sums and so the inequality of the preceding proposition is actually an equality. This is exactly the right idea and we give this situation a name.

Definition 6.6. *Let $f \in B(I;\mathbb{R})$ and let $L(f)$ and $U(f)$ denote their lower and upper Riemann-Darboux sums, respectively. We say that f is Riemann integrable (or Riemann-Darboux integrable) on I and write $f \in R(I;\mathbb{R})$ if $U(f) = L(f)$. In this case, the Riemann-Darboux integral of f is defined to be the number*

$$\int_a^b f\,dx = U(f) = L(f).$$

This number will also be denoted in the following (numerous) ways:

$$\int_a^b f = \int_a^b f(x)\,dx = \int_a^b f(t)\,dt = \int_I f(x)\,dx = \int_I f.$$

As suprema and infima can be difficult to compute, the remainder of this section is dedicated to establishing various conditions under which we can decide if a given function is integrable. Along the way, we will also establish a few basic properties of the integral. First, let's write down an ϵ-characterization of integrability due to Riemann.

Theorem 6.7 (Riemann's Condition for Integrability)**.** *Let $f \in B(I)$. Then $f \in R(I;\mathbb{R})$ if and only if the following conditions is satisfied:*

For each $\epsilon > 0$, there is a partition P_ϵ of I for which $U(f,P_\epsilon) - L(f,P_\epsilon) < \epsilon$.

Proof. We first suppose that f is Riemann integrable. By the ϵ-characterization of the supremum, let Q_1 be a partition for which $L(f) - L(f,Q_1) < \epsilon/2$. Similarly, by the characterization for infimum, let Q_2 be a partition of I for which $U(f,Q_2) - U(f) < \epsilon/2$. With these partitions in mind, we set $P_\epsilon = Q_1 \cup Q_2$ and observe that P_ϵ is a refinement of both Q_1 and Q_2. By Lemma 6.3, we have $L(f,P_\epsilon) \geq L(f,Q_1)$ and $U(f,P_\epsilon) \leq U(f,Q_2)$ and thus

$$U(f,P_\epsilon) - L(f,P_\epsilon) \leq U(f,Q_2) - L(f,Q_1) < U(f) + \epsilon/2 - (L(f) - \epsilon/2) = U(f) - L(f) + \epsilon.$$

Of course, because $f \in R(I)$, $U(f) = L(f)$ and so the above inequality shows that $U(f,P_\epsilon) - L(f,P_\epsilon) < \epsilon$.

Conversely, let's assume that the desired property holds. Let $\epsilon > 0$, and using the property select a partition P for which $U(f,P) - L(f,P) < \epsilon$. As $U(f)$ and $L(f)$ are constructed from infima and suprema respectively, we have

$$U(f) - L(f) \leq U(f,P) - L(f,P) < \epsilon.$$

In view of Proposition 6.5, we also have $U(f) - L(f) \geq 0$. Hence, to each $\epsilon > 0$, we have

$$0 \leq U(f) - L(f) < \epsilon.$$

We may therefor conclude that $U(f) = L(f)$ for the only number "lodged" between zero and every positive number is the number zero itself. □

Theorem 6.8. *Let $I = [a,b]$. If $f \in R(I;\mathbb{R})$ and h is a real-valued function which is continuous on the closure of the range of f, then the composition $h \circ f$ is Riemann-Darboux integrable on I, i.e., $h \circ f \in R(I;\mathbb{R})$.*

Proof. We shall establish integrability of the composition $h \circ f$ using Theorem 6.7 and, to this end, we fix $\epsilon > 0$.

We recall that Riemann-Darboux integrable functions are bounded by definition. Thus, the range of f is a bounded set and so its closure is compact by the Heine-Borel theorem. So, ϕ is a continuous function on a compact set and it is therefore bounded and uniformly continuous (By Theorem 4.19 in Rudin). Let $M' > 0$ be such that $|h \circ f(x)| < M'$ for all $x \in I$ and select $0 < \delta$ for which

$$|h(p) - h(q)| < \frac{\epsilon}{2(b-a)}$$

whenever $|p - q| < \delta$. In fact, we may select δ further so that

$$0 < \delta < \frac{\epsilon}{4M'} \qquad \text{so that} \qquad 2M'\delta < \frac{\epsilon}{2}.$$

Armed with this δ and thanks to Theorem 6.7, we may chose a partition P of I for which

$$U(f,P) - L(f,P) = \sum_{k=1}^{K} (M_k - m_k)\Delta x_k < \delta^2$$

Let's now consider analogous sums for the composition, $h \circ f$, with this partition P. For $k = 1, 2, \dots, K$, set

$$M_k' = \sup_{x_{k-1} \leq x \leq x_k} (h \circ f)(x) \qquad \text{and} \qquad m_k' = \inf_{x_{k-1} \leq x \leq x_k} (h \circ f)(x).$$

Define

$$G = \{k = 1, 2, \dots K \mid M_k - m_k < \delta\} \qquad \text{and} \qquad B = \{1, 2, \dots, K\} \setminus G = \{k = 1, 2, \dots, K \mid M_k - m_k \geq \delta\}.$$

For $k \in G$, we have

$$|f(x) - f(y)| \leq M_k - m_k < \delta$$

whenever $x, y \in [x_{k-1}, x_k]$. Thus by the uniform continuity of h,

$$M'_k - m'_k = \sup_{x_{k-1} \le x, y \le x_k} (h(f(x)) - h(f(y)) \le \sup_{x_{k-1} \le x, y \le x_k} |(h \circ f)(x) - (h \circ f)(y)| \le \frac{\epsilon}{2(b-a)}$$

whenever $k \in G$. Consequently,

$$\sum_{k \in G} (M'_k - m'_k)\Delta x_k \le \frac{\epsilon}{2(b-a)} \sum_{k \in G} \Delta x_k \le \frac{\epsilon}{2(b-a)}(b-a) = \frac{\epsilon}{2}.$$

Now, for $k \in B$, observe that

$$M'_k - m'_k = \sup_{x_{k-1} \le x \le x_k} h(f(x)) + \sup_{x_{k-1} \le x \le x_k} (-h(f(x))) \le M' + M' = 2M'$$

so

$$\delta \sum_{k \in B} (M'_k - m'_k)\Delta x_k \le 2M' \sum_{k \in B} \delta \Delta x_k \le 2M' \sum_{k \in B} (M_k - m_k)\Delta x_k \le 2M' \sum_{k=1}^{K} (M_k - m_k)\Delta x_k < 2M'\delta^2$$

and therefore

$$\sum_{k \in B} (M'_k - m'_k)\Delta x_k \le 2M'\delta < \frac{\epsilon}{2}$$

since we have chosen δ so that $0 < \delta < \epsilon/(4M')$. All together, we have

$$U(h \circ f, P) - L(h \circ f, P) = \sum_{k=1}^{K} (M'_k - m'_k)\Delta x_k = \sum_{k \in G} (M'_k - m'_k)\Delta x_k + \sum_{k \in B} (M'_k - m'_k)\Delta x_k < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus, $h \circ f \in R(I; \mathbb{R})$. ☐

With this result, we obtain two immediate corollaries. The first is proven by applying the Theorem 6.8 in the case that $h(x) = x^2$ and $h(x) = |x|$, both of which are continuous everywhere.

Corollary 6.9. *If $f \in R(I; \mathbb{R})$ where $I = [a, b]$, then f^2 and $|f|$ are both in $R(I; \mathbb{R})$.*

Corollary 6.10. *Continuous real-valued functions (on an interval $I = [a, b]$) are Riemann-Darboux integrable, i.e., $C^0(I; \mathbb{R}) \subseteq R(I; \mathbb{R})$.*

Proof. Using the result of the previous exercise, we know the identify function $\mathrm{Id}(x) = x$ is Riemann-Darboux integrable on $[a, b]$ (and, in fact, $\int_a^b x\,dx = (b-a)^2/2$). If f is any continuous function on $[a, b]$, then $f = f \circ \mathrm{Id} \in R(I; \mathbb{R})$ thanks to Theorem 6.8. ☐

As evidenced by the theorem above and its corollaries, the characterization given by Theorem 6.7 is very useful in theoretical arguments but it is sometimes hard to implement in practice. Our next result is one that is a little easier to implement and also involves so-called Riemann sums that you might remember from your first-year calculus course. First, let's precisely introduce the notion of Riemann sum.

Definition 6.11. *Let $I = [a, b]$ and $P = \{x_0, x_1, x_2, \dots, x_N\}$ be a partition of I. A set of points $\{x_1^*, x_2^*, \dots, x_N^*\}$ is said to be admissible for P if $x_{k-1} \le x_k^* \le x_k$ for $k = 1, 2, \dots, N$. Given a function f on I, a Riemann sum for f associated to P is a sum of the form*

$$\sum_{k=1}^{N} f(x_k^*)\Delta x_k$$

where the collection of points $\{x_1^*, x_2^*, \dots, x_N^*\}$ at which f is evaluated is admissible for P.

In this language, we have the following theorem of Darboux which characterizes integrability (and the integral) in terms of Riemann sums.

Theorem 6.12. *Let $f \in B(I;\mathbb{R})$. Then $f \in R(I;\mathbb{R})$ if and only if there is a number $\mathcal{I}$ such that, for every $\epsilon > 0$, there is a $\delta > 0$ such that*

$$\left|\mathcal{I} - \sum_{k=1}^{K} f(x_k^*)\Delta x_k\right| < \epsilon \tag{6.4}$$

whenever P is a partition of I with $\|P\| < \delta$ and $\{x_1^*, x_2^*, \dots, x_K^*\}$ is a set of points in I admissible for P. In this case,

$$\mathcal{I} = \int_a^b f.$$

The proof of this theorem is (necessarily and unavoidably) extremely technical. Though you should read the theorem's proof in detail, it might makes sense to skip it in your first reading – come back to it when you really want (and are ready) to think about the details.

Proof. We first prove that the condition (6.4) guarantees the integrability of f and that $\mathcal{I}$ is its integral. To this end, let $\epsilon > 0$ and, in view of (6.4), choose $\delta > 0$ such that, for any partition P with $\|P\| < \delta$,

$$\left|\mathcal{I} - \sum_{k=1}^{K} f(x_k^*)\Delta x_k\right| < \frac{\epsilon}{4}$$

whenever $\{x_0^*, x_2^*, \dots, x_K^*\}$ is admissible for P. Let's choose (and fix) some partition P with $\|P\| < \delta$ so that the above holds (for instance, you can simply choose a regular partition with sufficiently small increments) and write

$$P = \{x_0, x_1, \dots, x_K\}.$$

For each $k = 1, 2, \dots, K$, using the ϵ-characterizations of suprema and infima, let's choose two points x_k^* and y_k^* in $[x_{k-1}, x_k]$ with

$$M_k - \frac{\epsilon}{4(b-a)} < f(x_k^*) \qquad \text{and} \qquad f(y_k^*) < m_k + \frac{\epsilon}{4(b-a)}.$$

With these estimates, we have

$$\begin{aligned}
U(f,P) - L(f,P) &= \sum_{k=1}^{K} M_k\Delta x_k - \sum_{k=1}^{K} m_k\Delta x_k \\
&< \sum_{k=1}^{K}\left(f(x_k^*) + \frac{\epsilon}{4(b-a)}\right)\Delta x_k - \sum_{k=1}^{K}\left(f(y_k^*) - \frac{\epsilon}{4(b-a)}\right)\Delta x_k \\
&= \sum_{k=1}^{K} f(x_k^*)\Delta_k - \sum_{k=1}^{K} f(y_k^*)\Delta x_k + \sum_{k=1}^{K}\frac{2\epsilon}{4(b-a)}\Delta x_k \\
&\leq \left|\sum_{k=1}^{K} f(x_k^*)\Delta x_k - \sum_{k=1}^{K} f(y_k^*)\Delta x_k\right| + \frac{\epsilon}{2(b-a)}\sum_{k=1}^{K}\Delta x_k \\
&= \left|\sum_{k=1}^{K} f(x_k^*)\Delta x_k - \mathcal{I} + \mathcal{I} - \sum_{k=1}^{K} f(y_k^*)\Delta x_k\right| + \frac{\epsilon}{2} \\
&\leq \left|\sum_{k=1}^{K} f(x_k^*)\Delta x_k - \mathcal{I}\right| + \left|\mathcal{I} - \sum_{k=1}^{K} f(y_k^*)\Delta x_k\right| + \frac{\epsilon}{2} \\
&< \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{2} = \epsilon;
\end{aligned}$$

here, we have used the fact that the points $\{x_k^*\}$ and $\{y_k^*\}$ are both chosen to be admissible for P (so that (6.4) is valid). Thus, we have found a partition $P = P_\epsilon$ for which $U(f,P) - L(f,P) < \epsilon$ and so $f \in R(I;\mathbb{R})$ by virtue of

Theorem 6.7. By similar computations (which you should do!), we also find that

$$\int_a^b f - \mathcal{I} = U(f) - \mathcal{I} \leq U(f,P) - \mathcal{I} < \epsilon \qquad \text{and} \qquad \mathcal{I} - \int_a^b f = \mathcal{I} - L(f) \leq \mathcal{I} - L(f,P) < \epsilon.$$

In other words,

$$\left| \mathcal{I} - \int_a^b f \right| < \epsilon$$

and since $\epsilon > 0$ is arbitrary, we conclude

$$\mathcal{I} = \int_a^b f$$

as was asserted.

We now prove the "forward" direction. Assume that $f \in R(I;\mathbb{R})$, set

$$\mathcal{I} = \int_a^b f,$$

and let $\epsilon > 0$ be arbitrary but fixed. Our goal is to find a $\delta > 0$ for which (6.4) holds for every partition P with $\|P\| < \delta$ and every set of points admissible for P. First, using Theorem 6.7, let's select a partition

$$P_0 = \{y_0, y_1, \dots, y_{N-1}, y_N\}$$

of $[a,b]$ for which

$$U(f,P_0) - L(f,P_0) < \frac{\epsilon}{4}.$$

We note that, since $\mathcal{I} = U(f) = L(f)$ must be "lodged" between $U(f,P_0)$ and $L(f,P_0)$ we see that

$$|R - \mathcal{I}| < \frac{\epsilon}{2} \tag{6.5}$$

whenever R is a number with

$$L(f,P_0) - \frac{\epsilon}{4} < R \leq U(f,P_0).$$

With P_0 (and hence N) fixed, set

$$\delta = \min\left\{ \frac{\|P_0\|}{2}, \frac{\epsilon}{8NM} \right\}$$

where $M > 0$ is an upper bound for $|f|$ on I. It remains to show that this δ actually does what we need it to.

Let $P = \{x_0, x_1, \dots, x_K\}$ be any partition of $[a,b]$ with $\|P\| < \delta$ and let $\{x_1^*, x_2^*, \dots, x_K^*\}$ be a collection of points which is admissible for P. By the way that we choose $\delta \leq \|P_0\|/2$, P must be "finer" than P_0 (not that it is necessarily a refinement, but its increments are necessarily smaller. To see this precisely, set $k_0 = 0$, $k_N = K$ and, for $n = 1, 2, \dots, N-1$,

$$k_n = \min\left\{ k \mid x_k > y_n \right\}.$$

Observe,

$$\{x_0, x_1, \dots, x_{k_1-1}\} \subseteq [y_0, y_1],$$

$$\{x_{k_1}, x_{k_1+1}, \dots, x_{k_2-1}\} \subseteq [y_1, y_2],$$

and, in general, for $n = 1, 2, \dots, N$,

$$\{x_{k_{n-1}}, x_{k_{n-1}+1}, \dots, x_{k_n-1}\} \subseteq [y_{n-1}, y_n].$$

In this way, we have placed "most" of the subintervals for the partition P within subintervals of P_0; there are, at most, only a "small" number (N to be exact) of subintervals of P which straddle elements of P_0 – we'll deal with these separately. For now, observe that for $n = 1, 2, \dots, N$,

$$\begin{aligned}\sum_{k=k_{n-1}+1}^{k_n-1} \Delta x_k &= x_{k_n-1} - x_{k_{n-1}} \\ &= y_n - y_{n-1} + (x_{k_n-1} - y_n) + (y_{n-1} - x_{k_{n-1}}) \\ &= \Delta y_n + (x_{k_n-1} - y_n) + (y_{n-1} - x_{k_{n-1}}).\end{aligned}$$

By our choice of k_n, we have

$$0 \le y_n - x_{k_n-1} < x_{k_n} - x_{k_n-1} < \delta \qquad \text{and} \qquad 0 \le x_{k_{n-1}} - y_{n-1} < (x_{k_{n-1}} - x_{k_{n-1}-1}) < \delta$$

so that

$$\Delta y_n - 2\delta \le \sum_{k=k_{n-1}+1}^{k_n-1} \Delta x_k \le \Delta y_n \tag{6.6}$$

for $n = 1, 2, \dots, N$. Set

$$R = R(f, P, \{x^*\}) = \sum_{k=1}^{K} f(x_k^*)\Delta x_k$$

and observe that

$$\begin{aligned} R \;=\; & [f(x_1^*)\Delta x_1 + f(x_2^*)\Delta x_2 + \cdots + f(x_{k_1-1}^*)\Delta x_{k_1-1}] && + f(x_{k_1}^*)\Delta x_{k_1} \\ & + [f(x_{k_1+1}^*)\Delta x_{k_1+1} + f(x_{k_1+2}^*)\Delta x_{k_1+2} + \cdots + f(x_{k_2-1}^*)\Delta x_{k_2-1}] && + f(x_{k_2})\Delta x_k \\ & \vdots \\ & + [f(x_{k_{n-1}+1}^*)\Delta x_{k_{n-1}+1} + f(x_{k_{n-1}+2}^*)\Delta x_{k_{n-1}+2} + \cdots + f(x_{k_n-1}^*)\Delta x_{k_n-1}] && + f(x_{k_n}^*)\Delta x_{k_n} \\ & \vdots \\ & + [f(x_{k_{N-1}+1}^*)\Delta x_{k_{N-1}+1} + f(x_{k_{N-1}+2}^*)\Delta x_{k_{N-1}+2} + \cdots + f(x_{k_N-1}^*)\Delta x_{k_N-1}] && + f(x_{k_N}^*)\Delta x_{k_N}. \end{aligned}$$

Equivalently,

$$\begin{aligned} R &= \sum_{n=1}^{N} \sum_{k=k_{n-1}+1}^{k_n-1} f(x_k^*)\Delta x_k + \sum_{n=1}^{N} f(x_{k_n}^*)\Delta x_{k_n} \\ &=: R_1 + R_2. \end{aligned}$$

Let's estimate the inner summations in the first term above. Since $x_k^* \in [x_{k-1}, x_k] \subseteq [y_{n-1}, y_n]$ for all $k_{n-1}+1 \le k \le k_n - 1$, it follows from (6.6) and our definitions of M_n and m_n (as ingredients for the upper and lower Darboux sums for P_0) that

$$m_n(\Delta y_n - 2\delta) \le \sum_{k=k_{n-1}+1}^{k_n-1} m_n \Delta x_k \le \sum_{k=k_{n-1}+1}^{k=k_n-1} f(x_k^*)\Delta x_k \le \sum_{k=k_{n-1}+1}^{k=k_n-1} M_n \Delta x_k \le M_n \Delta y_n$$

for each $n = 1, 2, \dots, N$. Summing over n, we obtain

$$L(f, P_0) - 2\delta \sum_{n=1}^{N} m_n \le R_1 \le U(f, P_0)$$

and because

$$\left|\sum_{n=1}^{N} m_n\right| \leq N \max_{n=1,2,\ldots,N}\{|m_n|\} \leq NM,$$

this guarantees

$$L(f,P_0)-\frac{\epsilon}{4} < L(f,P_0)-2\delta MN \leq R_1 \leq U(f,P_0).$$

In view of (6.5), we conclude that

$$|\mathcal{I}-R_1| < \frac{\epsilon}{2}.$$

For the term R_2, we simply observe that

$$|R_2| = \left|\sum_{n=1}^{N} f(x_n)\Delta x_n\right| \leq MN\delta \leq \frac{\epsilon}{8}.$$

Combining the two preceding estimates, we find that

$$\left|\mathcal{I}-\sum_{k=1}^{K} f(x_k^*)\Delta x_k\right| = |\mathcal{I}-R| = |\mathcal{I}-R_1-R_2| \leq |\mathcal{I}-R_1|+|R_2| < \frac{\epsilon}{2}+\frac{\epsilon}{8} < \epsilon$$

as required. WOOF. □

Now that we've done the hard work of proving Darboux's theorem, let's see it bear fruit.

Corollary 6.13. *Let $I=[a,b]$ and consider the sequence of regular partitions $\{P_N\}$ given by*

$$P_N = \{x_k \in I \mid k=0,\ldots,N\} \qquad \textit{where} \qquad x_k = a+\frac{k(b-a)}{N}$$

for $k=0,1,2,\ldots N$. Also, suppose that, for each N, $\{x_1^*,x_2^*,\ldots,x_N^*\} \subseteq I$ is some choice[1] of points which is admissible for P_N. Then, if $f \in R(I;\mathbb{R})$,

$$\int_a^b f = \lim_{N\to\infty}\sum_{k=1}^{N} f(x_k^*)\Delta x_k = \lim_{N\to\infty}\frac{b-a}{N}\sum_{k=1}^{N} f(x_k^*).$$

In particular, this holds for any real-valued continuous function on I (in view of Corollary 6.10).

Proof. Let $\epsilon>0$ and, in view of Theorem 6.12, let $\delta>0$ be such that

$$\left|\int_a^b f - \sum_{k=1}^{K} f(x_k^*)\Delta x_k\right| < \epsilon$$

whenever $\|P\|<\delta$. By the Archimedean property, select N_0 for which $(b-a)/N_0<\delta$ and observe that $\|P_N\| = (b-a)/N \leq (b-a)/N_0 < \delta$ whenever $N \geq N_0$ so that

$$\left|\int_a^b f - \sum_{k=1}^{N} f(x_k^*)\Delta x_k\right| = \left|\int_a^b f - \sum_{k=1}^{N} f(x_k^*)\frac{b-a}{N}\right| < \epsilon.$$

□

Our next corollary of Darboux's theorem guarantees that $R(I;\mathbb{R})$ is a vector space over $\mathbb{R}$ and $f \mapsto \int_I f$ is a linear transformation from this vector space into $\mathbb{R}$ (i.e., it is a "linear functional").

Theorem 6.14. *Let $I=[a,b]$ and $f,g \in R(I;\mathbb{R})$. Then, for any real numbers α and β, $\alpha f+\beta g \in R(I;\mathbb{R})$ and*

$$\int_a^b \alpha f+\beta g = \alpha\int_a^b f + \beta\int_a^b g.$$

[1] This could be a choice of left or right endpoints, or midpoints.

Exercise 6.2:

Prove the theorem. Hint: Let $\epsilon > 0$. Choose δ_1 so that

$$\left| \int_a^b f - \sum_{k=1}^K f(x_k^*)\Delta x_k \right| < \frac{\epsilon}{2(|\alpha|+1)}$$

whenever P is a partition with $\|P\| < \delta_1$ and $\{x_1^*, x_2^*, \dots, x_k^*\}$ are admissible for the partition. Similarly, choose δ_2 so that

$$\left| \int_a^b g - \sum_{k=1}^K g(x_k^*)\Delta x_k \right| < \frac{\epsilon}{2(|\beta|+1)}$$

whenever P is a partition with $\|P\| < \delta_2$ and $\{x_1^*, x_2^*, \dots, x_k^*\}$ are admissible for P. Now, set $\delta = \min\{\delta_1, \delta_2\}$.

Exercise 6.3: When modifying a function doesn't change its integral

Let $I = [a,b]$ and suppose that $f \in R(I;\mathbb{R})$. Given $g : I \to \mathbb{R}$, set $D = \{x \in I \mid f(x) \neq g(x)\}$.

1. Using only the definitions and results in the present section of the notes, prove the following statement: If D is finite, then $g \in R(I;\mathbb{R})$ and
$$\int_a^b g = \int_a^b f.$$
2. Does the result above still hold if D is countably infinite? If so, prove it. If not, produce a counterexample (and work the details).

Exercise 6.4:

Let $I = [a,b]$ and let c and d be such that $a \leq c < d \leq b$ so that $J = [c,d] \subseteq I$. Define the so-called characteristic function

$$\mathbb{1}_J(x) = \begin{cases} 1 & c \leq x \leq d \\ 0 & \text{else} \end{cases} \tag{6.7}$$

of J. Prove the following statements:

1. If $f \in R(I;\mathbb{R})$, then $f \in R(J;\mathbb{R})$.
2. If $f \in R(I;\mathbb{R})$, then $f \cdot \mathbb{1}_J \in R(I;\mathbb{R})$.
3. If $f \in R(I;\mathbb{R})$, then
$$\int_c^d f(x)\,dx = \int_a^b f(x)\mathbb{1}_J(x)\,dx.$$

Exercise 6.5:

Use the above exercises to prove the following proposition. Also, using a diagram, explain why the proposition makes sense using "areas".

Proposition 6.15. *Let $I = [a,b]$ and $a \leq c \leq b$. If $f \in R(I;\mathbb{R})$, then $f \in R([a,c];\mathbb{R}) \cap R([c,b];\mathbb{R})$ and*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

We finish this section showing that the product of Riemann-integrable functions is Riemann integrable. Combining this with Theorem 6.14, this shows, in particular, that $R(I;\mathbb{R})$ is a commutative ring.

Proposition 6.16. *Let $f, g \in R(I;\mathbb{R})$, then their product fg is also a member of $R(I;\mathbb{R})$.*

Proof. Observe that

$$fg = \frac{1}{2}\left((f+g)^2 - f^2 - g^2\right)$$

and since sums (Theorem 6.14), squares (Corollary 6.9), and linear combinations (Theorem 6.14) of integrable functions are integrable, $fg \in R(I;\mathbb{R})$. □

6.2 The Riemann-Darboux Integral as a "signed" integral

When we think of doing integration (by approximations via Riemann sums), we think about summing up areas under rectangles as we move (along a partition) from a to b. The concept of "moving" is associated with an understanding that we have some direction in mind – from left to right. If you've taken a course in vector calculus, this interpretation coincides with the notion that "work" is calculated via a line integral along a path traveled from a point A to a point B. If we were to reverse that path, we would gain the energy lost doing that work. For this interpretation to make sense, we make the following convention.

Convention 6.17 (The Signed Integral). If $f \in R([a,b];\mathbb{R}$ where $a < b$, the integral of f from b to a is defined by

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx.$$

As we will see, this convention is one that will allow us to understand the interplay between integrals and derivatives a la the Fundamental theorem of calculus. With this convention, we obtain the following "generalization" of Proposition **??**

Theorem 6.18. *Let I be an interval and $f \in R(I;\mathbb{R})$. Then, for any numbers $a, b, c \in I$ (they do not need to have any specific order nor be endpoints), then*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Proof. By Exercise 6.4, we know that f is Riemann-Darboux integrable on every subinterval of I. Thus, to prove the theorem, it suffices to check the formula for all permutations (of orders) of a, b, and c. From Proposition 6.15, the formula clearly holds for $a < b < c$. Let's consider the case that $b < a < c$. To this end, we have

$$\int_b^c f(x)\,dx = \int_b^a f(x)\,dx + \int_a^c f(x)\,dx = -\int_a^b f(x)\,dx + \int_a^c f(x)\,dx$$

where we have used Proposition 6.15 and our convention. Rearranging and invoking the convention one more time, we have

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx - \int_b^c f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Checking all other cases is done similarly. □

6.3 The Riemann-Darboux integral for Complex-valued functions

Armed with the notions of integration and integrability for real-valued functions f on I, it is easy to generalize these to complex-valued functions.

Definition 6.19. *Let $I = [a, b]$ and consider a complex-valued function $f : I \to \mathbb{C}$. In this case f is necessarily of the form*

$$f(x) = u(x) + iv(x)$$

for $x \in I$ where $u, v : I \to \mathbb{R}$. We saw that f is Riemann integrable on I if u and v are Riemann integrable on I (i.e., $u, v \in R(I; \mathbb{R})$) and we define the integral of f on I to be the complex number

$$\int_I f(x)\,dx = \left(\int_I u(x)\,dx\right) + i\left(\int_I v(x)\,dx\right).$$

The set of complex-valued function on the interval I is denoted by $R(I; \mathbb{C})$. We will also use the notations

$$\int_I f = \int_a^b f = \int_a^b f(x)\,dx$$

to denote the integral of $f \in R(I; \mathbb{R})$.

Let's make a few notes concerning the above definition. First, the functions u and v are called the real and imaginary parts of f respectively. We'll often write $f = \mathrm{Re}(f) + i\,\mathrm{Im}(f)$ where $\mathrm{Re}(f) = u$ and $\mathrm{Im}(f) = v$. In the (special) case in which f is a real-valued function from I to $\mathbb{R}$, we can write $f = \mathrm{Re}(f) + i\,\mathrm{Im}(f) = \mathrm{Re}(f) + i0 = f + i0$ and so here

$$\int_I f = \int \mathrm{Re}(f) + i\int_I 0 = \int_I \mathrm{Re}(f)(x)\,dx + i0 = \int_I \mathrm{Re}(f)(x)\,dx$$

because the integral of the zero function is just 0. In this way we observe that the definition of the Riemann integral for complex-valued functions is an extension of the Riemann integral for real-valued functions (as it recaptures the real-valued version of the Riemann integral). For this reason, we will sometimes write $R(I) = R(I; \mathbb{C})$ and note that $R(I; \mathbb{R}) \subseteq R(I)$ by the above argument.

Now that we know what integrability means, it's high time to give some properties of the integral.

Theorem 6.20. *Let $I = [a, b] \subseteq \mathbb{R}$.*

1. *For any complex numbers α and β and any $f, g \in R(I)$, the linear combination $\alpha f + \beta g \in R(I)$ and*

$$\int_I (\alpha f + \beta g)) = \alpha\int_I f + \beta\int_I g.$$

   This says that $R(I)$ is a vector space over $\mathbb{C}$ and the integral (viewed as a function $f \to \int_I f$) is linear map from $R(I)$ to $\mathbb{C}$.

2. *If $f, g \in R(I)$, then the product $fg \in R(I)$.*

3. *Constant functions are Riemann-integrable and for any constant function $x \mapsto \alpha$ where $\alpha \in \mathbb{C}$,*

$$\int_I \alpha = \alpha(b - a).$$

4. *The set of continuous functions $C^0(I; \mathbb{C})$ are Riemann integrable. That is, $C^0(I; \mathbb{C}) \subseteq R(I)$.*

Proof. Exercise ☐

Exercise 6.6: here

In this exercise, you prove the real-valued analogue of the scalar multiplication portion of Item 1 of the proposition above. Throughout this exercise, c is a real number.

1. First, given a non-empty bounded set A of $\mathbb{R}$, we denote by cA the set of numbers of the form $c \cdot a$ where $a \in A$. That is, $cA = \{x \in \mathbb{R} : x = ca \text{ for } a \in A\}$. If $c > 0$, prove that
$$\sup cA = c \sup A \qquad \text{and} \qquad \inf cA = c \inf A.$$
2. If $c < 0$, formulate and prove an analogous statement for $\sup cA$ and $\inf cA$.
3. For the remainder of this exercise, $g : I \to \mathbb{R}$ will be an arbitrary bounded function. We will assume now that $c > 0$ and denote by cg the real-valued function on I defined by $(cg)(x) = cg(x)$ for $x \in I$. Use your result from Item 1 to prove that
$$U(cg, P) = cU(g, P) \quad \text{and} \quad L(cg, P) = cL(g, P).$$
for any partition P of I.
4. Continuing under the assumption that $c > 0$, prove that $U(cg) = c \cdot U(g)$ and $L(cg) = c \cdot L(g)$.
5. Use the item above to prove that, if $c > 0$, $g \in R(I)$ if and only if $cg \in R(I)$ and
$$c \int_I g = \int_I cg.$$
6. Comment on how the previous steps change if we allow c to be non-positive. In particular, is it still true that $cg \in R(I)$ if and only if $g \in R(I)$?

Another important property of the integral is captured by the following proposition.

Proposition 6.21. *Let $f \in R(I)$, then the function $|f| : I \to \mathbb{R}$ defined by*
$$|f|(x) = |f(x)| = \sqrt{(\mathrm{Re}(f(x))^2 + \mathrm{Im}(f(x))^2} \qquad \text{for } x \in I$$
is Riemann integrable and
$$\left| \int_I f \right| \leq \int_I |f|.$$

To prove the proposition, we will first need a lemma.

Lemma 6.22. *Let $h_1, h_2 \in R(I)$ be real-valued functions (i.e., $h_1, h_2 \in R(I; \mathbb{R})$) such that $h_1(x) \leq h_2(x)$ for all $x \in I$. Then*
$$\int_I h_1 \leq \int_I h_2.$$

Exercise 6.7:

Prove the lemma above. Hint: Start by showing that non-negative functions have non-negative integrals. Then use Item 1 of Theorem 6.14.

Proof. Let $f \in R(I)$ and write $f = u + iv \in R(I)$. By definition, we have that $u, v \in R(I; \mathbb{R})$ and
$$|f| = \sqrt{u^2 + v^2}.$$

Since the squares of real-valued integrable functions are integrable (Corollary 6.9), the sums of Riemann integrable functions are integrable (Theorem 6.14), and the square root (as a continuous function) applied to a non-negative Riemann integrable function is integrable by virtue of Theorem 6.8, we conclude that the complex modulus of f, $|f|$, is in $R(I; \mathbb{R})$.

Since $\int_a^b f$ is a complex number, the last result of Exercise 5.2 guarantees a $\theta \in (-\pi, \pi]$ for which

$$\left|\int_I f\right| = e^{-i\theta}\left(\int_I f\right).$$

In view of Item 1 of Theorem 6.20, this guarantees that

$$\left|\int_I f\right| = \int_I e^{-i\theta} f = \int_I \left(e^{-i\theta} f(x)\right)\,dx = \int_I \mathrm{Re}(e^{-i\theta} f(x))\,dx + i\int_I \mathrm{Im}(e^{-i\theta} f(x))\,dx.$$

As the left hand side of the above equation is purely real, this ensures that the purely imaginary part of the right hand side is zero and therefore

$$\left|\int_I f\right| = \int_I \mathrm{Re}(e^{-i\theta} f(x))\,dx.$$

Now, for each $x \in I$,

$$\mathrm{Re}(e^{-i\theta} f(x)) \leq \sqrt{(\mathrm{Re}(e^{i\theta} f(x)))^2 + (\mathrm{Im}(e^{-i\theta} f(x)))^2} = |e^{-i\theta} f(x)| = |f(x)|$$

where we have used the fact that $|zw| = |z||w|$ for complex numbers z, w. Thus, by Lemma 6.22, we have

$$\left|\int_I f\right| \leq \int_I \mathrm{Re}(e^{-i\theta} f(x))\,dx \leq \int_I |f(x)|\,dx = \int_I |f|$$

as desired. $\square$

6.4 The Fundamental Theorem of Calculus

We now have done enough to establish the famous fundamental theorems of calculus.

Theorem 6.23 (The Fundamental Theorem of Calculus, Part I)**.** *Let $f \in R(I;\mathbb{C})$ where $I = [a,b]$ and define $F : I \to \mathbb{C}$ by*

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

1. *F is Lipschitz in I and, in particular, $F \in C^0(I;\mathbb{C})$.*

2. *If f is continuous at $x_0 \in I$, then F is differentiable at x_0 and*

$$F'(x_0) = \frac{d}{dx}\int_a^x f(t)\,dt\Big|_{x=x_0} = f(x_0).$$

Proof. Since Riemann-Darboux integrable functions are bounded, let M be such that $|f(x)| \leq M$ for all $x \in I$. By Proposition 6.21 Lemma 6.22, we have

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

for $x \geq y$. From this, we see that F is Lipschitz. This proves the first item.

For the second item, let $\epsilon > 0$ and, given the continuity of f at x_0, let $\delta > 0$ be such that $|f(x) - f(x_0)| < \epsilon$ whenever $|x - x_0| < \delta$. With this, observe that

$$\begin{aligned}
\left| f(x_0) - \frac{F(x) - F(x_0)}{x - x_0} \right| &= \frac{1}{x - x_0} \left| f'(x_0)(x - x_0) - \int_{x_0}^{x} f(t)\, dt \right| \\
&= \frac{1}{x - x_0} \left| \int_{x_0}^{x} (f(x_0) - f(t))\, dt \right| \\
&= \frac{1}{|x - x_0|} \int_{x_0}^{x} |f(x_0) - f(t)|\, dt \\
&\leq \frac{1}{|x - x_0|} \int_{x_0}^{x} \epsilon\, dt \\
&= \epsilon
\end{aligned}$$

whenever $|x - x_0| < \delta$. $\square$

Theorem 6.24 (The Fundamental Theorem of Calculus, Part II)**.** *Let $I = [a, b]$. If $f \in R(I, \mathbb{C})$ and $F : [a, b] \to \mathbb{C}$ is a differentiable function for which $F'(x) = f(x)$ for $x \in [a, b]$. Then*

$$F(b) - F(a) = \int_a^b f(x)\, dx.$$

Proof. Let's first assume that F and f are real valued and let $\epsilon > 0$. By virtue of Theorem 6.12, select a partition $P = \{x_1, x_2, \ldots, x_K\}$ for which

$$\left| \int_a^b f(x)\, dx - \sum_{k=1}^{K} f(x_k^*) \Delta x_k \right| < \epsilon$$

whenever $\{x_1^*, x_2^*, \ldots, x_K^*\}$ is admissible for P. Because the given function F is differentiable on $[a, b]$, it is differentiable on every subinterval $[x_{k-1}, x_k]$ and, by the mean value theorem, for each $k = 1, 2, \ldots, K$, there is some $c_k \in [x_{k-1}, x_k]$ for which

$$F(x_k) - F(x_{k-1}) = F'(c_k)(x_k - x_{k-1}) = f(c_k)\Delta x_k$$

where we have used the hypothesis that $F' = f$. In particular, the collection $\{c_1, c_2, \ldots, c_K\}$ is admissible for P and so we have

$$\left| \int_a^b f(x)\, dx - \sum_{k=1}^{K} (F(x_k) - F(x_{k-1})) \right| < \epsilon.$$

Observe that the above sum is "telescoping" so that

$$\begin{aligned}
\sum_{k=1}^{K} (F(x_k) - F(x_{k-1})) &= F(x_K) - F(x_{K-1}) + F(X_{K-1}) - F(X_{K-2}) \\
&\qquad + F(X_{K-2}) + \cdots - F(x_2) \\
&\qquad\qquad + F(x_2) - F(x_1) + F(x_1) - F(x_0) \\
&= F(x_K) + 0 + 0 + \cdots 0 - F(x_0) \\
&= F(b) - F(a).
\end{aligned}$$

Consequently,

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

Since this is true for every $\epsilon > 0$, it must hold that

$$F(b) - F(a) = \int_a^b f(x)\, dx.$$

For the general result where F and f are complex-valued, we simply apply this argument to their real and imaginary parts and make use of Proposition 5.2 and the definition of the integral of complex-valued functions. $\square$

IBP

Our next proposition is often called the "change of variables formula". Because the proof is somewhat technical (and is actually best done in the context of the Riemann-Steiltjes integral), I have decided to prove only a special case.

Proposition 6.25 (Change of variables formula)**.** *Let $A < B$ and $a < b$ be real numbers and suppose that $h : [A, B] \to [a, b]$ is a strictly increasing function mapping $[A, B]$ onto $[a, b]$ with derivative $h' \in R([A, B])$. Also, let $f \in R([a, b])$. Then the function $x \mapsto (f \circ h)(x)h'(x) = f(h(x))h'(x)$ is integrable on $[A, B]$ and*

$$\int_a^b f(x)\,dx = \int_{[a,b]} f = \int_{[A,B]} (f \circ h) \cdot h' = \int_A^B f(h(x))h'(x)\,dx$$

Proof. We shall prove the theorem under the (slightly more restrictive) hypotheses that $f \in C^0(I)$ and $h \in C^1(I; \mathbb{R})$; the general proof is best done in the context of the Riemann-Steiltjes integral and a proof can be found in Rudin, Theorem 6.19. Since h and h' are real valued, it suffices to assume that f is also real valued, for the general result can be gotten by simply piecing real and imaginary parts together. Define $F : [a, b] \to \mathbb{R}$ by

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

and observe that, because f is continuous on $[a, b]$, F is differentiable on $[a, b]$ and $F' = f$ by the FTC1. Define $G : [A, B] \to \mathbb{R}$ by

$$G(y) = \int_A^y (f \circ h)(t)h'(t)\,dt$$

for $y \in [A, B]$ and observe (in view of our hypotheses) that

$$G'(y) = (f \circ h)(y)h'(y)$$

for $y \in [A, B]$ thanks to FTC1. With this, let's define $F'[a, b] \to \mathbb{R}$ by

$$\tilde{F}(x) = (G \circ h^{-1})(x) = G(h^{-1}(x))$$

for $x \in [a, b]$. Since $h \in C^1([a, b], \mathbb{R})$ and strictly increasing, the inverse function theorem guarantees that h^{-1} is differentiable on its domain and

$$(h^{-1})'(x) = \frac{1}{h'(y)} > 0$$

for all $y = h^{-1}(x) \in [A, B]$. Applying the Chain rule and FTC1, we conclude that $\tilde{F}$ is differentiable on $[a, b]$ and

$$\tilde{F}'(x) = G'(h^{-1}(x))\frac{1}{h'(y)} = f(h(h^{-1}(x))h'(h^{-1}(x))\frac{1}{h'(h^{-1}(x))} = f(x)$$

for $x \in [a, b]$. By the mean value theorem, it follows that F and $\widetilde{F}$ can differ only by a constant C, i.e.,

$$F(x) = \widetilde{F}(x) + C$$

for all $x \in [a, b]$. In particular,

$$C = F(a) - \widetilde{F}(a) = \int_a^a f(t)\,dt - \int_A^{h^{-1}(a)} (f \circ h)(t)h'(t)\,dt = 0 - \int_A^A (f \circ h)(t)h'(t)\,dt = 0 - 0 = 0$$

because $h^{-1}(a) = A$ and hence

$$\int_a^b f(t)\,dt = F(b) = \widetilde{F}(b) = \int_a^B (f \circ h)(t)h'(t)\,dt.$$

$\square$

It should be noted that the proposition above has a very beautiful generalization to integration in $\mathbb{R}^d$ in which the derivative h' is replaced by the Jacobean determinant of h's d-dimensional analogous. This generalization is an essential tool used in the theory of integration on manifolds.

6.4.1 Averages and the Mean Value Theorem for Integrals

If we think about the integral of a function over an interval as a sum, then it is reasonable to think of the integral divided by the length of the interval as its average. In fact, let's make this a definition:

Definition 6.26. *Let $f \in R(I;\mathbb{R})$. Then, for any interval $J = [c,d] \subseteq I$, the average value of f on J is the number*

$$\mathrm{Ave}_f(J) = \frac{1}{d-c}\int_c^d f(x)\,dx.$$

Example 6.1: Just some average examples

Let's compute some averages.

1. Consider $f : [0,1] \to \mathbb{R}$ given by $f(x) = x$. Using the result of your homework,

$$\mathrm{Ave}_f([0,1]) = \frac{1}{1-0}\int_0^1 x\,dx = 1\cdot\frac{1}{2} = \frac{1}{2}.$$

2. Consider the function $g : \mathbb{R} \to \mathbb{R}$ defined by

$$g(t) = \begin{cases} 1 & t \in (k,k+1] \text{ when } k \text{ is even} \\ -1 & t \in (k,k+1] \text{ when } k \text{ is odd} \end{cases}$$

It is not difficult to see that g is Riemann-Darboux integrable on any compact interval. For simplicity, let's determine its average over intervals of the form $J = [0,T]$. Denote by T_0 the largest integer with $T_0 \leq T$, i.e., $T_0 = \lfloor T \rfloor$. Then, by Theorem Addition,

$$\begin{aligned} \int_0^T g(t)\,dt &= \int_{T_0}^T g(t)\,dt + \sum_{k=0}^{T_0-1}\int_k^{k+1} g(t)\,dt \\ &= \int_{T_0}^T (-1)^{T_0}\,dt + \sum_{k=0}^{T_0-1}(-1)^k\,dt \\ &= (T-T_0)(-1)^{T_0} + \begin{cases} 1 & T_0 \text{ is odd} \\ 0 & T_0 \text{ is even} \end{cases}. \end{aligned}$$

Consequently,

$$\mathrm{Ave}_g([0,T]) = \frac{1}{T-0}\int_0^T g(t)\,dt = \frac{(T-T_0)(-1)^{T_0}}{T} + \begin{cases} \frac{1}{T} & T_0 \text{ is odd} \\ 0 & T_0 \text{ is even.} \end{cases}$$

In looking at the examples above, we ask: Can a function be equal to its average value? The first example above is certainly "yes" as $f(x) = 1/2$ when $x = 1/2$. The second, however, is no. note. Perhaps, it's not surprising that this has to do with continuity.

Theorem 6.27 (The mean value theorem for integrals)**.** *Let $f \in R(I;\mathbb{R})$. If f is continuous on I, then there exists*

$c \in I$ with

$$f(c) = \mathrm{Ave}_f(I) = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

Proof. By virtue of Lemma 6.4 and the integrability of f, we have

$$(b-a)\inf_{x\in I} f(x) \leq L(f,P) \leq \int_a^b f \leq U(f,P) \leq (b-a)\sup_{x\in I} f(x)$$

for every partition P of I. Consequently,

$$\inf_{x\in I} f(x) \leq \mathrm{Ave}_f([a,b]) = \frac{1}{b-a}\int_a^b f \leq \sup_{x\in I} f(x).$$

Since f is continuous, the extreme value theorem guarantees that the infimum and supremum above are attained on the interval I. With this observation, the above inequality says that $\mathrm{Ave}_f([a,b])$ is a real number sitting in between two values of f on the interval $[a,b]$ and hence there must be some $c \in [a,b]$ for which $f(c) = \mathrm{Ave}_f([a,b])$ thanks to the intermediate value theorem. $\square$

Exercise 6.8:

The mean value theorem for integrals furnishes another way to prove FTC, Part 1 in the special case that $f \in C^0(I;\mathbb{R})$. In particular, use the Mean Value Theorem for integrals to show that, if $f \in C^0([a,b];\mathbb{R}) \subseteq R(I;\mathbb{R})$, we have

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x).$$

Failure of the RD integral

Chapter 7

The Essence of Convergence

In this chapter, we discuss the convergence of functions. Specifically, we discuss the ways in which a sequence of functions converges (or does not converge) to some other function. As you saw when you first learned about Power series, it is really useful to approximate a given function (say e^x) by simple and easy-to-understand functions (say, the sequence of polynomials $1, 1+x, 1+x+2^2/2, 1+x+x^2/2+x^3/6, ...$) and developing a theory for doing so is our present goal. This theory is somewhat delicate and complicated. As we will see, there are many inequivalent ways (an infinite number) to define what it means for a sequence of functions to converge to another function – and each has a use that is important/applicable in some context (e.g., linear programming, solving differential equations, Fourier analysis, probability). Below, we introduce our first notion of convergence called "pointwise" convergence.

Definition 7.1. *Let (X, d) be a metric space and let $\{f_n\}$ be a sequence of complex-valued functions on X, i.e., $f_n : X \to \mathbb{C}$ for each $n = 1, 2, \dots, N$. Let $f : X \to \mathbb{C}$ be another function. We say that the sequence $\{f_n\}$ converges pointwise to f on X if, for each $x \in X$,*

$$\lim_{n\to\infty} f_n(x) = f(x).$$

The important thing to note about the above definition is that the x is chosen before the limit is taken. Stated with ϵ's and N's, the above definition is as follows:

The sequence of functions f_n converges to f pointwise on X if, for each $\epsilon > 0$ and $x \in X$, there is an $N \in \mathbb{N}$ (depending on both ϵ and x) for which

$$|f_n(x) - f(x)| < \epsilon \qquad \text{whenever} \qquad n \geq N.$$

Example 7.1:

In this example, we consider a sequence of real-valued functions converging pointwise on the interval $I = [0, 1]$. For each natural number n, define $f_n : I \to \mathbb{R} \subseteq \mathbb{C}$ by

$$f_n(x) = x^n$$

for $x \in I$ and $n \in \mathbb{N}$. We observe that, for $0 \leq x < 1$,

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} x^n = 0$$

and, for $x = 1$,

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} 1^n = 1.$$

Thus, our sequence of functions converges uniformly to the function $f : I \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & 0 \leq x < 1 \\ 1 & x = 1 \end{cases}$$

for $x \in I$. The graphs of f_n are illustrated for $n = 1, 2, \ldots, 20$ in Figure 7.1.

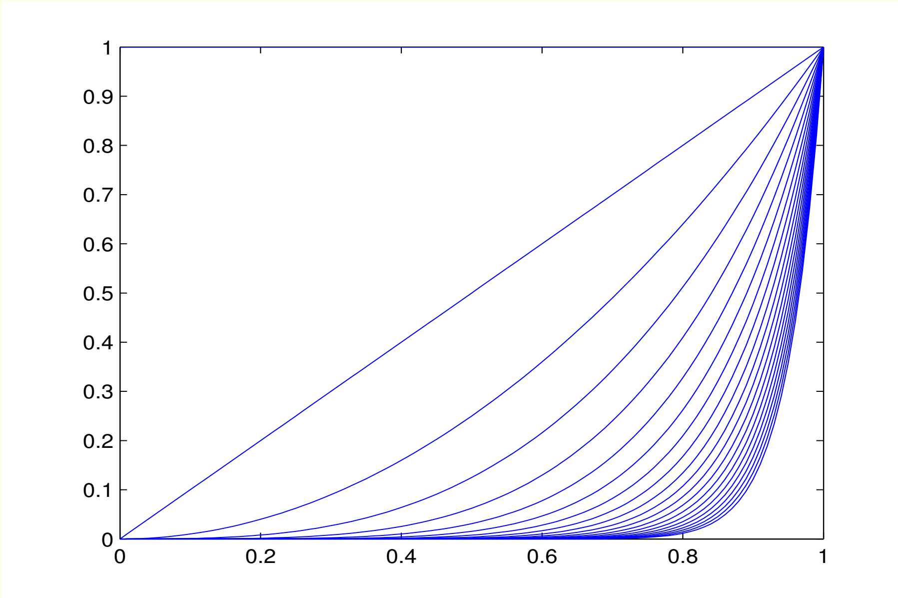

Figure 7.1: A famous picture: The graphs of $f_n(x) = x^n$ for $n = 1, 2, \ldots, 20$.

It is important to note that each function f_n is continuous on I, however, the limit function f is not continuous on I. This illustrates that nice properties like continuity can be "broken" under taking pointwise limits.

A much stronger notion of convergence is captured by the following definition.

Definition 7.2. *Let $\{f_n\}$ be a sequence of complex-valued functions on X. Let $f : X \to \mathbb{C}$ be another complex-valued function on X. We say that the sequence $\{f_n\}$ converges uniformly to f on X if, for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ for which*

$$|f_n(x) - f(x)| < \epsilon \quad \textit{whenever} \quad x \in I \ \textit{ and } \ n \geq N.$$

In contrast to the definition of pointwise convergence, the definition of convergence requires that the integer N depend only on ϵ and be independent of $x \in X$. This notion is illustrated in Figure 7.2. In the figure, we see the graph of a real-valued function f (in black) in the center of a "band" of radius ϵ (in red). For a sequence of functions $\{f_n\}$ to converge uniformly to f (on an interval) means that, for sufficiently large n, the graph of f_n is completely contained in the band of radius ϵ surrounding f; the blue line is an example of the graph of one such f_n.

We further illustrate this definition with some examples.

Example 7.2:

Consider the sequence $\{f_n\}$ of functions defined on the interval $I = [-\pi, \pi]$ by

$$f_n(x) = \cos(x/n) - 1/2$$

for $x \in I$ and $n \in \mathbb{N}$. The graphs of f_n are illustrated for $n = 1, 2, \ldots 10$ in Figure 7.3.

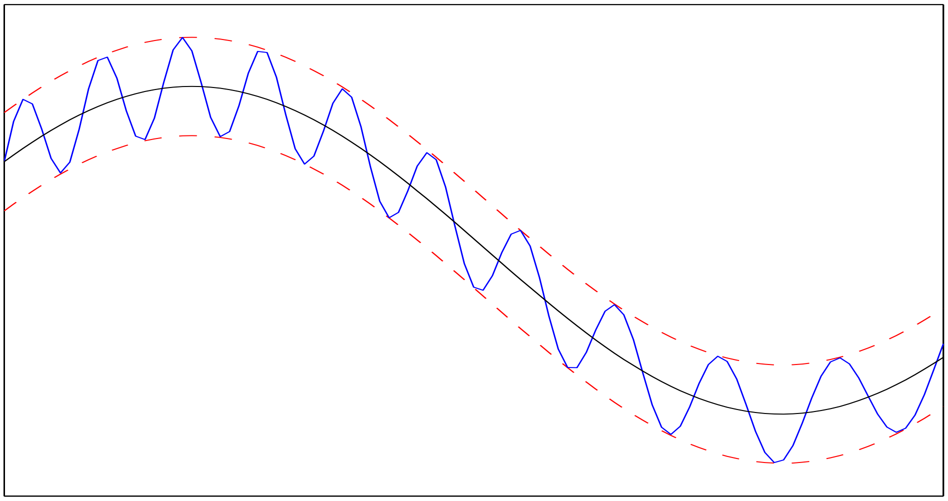

Figure 7.2: An illustration of uniform convergence

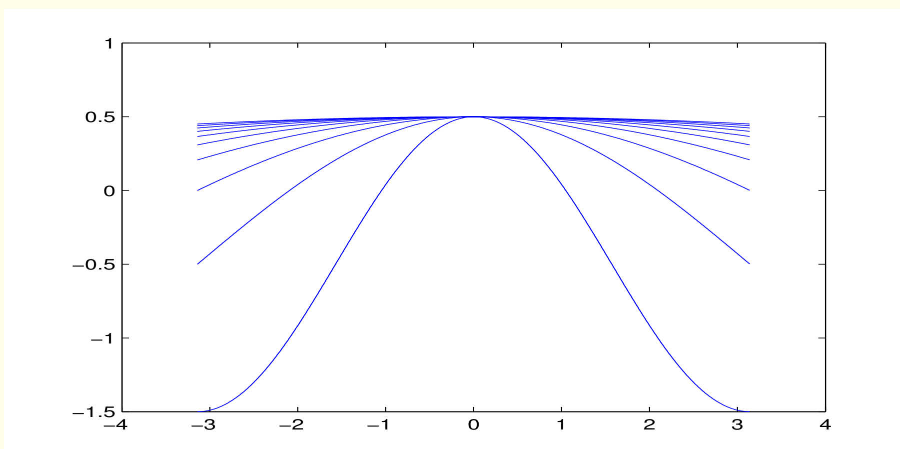

Figure 7.3: The graphs of $f_n(x) = \cos(x/n) - 1/2$ for $n = 1, 2, \ldots, 10$.

The figure suggests that the sequence $\{f_n\}$ converges to the constant function $f(x) = 1/2$ as $n \to \infty$. Let's prove that, not only does it converge to $f(x) = 1/2$, it does so uniformly.
Let $\epsilon > 0$ and select $N \in \mathbb{N}$ such that $N > \pi/\sqrt{\epsilon}$. Recalling the inequality for cosine,

$$|\cos(\theta) - 1| \leq |\theta^2| \qquad \text{for all } \theta \in \mathbb{R}$$

which can be gotten from the mean value theorem or the racetrack principle, we observe that, for any $n \geq N$ and $x \in I = [-\pi, \pi]$,

$$|f_n(x) - f(x)| = |\cos(x/n) - 1/2 - 1/2| = |\cos(x/n) - 1| \leq \frac{x^2}{n^2} \leq \frac{\pi^2}{n^2} < \epsilon$$

because $n^2 \geq N^2 > \pi^2/\epsilon$. The careful reader should note that the above estimate holds for all $x \in I$ and for all $n \geq N$ (and not for a particular x). We have shown that the sequence $\{f_n\}$ converges uniformly to $f(x) = 1/2$.

Exercise 7.1:

Given an interval I, we recall the supremum norm defined, for $f : I \to \mathbb{C}$ by

$$\|f\|_\infty = \sup_{x \in I} |f(x)|.$$

In this exercise, you will prove that $\|\cdot\|_\infty$ is a *bona fide* norm on the space of bounded complex-valued functions on I.

1. Prove that, for any pair of bounded functions function f and g,

$$\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty.$$

2. Prove that, for each complex number α and bounded function $f : I \to \mathbb{C}$,

$$\|\alpha f\|_\infty = |\alpha| \|f\|_\infty$$

where $|\alpha|$ is the complex modulus of α.

3. Prove that, for a bounded function f, $\|f\|_\infty = 0$ if and only if $f(x) = 0$ for all $x \in I$.
4. Given a sequence $\{f_n\}$ of bounded complex-valued functions on I and $f : I \to \mathbb{C}$, prove that the sequence $\{f_n\}$ converges uniformly to f if and only if

$$\lim_{n \to \infty} \|f_n - f\|_\infty = 0.$$

As the notion of "Cauchy sequence" is essential for the convergence for complex-numbers and, in fact, provides a characterization for convergence as you proved in Homework 1, we have a similar Cauchy property for functions which characterizes uniform convergence. This characterization is outlined in the following theorem.

Theorem 7.3. *Let $\{f_n\}$ be a sequence of complex-valued functions on an interval $I \subseteq \mathbb{R}$. The sequence $\{f_n\}$ converges uniformly (to some function f) on I if and only if it satisfies the following property:*

(UC) For all $\epsilon > 0$, there exists a natural number N such that

$$|f_n(x) - f_m(x)| < \epsilon \quad \text{whenever } x \in I \text{ and } n, m \geq N.$$

The equivalent property (UC) is called the Uniform Cauchy condition. Any sequence of functions $\{f_n\}$ satisfying the condition is said to be uniformly Cauchy on I.

Proof. Let us first assume that $\{f_n\}$ converges uniformly to a function f on I. Let $\epsilon > 0$ and by our supposition let N be a natural number for which

$$|f_n(x) - f(x)| < \epsilon/2$$

for all $n \geq N$ and $x \in I$. Then, for any $n, m \geq N$, we have

$$|f_n(x) - f_m(x)| = |f_n(x) - f(x) + f(x) - f_m(x)| \leq |f_n(x) - f(x)| + |f(x) - f_m(x)| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all $x \in I$. Thus the sequence $\{f_n\}$ is uniformly Cauchy on I.

Conversely, let's assume that the sequence $f_n(x)$ is uniformly Cauchy on I. This implies, in particular, that $\{f_n(x)\}$ is a Cauchy sequence of complex numbers for each $x \in I$. Because all Cauchy sequences of complex numbers converge, for each $x \in I$, the limit $\lim_{n\to\infty} f_n(x)$ exists and we will denote its value by $f(x)$, which is just a complex number. In this way, we produce a function $f : I \to \mathbb{C}$ simply by identifying each x with the value of the limit $\lim_{n\to\infty} f_n(x)$, i.e., defining

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for each $x \in I$. So now we have a candidate (f) for the uniform limit. It remains to show that our sequence, in fact, converges uniformly to this f. To see this, we let $\epsilon > 0$ and choose a natural number N for which

$$|f_n(x) - f_m(x)| < \frac{\epsilon}{2}$$

for all $n, m \geq N$ and $x \in I$. Now, let $x \in I$ and $n \geq N$ be arbitrary (but fixed). The convergence of the numerical sequence $\{f_n(x)\}$ guarantees that there is some natural number $N_x \geq N$ for which

$$|f_m(x) - f(x)| < \frac{\epsilon}{2}$$

whenever $m \geq N_x$. In particular, this works when $m = N_x \geq N$ and so

$$|f_n(x) - f(x)| = |f_n(x) - f_{N_x}(x) + f_{N_x}(x) - f(x)| \leq |f_n(x) - f_{N_x}(x)| + |f_{N_x}(x) - f(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus, to each $\epsilon > 0$, we have found a natural number N for which

$$|f_n(x) - f(x)| < \epsilon$$

whenever $x \in I$ and $n \geq N$. Therefore, $\{f_n\}$ converges uniformly on I (to f). □

Corollary 7.4. *Let $B = B(X;\mathbb{C})$ denote the set of bounded complex-valued function on X and define*

$$d_\infty(f, g) = \|f - g\|_\infty = \sup_{x\in X} |f(x) - g(x)|$$

for $f, g \in B$. Then (B, d_∞) is a complete metric space.

Proof. In the previous exercise 7.1, you showed that $\|\cdot\|_\infty$ defined a norm on B and, as each norm defines a metric in precisely the way above, we conclude that (B, d_∞) is a metric space. Let $\{f_n\}$ be a Cauchy-sequence in in this metric, i.e., for every $\epsilon > 0$, there exists N for which

$$d_\infty(f_n, f_m) = \sup_{x\in X} |f_n(x) - f_m(x)| < \epsilon$$

whenever $n, m \geq N$. In particular,

$$|f_n(x) - f_m(x)| \leq d_\infty(f_n, f_m) < \epsilon$$

for all $x \in X$ and $n, m \geq N$. Hence, $\{f_n\}$ is uniformly Cauchy. By the Theorem 7.3, $\{f_n\}$ is uniformly convergent to some $f : X \to \mathbb{C}$ and, in view of the Exercise 7.1,

$$\lim_{n\to\infty} d_\infty(f_n, f) = \lim_{n\to\infty} \|f_n - f\| = 0.$$

It simply remains to show that $f \in B$. To this end, let $\epsilon = 1$ and let N be such that

$$d_\infty(f_n, f) = \|f_n - f\|_\infty < 1$$

whenever $n \geq N$. In particular, we have

$$\sup_{x\in X} |f(x)| = \|f\|_\infty \leq \|f - f_N\|_\infty + \|f_N\|_\infty < 1 + \|f_N\|_| infty < \infty$$

where we have used the fact that $f_N \in B$ and hence $\|f_N\|_\infty$ is finite. □

Theorem 7.3 extremely useful when one has a sequence of nice functions (which is uniformly Cauchy) but has no obvious candidate for the uniform limit. Here, of course, infinite series comes to mind.

Definition 7.5. *Let $\{f_n\}$ be a sequence of complex-valued functions on I. The (formal) sum $\sum_n f_n$ is called a series of functions. To investigate the convergence of $\sum_n f_n$, we define, for each $N = 1, 2, \dots,$*

$$S_N(x) = \sum_{n=1}^{N} f_n(x) \qquad \text{for } x \in I.$$

The functions $S_1, S_2, \dots,$ form a sequence of complex-valued functions on I, $\{S_N\}$, called the sequence of partial sums for the series $\sum_n f_n$. If, for each $x \in I$, the limit

$$\lim_{N\to\infty} S_N(x)$$

exists, we say that the series $\sum_n f_n$ converges on I. In this case, the limit is a function $S : I \to \mathbb{R}$ defined by

$$S(x) = \lim_{N\to\infty} S_N(x) = \lim_{N\to\infty} \sum_{n=1}^{N} f_n(x)$$

and we write

$$\sum_{n=1}^{\infty} f_n(x) = S(x)$$

to denote this function, called the sum of the series. We say that the series $\sum_n f_n$ converges uniformly on I if its sequence of partial sums $\{S_N\}$ converges uniformly on I to the sum of the series.

As with numerical series, one can often learn that a series converges without ever knowing its sum. For instance, the integral test from calculus shows that the series of numbers

$$\sum_{n=1}^{\infty} \frac{1}{n^3}$$

converges (this is p-series for $p = 3$). Though it can be approximated to any degree of accuracy, its sum it unknown. With this in mind, it is important to have various test for series (uniform) convergence without knowing the limit. The following corollary of Theorem 7.3 gives us exactly this.

Corollary 7.6 (Uniform Cauchy Criterion)**.** *Let $\{f_n\}$ be a sequence of complex-valued functions on I and consider the series $\sum_n f_n$. The series $\sum_n f_n$ converges uniformly on I if and only if the following property is satisfied:*

For all $\epsilon > 0$, there is a natural number N for which

$$\left| \sum_{k=n}^{k=m} f_k(x) \right| < \epsilon$$

for all $x \in I$ and $m \geq n \geq N$. This property is called the Uniform Cauchy Criterion for the series $\sum_n f_n$.

Exercise 7.2:

In this exercise, you will prove Corollary 7.6 and then use the corollary to establish sufficient conditions for the absolute convergence of power series – things you will remember from calculus (M122).

1. Using Theorem 7.3, prove Corollary 7.6.

2. If a series $\sum_n f_n$ of functions $\{f_m\}$ converges uniformly on I, prove that $\{f_n\}$ converges uniformly to the zero function on I.

3. For the remainder of this exercise, we fix a positive constant M and define $I = [-M, M] \subseteq \mathbb{R}$. Given a sequence of complex-numbers $\{c_n\}$, consider the sequence of complex-valued functions $\{f_n\}$ on I defined by
$$f_n(x) = \frac{c_n}{n!}x^n$$
for $x \in I$. If the sequence $\{c_n\}$ is bounded, i.e., $\sup_{n\in\mathbb{N}} |c_n| < \infty$, use Corollary 7.6 (and no other convergence test) to prove that the series
$$\sum_{n=1}^{\infty} \frac{c_n}{n!}x^n$$
converges uniformly on I.

4. Let $f : I \to \mathbb{C}$ be infinitely differentiable and assume that $\sup_{n=0,1,\dots} |f^{(n)}(0)| < \infty$; here $f^{(n)}(0)$ is the n^{th}-derivative of f at 0. Use the previous item to prove that the series
$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n$$
converges uniformly on I . This series is called the Maclaurin series for f. (Your proof here should be approximately one sentence).

5. Looking back at Item 3, find a condition on the sequence $\{c_k\}$ which is less restrictive than boundedness and which still guarantees that the series
$$\sum_{n=1}^{\infty} \frac{c_n}{n!}x^n$$
converges uniformly on I. Hint: You should take a look at Stirling's formula (which you can take for granted as long as you interpret the formula/approximation correctly). If you're interested, a nice proof of Stirling's formula can be found in Exercise 5 of Homework 2 for my Math 122 class.

7.0.1 Properties of Uniform Convergence

In this short subsection, we discuss some properties preserved under uniform convergence. Specifically, we focus on continuity and integration. Let's consider a couple of examples.

Example 7.3:

Given $0 < \delta < 1$, let $I_\delta = [-1 + \delta, 1 - \delta]$ and consider the series
$$\sum_{n=0}^{\infty} x^n$$
for $x \in I_\delta$. We claim that this series converges uniformly on I_δ to the function
$$f(x) = \frac{1}{1 - x}. \tag{7.1}$$

To see this, we first observe that the partial sums $\{S_N\}$ satisfy the formula

$$S_N(x) = \sum_{n=0}^{N} x^n = \frac{1-x^{N+1}}{1-x}$$

for $x \in I_\delta$. The validity of this formula can be seen by multiplying both sides by $1-x$ and simplifying. To see that this series converges uniformly, let $\epsilon > 0$ and choose M to be a natural number for which $M > \ln(\epsilon\delta)/\ln(1-\delta)$. For any $x \in I_\delta$ and $N \geq M$, observe that

$$|f(x) - S_N(x)| = \left|\frac{1}{1-x} - \frac{1-x^{N+1}}{1-x}\right| = \frac{|x|^{N+1}}{|1-x|} \leq \frac{(1-\delta)^{N+1}}{\delta} < \epsilon$$

where we have used the fact that $N+1 > M \geq \ln(\epsilon\delta)/\ln(1-\delta)$. Therefore, we have proved that this series converges uniformly to f. I encourage you to show that this series converges uniformly using only Corollary 7.6 (and not making reference to f).

An important thing to note about the above example is that, each $S_N(x)$ is continuous and the limit function $f(x) = 1/(1-x)$ is also continuous on the interval I_δ, a fact that was also true in the preceding example. This stands in contrast to the Example 7 in which the limit function failed to be continuous. As it turns out, this is a key difference between pointwise convergence and uniform convergence. This is detailed in the following theorem, whose proof Is needed, Evan!

Theorem 7.7. *Let $\{f_n\}$ be a sequence of complex-valued functions on I and suppose that $\{f_n\}$ converges uniformly to a function $f : I \to \mathbb{C}$. If each function f_n is continuous, i.e., $\{f_n\} \subseteq C^0(I;\mathbb{C})$, then f is necessarily a continuous function.*

Proof. To show that f is continuous on I, we must show that, for each $x \in I$ and $\epsilon > 0$, there is a $\delta > 0$ for which $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. To this end, let $x \in I$ and $\epsilon > 0$ be fixed. Since $\{f_n\}$ converges uniformly to f on I, let N be such that $|f_n(y) - f(y)| < \epsilon/3$ whenever $n \geq N$ and $y \in I$. In particular, $|f_N(y) - f(y)| < \epsilon/3$ for all $y \in I$. Now, because f_N is continuous on I, it is continuous at $x \in I$ and so there is a $\delta > 0$ for which $|f_N(x) - f_N(y)| < \epsilon/3$ whenever $|x - y| < \delta$. Thus, for $y \in I$ with $|x - y| < \delta$, we have

$$|f(x) - f(y)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}.$$

□

Let's explore some other important properties of uniform convergence. Our next result shows that uniform convergence plays nicely with the Riemann-Darboux integral.

Theorem 7.8. *Let $\{f_n\}$ be a sequence of complex-valued functions which converges uniformly to a function $f : I \to \mathbb{C}$; here, $I = [a,b]$. If each function f_n is Riemann-integrable, i.e., $\{f_n\} \subseteq R(I)$, then f is Riemann-integrable and*

$$\lim_{n\to\infty} \int_I |f_n - f| = 0.$$

Further

$$\lim_{n\to\infty} \int_I f_n = \int f.$$

Proof. We first show that the limit f is Riemann-integrable by showing its real and imaginary parts, u and v are Riemann-integrable. For each n, denote by u_n and v_n the real and imaginary parts of f_n respectively. We will show that u and v are Riemann integrable by appealing to the $\epsilon - P$ characterization, Theorem 6.7. Let's first focus on the real parts $\{u_n\}$ and u. Let $\epsilon > 0$ and, by the uniformly convergence of $\{f_n\}$, let N be a natural number for which

$$|u_n(x) - u(x)| \leq \sqrt{(u_n(x) - u(x))^2 + (v_n(x) - v(x))^2} = |f_n(x) - f(x)| < \epsilon/4(b-a)$$

for all $x \in I$ and $n \geq N$. In particular, upon setting $u_0 = u_N$, this yields the inequality

$$u_0(x) - \frac{\epsilon}{4(b-a)} < u(x) < u_0(x) + \frac{\epsilon}{4(b-a)} \tag{7.2}$$

for all $x \in I$. This inequality implies that u is bounded on the interval I in view of our hypothesis that $u_0 = u_N \in R(I)$. By virtue of Theorem 6.7, let P be a partition of I for which $U(u_0, P) - L(u_0, P) < \epsilon/2$. For this partition, the inequality (7.2) guarantees that

$$\begin{aligned} U(u,P) &= \sum_n \left(\sup_{x_{n-1} \leq x \leq x_n} u(x) \right))(x_n - x_{n-1}) \\ &\leq \sum_n \left(\sup_{x_{n-1} \leq x \leq x_n} u_0(x) + \frac{\epsilon}{4(b-a)} \right) (x_n - x_{n-1}) \\ &\leq U(u_0, P) + \sum_n \frac{\epsilon}{4(b-a)} (x_n - x_{n-1}) \\ &\leq U(u_0, P) + \frac{\epsilon}{4}. \end{aligned}$$

Similarly, the inequality (7.2) guarantees the analogous lower estimate

$$L(u_0, P) - \frac{\epsilon}{4} \leq L(u, P)$$

Together, these estimates guarantees that

$$U(u,P) - L(u,P) \leq U(u_0, P) - L(u_0, P) + \frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

and from this we can conclude that $u \in R(I)$. A completely analogous argument shows that $v \in R(I)$ and so, by the definition of Riemann-integrability for complex-valued functions, the limit function $f \in R(I)$.

Let us now prove the statements concerning the limit $\lim_{n\to\infty} \int_I |f_n - f|$. In view of the definition of the L^∞-norm, we have

$$|f_n(x) - f(x)| \leq \|f_n - f\|_\infty$$

for all $x \in I$ and $n \in \mathbb{N}$. In view of Lemma 6.22, we have

$$0 \leq \int_I \leq |f_n(x) - f(x)|\, dx \leq \int_I \|f_n - f\|_\infty \, dx = (b-a)\|f_n - f\|_\infty.$$

Thus, by virtue of Exercise 9 and the squeeze theorem, the preceding inequality shows that

$$\lim_{n\to\infty} \int_I |f_n - f| = 0$$

because $\|f_n - f\|_\infty \to 0$ as $n \to \infty$.

Finally, by virtue of Theorem 6.20 and Proposition 6.21, we have

$$\left| \int_I f_n - \int_I f \right| = \left| \int_I (f_n - f) \right| \leq \int_I |f_n - f|$$

for all n. Another appeal to the squeeze theorem (and the preceding limit) guarantees that

$$\lim_{n\to\infty} \int_I f_n = \int_I f.$$

$\square$

Corollary 7.9. *Let $\{f_n\}$ be a sequence of complex-valued functions on $I=[a,b]$ and suppose that the series $\sum_{n=0}^\infty f_n$ converges uniformly on I. If each f_n is Riemann-integrable, then the sum of the series is Riemann-integrable and*

$$\int_I \sum_{n=0}^\infty f_n = \sum_{n=0}^\infty \int_I f_n.$$

Proof. The hypothesis that $\sum_{n=0}^\infty f_n$ converges uniformly means that the sequence of partial sums $\{S_N\}$ defined by

$$S_N(x)=\sum_{n=0}^N f_n(x)$$

for $x\in I$ converges uniformly on I. Also, the supposition that each f_n is Riemann-integrable guarantees that each partial sum is Riemann-integrable in view of Theorem 6.20. By the (finite) linearity of the integral, we have

$$\int_I S_N=\sum_{n=0}^N\int_I f_n$$

for each natural number N. Thus, an appeal to the preceding theorem guarantees that the limit $\sum_{n=0}^\infty f_n = \lim_{N\to\infty} S_N$ is Riemann-integrable and

$$\int_I\sum_{n=0}^\infty f_n=\int_I \lim_{N\to\infty} S_N=\lim_{N\to\infty}\int_I S_N=\lim_{N\to\infty}\sum_{n=0}^N\int_I f_n;$$

in particular, the limit on the right exists. Of course, this is what it means for the series of the numbers $\int_I f_n$ to converge and so we have

$$\int_I\sum_{n=0}^\infty f_n=\lim_{N\to\infty}\sum_{n=0}^N\int_I f_n=\sum_{n=0}^\infty\int_I f_n.$$

□

7.0.2 The Weierstrass M-test

We've been developing the theory of uniform convergence for sequences of functions. Along the way, we've proved some results about the uniform convergence of series of functions, the most important of which is Corollary 7.6. This corollary showed that a series is uniformly convergent if and only if it satisfies the Uniform Cauchy Criterion. As you saw in Exercise 10, while this criterion/condition is very useful, it is not terribly easy to apply. Our main result of this section, the M-test of Weierstrass, gives an relatively straightforward condition guaranteeing that a given series converges uniformly. We will then amass some facts following from this result which will be used in our study of Fourier series.

Theorem 7.10 (The Weierstrass M-test)**.** *Let $I=[a,b]$ be an interval and consider a sequence of bounded complex-valued functions $\{f_n\}$ on I. For each $n\in\mathbb{N}$, set*

$$M_n=\|f_n\|_\infty=\sup_{x\in I}|f_n(x)|.$$

If the series $\sum_{n=1}^\infty M_n$ converges, then the series $\sum_{n=1}^\infty f_n$ converges uniformly on I.

Before giving the proof, observe that the series $\sum_{n=1}^\infty M_n$ is a series of non-negative numbers and determining the convergence of this series is the subject matter of introductory calculus. This is usually an easier condition to verify that the Cauchy criterion.

Proof. We will verify that the Cauchy criterion (Corollary 7.6) is satisfied for the series $\sum_n f_n$. To this end, let $\epsilon > 0$. Given that $\sum_n M_n$ converges, its partial sums are necessarily a Cauchy sequence and so there must be some natural number N for which

$$\sum_{k=n}^{m} M_k \leq \sum_{k=n-1}^{m} M_k = \sum_{k=1}^{m} M_k - \sum_{k=1}^{n} M_k = \left|\sum_{k=1}^{m} M_k - \sum_{k=1}^{n} M_k\right| < \epsilon$$

whenever $m \geq n \geq N$. Here, we have used the fact that $M_k \geq 0$ for all k. Observe now that, for any $x \in I$ and $m \geq n \geq N$, the triangle inequality guarantees that

$$\left|\sum_{k=n}^{m} f_k(x)\right| \leq \sum_{k=n}^{m} |f_k(x)| \leq \sum_{k=n}^{m} \|f_k\|_\infty = \sum_{k=n}^{m} M_k < \epsilon,$$

as desired. $\square$

Following directly from Theorems 7.10 and 7.7 and Corollary 7.9, we obtain the following corollary.

Corollary 7.11. *Let I be an interval and let $\{f_k\}$ be a sequence of complex-valued functions on I, i.e., $\{f_k\} \subseteq C(I)$. For each $n \in \mathbb{N}$, set*

$$M_n = \|f_n\|_\infty = \sup_{x \in I} |f_n(x)|.$$

If the series $\sum_{n=1}^{\infty} M_n$ converges, then $\sum_{n=1}^{\infty} f_n$ converges uniformly on I and its sum

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

is a continuous function on I, i.e., $f \in C(I)$. Further,

$$\int_I f = \int_I \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n.$$

Proof. The statement regarding uniform convergence follows directly from Theorem 7.10. Because f_n is continuous for each n, the partial sums $\{S_n\}$ are necessarily continuous functions on I. The uniform convergence of the series is the statement that the partial sums converge uniformly to the sum of the series and so, by virtue of Theorem 7.7, the sum f is necessarily continuous on I. Finally, upon noting that $\{f_n\} \subseteq C(I) \subseteq R(I)$, an appeal to Corollary 7.9 gives the final statement immediately. $\square$

Exercise 7.3: T

he Weierstrass M-test says that the "M condition", i.e., the condition that $\sum_{n=1}^{\infty} M_n$ converges, is a sufficient condition for the uniform convergence of the series $\sum f_n$. This is in contrast to Corollary 7.6 which gives a condition both necessary and sufficient for uniform convergence. Show that that "M condition" (of the Weierstrass M-test) is not necessary for convergence. That is, find a sequence of functions $\{f_n\}$ on an interval I for which $\sum_{n=1}^{\infty} f_n$ converges uniformly yet $\sum_{n}^{\infty} M_n = \infty$ for $M_n = \|f_n\|_\infty$. Hint: A nice example can be produced which is an alternating series. Feel free to use results from introductory calculus (such as the alternating series test).

7.0.3 Defining Convergence with the Integral: A glimpse at Lebesgue norms

As the supremum norm $\|\cdot\|_\infty$ allows us to measure the "size" of a function bounded function (and with it you were able to characterize uniform convergence), the integral also allows us to measure the "size" of a function by integrating its absolute value. Measuring the size of functions with the integral turns out to be a very fruitful activity. To formalize things, I will take this opportunity to introduce a class of "norms" on functions, called the Lebesgue norms or the L^p norms, of which the supremum norm is an important example. To this end, we fix an interval I and, for each $1 \leq p < \infty$, we define the $L^p(I)$ norm of a function $f \in R(I)$ by

$$\|f\|_p = \left(\int_I |f(x)|^p \, dx\right)^{1/p}.$$

For $p = \infty$, we have as before

$$\|f\|_p = \|f\|_\infty = \sup_{x \in I} |f(x)|$$

for $f \in R(I)$. For each $1 \leq p \leq \infty$, each L^p norm gives us a different way to measure the "size" of a function. Let's accumulate some facts about these norms.

Proposition 7.12. *Given an interval I and $1 \leq p \leq \infty$, let $\|\cdot\|_p$ denote the $L^p(I)$ norm defined above. Then, for any $f, g \in R(I)$ and $\alpha \in \mathbb{C}$, we have*

1.

$$\|f\|_p \geq 0$$

2.

$$\|\alpha f\|_p = |\alpha| \|f\|_p$$

3.

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p$$

Truthfully, the above proposition only guarantees that $\|\cdot\|_p$ is a so-called *semi-norm* on $R(I)$ because there are non-zero functions $f \in R(I)$ for which $\|f\|_p = 0$.

Proof. As you have already shown that these properties hold when $p = \infty$ (Exercise 9), we shall assume that $1 \leq p < \infty$. Now, because the integral of a non-negative function is non-negative, the validity of Item 1 is clear. Also, for $f \in R(I)$ and $\alpha \in \mathbb{C}$,

$$\|\alpha f\|_p^p = (\|\alpha f\|_p)^p = \int_I |\alpha f(x)|^p \, dx = \int_I |\alpha|^p |f(x)|^p \, dx = |\alpha|^p \int_I |f(x)|^p \, dx$$

from which we immediately obtain Item 2. It remains to prove Item 3, also called Minkowski's inequality. This inequality is most easily obtained using the machinery of measure theory, though our proof here only relies on the convexity of the function $\mathbb{C} \ni z \mapsto |z|^p$, a fact which can be established using only elementary calculus.

To this end, we first assume show that, if $h_1, h_2 \in R(I)$ are such that $\|h_1\|_p, \|h_2\|_p \leq 1$, then, for any $0 \leq t \leq 1$, $\|th_1 + (1-t)h_2\|_p \leq 1$. This is equivalently the statement that the unit ball

$$B_p = \{h \in R(I) : \|h\|_p \leq 1\}$$

is a convex set. Let us fix $0 \leq t \leq 1$ and $h_1, h_2 \in B_p$ and observe that the convexity of the map $z \mapsto |z|^p$ guarantees that

$$|th_1(x) + (1-t)h_2(x)|^p \leq t|h_1(x)|^p + (1-t)|h_2(x)|^p$$

for all $x \in I$. I'll make note that the convexity used here for complex numbers is also called the supporting hyperplane property and can be understood geometrically as the graph of the function $|z|^p$ always living below its secant lines/planes. In view of this inequality, the monotonicity of the integral guarantees that

$$\int_I |th_1(x) + (1-t)h_2(x)|^p \, dx \leq t \int_I |h_1(x)|^p \, dx + (1-t)|h_2(x)|^p \, dx$$

or equivalently

$$\|th_1 + (1-t)h_2\|_p^p \leq t\|h_1\|_p^p + (1-t)\|h_2\|_p^p.$$

Recalling that $\|h_1\|_p \leq 1$ and $\|h_2\|_p \leq 1$, we conclude that

$$\|th_1 + (1-t)h_2\|_p^p \leq t \cdot 1 + (1-t) \cdot 1 = 1$$

and so $\|th_1 + (1-t)h_2\|_p \leq 1$, as was asserted.

We now get to the task at hand. Let $f, g \in R(I)$ and we shall assume that $\|f\|_p$ and $\|g\|_p$ are non-zero (treating these trivial cases is much more simple). We write

$$\frac{f+g}{\|f\|_p + \|g\|_p} = \frac{\|f\|_p}{\|f\|_p + \|g\|_p} \frac{f}{\|f\|_p} + \frac{\|g\|_p}{\|f\|_p + \|g\|_p} \frac{g}{\|g\|_p} = t\frac{f}{\|f\|_p} + (1-t)\frac{g}{\|g\|_p}$$

where $t = \|f\|_p/(\|f\|_p + \|g\|_p)$ is a number between 0 and 1. By virtue of Item 2, both $h_1 = f/\|f\|_p$ and $h_2 = g/\|g\|_p$ have L^p norm 1. In view of the property proved in the preceding paragraph, we conclude that

$$\left\| \frac{f+g}{\|f\|_p + \|g\|_p} \right\|_p = \|th_1 + (1-t)h_2\|_p \leq 1.$$

Therefore, a final appeal to Item 2 gives the inequality

$$\frac{1}{\|f\|_p + \|g\|_p} \|f+g\|_p \leq 1$$

from which the desired result follows without trouble. $\square$

With these norms and this way of measuring functions, we can define new notions of convergence. To this end, given a sequence of functions $\{f_n\} \subseteq R(I)$ and $f \in R(I)$, we say that $\{f_n\}$ *converges to f in $L^p(I)$ or with respect to the L^p norm* if

$$\lim_{n\to\infty} \|f_n - f\|_p = 0.$$

There are three L^p norms that will be of particular interest for us, $p = 1$, 2 and ∞. In the case that $p = 2$, there is an additional structure with which you are already familiar from linear algebra, the inner product (a generalization of the dot product). For integrable functions f and g, we define the L^2 inner product of f and g to be the number

$$\langle f, g \rangle = \int_I f(x)\overline{g(x)}\, dx.$$

As it is easy to verify using properties of the integral, the $L^2(I)$ inner product satisfies the following properties:

1. $$\langle f, g \rangle = \overline{\langle g, f \rangle} \qquad \text{for } f, g \in R(I)$$

2. $$\langle \alpha f + \beta h, g \rangle = \alpha \langle f, g \rangle + \beta \langle h, g \rangle \qquad \text{for } f, g, h \in R(I) \text{ and } \alpha, \beta \in \mathbb{C}.$$

3. $$\langle g, \alpha f + \beta h \rangle = \overline{\alpha} \langle g, f \rangle + \overline{\beta} \langle g, h \rangle \qquad \text{for } f, g, h \in R(I) \text{ and } \alpha, \beta \in \mathbb{C}.$$

We also notice, that the L^2 inner product recaptures the L^2 norm:

$$\|f\|_2 = \left(\int_I |f(x)|^2\, dx \right)^{1/2} = \left(\int_I f(x)\overline{f(x)}\, dx \right)^{1/2} = \sqrt{\langle f, f \rangle}$$

for $f \in R(I)$. An extremely important property of the L^2 inner product is captured by the following theorem.

Theorem 7.13 (The Cauchy-Schwarz Inequality)**.** *For any $f, g \in R(I)$,*

$$|\langle f, g \rangle| \leq \|f\|_2 \|g\|_2$$

Proof. Let's first assume that $h_1, h_2 \in R(I)$ have $\|h_1\|_2 = \|h_2\|_2 = 1$. We observe that, for any $x \in I$,

$$0 \leq (|h_1(x)| - |h_2(x)|)^2 = \left(|h_1(x)|^2 + |h_2(x)|^2 - 2|h_1(x)||h_2(x)|\right).$$

Therefore

$$|h_1(x)||h_2(x)| \leq \frac{|h_1(x)|^2}{2} + \frac{|h_2(x)|^2}{2}$$

for all $x \in I$. By virtue of Proposition 6.21, the preceding inequality shows that

$$\begin{aligned}
|\langle h_1, h_2\rangle| &= \left|\int_I h_1(x)\overline{h_2(x)}\,dx\right| \\
&\leq \int_I |h_1(x)||h_2(x)|\,dx \\
&\leq \frac{1}{2}\int_I |h_1(x)|^2\,dx + \frac{1}{2}\int_I |h_2(x)|^2\,dx \\
&\leq \frac{1}{2}\|h_1\|_2^2 + \frac{1}{2}\|h_2\|_2^2 = 1.
\end{aligned}$$

Thus $|\langle h_1, h_2\rangle| \leq 1$ whenever $h_1, h_2 \in R(I)$ have unit L^2-norm. Now, given any $f, g \in R(I)$ with non-zero L^2 norms, we observe that $h_1 = f/\|f\|_2$ and $h_2 = g/\|g\|_2$ have $\|h_1\|_2 = \|h_2\|_2 = 1$ and so by the properties of the L^2 inner product outlined above

$$|\langle f, g\rangle| = \|f\|_2\|g\|_2\left|\left\langle \frac{f}{\|f\|_2}, \frac{g}{\|g\|_2}\right\rangle\right| = \|f\|_2\|g\|_2|\langle h_1, h_2\rangle| \leq \|f\|_2\|g\|_2$$

as desired.

Finally, let us assume that $\|f\|_2 = 0$ or $\|g\|_2 = 0$. In this final case, our job is to show that $\langle f, g\rangle = 0$ because the right-hand side of the Cauchy-Schwarz inequality is zero. Without loss of generality we assume that $\|g\|_2 = 0$ and observe that, for all $t \in \mathbb{R}$,

$$\begin{aligned}
\|f + tg\|_2^2 &= \langle f + tg, f + tg\rangle = \langle f, f\rangle + \langle f, tg\rangle + \langle tg, f\rangle + \langle tg, tg\rangle \\
&= \|f\|_2^2 + \langle f, tg\rangle + \overline{\langle f, tg\rangle} + t^2\|g\|_2^2 \\
&= \|f\|_2^2 + 2\operatorname{Re}(\langle f, tg\rangle) + 0 \\
&= \|f\|_2^2 + 2t\operatorname{Re}(\langle f, g\rangle)
\end{aligned}$$

where we have used the fact that t is real and $z + \overline{z} = 2\operatorname{Re} z$ for any complex number z (this is something you should check). In view of the equation above, we have

$$0 \leq \|f\|_2^2 + 2t\operatorname{Re}(\langle f, g\rangle)$$

for all $t \in \mathbb{R}$. I claim that this inequality implies that $\operatorname{Re}(\langle f, g\rangle) = 0$. If $\operatorname{Re}(\langle f, g\rangle) \neq 0$, then setting $t = -(\|f\|_2^2 + 1)/\operatorname{Re}(\langle f, g\rangle)$ in the above inequality yields

$$0 \leq \|f\|_2^2 + 2\left(-\frac{\|f\|_2^2 + 1}{\operatorname{Re}(\langle f, g\rangle)}\right)\operatorname{Re}(\langle f, g\rangle) = \|f\|_2^2 - 2\|f\|_2^2 - 2 = -(\|f\|_2^2 + 2)$$

which is impossible because $\|f\|_2^2 + 2 \geq 2 > 0$. From this we conclude that $\operatorname{Re}(\langle f, g\rangle) = 0$. An analogous argument (done by expanding $\|f + itg\|_2^2$) shows that $\operatorname{Im}(\langle f, g\rangle) = 0$. All together, we conclude that $\langle f, g\rangle = 0$. $\square$

There are many generalizations of the Cauchy-Schwarz inequality that turn out to be useful for Fourier analysis. The following one, which we give without proof, is called Hölder's inequality [**?**] . The theorem essentially says that the integral of a product of functions f and g is bounded above in absolute value by the L^p norm of f and the L^q norm of g where $1 \leq p, q \leq \infty$ are such that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Such a pair p and q are said to be conjugate exponents and here we assume the convention that $1/\infty = 0$. So, for example $p = 2$ and $q = 2$ are conjugate exponents. Also $p = 1$ and $q = \infty$ are conjugate exponents.

Theorem 7.14 (Hölder's inequality)**.** *Let $1 \leq p, q \leq \infty$ be conjugate exponents. Then, for any $f, g \in R(I)$, the product fg is integrable and*

$$\left|\int_I f(x)g(x)\,dx\right| \leq \|f\|_p\|g\|_q.$$

Exercise 7.4: T

hough we've already proven the triangle inequality for the L^p norm (also called the Minkowski inequality), please show that the triangle inequality

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p$$

is a consequence of Hölder's inequality (and thus the latter is more "fundamental"). Hint: First observe that $|f(x)+g(x)|^p \leq |f(x)+g(x)|^{p-1}(|f(x)|+|g(x)|)$ for all x. Then apply Hölder's inequality to the terms on the right-hand side.

As an application of Hölder's inequality, we have the following theorem which gives a relationship to convergence between L^p norms.

Theorem 7.15. *Let $I = [a,b]$ be a bounded interval and let $\{f_n\}$ be a sequence in $R(I)$. Also, let $f \in R(I)$. Given any $1 \leq r \leq s \leq \infty$, if*

$$\lim_{n\to\infty} \|f_n - f\|_s = 0 \quad \textit{then} \quad \lim_{n\to\infty}\|f_n - f\|_r = 0.$$

If you take a course in measure theory, you will learn that this result depends critically on the fact that $I = [a,b]$ is a bounded interval . Before giving the proof (taking Hölder's inequality for granted), we note that it implies the following statement (as a special case).

$$\text{If} \quad \lim_{n\to\infty}\|f_n - f\|_\infty = 0 \quad \text{then} \quad \lim_{n\to\infty}\|f_n - f\|_1 = \lim_{n\to\infty}\int_I |f_n(x) - f(x)|\,dx = 0.$$

This statement should be familiar as it recaptures Theorem 7.8 in view of the correspondence between uniform convergence and convergence in the L^∞ norm. Now let's prove the theorem.

Proof. Fixing $1 \leq r \leq s$, set $p = s/r$ and observe that $p \geq 1$. In the case that $r = s = \infty$, the assertion is obvious. We therefore assume that $r < \infty$ and, in view of Hölder's inequality, we obtain

$$\|f_n - f\|_r^r = \int_I |f_n(x) - f(x)|^r\,dx = \int_I |f_n(x) - f(x)|^r \cdot 1\,dx \leq \|(f_n - f)^r\|_p\|1\|_q \tag{7.3}$$

where q is the conjugate exponent to p and 1 is the constant function. If $p = \infty$, necessarily $s = \infty$, $q = 1$ and we have

$$\|(f_n - f)^r\|_p = \sup_{x\in I}|f_n(x) - f(x)|^r = \|f_n - f\|_\infty^r. \tag{7.4}$$

In this case, combining the two preceding inequalities guarantee that

$$\|f_n - f\|_r^r \leq \|f_n - f\|_\infty^r\|1\|_1 = \|f_n - f\|_\infty^r|b-a|$$

or, equivalently,

$$\|f_n - f\|_r \leq (b-a)^{1/r}\|f_n - f\|_\infty.$$

If $p < \infty$, we note that

$$\begin{aligned}
\|(f_n - f)^r\|_p &= \left(\int_I (|f_n(x) - f(x)|^r)^p\,dx\right) \\
&= \left(\int_I |f_n(x) - f(x)|^{pr}\,dx\right)^{1/p} \\
&= (\|f_n - f\|_s^s)^{1/p} = \|f_n - f\|_s^{s/p} = \|f_n - f\|_s^r
\end{aligned}$$

where we have used the fact that $pr = s$ and $s/p = r$. Combining this with (7.3) yields

$$\|f_n - f\|_r^r \leq \|f_n - f\|_s^r \|1\|_q = \|f_n - f\|_s^r \|1\|_q$$

and therefore

$$\|f_n - f\|_r \leq \|f_n - f\|_s \|1\|_q^{1/r}.$$

Finally, noting that

$$\|1\|_q = \begin{cases} \left(\int_I 1^q\right)^{1/q} = (b-a)^{1/q} & q < \infty \\ 1 & q = \infty \end{cases} = (b-a)^{1/q}$$

(as long as we interpret $1/\infty = 0$, we have

$$\|f_n - f\|_r \leq \|f_n - f\|_s (b-a)^{1/rq} = (b-a)^{\left(\frac{1}{r}-\frac{1}{s}\right)} \|f_n - f\|_s \tag{7.5}$$

where we have used the fact that $\frac{1}{r} = \frac{1}{rp} + \frac{1}{rq} = \frac{1}{s} + \frac{1}{rq}$. Combining both cases (7.3) and (7.5) (and using the conventions that $1/0 = \infty$ and $1/\infty = 0$, we obtain

$$\|f_n - f\|_r \leq (b-a)^{\left(\frac{1}{r}-\frac{1}{s}\right)} \|f_n - f\|_s$$

whenever $1 \leq r \leq s$. Finally, if the sequence $\{f_n\}$ has $\lim_{n\to\infty} \|f_n - f\|_s = 0$, the preceding inequality guarantees that $\lim_{n\to\infty} \|f_n - f\|_r = 0$. □

Example 7.4: T

o illustrate the preceding theorem, let's construct a sequence of functions which converge to the zero function with respect to the L^s norm for "small" s but diverge in the L^s norm for "large" s. To this end, set $I = [-1, 1]$ and fix $0 < a \leq \infty$. For each $n \in \mathbb{N}$, define

$$f_n(x) = n^{1/a} e^{-n|x|} \quad \text{for} \quad -1 \leq x \leq 1.$$

We are assuming the convention that $n^{1/a} = n^0 = 1$ when $a = \infty$. Figure 7.4 illustrates f_2 and f_{10} in the case that $a = 1$.

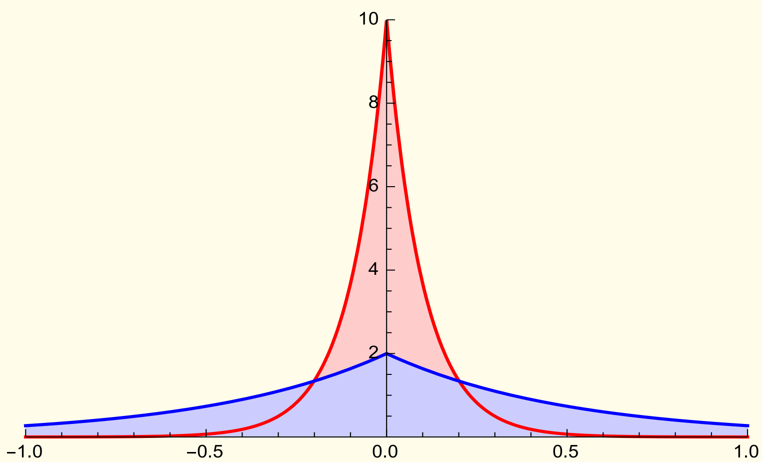

Figure 7.4: The graphs of f_2 and f_{10} when $a = 1$.

A study of this particular sequence of functions provides a nice way to understand which factors contribute to the L^s norm of a function. For this sequence f_n, for a value of $a < \infty$, we see that the peaks at $f_n(x)$ (which happen at $x = 0$) grow unboundedly while the graphs become more and more narrow as $n \to \infty$. In terms of area under the graph, which is the essential contributor to the L^s norms, this can be seen as a competition between growing height and shrinking width. Let's nail things down precisely.

As suggested by the figure, it is easily verified that, for each n, f_n is continuous on the interval I, i.e., $\{f_n\} \subseteq C(I)$, and therefore $\{f_n\}$ is a sequence of Riemann integrable functions. Let's compute the $L^s(I)$ norms of this sequence: For $s = \infty$, we have

$$\|f_n\|_s = \|f_n\|_\infty = \sup_{x\in I} |f_n(x)| = n^{1/a}.$$

for each $n \in \mathbb{N}$. For $1 \leq s < \infty$, we have

$$\begin{aligned}
\|f_n\|_s &= \left(\int_I |f_n(x)|^s\,dx\right)^{1/s}\\
&= \left(\int_{-1}^{1} n^{s/a} e^{-sn|x|}\,dx\right)^{1/s}\\
&= n^{1/a}\left(2\int_0^1 e^{-snx}\,dx\right)^{1/s}\\
&= n^{1/a}2^{1/s}\left(\frac{e^{-snx}}{-sn}\Big|_{x=0}^{x=1}\right)^{1/s}\\
&= n^{1/a}\left(\frac{2}{sn}\right)^{1/s}\left(1-e^{-sn}\right)^{1/s}\\
&= n^{(1/a-1/s)}\left(\frac{2}{s}\right)^{1/s}\left(1-\frac{1}{e^{sn}}\right)^{1/s}
\end{aligned}$$

for each $n \in \mathbb{N}$. We therefore have the following behavior: if $s < a$, then $1/a - 1/s < 0$ (where we can't have $s = \infty$) and so

$$\lim_{n\to\infty} \|f_n - 0\|_s = \lim_{n\to\infty} \|f_n\|_s = \lim_{n\to\infty} n^{1/a-1/s}(2/s)^{1/s}(1-1/e^{sn})^{1/s} = 0 \cdot (2/s)^{1/s} \cdot 1 = 0.$$

Consequently, if $s < a$, $\{f_n\}$ converges to the zero function with respect to the $L^s(I)$ norm. If $s \geq a$, then, for $s = \infty$,

$$\lim_{n\to\infty} \|f_n - 0\|_s = \lim_{n\to\infty} n^{1/a} = \infty$$

and, for $s < \infty$ $1/a - 1/s \geq 0$,

$$\lim_{n\to\infty} \|f_n - 0\|_s = \lim_{n\to\infty} n^{(1/a-1/s)}(2/s)^{1/s}(1-1/e^{sn})^{1/s} = \begin{cases} \infty & a < s \\ (2/s)^{1/s} & a = s. \end{cases}.$$

In other words, the sequence $\{f_n\}$ converges to 0 for all $s < a$ (all small s) and does not converge to 0 for all $s \geq a$ (all large s). In particular, upon fixing $s < a$, if $r \leq s$, then $\{f_n\}$ converges to zero in both L^s and L^r norms. If $r > s$, then it is possible to $\{f_n\}$ to not converge to zero in the L^r norm (namely, when $r \geq a$) while still converging to zero in the L^s norm. As it must be, this is consistent with the preceding theorem.

Bibliography