Real Analysis

Supplementary course notes for MA338

Evan Randles

Draft date May 14, 2026

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 [5], *The Way of Analysis* by R. S. Strichartz [6], *Mathematical Analysis* by T. M. Apostol [1], and *An Introduction to Analysis* by W. R. Wade [7]. 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 Ciaran Groom, Grace Hillis, Cole Hoffman, Simge Ildes, Lora LaRochelle, Dante Lave, Lucas Lefebvre, Jasper Loverude, Alexander Lyon, Grace Moberg, Tava, Proctor, Ewan Ward, Abel Yan, and Zhazira Koldasbay. Of course, I am responsible for the errors that inevitably remain.

Chapter 1

The Real Numbers

1.1 Sups, Infs and all that

In this short section we discuss the essential properties of real numbers that will we need in our study of analysis. Though most of our studies will involve complex-valued things (numbers, functions, etc.), it is essential that we understand some basic objects of mathematical analysis on the real line – these objects provide the foundation upon which the notions of approximation (e.g., distance and convergence) is built.

Definition 1.1 (Bounded Sets)**.** *Let A be a subset of real numbers.*

1. *We say that the set A* ***is bounded above*** *if there exists a real number M for which*

$$x \leq M \qquad \text{for all} \qquad x \in A.$$

   In this case, M is said to be an ***upper bound*** *for the set A. We also say that textbfA is bounded above by M.*

2. *We say that the set A* ***is bounded below*** *if there exists a real number N for which*

$$N \leq x \qquad \text{for all} \qquad x \in A.$$

   In this case N is said to be a ***lower bound*** *for the set A. We also say that A is* ***bounded below*** *by N.*

3. *We say that the set A is* ***bounded*** *if it is bounded above and below.*

In light of the preceding definition, we observe that a bounded set A can be contained inside a "closed" interval, i.e., if A is bounded then there are constants N and M for which $A \subseteq [N, M]$. This is an important topological notion that extends well beyond the real line. The following exercise expands on this.

Exercise 1.1:

Given a real number a and a positive real number r, we define the (open) ball centered at a with radius r as the set

$$B_r(a) = \{x \in \mathbb{R} : |x - a| < r\} = (a - r, a + r).$$

Let A be a set of real numbers. Prove that the following are equivalent.

1. A is bounded.
2. There exists $r > 0$ for which $A \subseteq B_r(0)$.
3. There exists $a \in \mathbb{R}$ and $r > 0$ for which $A \subseteq B_r(a)$.

Hint: Prove the implication $(1) \rightarrow (2) \rightarrow (3) \rightarrow (1)$.

Perhaps the most important notion for (bounded) sets of real numbers is captured by the following definition.

Definition 1.2 (Supremum)**.** *Let A be a set of real numbers which is bounded above. Suppose that there exists a number S with the following two properties:*

i. S is an upper bound of A.

ii. For any upper bound M of A, $s \leq M$.

*Then S is called the **supremum** (or least upper bound) of the set A and we write $S = \sup A$.*

The careful reader should note that I've used the definite article "the" in the above definition. This language appears to imply (and, in fact, does) that, given a set A which is bounded above, that there can be at most one number called the supremum (of A). Let's justify this.

Lemma 1.3. *Let A be a set of real numbers which is bounded above and suppose that S_1 and S_1 are numbers which both satisfy Conditions (i) and (ii) of the preceding definition. Then $S_1 = S_2$.*

Proof. In view of our hypothesis, S_1 and S_2 are both upper bounds for A. Thus, given that condition (ii) is satisfied for S_1, we have $S_1 \leq S_2$ (because S_2 is an upper bound for A). Similarly, given that (ii) is satisfied for S_2, $S_2 \leq S_1$. Thus, $S_1 \leq S_2$ and $S_2 \leq S_1$ whence $S_1 = S_2$. □

In light of the above result, to each $A \subseteq \mathbb{R}$ which is bounded above, there is at most one supremum for A. The natural thing to ask is the following. Must such a set A have a supremum at all? Though we will not worry about it in this course, the answer to this question is deeply rooted in the construction of the real numbers. We will therefore take the following for granted.

Theorem 1.4 (The completeness axiom, Theorem 1.11 of [5])**.** *Suppose A is a non-empty subset of $\mathbb{R}$ which is bounded above. Then*

$$S = \sup A$$

exists.

In the case that a non-empty set A is not bounded above, we assume the common convention and write

$$\sup A = \infty.$$

Here (and being consistent with Theorem 1.4), $\sup A$ is only understood to exists in an extended sense (as an extended real number). Here we will always assume a clear distinction between real numbers and ∞. It is still however instructive to have $\sup A$ be able to take the "value" of ∞ when A is unbounded.

It will be useful for us to have a characterization of the supremum of a bounded set A. This condition, which we will make use of often, is the subject of the following proposition.

Proposition 1.5 (Characterization of Supremum)**.** *Let A be a non-empty subset of real numbers which is bounded above and let S be an upper bound of A. Then S is the supremum of A if and only if the following condition is satisfied.*

For each $\epsilon > 0$, there is an element $x \in A$ for which $x > S - \epsilon$.

This condition is illustrated in Figure 1.1.

Proof. Suppose that $S = \sup A$ and let $\epsilon > 0$. We observe that $S - \epsilon < S$. Since S is the supremum (the least upper bound) of A, any number strictly less than S cannot be an upper bound of A and so $S - \epsilon$ cannot be an upper bound of A. Since A is non-empty, we must therefore have some $x \in A$ for which $S - \epsilon < x$, just as we aimed to show.

Conversely, assume that S is an upper bound of A satisfying the property that, for all $\epsilon > 0$ there is an element $x \in A$ for which $x > S - \epsilon$. Let M be another upper bound of A. In the case that $M < S$, we set $\epsilon = S - M > 0$ and observe that, by the supposition, there is an $x \in A$ for which $S - \epsilon < x$. We note however that $S - \epsilon = S - (S - M) = M$ and thus there is an $x \in A$ for which $x > M$ showing that M cannot be an upper bound for A, a contradiction. Hence $S \leq M$ and so $S = \sup A$. □

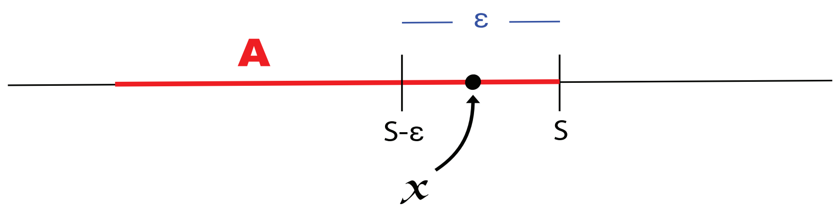

Figure 1.1: An illustration of the characterizing condition in Proposition 1.5

To see the utility of the proposition above, let us apply the proposition in the case where A is a subinterval of the real line. As you expect, the least upper bound of the interval should be its larger end point. Though these seems obvious, the claim requires some checking which we do in the following example.

Example 1.1:

Let $a < b$ be real numbers and let I be an interval of the form $[a, b]$, $(a, b]$, $[a, b)$ or (a, b). Then

$$\sup I = b.$$

To prove the statement above, let $\epsilon > 0$ and observe that $b - \epsilon < b$. Let

$$x = \max\left\{b - \frac{b-a}{2}, b - \frac{\epsilon}{2}\right\}$$

and observe that, by definition,

$$a = b - (b - a) \leq b - \frac{b-a}{2} \leq x < b$$

and

$$b - \epsilon \leq b - \frac{\epsilon}{2} \leq x < b.$$

Thus x is a member of the interval I and is such that $b - \epsilon < x$. Thus, to each $\epsilon > 0$ we have produced an $x \in I$ for which $b - \epsilon < x$. By virtue of Proposition 1.5, we conclude that $b = \sup I$.

Of course, we have notion parallel to the supremum for sets which are bounded below.

Definition 1.6. *Let A be a set of real numbers which is bounded below. Suppose that there exists a number I with the following two properties:*

i. I is a lower bound of A.

ii. For any lower bound N of A, $N \leq I$.

*Then I is called the **infimum** (or greatest lower bound) of the set A and we write $I = \inf A$.*

As with the definition of supremum, a set which is bounded below has at most one infimum and the justification of this parallels (almost exactly) the proof of Lemma 1.3. The question of existence is dealt with by the following corollary to Theorem 1.4 and whose proof I'll leave as an exercise.

Corollary 1.7. *Suppose A is a non-empty subset of $\mathbb{R}$ which is bounded below. Then*

$$I = \inf A$$

exists.

Exercise 1.2:

Prove Corollary 1.7 by completing the following steps.

1. Given a non-empty subset A which is bounded below, define the set

$$(-A) = \{x \in \mathbb{R} : x = -a \;\text{ for some }\; a \in A\}.$$

   Prove that $-A$ is bounded above.

2. By making an appeal to Theorem 1.4, let $S = \sup(-A)$. Prove that $-S = \inf A$ and so the infimum of A exists as the corollary asserts.

As we did for sets which were unbounded from above, if a non-empty set A is not bounded below, we will assume the standard convention and write

$$\inf A = -\infty.$$

Analogous to Proposition 1.5, we have the following proposition for infima.

Proposition 1.8. *Let A be a non-empty set of real numbers which is bounded below and let I be a lower bound for A. Then $I = \inf A$ if and only if the following condition is satisfied. For each $\epsilon > 0$, there exists $x \in A$ for which $x < I + \epsilon$.*

Exercise 1.3:

Prove the proposition above. You may prove it directly (as I did for Proposition 1.5) or you may prove it using the idea used the prove Corollary 1.7, i.e., by considering the set $-A$.

Chapter 2

Point-set Topology

Chapter 3

Numerical Sequence and Series

Chapter 4

Continuity

Chapter 5

Differentiation

In this chapter, we study the derivative. As you recall from calculus, the derivative (of a real-valued function of a single real variable) is a number associated to the "instantaneous" slope of the graph of the function, one which defines the "tangent line". In our study here, we begin by taking this perspective. Though a careful study, we will see that the notion of "tangent", at least as we saw in in our first calculus course, is better replaced by the concept of "best linear/affine approximation". This new concept, as we will see, is key to understanding differentiability of multivariate functions.

5.1 Real-valued functions of a real variable

Throughout this section, we will study real-valued functions whose domain is a subset of real numbers containing (at least) an interval I. Unless further specified, $I = [a, b], (a, b], [a, b),$ or (a, b) where $-\infty \leq b < a \leq \infty$.

Definition 5.1. *Let f be a real-valued function defined on an interval I. For $x_0 \in I$, we define the* ***difference quotient*** *of f at x_0 by*

$$\Delta_f(x; x_0) = \frac{f(x) - f(x_0)}{x - x_0}$$

which is a real-valued function having $I \setminus x_0$ as its domain. We say that f is ***differentiable*** *at x_0 provided that the limit*

$$\lim_{x \to x_0} \Delta_f(x; x_0)$$

exists. If so, the value of this limit is called the ***derivative*** *of f at x_0 and denoted by*

$$f'(x_0) = \lim_{x \to x_0} \Delta_f(x; x_0).$$

Remark 5.2. Let's make a couple of remarks about the above definition.

a. The derivative is also denoted by $\frac{df}{dx}(x_0)$.

b. Regardless of the choice of $x_0 \in I$, the fact that I is an interval guarantees that x_0 is an accumulation point of the domain of $x \mapsto \Delta_f(x : x_0)$ and hence it makes sense to talk about the limit. Note

c. In the case that x_0 is the endpoint of I, e.g., $x_0 = a$ or $x_0 = b$ for $I = [a, b]$, the concept of the derivative as defined above makes perfect sense but the limit is necessarily one sided. Some of the literature separates these endpoint cases, interpreting them as one-sided derivatives. The above definition works without a need to make such a distinction. On the other hand, if x_0 is interior to I, the above definition coincides with what you (most likely) saw in calculus.

Example 5.1: Monomials

For every natural number n and real constant C, the monomial $m_n(x) = Cx^n$ is differentiable at each point $x_0 \in \mathbb{R}$ and $m_n'(x_0) = Cnx_0^{n-1}$. In particular, the derivative of the constant function C is 0.

Proof. Fix $x_0 \in \mathbb{R}$. Let's first work the cases for $n = 0$ and $n = 1$. When $n = 0$, we have (by convention),

$$m_n(x) = Cx^0 = C$$

for all $x \in \mathbb{R}$. Thus

$$\Delta_{m_0}(x; x_0) = \frac{C - C}{x - x_0} = \frac{0}{x - x_0} = 0.$$

In particular, $m_n = m_0$ is differentiable at x_0 and

$$m_0'(x_0) = \lim_{x \to x_0} \Delta_{m_0}(x; x_0) = \lim_{x \to x_0} 0 = 0 = C \cdot 0 \cdot x_0^{-1} = Cnx_0^{n-1}$$

where we are taking the convention that the product of 0 and x_0^{-1} is 0 even when $x_0 = 0$. Now, for $n = 1$,

$$\Delta_{m_1}(x; x_0) = \frac{Cx^1 - Cx_0^1}{x - x_0} = C\frac{x - x_0}{x - x_0} = C$$

so that m_1 is differentiable at x_0 and

$$m_1'(x_0) = \lim_{x \to x_0} \Delta_{m_1}(x; x_0) = \lim_{x \to x_0} C = C = C \cdot 1 \cdot x_0^{1-1} = Cnx_0^{n-1}.$$

Finally, for $n > 1$, observe that

$$x^n - x_0^n = (x - x_0)(x^{n-1} + x^{n-2}x_0^1 + \cdots + x^1 x_0^{n-2} + x_0^{n-1})$$

for all x, a formula that is readily verified (just check it!). We remark that there are (necessarily) n terms in the second multiplicand above. From this, we see that

$$\Delta_{m_n}(x; x_0) = \frac{Cx^n - Cx_0^n}{x - x_0} = C(x^{n-1} + x^{n-2}x_0^1 + \cdots + x^1 x_0^{n-2} + x_0^n)$$

for $x \neq x_0$. Using the fact that polynomials are continuous, we immediately find that m_n is differentiable at x_0 and

$$\begin{aligned}
m_n'(x_0) &= \lim_{x \to x_0} \Delta_{m_n}(x; x_0) \\
&= \lim_{x \to x_0} C(x^{n-1} + x^{n-2}x_0^1 + \cdots + x^1 x_0^{n-2} + x_0^n) \\
&= C(x_0^{n-1} + x_0^{n-2}x_0^1 + \cdots + x_0^1 x_0^{n-2} + x_0^{n-1}) \\
&= C(x_0^{n-1} + x_0^{n-1} + \cdots + x_0^{n-1} + x_0^{n-1}) \\
&= Cnx_0^{n-1}.
\end{aligned}$$

□

Exercise 5.1:

Below is a list of functions mapping real numbers to real numbers (though sometimes the domain will be a proper subset of real numbers). For each function:

a. Determine the function's natural domain.

b. Identify the difference quotient and its domain.

c. Using the definition in terms of the difference quotient, determine where (within its domain) the function is differentiable, in this case, determine the value of the derivative.

d. Using the definition in terms of the difference quotient, determine where the function is not differentiable within its domain.

Of course, your arguments should be rigorous.

1. Given $m, b \in \mathbb{R}$, consider the "affine" function $T(x) = mx + b$ for $x \in \mathbb{R}$.
2. For $C \neq 0$ and $\alpha \in \mathbb{Z}$, $p(x) = Cx^{\alpha}$.
3. $x \mapsto \sqrt{x}$.
4. $x \mapsto |x|$.

Let's now start to establish some basic results about derivatives. Our first asserts that, if function is differentiable at a point, it must also be continuous there.

Proposition 5.3. *Let f be a real-valued function defined on an interval I and let $x_0 \in I$. If x_0 is differentiable at x_0, then f is continuous at x_0.*

Proof. For $x \neq x_0$, we see that
$$f(x) - f(x_0) = \Delta_f(x; x_0)(x - x_0).$$
When f is differentiable at x_0, $\lim_{x \to x_0} \Delta_f(x; x_0) = f'(x_0)$ and so using the algebra of limits, we have
$$\lim_{x \to x_0} (f(x) - f(x_0)) = \lim_{x \to x_0} (\Delta_f(x; x_0)(x - x_0)) = f'(x_0) \lim_{x \to x_0} (x - x_0) = f'(x_0) \cdot 0 = 0.$$
Equivalently,
$$f(x_0) = \lim_{x \to x_0} f(x)$$
and so f is continuous at x_0. □

Our next result is akin to the algebra of limits and it gives us some basic rules for computing derivatives of functions formed by sums, products and quotients or differentiable functions. In particular, given that we know a bunch functions (affine functions, monomials, etc) that are differentiable, we can use the proposition to build more differentiable functions.

Proposition 5.4. *Let f and g be real-valued functions defined (at least) on a common interval I. If, for $x_0 \in I$, f and g are both differentiable at x_0, then so are $f + g$, fg, and f/g where the last requires that $g(x_0) \neq 0$. Further, we have*

1.
$$(f + g)'(x_0) = f'(x_0) + g'(x_0)$$

2.
$$(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$$

and

3.
$$(f/g)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{(g(x_0))^2}.$$

These are commonly called the sum, product, and quotient rules.

Proof. We shall prove the product rule and leave the sum and quotient rules as an exercise. Note

For $x \in I$, observe that

$$(fg)(x) - (fg)(x_0) = f(x)g(x) - f(x_0)g(x_0) = (f(x) - f(x_0))g(x) + f(x_0)(g(x) - g(x_0))$$

where we have simply added 0 by adding and subtracting $f(x_0)g(x)$. Thus, for $x \neq x_0$,

$$\Delta_{fg}(x;x_0) = \left(\frac{f(x) - f(x_0)}{x - x_0}\right) g(x) + f(x_0)\left(\frac{g(x) - g(x_0)}{x - x_0}\right) = \Delta_f(x;x_0)g(x) + f(x_0)\Delta_g(x;x_0).$$

Because g is differentiable at x_0, Proposition 5.3 guarantees that $\lim_{x\to x_0} g(x) = g(x_0)$. With this, the algebra of limits and our assumption that f and g are both differentiable at x_0 gives us that

$$(fg)'(x_0) = \lim_{x\to x_0} \Delta_{fg}(x;x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0),$$

as was asserted. □

Corollary 5.5. *Let $a_0, a_2, \ldots, a_n$ be real numbers and define the real-valued polynomial function $p : \mathbb{R} \to \mathbb{R}$ by*

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

for $x \in \mathbb{R}$. Then p is differentiable at each $x \in \mathbb{R}$ and

$$p'(x) = a_1 + 2a_2x + 3a_3x^2 + \cdots a_nnx^{n-1}.$$

The above corollary is proved by using our knowledge of monomials (and their derivatives) and repeatedly applying the sum rule from Proposition 5.4. We now turn to the "chain rule" which gives us sufficient conditions under which compositions of functions can be differentiated. If we think about the derivative of a function measuring instantaneous change, the chain rule can be interpreted as saying that this change is propagated (multiplicatively) under composition note here

Theorem 5.6. *Let f and g be real valued functions with f defined on an interval I, g defined on an interval J and $f(I) \subseteq J$ and define $g \circ f : I \to \mathbb{R}$ defined by*

$$(g \circ f)(x) = g(f(x))$$

for $x \in I$. If f is differentiable at $x_0 \in I$ and g is differentiable at $y_0 = f(x_0) \in J$, then $g \circ f$ is differentiable at x_0 and

$$(g \circ f)'(x_0) = g'(y_0)f'(x_0) = g'(f(x_0))f'(x_0).$$

A straightforward proof using difference quotients can be found in [5]; see Theorem 5.5. We will give a distinct proof of this theorem which makes use of the interpretation of the derivative developed in the next subsection. I, personally, feel that the tack taken here is more illustrative.

Example 5.2:

Needed.

Exercise 5.2:

Use the chain rule (and the definition of the derivative) to differentiate the following functions (including saying where they are differentiable and finding their derivative everywhere it is possible).

1.
$$h(x) = \begin{cases} x^2 \cos(1/x) & x \neq 0 \\ 0 & x = 0. \end{cases}$$

2.
$$j(x) = e^{e^{e^{e^{e^x}}}}$$

Hint: Though we have not proved it, you may assume that the exponential and cosine functions have the derivative you know them to have.

5.2 Really, what is the derivative?

Theorem 5.7. *Let I be an interval in $\mathbb{R}$, $f : I \to \mathbb{R}$, and $x \in I$. Then f is differentiable at x if and only if there is a linear function $L : \mathbb{R} \to \mathbb{R}$ (which is necessarily of the form $L(h) = Dh$ where $D = (m)$ is a 1×1 matrix with its single entry $m \in \mathbb{R}$) for which*

$$f(x+h) - f(x) - L(h) = \mathcal{E}(h)\,|h| \tag{5.1}$$

where $\mathcal{E}$ is a real-valued function having the property that

$$\lim_{h \to 0} \mathcal{E}(h) = 0. \tag{5.2}$$

In this case, $L(h) = Dh$ where $D = (f'(x))$.

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

In what follows, I will discuss the continuity and differentiability of functions mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. To make things as clear as possible, points/vectors in $\mathbb{R}^n$ (or $\mathbb{R}^m$) will be expressed henceforth as column vectors and, when not written out in components/coordinates, I will write them in boldface. For instance,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

and, further, the (Euclidean) norm of this vector is

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

We should remark that, in $\mathbb{R}^n$, the Euclidean metric $d = d_2 = d_{\mathbb{R}^n}$ is given by $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. On Friday, we investigated differentiability for functions mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. Recall that, every linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ can be written (uniquely) in the form

$$L(\mathbf{x}) = D\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \text{for} \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$$

where D is the $m \times n$ matrix having entries a_{ij} as indicated above. Here is the definition of differentiability:

Definition 5.8. *Given an open set $\mathcal{O} \subseteq \mathbb{R}^n$, let $f : \mathcal{O} \to \mathbb{R}^m$; here n and m are positive integers. Given a point $\mathbf{x}_0 \in \mathcal{O}$, we say that f is differentiable at $\mathbf{x}_0$ if there is a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ for which*

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - L(\mathbf{h}) = \mathcal{E}(\mathbf{h})\|\mathbf{h}\| \tag{5.3}$$

where $\mathcal{E}$ (which depends on $\mathbf{x_0}$ and f) is an $\mathbb{R}^m$-valued function having the property that

$$\lim_{\mathbf{h}\to\mathbf{0}} \mathcal{E}(\mathbf{h}) = \mathbf{0}. \tag{5.4}$$

*In this case, the **derivative** of f at $\mathbf{x}_0$ (or **Jacobian derivative** or **Total Derivative**) is the $m \times n$ matrix $D = Df(\mathbf{x}_0)$ for which $L(\mathbf{h}) := D\mathbf{h}$ for $\mathbf{h} \in \mathbb{R}^n$.*

Before moving on to more exercises, let's make two remarks.

Remark 5.9. If we translate what the limit means in terms of Euclidean metrics on $\mathbb{R}^n$ and $\mathbb{R}^m$, the limit (5.4) is equivalently the statement: For every $\epsilon > 0$, there is a $\delta > 0$ for which

$$\|\mathcal{E}(\mathbf{h})\| < \epsilon \qquad \text{whenever} \qquad \|\mathbf{h}\| < \delta;$$

here, you have to a little careful because the first appearance of $\|\cdot\|$ means the norm on $\mathbb{R}^m$ and the second is the norm on $\mathbb{R}^n$. If it's helpful to you, we established an equivalence between the 1, 2, and ∞ metrics/norms on Homework 2 that you can freely use.

Remark 5.10. In the definition above, I use the definite article in the phrase ".. .the $m \times n$ matrix..." As we discussed in Homework 1, when we do this, we should make sure that such a matrix is indeed unique. This is the subject of the second exercise below. Outside of that exercise, you can take it for granted.

Exercise 5.3:

Let's verify differentiability for some functions using the definition above. In what follows, I give three functions. For each function, I will identify a point in the functions domain and a candidate for its derivative at that point. Please use the definition to show that each function is differentiable at the indicated point.

1. $f : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$f(\mathbf{x}) = \begin{pmatrix} xyz \\ x+z \end{pmatrix} \qquad \text{for} \qquad \mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3.$$

   Here

$$\mathbf{x}_0 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \qquad \text{and} \qquad D = \begin{pmatrix} 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

2. $g : \mathbb{R} \to \mathbb{R}^2$ defined by

$$g(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}$$

   for $t \in \mathbb{R}$. The point at which I'd like you to investigate is an arbitrary point $t_0 \in \mathbb{R}$ and, for this point, the candidate for the derivative is

$$D = \begin{pmatrix} -\sin(t_0) \\ \cos(t_0) \end{pmatrix}.$$

   Here, you may freely use any trigonometric identities and the following two facts: $|\cos(u) - 1| \le u^2$ and $|\sin(u) - u| \le u^3$ whenever $|u| \le 1$. We will prove these at some point.

3. Let $h : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$h\begin{pmatrix} x \\ y \end{pmatrix} = xy \qquad \text{for} \qquad \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

Here, for some $x_0, y_0 \in \mathbb{R}$,

$$\mathbf{x}_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \qquad \text{and} \qquad D = \begin{pmatrix} y_0 & x_0 \end{pmatrix}.$$

Exercise 5.4:

In this exercise, we shall prove that our derivatives (and affine approximations) to differentiable functions are unique. An affine function T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is, by definition, a function of the form

$$T(\mathbf{x}) = L(\mathbf{x}) + \mathbf{b}$$

where $L(\mathbf{x}) = D\mathbf{x}$ is a linear function (characterized by the $m\times n$ matrix D) and $\mathbf{b} \in \mathbb{R}^m$. Using this language of affine functions, we can state differentiability of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ as follows: f is differentiable at a point $\mathbf{x}_0$ if there is an affine function $T : \mathbb{R}^n \to \mathbb{R}^m$ for which

$$f(\mathbf{x}_0 + \mathbf{h}) = T(\mathbf{h}) + \mathcal{E}(\mathbf{h})\|\mathbf{h}\| \tag{5.5}$$

for $\mathbf{h} \in \mathbb{R}^n$ where $\lim_{h\to\mathbf{0}} \mathcal{E}(\mathbf{h}) = \mathbf{0}$. Prove that, if f is differentiable at $\mathbf{x}_0$, then there is one and only one affine function T for which (5.5) holds. *Hint: Differentiability should give you one affine function. After that, your job is to assume that, if you have any two such functions, they must be equal. The following lemma might be helpful. If you use it, give a proof.*

Lemma 5.11. *Let $M : \mathbb{R}^n \to \mathbb{R}^m$ be linear. Then M is not the zero transformation if and only if there is some vector $\mathbf{v} \in \mathbb{R}^n$ with unit length (i.e., $\|\mathbf{v}\| = 1$) for which $M(\mathbf{v}) \neq \mathbf{0}$.*

Exercise 5.5:

Let $\mathcal{O}$ be open in $\mathbb{R}^n$ and let $f : \mathcal{O} \to \mathbb{R}$. Given a point $\mathbf{x}_0 \in \mathcal{O}$ and an index $j \in \{1, 2, \dots, n\}$, the partial derivative of f with respect to the variable x_j at $\mathbf{x}_0$ is defined by

$$\partial_j f(\mathbf{x}_0) = \frac{\partial f}{\partial x_j}(\mathbf{x}_0) = \lim_{h\to 0} \frac{f(\mathbf{x}_0 + h\mathbf{e}_j) - f(\mathbf{x}_0)}{h}$$

provided this limit exists; here, $\mathbf{e}_j \in \mathbb{R}^n$ is the unit vector with 1 in the jth entry and zeros everywhere else, i.e.,

$$\mathbf{e_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{pmatrix}, \mathbf{e_2} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

1. Directly using the definition, compute all (first-order) partial derivatives of the functions (here, I'm using the convention that $x_1 = x$, $x_2 = y$ and $x_3 = z$):

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto xyz \qquad \text{and} \qquad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto x + z.$$

2. Let $\mathcal{O} \subseteq \mathbb{R}^n$ and $f : \mathcal{O} \to \mathbb{R}$. Given a point $\mathbf{x}_0 \in \mathcal{O}$ and an index $j \in \{1, 2, \dots, n\}$, define

$$g_j(t) = f(\mathbf{x}_0 + t\mathbf{e}_j).$$

Show that g_j is differentiable at 0 if and only if the partial derivative of f with respect to x_j at $\mathbf{x}_0$ exists and, in this case,

$$g_j'(0) = \frac{\partial f}{\partial x_j}(\mathbf{x}_0).$$

3. It you look closely you'll see that g_j above is the composition of functions, the inner such function being vector-valued and differentiable everywhere. Use the chain rule to prove the following theorem:

Theorem 5.12. *Let $f : \mathcal{O} \to \mathbb{R}$ be as above. If f is differentiable at $\mathbf{x}_0$, then all of its (first-order) partial derivatives $\partial f/\partial x_1$, $\partial f/\partial x_2$, ... $\partial f/\partial x_n$ exists at $\mathbf{x}_0$ and, moreover,*

$$Df(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \end{pmatrix}$$

We'll have a more general version of this in class on Monday.

4. Go back and confirm that the above formula (which determines the derivative matrix) is consistent with the proposed derivative matrices in Exercise 4.

5.4 Mean Value Theorems and L'Hôpital's rule.

All Needed

5.5 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 5.13. *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,*

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

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

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.*

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., $f = (\operatorname{Re}(f), \operatorname{Im}(f))^\top$. 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 5.14. *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 $u = \operatorname{Re}(f)$ and $v = \operatorname{Im}(f)$ 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 5.6:

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 5.7:

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 $\operatorname{Cis} : \mathbb{R} \to \mathbb{C}$ by

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

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

1. Show that $|\operatorname{Cis}(\theta)| = 1$ for all $\theta \in \mathbb{R}$.

2. Show that $\operatorname{Cis}(\theta_1 + \theta_2) = \operatorname{Cis}(\theta_1)\operatorname{Cis}(\theta_2)$.

3. Using the previous proposition, show that Cis is differentiable at $\theta_0 = 0$ and $\operatorname{Cis}'(0) = i$.

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.

5.5.1 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 5.15. *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 $P = \{x_0, x_1, x_2, \dots, x_K\}$ of I such that*

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

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

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

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

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

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 $k = 1, 2, \dots, K$.

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

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

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

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

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

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

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,*

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

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:*

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

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. $\square$

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,

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

whenever $k \in G$. Consequently,

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

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

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

so

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

and therefore

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

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 $P = \{x_0, x_1, x_2, \ldots, x_N\}$ be a partition of I. A set of points $\{x_1^*, x_2^*, \ldots, x_N^*\}$ is said to be admissible for P if $x_{k-1} \leq x_k^* \leq x_k$ for $k = 1, 2, \ldots, N$. 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 $\{x_1^*, x_2^*, \ldots, x_N^*\}$ 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 do.

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 $n = 1, 2, \dots, N-1$,

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

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
$$\left|\sum_{n=1}^{N} m_n\right| \leq N \max_{n=1,2,\dots,N}\{|m_n|\} \leq NM,$$
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*
$$P_N = \{x_k \in I \mid k = 0,\dots,N\} \qquad \textit{where} \qquad x_k = a + \frac{k(b-a)}{N}$$
for $k = 0,1,2,\dots N$. Also, suppose that, for each N, $\{x_1^*, x_2^*, \dots, x_N^*\} \subseteq I$ 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 Archimedian 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 $g \in B(I)$, 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, $fg \in R(I;\mathbb{R})$. $\square$

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 6.15.

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. $\square$

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.

For a complex-valued function f, applying Theorem 6.12 to f's real an imaginary parts quickly yields the following characterization of integrability.

Theorem 6.20 (Darboux's Characterization of Integrability for Complex-Valued Functions)**.** *Let $f \in B(I; \mathbb{C})$ where $I = [a, b]$. Then $f \in R(I; \mathbb{C})$ if and only if there is a complex number $\mathcal{I}$ such that, for every $\epsilon > 0$, there is $\delta > 0$ such that*

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

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,

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

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

Theorem 6.21. *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

$$U(cg,P) = cU(g,P) \qquad \text{and} \qquad L(cg,P) = cL(g,P).$$

   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 theorem.

Theorem 6.22. *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 theorem, we will first need a lemma.

Lemma 6.23. *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 5.7 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 6.21, this guarantees that

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

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

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

Now, for each $x \in I$,

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

where we have used the fact that $|zw| = |z||w|$ for complex numbers z, w. Thus, by Lemma 6.23, we have

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

as desired. $\square$

6.4 The Fundamental Theorem of Calculus

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

Theorem 6.24 (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 Theorem 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 $|x - x_0| < \delta$. □

Theorem 6.25 (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 6.20, select a partition $P = \{x_1, x_2, \dots, x_K\}$ for which

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

whenever $\{x_1^*, x_2^*, \dots, x_K^*\}$ 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 $k = 1, 2, \dots, K$, 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 $\{c_1, c_2, \dots, c_K\}$ 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

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

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 5.14 and the definition of the integral of complex-valued functions. □

Applying Theorem 6.25 to the product FG and using linearity of the integral immediately gives the following corollary.

Corollary 6.26 (Integration by parts)**.** *Let F and G be complex-valued differentiable functions on $I = [a,b]$. If $F' = f$ and $G' = g$ are both in $R(I:\mathbb{C})$, then*

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

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 6.27 (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 6.28. *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*

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

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,

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

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 6.29 (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. □

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.

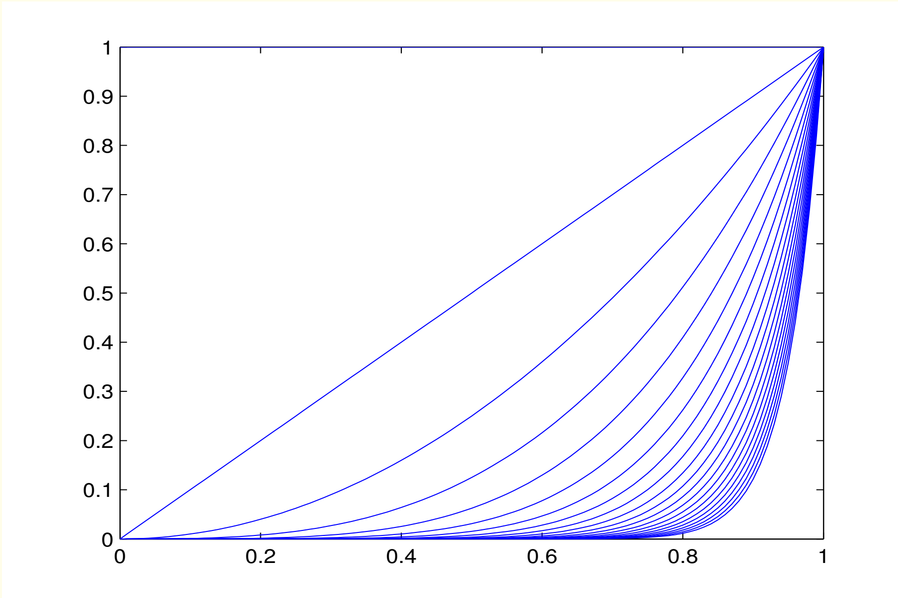

Figure 7.1: A famous picture: The graphs of $f_n(x) = x^n$ for $n = 1, 2, \ldots, 20$.

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*

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

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.

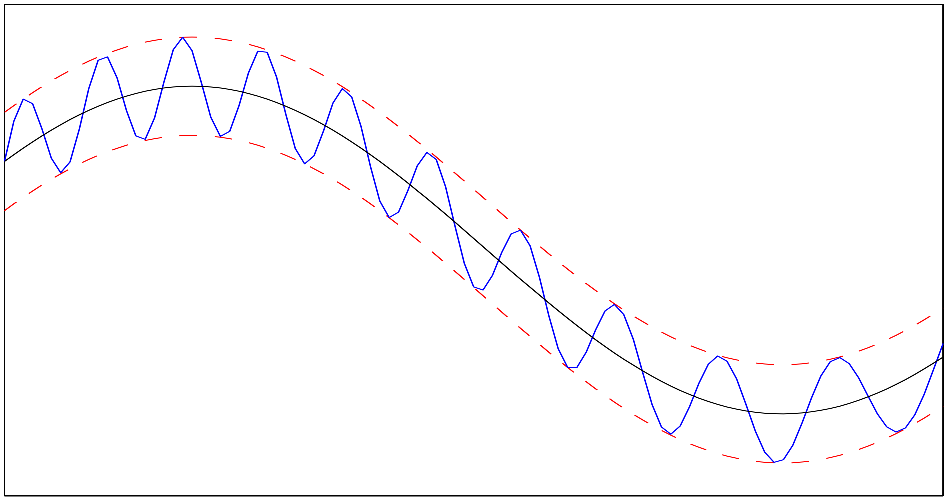

Figure 7.2: An illustration of uniform convergence

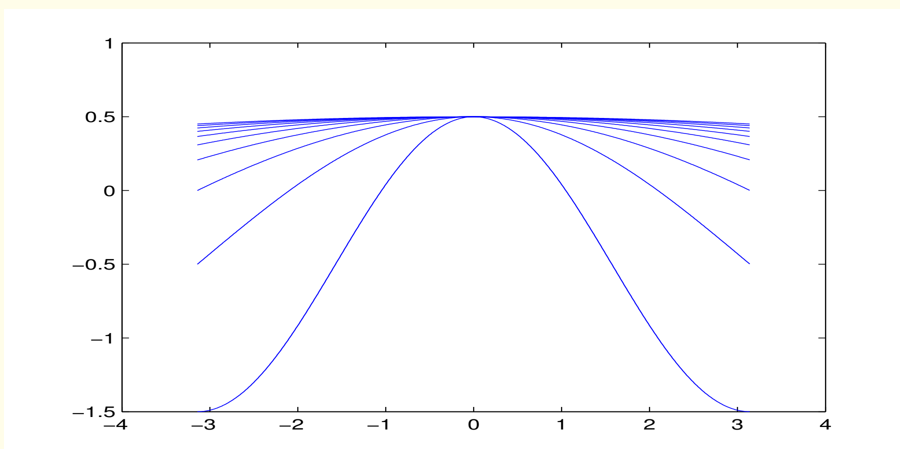

Figure 7.3: The graphs of $f_n(x) = \cos(x/n) - 1/2$ for $n = 1, 2, \ldots, 10$.

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 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 a set X. The sequence $\{f_n\}$ converges uniformly (to some function f) on X if and only if it satisfies the following property:*

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

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

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 X.

Proof. Let us first assume that $\{f_n\}$ converges uniformly to a function f on X. 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 $x \in X$. 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 $x \in X$. Thus the sequence $\{f_n\}$ is uniformly Cauchy on X.

Conversely, let's assume that the sequence $f_n(x)$ is uniformly Cauchy on X. This implies, in particular, that $\{f_n(x)\}$ is a Cauchy sequence of complex numbers for each $x \in X$. Because all Cauchy sequences of complex numbers converge, for each $x \in X$, 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 $f : X \to \mathbb{C}$ 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 $x \in X$. So now we have a candidate, namely 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 $x \in X$. Now, let $x \in X$ 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 $x \in X$ and $n \geq N$. Therefore, $\{f_n\}$ converges uniformly on X (to f). □

THIS SHOULD BE AN EXERCISE

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 8.9, $\{f_n\}$ 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 8.9 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 X be a set and $\{f_n\}$ be a sequence of complex-valued functions on X. The (formal) sum $\sum_n f_n$ is called a series of the functions $\{f_n\}$. 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 X, $\{S_N\}$, called the ***sequence of partial sums for the series*** *$\sum_n f_n$. If, for each $x \in X$, the limit*

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

exists, we say that ***the series $\sum_n f_n$ converges pointwise on X****. 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 X*** *if its sequence of partial sums $\{S_N\}$ converges uniformly on X 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 8.9 gives us exactly this.

Corollary 7.6 (Uniform Cauchy Criterion)**.** *Let $\{f_n\}$ be a sequence of complex-valued functions on a set X and consider the series $\sum_n f_n$. The series $\sum_n f_n$ converges uniformly on X 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 $x \in X$ 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 likely remember from calculus.

1. Using Theorem 8.9, prove Corollary 7.6.

2. If a series $\sum_n f_n$ of functions $\{f_m\}$ converges uniformly on an interval 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 $\{c_n\}$ 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 produced by working out the details of Exercise 20 in Chapter 8 of Rudin.

The following theorem gives an easy sufficient condition for a series to converge uniformly. NOTE here Evan

Theorem 7.7 (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
$$\sum_{k=n}^{m} M_k \le \sum_{k=n-1}^{m} M_k = \sum_{k=1}^{m} M_k - \sum_{k=1}^{n} M_k = \left|\sum_{k=1}^{m} M_k - \sum_{k=1}^{n} M_k\right| < \epsilon$$

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. □

Exercise 7.3:

The 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).

7.1 Properties of Uniform Convergence

In this section, we discuss some properties preserved under uniform convergence. Specifically, we focus on continuity and integration. Let's start with an example.

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 Theorem 7.7 (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 stated in the context of complex-valued functions on a metric space X.

Theorem 7.8. *Let $\{f_n\}$ be a sequence of complex-valued functions on a metrix space X and suppose that $\{f_n\}$ converges uniformly to a function $f : X \to \mathbb{C}$. If each function f_n is continuous, i.e., $\{f_n\} \subseteq C^0(X;\mathbb{C})$, then f is necessarily continuous.*

Proof. To show that f is continuous on X, we must show that, for each $x_0 \in X$ and $\epsilon > 0$, there is a $\delta > 0$ for which $|f(x_0) - f(x)| < \epsilon$ whenever $d(x_0, x) < \delta$. To this end, let $x_0 \in X$ and $\epsilon > 0$ be fixed. Since $\{f_n\}$ converges uniformly to f on X, let N be such that $|f_n(x) - f(x)| < \epsilon/3$ whenever $n \geq N$ and $x \in X$. In particular, $|f_N(x) - f(x)| < \epsilon/3$ for all $x \in X$. Now, because f_N is continuous on X, it is continuous at $x_0 \in I$ and so there is a $\delta > 0$ for which $|f_N(x_0) - f_N(x)| < \epsilon/3$ whenever $d(x_0, x) < \delta$. Thus, for $x \in X$ with $d(x_0, x) < \delta$, we have

$$|f(x_0) - f(x)| \leq |f(x_0) - f_N(x_0)| + |f_N(x_0) - f_N(x)| + |f_N(x) - f(x)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}.$$

□

It is instructive to translate the above statement about continuity and uniform convergence into the context of limits. It says that, if $\{f_n\}$ converges uniformly on X then, for each $x_0 \in X$,

$$\lim_{n\to\infty} \lim_{x\to x_0} f_n(x) = \lim_{x\to x_0} \lim_{n\to\infty} f_n(x);$$

in other words, we can exchange the order of limits. This is a rare property and important theme in analysis. We are very often concerned with understanding when such a property holds, for it usually makes life easy. For example, Fubini's theorem (from multivariable calculus) is an example of such a result. So is the following result concerning exchanging limits and integrals.

Note

Theorem 7.9. *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 6.23, we have

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

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 Theorems 6.21 and 6.22, 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 7.10. *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 6.21. 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

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

□

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,

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

We end this section with a result concerning uniform convergence and differentiation.

Theorem 7.12. *Let $\{f_n\}$ be a sequence of complex-valued differentiable function on $I = [a, b]$. If $\{f_n'\}$ converges uniformly on I and, for at least one $x_0 \in I$, $\{f_n(x_0)\}$ is a convergent sequence of complex numbers, then the sequence $\{f_n\}$ is itself converges uniformly on I to a differentiable function f and*

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

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

Proof. We first show that $\{f_n\}$ converges uniformly on I. To this end, let's fix $\epsilon > 0$. In view of the convergence of $\{f_n(x_0)\}$ and, by virtue of Theorem 8.9 and the assumed uniform convergence of $\{f_n'\}$, we may select N so that, whenever $m, n \geq N$, $|f_n(x_0) - f_m(x_0)| < \epsilon/2$ and

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

for all $x \in I$. Now, for any $n, m \geq N$ and $x \in I$, the mean value theorem applied to the difference $f_n - f_m$ guarantees c between x and x_0 so that

$$f_n(x) - f_m(x) = (f_n(x_0) - f_m(x_0)) + (f_n'(c) - f_m'(c))(x - x_0)$$

and therefore

$$\begin{aligned} |f_n(x) - f_m(x)| &\leq |f_n(x_0) - f_m(x_0)| + |f_n'(c) - f_m'(c))| \, |x - x_0| \\ &\leq \frac{\epsilon}{2} \, |f_n'(c) - f_m'(c)| \, (b - a) \\ &= < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Hence, the sequence $\{f_n\}$ is uniformly Cauchy and therefore uniformly convergent by Theorem 8.9; we shall denote its limit by f.

Our final tasks is to show that f is itself differentiable on I and $f'(x_0) = \lim_{n\to\infty} f'(x_0)$ for all $x_0 \in I$. To this end, the following construction is useful: Fix $x_0 \in I$ and, for each $n \in \mathbb{N}$, set

$$\delta_n(x; x_0) = \begin{cases} \frac{f_n(x) - f_n(x_0)}{x - x_0} & x \neq x_0 \\ f_n'(x_0) & x = x_0 \end{cases}.$$

By construction, it is clear that each δ_n is continuous on I. We claim that δ_n is uniformly convergent on I.

To see this, let's fix $\epsilon > 0$ and, using the uniform convergence of $\{f_n'\}$, let N be such that $|f_n'(x) - f_m'(x)| < \epsilon$ for every $x \in I$. By using the construction in the previous paragraph, we have, for any $x \in I$

$$|\delta_n(x; x_0) - \delta_m(x; x_0)| = |f_n'(c) - f_m'(c)| < \epsilon$$

whenever $n, m \geq N$; here, for each n, m and x, c exists (and lives between x and $x + x_0$) thanks to the mean value theorem. This proves our claim.

With the claim in hand, we see that δ_n is a uniformly convergent sequence of continuous functions which converges uniformly to to a continuous function by Theorem 7.8. Consequently,

$$\lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{n\to\infty} \delta_n(x_0; x_0) = \lim_{n\to\infty} f_n'(x_0);$$

In particular, f is differentiable at x_0 and $f'(x_0) = \lim_{n\to\infty} f_n'(x_0)$. □

Exercise 7.4:

Consider the so-called trigonometric series

$$f(x) = \sum_{n=1}^{\infty} \frac{\cos(nx)}{n^2}.$$

Prove that this series is uniformly convergent and f must be continuous. Further, for each $m \in \mathbb{Z}$, compute

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-imx} \, dx.$$

You should make sure you justify every step using results of the two preceding sections. Note: It might be useful to recall that $\cos(nx) = (e^{inx} + e^{-inx})/2$.

7.2 The space $\mathscr{C}$ and theorems of Weierstrass and Stone

In this section, we study a particular type of metric space whose elements are continuous functions. Though this will, at first, appear to us like any old (class of) example(s), it turns out that the space plays a central role in a large area of functional analysis called spectral theory (much of this is thanks to the pioneering work of J. von Neumann and I. N. Gelfand). The applications of this area of mathematics are legion and include, in particular, the mathematics of quantum mechanics. Without further ado, let X be a metric space and denote by $\mathscr{C}(X)$ the set of complex-valued, bounded, and continuous functions on X, i.e.,

$$\mathscr{C}(X) = C^0(X;\mathbb{C}) \cap B(X;\mathbb{C}).$$

We remark that $\mathscr{C}(X) = C^0(X;\mathbb{C})$ whenever X is compact because continuous functions on compacts sets are necessarily bounded. In any case, we equip $\mathscr{C}(X)$ with the supremum norm

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

and associated metric

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

for $f,g \in \mathscr{C}(X)$. As the following theorem, which follows immediately from Corollary 7.4 and Theorem 7.8, shows that this space is quite nice topologically speaking.

Theorem 7.13. *$(\mathscr{C}(X), d_\infty)$ is a complete metric space.*

Algebraically, $\mathscr{C}(X)$ is also quite interesting and the dedicated reader is invited to verify that $\mathscr{C}(X)$ is a vector space over $\mathbb{C}$ and, in fact, is a commutative unital algebra (over $\mathbb{C}$)! To get some understanding of this space and what's in it, let's focus our attention to the case in which $X = I = [a,b]$ and we'll discuss the functions (points!) living in $\mathscr{C}(I)$.

Example 7.4: Examples of continuous functions

Let's fix $I = [a,b]$ for finite a and b with $a < b$.

1. A polynomial on I is a function $p : I \to \mathbb{C}$ with the rule

$$p(x) = c_0 + c_1 x + \cdots + c_n x^n$$

where $c_0, c_1, \dots, c_n$ are complex numbers. We know that each such polynomial is a continuous function and so

$$\mathcal{P}(I) = \{p : I \to \mathbb{C} : p(x) = c_0 + c_1 x + \cdots + c_n x^n \text{ for } c_0, c_1, \dots, c_n\} \subseteq \mathscr{C}(I).$$

In fact, the reader should check that $\mathcal{P}(I)$ is itself a vector space and is, in fact, a subspace of $\mathscr{C}(I)$.

2. We say that a function $f : I \to \mathbb{C}$ is analytic on I if, for each $x_0 \in I$, f has a power series expansion centered at x_0 with positive radius of convergence,

$$f(x) = \sum_{n=0}^\infty c_n (x - x_0)^n.$$

In this case, it is easy to prove that f is infinitely differentiable on I and, in fact, for each x_0, $c_n = f^{(n)}(x_0)/n!$ so f's power series expansion is simply its Taylor series expansion

$$f(x) = \sum_{n=0}^\infty \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n.$$

In particular, each analytic function f is continuous on I. Denoting $\mathcal{A}(I)$ by the collection of analytic functions on I, we have $\mathcal{A}(I) \subseteq \mathscr{C}(I)$.

3. For any $n \geq 1$, the set of n-times continuously differentiable functions, $C^n(I;\mathbb{C})$ on I is a subset (vector subspace) of $\mathscr{C}(I)$. As the intersection of vector subspaces is also a subspace, we have

$$C^\infty(I;\mathbb{C}) = \bigcup_{n=1}^{\infty} C(I;\mathbb{C}) \subseteq \mathscr{C}(I);$$

these are the so-called smooth functions on I.

4. All together, we have

$$\mathcal{P}(I) \subseteq \mathcal{A}(I) \subseteq C^\infty(I:\mathbb{C}) \subseteq C^{n+1}(I;\mathbb{C}) \subseteq C^n(I;\mathbb{C}) \subseteq \cdots \subseteq C^1(I;\mathbb{C}) \subseteq \mathscr{C}(I).$$

Each set inclusion above is, in fact, proper (i.e., none are equalities of sets). For instance, you will have likely seen in a previous calculus class that

$$f(x) = \begin{cases} e^{1/x^2} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

is infinitely differentiable on any interval and so, in particular, $f \in C^\infty([-1,1])$. It is fairly easy to see, however, that f has no convergent power series representation at $x_0 = 0$ and hence $f \notin \mathcal{A}([-1,1])$. Below, we focus on the proper inclusion $C^1(I;\mathbb{C}) \subseteq \mathscr{C}(I)$ and discuss, in particular, how "big" of a subset C^1 is relative to $\mathscr{C}$.

As mentioned above, the inclusion $C^1(I;\mathbb{C}) \subseteq \mathscr{C}(I)$ is proper. For instance, the absolute value function fails to be differentiable at zero. More generally, on any interval, one can tie together linear functions in a (piecewise) way that makes an everywhere continuous function which fails to be differentiable at a finite number of (break) points; this is the idea of splines. Functions defined this way are incredibly useful in practice and, while suffering the small annoyance of failing to be differentiable at some places, they are still differentiable almost everywhere[1] and therefore easy to analyze. This was the state of mathematics in the nearly nineteenth century: Continuous functions were simply functions with unbroken graphs but with, possibly a small (say finite or countable) number of sharp cusps at which the function failed to be differentiable. Things were more or less copacetic until a monster reared its head.

Example 7.5: Weierstrass's Function

For $0 < a < 1$ and $b \in \mathbb{R}$ for which $ab \geq 1$, consider

$$W_{a,b}(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x)$$

defined (at least formally) for $x \in \mathbb{R}$. For the summands $w_n(x) = a^n \cos(b^n \pi x)$, it is easy to see that

$$M_n = \sup_x |w_n(x)| = a^n$$

and certainly $\sum_n a^n < \infty$. Thus, by the Weierstrass M-test and Theorem 7.8, the above series converges absolutely and uniformly to an everywhere continuous function $W_{a,b}(x)$.

Theorem 7.14 (Hardy, 1916)**.** *$W_{a,b}(x)$ is nowhere differentiable.*

Weierstrass first established this result in 1875 under the condition that b was a positive odd integer and $ab \geq 1 + 3\pi/2$; the general result (with $ab \geq 1$) was proved by G. H. Hardy in 1916 [3]. For a straightforward proof of the non-differentiability (where a^n and b^n are replaced by $1/n!$ and $(n!)^2$, respectively), see Chapter 11 of [4]. A plot of this function for $b = 3$ and $a = 1/2$ is illustrated in Figure 7.4

[1] This is a technical term that you will lean in MA439

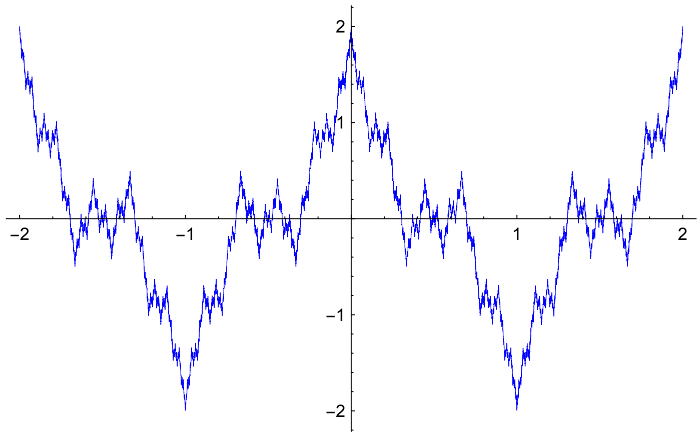

Figure 7.4: Blah

As you might have felt reading the above example, just the very existence of the Weierstrass function made folks (mathematicians) very uncomfortable. The situation actually is much "worse" as can be seen by a result of N. Wiener:

Theorem 7.15 (N. Wiener, 1923)**.** *There exists a sample space/set Ω, a probability (measure) $\mathbb{P}$ on Ω, and a function $X : [0,\infty) \times \Omega \to \mathbb{R}$, called a Brownian motion, for which*

$$\mathbb{P}(\{\omega \in \Omega : t \mapsto X_t(\omega) \text{ is continuous but nowhere differentiable}\}) = 1.$$

What is stated above is actually a corollary of a more important theorem of Wiener that established the existence of Brownian motion, a stochastic process hypothesized in 1905 by A. Einstein. A short history of Brownian motion is given in [**?**], along with the history of contributions of T. N. Thiele and L. Bachelier whose works predated Einstein but were not widely recognized in their times.

Qualitatively, what we can take away from Wiener's theorem is this: Most continuous functions are nowhere differentiable. In other words, Weierstrass' monster is a normal thing and not a pathology. This uncomfortable state of affairs is actually not new to us as students of real analysis. In our discussion of real numbers, we learned that real numbers are very hard to understand. In fact, if you tasked every human who has ever lived with the job of creating/labeling/describing one new real number in each minute (or second) of their lives, the collection of all real numbers produced this way (and all the knowledge created) would represented a null fraction of all real numbers, even if human history were to go on forever. We did find solace in the fact that we can approximate every real number by a rational one and rational numbers are object that we do (at least in some sense) understand. In short, we proved that $\mathbb{Q}$ is dense in $\mathbb{R}$. Taking by analogy the fact that we understand polynomials, we might ask: Is the set of polynomials dense in $\mathscr{C}(I)$ in the metric d_∞? Here, Weierstrass is now a hero in our story.

Theorem 7.16 (Weierstrass, 1875)**.** *Let $I = [a,b]$ with a and b (finite) real numbers with $a < b$. Given any $f \in \mathscr{C}(I)$ and $\epsilon > 0$, there exists $p(x) \in \mathcal{P}(I)$ for which*

$$d_\infty(f,p) = \sup_{x\in I} |f(x) - p(x)| < \epsilon.$$

In other words, $\mathcal{P}(I)$ is dense in $\mathscr{C}(I)$ with respect to the d_∞ metric.

Proof. Let's first prove a special case: Take $f \in \mathscr{C}([0,1])$ with $f(0) = f(1) = 0$. We extend f to all of $\mathbb{R}$ by specifying that $f(x) = 0$ whenever $x \le 0$ and $1 \le x$. Denoting this extension by f, by an abuse of notation, it is clear that this function is uniformly continuous on all of $\mathbb{R}$. For each $n \in \mathbb{N}_+$, let

$$Q_n(t) = c_n(1-t^2)^n$$

where c_n is chosen so that

$$\int_{-1}^{1} Q_n(t)\,dt = c_n \int_{-1}^{1} (1-t^2)^n\,dx = 1.$$

By a simple calculus argument, it is easy to see that $(1-t^2)^n \geq 1 - nt^2$ whenever $0 \leq t \leq 1$. Consequently,

$$1 = c_n \int_{-1}^{1} (1-t^2)^n\,dt = 2c_n \int_{0}^{1} (1-t^2)^n \geq 2c_n \int_{0}^{1/\sqrt{n}} (1-t^2)^n\,dt \geq 2c_n \int_{0}^{1/\sqrt{n}} 1 - nx^2\,dx = \frac{4c_n}{3\sqrt{n}}$$

and so it follows that

$$0 \leq c_n \leq \sqrt{n} \tag{7.3}$$

for all $n \in \mathbb{N}_+$. With this in hand, let's take $\epsilon > 0$ and, given the uniform continuity of f, let $\delta > 0$ for which $|f(x) - f(y)| < \epsilon/2$ whenever $|x - y| < \delta$. For $n \in \mathbb{N}_+$ and $0 \leq x \leq 1$, put

$$P_n(x) = \int_{-1}^{1} f(x+t)Q_n(t)\,dt = \int_{-x}^{1-x} f(x+t)Q_n(t)\,dt$$

where we have noted that $f(x+t) = 0$ whenever $x + t \leq 0$ or $x + t \geq 1$. By making a change of variables, we see that

$$P_n(x) = \int_{0}^{1} f(t)Q_n(x-t)\,dt$$

and since $Q_n(x-t)$ is a polynomial, the integration must produce a polynomial in x (with coefficients that depend on f and n). In other words, each $P_n(x)$ is a polynomial in x. Observe that

$$\begin{aligned}
P_n(x) - f(x) &= \int_{-1}^{1} f(x+t)Q_n(t)\,dt - f(x)\int_{-1}^{1} Q_n(t)\,dt \\
&= \int_{-1}^{1} \left(f(x+t) - f(x)\right) Q_n(t)\,dt \\
&= \int_{-1}^{-\delta} \left(f(x+t) - f(x)\right) Q_n(t)\,dt + \int_{-\delta}^{\delta} \left(f(x+t) - f(x)\right) Q_n(t)\,dt \\
&\qquad + \int_{\delta}^{1} \left(f(x+t) - f(x)\right) Q_n(t)\,dt
\end{aligned}$$

for $n \in \mathbb{N}_+$ and $0 \leq x \leq 1$. Combining the first and third integrals above, we will simplify this as

$$P_n(x) - f(x) = \int_{-\delta}^{\delta} \left(f(x+t) - f(x)\right) Q_n(t)\,dt + \int_{\delta \leq |t| \leq 1} \left(f(x+t) - f(x)\right) Q_n(t)\,dt$$

for $n \in \mathbb{N}_+$ and $0 \leq x \leq 1$. Using our choice of δ and the triangle inequality, we have

$$\begin{aligned}
|P_n(x) - f(x)| &\leq \left|\int_{-\delta}^{\delta} \left(f(x+t) - f(x)\right) Q_n(t)\,dt\right| + \left|\int_{\delta \leq |t| \leq 1} (f(x+t) - f(x))Q_n(t)\,dt\right| \\
&= \int_{-\delta}^{\delta} |f(x+t) - f(x)|\,|Q_n(t)|\,dt + \int_{\delta \leq |t| \leq 1} |f(x+t) - f(x)|\,|Q_n(t)|\,dt \\
&= \int_{-\delta}^{\delta} \frac{\epsilon}{2} Q_n(t)\,dt + \int_{\delta \leq |t| \leq 1} 2M Q_n(t)\,dt \\
&\leq \frac{\epsilon}{2}\int_{-1}^{1} Q_n(t)\,dt + 4M \int_{\delta}^{1} Q_n(t)\,dt \\
&= \frac{\epsilon}{2} + 4M \int_{\delta}^{1} Q_n(t)\,dt
\end{aligned}$$

for $n \in \mathbb{N}_+$ and $0 \leq x \leq 1$ where we have used the fact that the polynomials Q_n are non-negative, even/symmetric, and have total integral 1. Upon noting that

$$Q_n(t) = c_n(1-t^2)^n \leq c_n(1-\delta^2)^n$$

for $\delta \leq t \leq 1$, we have

$$\int_\delta^1 Q_n(t)\,dt \leq \int_\delta^1 c_n(1-\delta^2)^n\,dt = c_n(1-\delta^2)^n(1-\delta) = \frac{c_n}{1+\delta}(1-\delta^2)^{n+1} \leq \frac{\sqrt{n}}{1+\delta}(1-\delta^2)^{n+1}$$

for $n \in \mathbb{N}_+$ where we have made use of (7.3). Putting our estimates together, we find that

$$|P_n(x) - f(x)| \leq \frac{\epsilon}{2} + \frac{4M\sqrt{n}}{1+\delta}(1-\delta^2)^{n+1} = \frac{\epsilon}{2} + C\sqrt{n}r^n$$

for all $n \in \mathbb{N}_+$ and $0 \leq x \leq 1$ where $C = 4M/(1+\delta) > 0$ and $r = 1-\delta^2 \in (0,1)$. Since, $\sqrt{n}r^n \to 0$ as $n \to \infty$ (Geometric sequences dominate polynomials/roots), we can find $N \in \mathbb{N}_+$ for which the latter term is less than $\epsilon/2$ for $n = N$. Thus, for $p = P_N$ (or $p = P_n$ for $n \geq N$), we have

$$\sup_{0\leq x\leq 1} |P_N(x) - f(x)| \leq \frac{\epsilon}{2} + C\sqrt{N}r^{N+1} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

which completes the proof of the special case.

For the general case, we take $f \in \mathscr{C}([a,b])$ and define

$$g(t) = f(bt + a(1-t)) - f(a) - (f(b) - (f(a))t$$

for $0 \leq t \leq 1$. By construction, g is continuous, satisfies $g(0) = g(1) = 0$ and

$$f(x) = g\left(\frac{x-a}{b-a}\right) + f(a) + \frac{f(b)-f(a)}{b-a}(x-a)$$

for $a \leq x \leq b$ (here, we have made the substitution $t = (x-a)/(b-a)$. Now, for $\epsilon > 0$, our argument for the special case gives us a polynomial $q(t)$ for which

$$\sup_{0\leq t\leq 1} |g(t) - q(t)| < \epsilon.$$

This, in turn, gives us the polynomial

$$p(x) = q\left(\frac{x-a}{b-a}\right) + f(a) + \frac{f(b)-f(a)}{b-a}(x-a)$$

and, with this, we observe

$$d_\infty(f,p) = \sup_{a\leq x\leq b} |f(x) - p(x)| = \sup_{a\leq x\leq b}\left|g\left(\frac{x-a}{b-a}\right) - q\left(\frac{x-a}{b-a}\right)\right| = \sup_{0\leq t\leq 1}|g(t) - q(t)| < \epsilon.$$

$\square$

Exercise 7.5: Bernstein's Approximation

Another approach to proving the Weierstrass approximation theorem is to replace the the integral in the above proof by a sum and the polynomials Q_n by the so-called Bernstein polynomials. For $n \in \mathbb{N}_+$ and $k = 0, 1, 2, \ldots, n$, define

$$B_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}$$

for $0 \leq x \leq 1$.

1. Prove that, for each n and $0 \leq x \leq 1$,
$$\sum_{k=0}^{n} B_{n,k}(x) = 1$$
and
$$\sum_{k=0}^{n} \left(\frac{k}{n} - x\right)^2 B_{n,k}(x) = x\frac{1-x}{n}.$$

2. Now let $f : [0,1] \to \mathbb{C}$ be uniformly continuous and let $\epsilon > 0$. Let $M = \sup_x |f(x)|$ and take $\delta > 0$ for which $|f(x) - f(y)| < \epsilon/2$ whenever $|x - y| < \delta$. Prove the following: For a fixed $n \in \mathbb{N}_+$,
$$|f(k/n) - f(x)| \leq \max\left\{\frac{\epsilon}{2}, \frac{2M}{\delta^2}\left(\frac{k}{n} - x\right)^2\right\} \leq \frac{\epsilon}{2} + \frac{2M}{\delta^2}\left(\frac{k}{n} - x\right)^2$$
whenever $0 \leq x \leq 1$ and $k = 0, 1, \dots, n$. Hint: If k/n and x are close, so are $f(x)$ and $f(k/n)$. When k/n and x are not close, e.g., $|k/n - x| \geq \delta$, $|f(x) - f(k/n)| \leq 2M$.

3. For our f as above and $n \in \mathbb{N}_+$, consider the polynomial
$$P_n(x) = \sum_{k=0}^{n} f(k/n) B_{n,k}(x)$$
defined for $0 \leq x \leq 1$. Use the two previous items to prove that
$$|P_n(x) - f(x)| \leq \frac{\epsilon}{2} + \frac{M}{2\delta^2 n}$$
for $n \in \mathbb{N}_+$ and $0 \leq x \leq 1$.

4. By completing the argument above, argue why this yields another proof of the Weierstrass approximation theorem (for $f \in \mathscr{C}([0,1])$).

5. Implement this approximation algorithm for the following functions (and produce computer illustrations/graphs) for $n = 2, 4, 8$ and 16, at least:
   - $f(x) = \cos(\pi x)$
   - $f(x) = |x - 1/2|$
   - $$f(x) = \sum_{k=0}^{10} \frac{1}{2^k} \cos(3^k \pi x).$$

Definition 7.17. *A family $\mathcal{A}$ of complex functions on a set X is said to be an algebra if*

1. *$f + g \in \mathcal{A}$ whenever $f, g \in \mathcal{A}$.*
2. *$fg \in \mathcal{A}$ whenever $f, g \in \mathcal{A}$.*
3. *$\alpha f \in \mathcal{A}$ whenever $f \in \mathcal{A}$ and $\alpha \in \mathcal{C}$.*

Additionally, we say that $\mathcal{A}$ is self-adjoint if $f^*(x) = \overline{f(x)} \in \mathcal{A}$ whenever $f \in \mathcal{A}$. We say that $\mathcal{A}$ separates points if, to each distinct points $x_1, x_2 \in X$, there is $f \in \mathcal{A}$ for which $f(x_1) \neq f(x_2)$. Finally, we say that $\mathcal{A}$ vanishes at no point of X if, for each $x \in X$ there is some $f \in \mathcal{A}$ for which $f(x) \neq 0$.

Theorem 7.18 (Stone-Weiesrstrass, Stone 1937). *If X is compact and $\mathcal{A}$ is an algebra of functions on a metric space X which is self-adjoint, vanishes at no point of X and separates points, then $\mathcal{A}$ is dense in $\mathscr{C}(X)$.*

7.3 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 7.19. *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 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

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

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

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

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

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 7.20 (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 Theorem 6.22, 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 7.21 (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 7.6: Minkowski via Hölder

Though 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 7.22. *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 7.9 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

$$\|f_n - f\|_r^r = \int_I |f_n(x) - f(x)|^r\, dx = \int_I |f_n(x) - f(x)|^r \cdot 1\, dx \leq \|(f_n - f)^r\|_p \|1\|_q \tag{7.4}$$

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

$$\|(f_n - f)^r\|_p = \sup_{x\in I} |f_n(x) - f(x)|^r = \|f_n - f\|_\infty^r. \tag{7.5}$$

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 (7.4) 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

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

we have

$$\|f_n - f\|_r \leq \|f_n - f\|_s (b-a)^{1/rq} = (b-a)^{\left(\frac{1}{r} - \frac{1}{s}\right)}\|f_n - f\|_s \tag{7.6}$$

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 (7.4) and (7.6) (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 7.6:

To 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 7.5 illustrates f_2 and f_{10} in the case that $a = 1$.

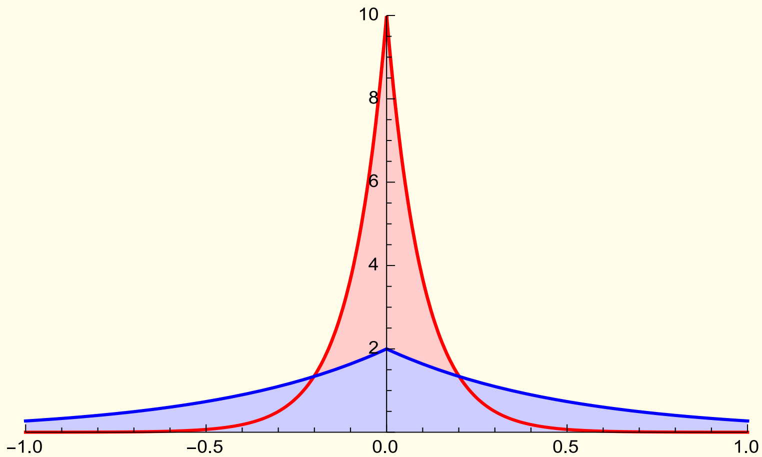

Figure 7.5: The graphs of f_2 and f_{10} when $a = 1$.

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.

Chapter 8

Fourier Series

8.1 Life on the circle

Our analysis thus far has focused on complex-valued functions defined on intervals of the form $I = [a, b]$. In our study of Fourier series, we will study complex-valued functions on the "circle" or on the torus. As we see below, this is just a fancy way of talking about complex-valued and 2π-periodic functions – such objects have several equivalent descriptions. As a first introduction, consider the unit circle

$$S^1 = \left\{ z = a + ib \in \mathbb{C} : |z| = \sqrt{a^2 + b^2} = 1 \right\}$$

in the complex plane $\mathbb{C}$. As a direct consequence of Exercise 5.7, S^1 is given by

$$S^1 = \{e^{i\theta} : \theta \in \mathbb{R}\}.$$

In fact, the exercise states that

$$S^1 = \{e^{i\theta} : \theta \in (-\pi, \pi]\}$$

and this description is one to one in the sense that the function $\theta \mapsto e^{i\theta}$ bijectively maps $(-\pi, \pi]$ onto the unit circle S^1. The following proposition gives an equivalence between 2π-periodic functions, functions on the interval $(-\pi, \pi]$ and functions on the circle.

Proposition 8.1. *There is a one-to-one-to-one correspondence between the complex-valued functions on S^1, the complex-valued functions on $(-\pi, \pi]$ and the complex-valued and 2π-periodic functions on $\mathbb{R}$. This correspondence is given by*

$$\tilde{f}(\theta) = F(e^{i\theta}) = F(z) = F(e^{ix}) = f(x)$$

where $z = e^{i\theta} = e^{ix}$ for function $F : S^1 \to \mathbb{C}$, $\tilde{f} : (-\pi, \pi] \to \mathbb{C}$ and $f : \mathbb{R} \to \mathbb{C}$, the latter of which is 2π-periodic.

As the proof of the proposition is not terribly illuminating, I'll omit it. Figure 8.1 illustrates this correspondence and the main idea of the proposition. In view of proposition, we will focus our attention on complex-valued and 2π-periodic functions which we will usually denote by f. These functions will often be said to be functions on the torus $\mathbb{T}$, however, this is a slight abuse of language and notation, both of which are justified by the proposition. Precisely, the torus $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$ which can be recognized as the quotient of the additive groups $\mathbb{R}$ and $2\pi\mathbb{Z}$. In fact, one can also recognize $\mathbb{T}$ by the interval $(-\pi, \pi]$. This association is given by the fact that there is a one-to-one correspondence between the interval $(-\pi, \pi]$ and $\mathbb{R}/(2\pi\mathbb{Z})$. You won't however need to worry about this construction nor what a quotient group is. You can think of this reference as merely cultural (where I mean the culture of harmonic analysis in the lens of topological group theory). The essential thing you'll need to understand, and to work with, is the notion of 2π-periodic functions. Just know that there is a lot of important group theory going on in the background.

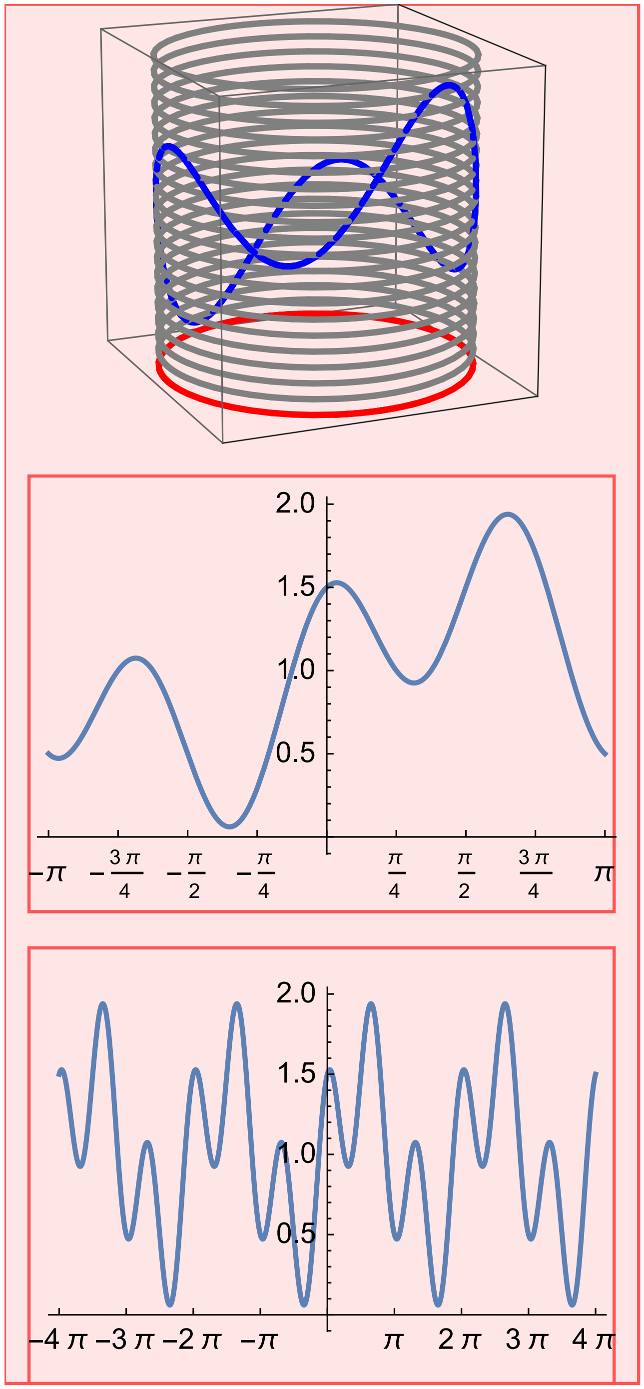

Figure 8.1: An illustration of the correspondence between functions on the circle, functions on $(-\pi, \pi]$ and 2π-periodic functions

8.2 Integration and Differentiation on $\mathbb{T}$

Okay, now let's talk about integration on $\mathbb{T}$, that is, the integration of 2π periodic functions.

Definition 8.2. *Given a function $f:\mathbb{R}\to\mathbb{C}$, we say that f is Riemann-integrable on the torus $\mathbb{T}$ (or integrable on the circle) if f is 2π periodic and Riemann-integrable on the interval $[-\pi,\pi]$. In this case we write $f\in R(\mathbb{T})$ and define*

$$\int_{\mathbb{T}} f=\int_{[-\pi,\pi]} f(x)\,dx;$$

this is called the integral of f over $\mathbb{T}$. We will also write

$$\int_{\mathbb{T}} f(x)\,dx=\int_{\mathbb{T}} f.$$

As the integral of f on $\mathbb{T}$ is defined in terms of the Riemann integral on the interval $[-\pi,\pi]$, it's easy to see that all of our preceding results on the Riemann-Darboux integral (and all related tests, theorems, etc) apply to this integral too, with only minor (and obvious) changes in notation. Instead of restating every result in this new notation, we will simply refer to the original results and ask the reader to make appropriate changes to notation and context.

Remark 8.3 (A word of caution). It is common for authors to define the integral of a function f over $\mathbb{T}$ instead by the number

$$\frac{1}{2\pi}\int_{[-\pi,\pi]} f(x)\,dx.$$

This convention (which we do not use) has some advantages in Fourier analysis as one doesn't have to carry around 2π everywhere and so formulas/statements become simpler. Though I do like this convention, it is slightly less transparent and so I've decided to avoid it. In any case, in your readings of other texts/notes, you should watch out as it is often not clear which convention is being used.

One result that is special for the integral of a function f over $\mathbb{T}$, is that it can be computed by integrating f over any interval of length 2π, not just the interval $[-\pi,\pi]$. This fact, captured by the following proposition, relies essentially on our requirement that f is 2π-periodic.

Proposition 8.4. *Let $f:\mathbb{R}\to\mathbb{C}$ be 2π-periodic. Then $f\in R(\mathbb{T})$ if and only if, $f\in R(I)$ for some interval I of length 2π, i.e., $I=[a,a+2\pi]$ for some $a\in\mathbb{R}$. For any such f, we have*

$$\int_{\mathbb{T}} f(x)\,dx=\int_I f(x)\,dx.$$

In other words, integrating f over any interval of length 2π gives the same number.

Exercise 8.1:

Suppose that $f:\mathbb{R}\to\mathbb{C}$ is 2π-periodic. With Proposition 6.27 in mind, do the following.

1. Prove the following statement. If $f\in R([a,b])$ then, for any $n\in\mathbb{Z}$, $f\in R([a+2\pi n,b+2\pi n])$ and

$$\int_a^b f=\int_{a+2\pi n}^{b+2\pi n} f.$$

2. Prove the following statement. If $f\in R([-\pi,\pi])$, then, for any $x_0\in\mathbb{R}$, the function $x\mapsto f(x+x_0)$ is Riemann integrable on $[-\pi,\pi]$ and

$$\int_{-\pi}^{\pi} f(x)\,dx=\int_{-\pi}^{\pi} f(x+x_0)\,dx.$$

3. Prove Proposition 8.4.

In view of Definition 8.2, $R(\mathbb{T})$ is the set of Riemann integrable functions on $\mathbb{T}$. Structurally, this set can be recognized a vector space (over $\mathbb{C}$) of complex-valued functions under the usual notion of function addition. Also, $f \mapsto \int_{\mathbb{T}} f$ is an important linear map from this vector space into the complex plane. Let's introduce some other important spaces of functions, all of which turn out to be subspaces of $R(\mathbb{T})$.

Definition 8.5. *Let $f : \mathbb{R} \to \mathbb{C}$ be 2π periodic.*

1. *We say that f is continuous on $\mathbb{T}$ if f is continuous on $\mathbb{R}$. Here we write that $f \in C(\mathbb{T})$ or $f \in C^0(\mathbb{T})$.*
2. *We say that f is once continuously differentiable on $\mathbb{T}$ if f is differentiable on $\mathbb{R}$ with continuous derivative $f' \in C^0(\mathbb{T})$. In this case we write $f \in C^1(\mathbb{T})$.*
3. *More generally, we say that f is n-times continuously differentiable on $\mathbb{T}$ if $f \in C^{n-1}(\mathbb{T})$ and the $(n-1)$th derivative of f, $f^{(n-1)}$, is differentiable and its derivative $f^{(n)} := (f^{(n-1)})' \in C(\mathbb{T})$. In this case, we write $f \in C^n(\mathbb{T})$.*
4. *We say that f is smooth on $\mathbb{T}$ if $f \in C^n(\mathbb{T})$ for all n. In this case we write $f \in C^\infty(\mathbb{T})$.*

Exercise 8.2:

To clarify the preceding definition, this exercise asks you to work out some details. First, let's fix something that the definition sweeps under the carpet, so to speak.

1. Given a 2π-periodic function f, it makes sense to ask its derivative f' exists and is continuous which is what we usually mean by continuously differentiable. The definition above requires, however, that $f' \in C^0(\mathbb{T})$ and so it requires the additional condition that f' is also a 2π periodic function. To clear this up, prove the following statement:

   If $f : \mathbb{R} \to \mathbb{C}$ is 2π-periodic and differentiable with derivative $f' : \mathbb{R} \to \mathbb{C}$. Then f' is also 2π-periodic.

Given a 2π-periodic function $f : \mathbb{R} \to \mathbb{C}$, for $f \in C^1(\mathbb{T})$ the definition above requires that f is differentiable on all of the real line $\mathbb{R}$ and, moreover, that f' is continuous. For the next item, you will prove that this requirement is unnecessarily strong.

2. Prove that $f \in C^1(\mathbb{T})$ if and only if f is continuously differentiable (meaning it is differentiable and has continuous derivative) on some interval $I = (a, b)$ where $b - a > 2\pi$.
3. Show that the equivalent condition of the above statement cannot be relaxed. That is, find a function $f : \mathbb{R} \to \mathbb{C}$ which is 2π-periodic and for which f is continuous differentiable on some interval $I = (a, b)$ for which $b - a = 2\pi$, yet $f \notin C^1(\mathbb{T})$.

As continuous functions are integrable and differentiable functions are continuous (and so on and so forth), we obtain

$$C^\infty(\mathbb{T}) \subseteq \cdots \subseteq C^n(\mathbb{T}) \subseteq \cdots \subseteq C^1(\mathbb{T}) \subseteq C^0(\mathbb{T}) \subseteq R(\mathbb{T})$$

where each of these containments is proper. All of the above sets are, in fact, vector spaces over $\mathbb{C}$ under the usual notion of function addition as we previously discussed and, further, all of above relations remain true when one replaces "$\subseteq$" with "$\leq$" where $V \leq W$ means V is a subspace of W.

8.3 Pointwise and Uniform convergence of functions on $\mathbb{T}$

Throughout the next several sections, we explore several notions of convergence of Fourier series for functions on $\mathbb{T}$. To this end, this subsection is dedicated to stating (and restating) several notions and results pertaining to the convergence of functions and series of functions on $\mathbb{T}$. The functions (and sequences of functions) we study will be taken to be, at worst, Riemann-integrable on $\mathbb{T}$.

Consider a sequence of functions $\{f_n\}_{n=1}^\infty \subseteq R(\mathbb{T})$ and another function $f \in R(\mathbb{T})$. Given a set $E \subseteq \mathbb{R}$, we say that $\{f_n\}$ *converges to* f *on* E if, for each $x \in E$, the sequence of complex numbers $\{f_n(x)\}$ converges to $f(x)$, i.e., $\lim_{n\to\infty} f_n(x) = f(x)$. As all functions in sight are 2π-periodic, $\lim_{n\to\infty} f_n(x) = f(x)$ if and only if $\lim_{n\to\infty} f_n(x + 2\pi k) = f(x + 2\pi k)$ for each $k \in \mathbb{Z}$. In other words, $\{f_n\}$ converges to f on E if and only if $\{f_n\}$ converges on the set

$$\bigcup_{k\in\mathbb{Z}} (E + 2\pi k) = \{x \in \mathbb{R} : x = y + 2\pi k \text{ for some } y \in E \text{ and } k \in \mathbb{Z}\}.$$

In the special case that $\{f_n\}$ converges to f on $\mathbb{R}$ or, equivalently, on any interval of length at least 2π, we say that $\{f_n\}$ converges pointwise to f on $\mathbb{T}$. In other words, $\{f_n\}$ converges pointwise to f on $\mathbb{T}$ if, for each $x \in \mathbb{R}$ and $\epsilon > 0$, there exists a natural number N for which

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

whenever $n \geq N$. Just as we did on an interval I,

Definition 8.6. *Let $\{f_n\}$ be a sequence of functions in $R(\mathbb{T})$ and let $f \in R(\mathbb{T})$. We say that $\{f_n\}$ converges uniformly to f on $\mathbb{T}$ if, for each $\epsilon > 0$, there exists a natural number N for which*

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

for all $n \geq N$ and $x \in \mathbb{R}$.

Of course, the above notion of uniform convergence was the same as it was for uniform convergence of functions on $\mathbb{R}$. The difference here is that the functions of interest are 2π periodic. In fact, it's easy to see that $\{f_n\}$ converges uniformly to f on $\mathbb{T}$ if and only if $\{f_n\}$ converges uniformly to f on any interval I of length at least 2π. As we saw before, uniform convergence on $\mathbb{T}$ is also captured by the sup norm. Let's define it in this context.

Definition 8.7. *Given any $f \in R(\mathbb{T})$ (which is necessarily bounded), define*

$$\|f\|_\infty = \sup_{\mathbb{T}} |f(x)|$$

where $\sup_{\mathbb{T}} |f(x)| := \sup\{|f(x)| : x \in I\}$ given any interval $I \subseteq \mathbb{R}$ of length at least 2π.

In this language and by virtue of our previous results, we have the following proposition.

Proposition 8.8. *Let $\{f_n\} \in R(\mathbb{T})$ and let $f \in R(\mathbb{T})$. Then $\{f_n\}$ converges uniformly to f on $\mathbb{T}$ if and only if*

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

If either of the equivalent conditions of the above proposition is met, $\{f_n\}$ is also said to converge to f with respect to the $L^\infty(\mathbb{T})$ norm – a notion we will discuss more fully in the next subsection. In the context of the torus $\mathbb{T}$, we restate Theorem 8.9:

Theorem 8.9. *Let $\{f_n\}$ be a sequence of functions in $R(\mathbb{T})$. Then $\{f_n\}$ converges uniformly on $\mathbb{T}$ if and only if the following condition is satisfied:*
(UC) For all $\epsilon > 0$, there exists a natural number N such that

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

Let's also recapture Theorems 7.8 and 7.9 in the context of the torus $\mathbb{T}$.

Theorem 8.10. *Let $\{f_n\}$ be a sequence of functions in $R(\mathbb{T})$ and suppose that $\{f_n\}$ converges uniformly on $\mathbb{R}$ to a function $f : \mathbb{R} \to \mathbb{C}$. Then the following properties hold:*

1. *The limit f is Riemann integrable on $\mathbb{T}$, i.e., $f \in R(\mathbb{T})$, and the sequence $\{f_n\}$ converges uniformly to f on $\mathbb{T}$.*

2. *If $\{f_n\} \subseteq C(\mathbb{T})$, then $f \in C(\mathbb{T})$.*

3. *We have*
$$\lim_{n\to\infty}\int_{\mathbb{T}} f_n = \int_{\mathbb{T}} \lim_{n\to\infty} f_n = \int_{\mathbb{T}} f.$$

Let's now talk about series on $\mathbb{T}$. Given a sequence $\{f_n\} \in R(\mathbb{T})$, we investigate the corresponding series $\sum_n f_n$. For each natural number N, we define the Nth partial sum S_N of the series $\sum_n f_n$ by

$$S_N(x) = \sum_{n=1}^{N} f_n(x)$$

for $x \in \mathbb{R}$. It is clear that $\{S_N\}$ is a sequence of Riemann-integrable functions on $\mathbb{T}$ and we can ask whether or not it has a limit. If, for $x \in \mathbb{R}$,

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

exists, we say that the series $\sum_n f_n$ converges at x and define the sum of the series to be the number

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

If the sequence of partial sums S_N converges at each $x \in \mathbb{R}$, we say that the series $\sum_{n=1}^{\infty} f_n$ converges pointwise on $\mathbb{T}$ and write
$$\sum_{n=1}^{\infty} f_n(x) = \lim_{N\to\infty} S_N(x)$$

for each $x \in \mathbb{R}$. Further, we say that the series $\sum_{n=1}^{\infty} f_n$ converges uniformly on $\mathbb{T}$ if the sequence $\{S_N\}$ of partial sums converges uniformly on $\mathbb{T}$. Rephrasing our work from the previous section, we can state the Weierstrass M-test in the context of $\mathbb{T}$.

Theorem 8.11. *Let $\{f_n\}$ be a sequence of functions in $R(\mathbb{T})$ and, for each $n \in \mathbb{N}$, let $M_n = \|f_n\|_\infty$. If the numerical series $\sum_{n=1}^{\infty} M_n$ converges, then the following statements hold:*

1. *The series $\sum_{n=1}^{\infty} f_n$ converges uniformly on $\mathbb{T}$.*

2. *We have*
$$\int_{\mathbb{T}} \sum_{n=1}^{\infty} f_n(x)\,dx = \sum_{n=1}^{\infty}\int_{\mathbb{T}} f_n(x)\,dx.$$

3. *If, additionally, $\{f_n\} \subseteq C(\mathbb{T})$, the sum of the series*
$$\sum_{n=1}^{\infty} f_n(x) = \lim_{N\to\infty}\sum_{n=1}^{N} f_n(x)$$
is also a member of $C(\mathbb{T})$.

8.3.1 The Lebesgue norms on $\mathbb{T}$ and the $L^2(\mathbb{T})$ inner product

We begin this section by introducing the Lebesgue norms on the torus $\mathbb{T}$. These norms, denoted by $\|\cdot\|_p$ for $1 \leq p \leq \infty$, give a notion of "size" to each function in $R(\mathbb{T})$ and, in fact, we've already seen some beautiful properties of the $\|\cdot\|_\infty$ norm studied in the preceding section. As we did on subintervals of the real line, we will pay special attention to the $p = 2$ case as it is captured by the rich structure of the so-called L^2 inner product.

Warning: The notation in this subsection will parallel that of Subsection 7.3; however, the definitions will all differ by a factor of $(2\pi)^{1/p}$. These changes are only to simplify notation and won't affect anything essentially. In any case, watch out for the 2π!

Definition 8.12. *1. For $1 \leq p < \infty$, we define the $L^p = L^p(\mathbb{T})$ norm of a function $f \in R(\mathbb{T})$ by*

$$\|f\|_p = \left(\frac{1}{2\pi} \int_{\mathbb{T}} |f(x)|^p \, dx \right)^{1/p}.$$

2. If $p = \infty$, we define the $L^p = L^\infty = L^\infty(\mathbb{T})$ of $f \in R(\mathbb{T})$ by

$$\|f\|_\infty = \sup_{\mathbb{T}} |f(x)|$$

where $\sup_{\mathbb{T}} |f(x)| := \sup\{|f(x)| : x \in I\}$ given any subinterval $I \subseteq \mathbb{R}$ of length at least 2π, a notion which makes sense precisely because f is 2π-periodic.

3. Let $1 \leq p \leq \infty$. Given a sequence of function $\{f_n\} \subseteq R(\mathbb{T})$ and another function $f \in R(\mathbb{T})$, we say that $\{f_n\}$ converges to f with respect to the $L^p = L^p(\mathbb{T})$ norm if

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

Tracking carefully the appearance of 2π in the above definitions, an appeal to the results of Subsection 7.3 yields the following results about the L^p norm on $\mathbb{T}$.

Proposition 8.13. *For any $f, g \in R(\mathbb{T})$ and $\alpha \in \mathbb{C}$, we have*

1.

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

2.

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

3.

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

Proposition 8.14 (Hölder's inequality)**.** *Let $1 \leq p \leq q \leq \infty$ be conjugate exponents, i.e.,*

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

For any $f, g \in R(\mathbb{T})$,

$$\left| \frac{1}{2\pi} \int_{\mathbb{T}} f(x) g(x) \, dx \right| \leq \|f\|_p \|g\|_q.$$

Proposition 8.15. *Given any $1 \leq r \leq s \leq \infty$, if a sequence $\{f_n\} \subseteq R(\mathbb{T})$ converges to $f \in R(\mathbb{T})$ with respect to the $L^s(\mathbb{T})$ norm, then $\{f_n\}$ converges to f with respect to the $L^r(\mathbb{T})$ norm. In particular, if $\{f_n\}$ converges to f uniformly, i.e.,*

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

then, for every $1 \leq p \leq \infty$, $\{f_n\}$ converges to f with respect to the $L^p(\mathbb{T})$ norm, i.e.,

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

We now turn our focus to the case in which $p = 2$. Here, the norm $\|\cdot\|_2$ is captured by the following inner product.

Definition 8.16. *Given $f, g \in R(\mathbb{T})$, we defined the $L^2 = L^2(\mathbb{T})$ inner product of f and g to be the complex number*

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{\mathbb{T}} f(x)\overline{g(x)}\, dx$$

Following directly from the defintion, we see that the $L^2(\mathbb{T})$ inner product characterizes the $L^2(\mathbb{T})$ norm. That is, for each $f \in R(\mathbb{T})$,

$$\|f\|_2 = \sqrt{\langle f, f, \rangle}.$$

Let's summarize some properties of the L^2 inner product.

Proposition 8.17. *Given any $f, g, h \in R(\mathbb{T})$ and $\alpha \in \mathbb{C}$, we have:*

1. $$\langle f, g\rangle = \overline{\langle g, f\rangle}$$
2. $$\langle \alpha f, g\rangle = \alpha\langle f, g\rangle$$
3. $$\langle f, \alpha g\rangle = \overline{\alpha}\langle f, g\rangle$$
4. $$\langle f + g, h\rangle = \langle f, h\rangle + \langle g, h\rangle$$
5. $$\langle f, g + h\rangle = \langle f, g\rangle + \langle f, h\rangle$$

Two functions $f, g \in \mathbb{R}(\mathbb{T})$ are said to be *orthogonal* (with respect to the L^2 inner product) if $\langle f, g\rangle = 0$. In this case we write $f \perp g$. Using this language, we state three more crucial properties of the $L^2(\mathbb{T})$ inner product and norm.

Proposition 8.18. *Given $f, g \in R(\mathbb{T})$, we have the following statements:*

1. *If $f \perp g$, then*
$$\|f + g\|_2^2 = \|f\|_2^2 + \|g\|_2^2.$$
2. $$\left|\frac{1}{2\pi}\int_{\mathbb{T}} f(x)g(x)\, dx\right| \leq \|f\|_2\|g\|_2$$
3. $$\|f + g\|_2 \leq \|f\|_2 + \|g\|_2$$

Proof. We prove only the first property as the others follow from our results of Subsection 7.3. Fix $f, g \in R(\mathbb{T})$ and suppose that $f \perp g$. Then, given the properties of the inner product summarized in the preceding proposition,

$$\begin{aligned} \|f+g\|_2^2 &= \langle f+g, f+g\rangle \\ &= \langle f, f+g\rangle + \langle g, f+g\rangle \\ &= \langle f,f\rangle + \langle f,g\rangle + \langle g,f\rangle + \langle g,g\rangle \\ &= \|f\|_2^2 + \langle f,g\rangle + \overline{\langle f,g\rangle} + \|g\|_2^2 \\ &= \|f\|_2^2 + 0 + 0 + \|g\|_2^2 \\ &= \|f\|_2^2 + \|g\|_2^2. \end{aligned}$$

□

Example 8.1:

For any $n \in \mathbb{Z}$, the function $x \to e^{inx}$ is $C^\infty(\mathbb{T})$ and therefore Riemann integrable on $\mathbb{T}$. We have, for each $n, m \in \mathbb{Z}$,

$$\langle e^{inx}, e^{imx}\rangle = \frac{1}{2\pi}\int_{\mathbb{T}} e^{inx}e^{-imx}\,dx = \begin{cases} 1 & n = m \\ 0 & m \neq n \end{cases}$$

thanks to the fundamental theorem of calculus. For this reason, $e^{inx} \perp e^{imx}$ whenever $n \neq m$. Such a collection $\{e^{inx}\}_{n\in\mathbb{Z}}$ is said to be an *orthonormal system*. The word normal comes from the fact that that $L^2(\mathbb{T})$ norm of each function e^{inx} is one.

8.4 Fourier coefficients and the uniform theory

We now (and finally) begin our study of Fourier series. Our main goal is to approximate a given function $f \in R(\mathbb{T})$ by a sequence of trigonometric polynomials, called Fourier polynomials. A trigonometric polynomial of order N is, by definition, a function of the form

$$P_N(x) = \sum_{n=-N}^{N} c_n e^{inx}$$

for $x \in \mathbb{R}$ where $\{c_n\}$ are complex numbers. It is clear that each such polynomial P_N is a member of $C^\infty(\mathbb{T})$ and hence a member of $R(\mathbb{T})$. Our goal is then to find a sequence of polynomials P_N which approximate f in some sense (pointwise, uniform, with respect to L^2, etc.). As we will see shortly, the following definition provides a good start.

Definition 8.19. *Given $f \in R(\mathbb{T})$, define*

$$\hat{f}(n) = \langle f, e^{inx}\rangle = \frac{1}{2\pi}\int_{\mathbb{T}} f(x)e^{-inx}\,dx$$

for each $n \in \mathbb{Z}$. For each $N \in \mathbb{N}$, we define the Fourier polynomial of order N by

$$S_N(x) = S_N(f)(x) = \sum_{n=-N}^{N} \hat{f}(n)e^{inx}$$

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

The following theorem states that the Fourier coefficients determine a function at each point of continuity. This was proved earlier in the course; for a proof see Theorem 2.1 of the course textbook. We will recapture the result shortly. I don't like that this is here. Perhaps instead we should do some examples here to show that the Fourier polynomials tell us something about f?

Proposition 8.20. *Suppose that, for $f, g \in R(\mathbb{T})$, $\hat{f}(n) = \hat{g}(n)$ for all $n \in \mathbb{Z}$. If f and g are continuous at $x \in \mathbb{R}$, then $f(x) = g(x)$. If, additionally, $f, g \in C(\mathbb{T})$, then $f = g$.*

With the above definition in mind, we are interested in whether or not the Fourier polynomials of f approximate f in any sense. The natural thing to do is then to take $N \to \infty$ and this leads us to the notion of Fourier series.

Definition 8.21. *given a function $f \in R(\mathbb{T})$, its Fourier series is the formal expression*

$$\sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx} = \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{inx}$$

is called the Fourier series of f. We write

$$f \sim \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}$$

To investigate the convergence of this series, we must investigate the convergence of the Fourier polynomials S_N which are, by definition, partial sums of the Fourier series for f. That is, we say that the Fourier series for f converges at $x \in \mathbb{R}$ if $\lim_{N\to\infty} S_N(x)$ exists. In this case,

$$\sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx} = \lim_{N\to\infty} \sum_{n=-N}^{N} \hat{f}(n)e^{inx}$$

is the sum of the series. To refine our discussion above, our main goal is to ask: When and in what sense is

$$f(x) = \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}?$$

In view of the Weierstrass M-test, we can give a sufficient condition for the uniform convergence of Fourier series in terms of the summability of the numerical series $\{\hat{f}(n)\}$.

Theorem 8.22. *Let $f \in C(\mathbb{T})$ and let $\{\hat{f}(n)\}_{n\in\mathbb{Z}}$ be the Fourier coefficients of f. If the series $\sum_{n\in\mathbb{Z}} |\hat{f}(n)|$ converges, then the Fourier series of f, $\sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}$, converges uniformly to f.*

Proof. We observe that

$$|\hat{f}(n)e^{inx}| = |\hat{f}(n)||e^{inx}| = |\hat{f}(n)|$$

for all $n \in \mathbb{Z}$ and $x \in \mathbb{R}$. In other words, the summands $g_n(x) = \hat{f}(n)e^{inx}$ satisfy $|g_n(x)| \le M_n = |\hat{f}(n)|$ for all $n \in \mathbb{Z}$ and $x \in \mathbb{R}$. In view of our supposition that the series $\sum |\hat{f}(n)| = \sum M_n$ converges, the Weierstrass M-test guarantees that the Fourier series converges uniformly to some complex-valued function g on $\mathbb{R}$. Further, because each summand $\hat{f}(n)e^{inx}$ is continuous, g is continuous. Also, observe that

$$g(x + 2\pi) = \lim_{N\to\infty} \sum_{n=-N}^{N} \hat{f}(n)e^{in(x+2\pi)} = \lim_{N\to\infty} \sum_{n=-N}^{N} \hat{f}(n)e^{inx} = g(x)$$

for all $x \in \mathbb{R}$ and so g is 2π-periodic. We therefore conclude that $g \in C(\mathbb{T})$. It remains to show that $g = f$. To this end, we compute the Fourier coefficients of g. Observe that, for any $m \in \mathbb{Z}$,

$$g(x)e^{-imx} = e^{-imx} \lim_{N\to\infty} \sum_{n=-N}^{N} \hat{f}(n)e^{inx} = \lim_{N\to\infty} \sum_{n=-N}^{N} \hat{f}(n)e^{inx}e^{-imx}$$

for all $x \in \mathbb{R}$. In fact, the Weierstrass M-text applied to the summands $n \mapsto \hat{f}(n)e^{inx}e^{-imx}$ shows that the partial sums on the right hand side converge uniformly to $g(x)e^{-imx}$. Since each summand is Riemann integrable (it is continuous) and the series converges uniformly, we may integrate term-by-term. For each $m \in \mathbb{Z}$, we have

$$\hat{g}(m) = \langle g, e^{imx} \rangle = \frac{1}{2\pi} \int_{\mathbb{T}} g(x)e^{-imx}\,dx = \sum_{n\in\mathbb{Z}} \frac{1}{2\pi} \int_{\mathbb{T}} (\hat{f}(n)e^{inx}e^{-imx})\,dx = \sum_{n\in\mathbb{Z}} \hat{f}(n)\langle e^{inx}, e^{imx} \rangle = \hat{f}(m)$$

where we have used the fact that $\langle e^{inx}, e^{imx}\rangle = 1$ when $n = m$ and 0 otherwise. Therefore $\hat{f}(n) = \hat{g}(n)$ for all $n \in \mathbb{Z}$ and, in view of the preceding proposition $f = g$. In other words,

$$f(x) = \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}$$

where the convergence is uniform for $x \in \mathbb{R}$. □

I've always found the result above somewhat unsatisfying, though it is powerful as we will shortly see. The reason I find it unsatisfying is because its hypotheses are stated in terms of the Fourier coefficients of f and so, to apply the theorem, one has to compute the Fourier coefficients of f and then ask if they are absolutely summable. One would like to instead have hypotheses stated in terms of f itself. In any case, the theorem allows us to prove the following result.

Corollary 8.23. *If $f \in C^2(\mathbb{T})$, then the Fourier series for f, $\sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}$, converges uniformly to f on $\mathbb{R}$.*

Proof. Let $f(x) = u(x) + iv(x)$ where u and v are real-valued 2π periodic functions which are, by hypothesis, twice continuously differentiable on $\mathbb{R}$. We have $f'(x) = u'(x) + iv'(x)$ and $f''(x) = u''(x) + iv''(x)$ for $x \in \mathbb{R}$. I claim that, for all non-zero $n \in \mathbb{Z}$,

$$\hat{f}(n) = \frac{-1}{2\pi n^2}\int_{\mathbb{T}} f''(x)e^{-inx}\,dx.$$

To see this, we apply the complex-version of integration by parts (which works in view of the FTC you proved in Homework 1). This is the statement: If u and v are complex-valued functions differentiable on the interval $[a, b]$ with derivatives u' and v',

$$\int_{[a,b]} u(x)v'(x)\,dx = u(x)v(x)\Big|_a^b - \int_{[a,b]} u'(x)v(x)\,dx$$

where $u(x)v(x)\big|_a^b := u(b)v(b) - u(a)v(a)$. So, upon taking $u = f$ and $v'(x) = e^{-inx}$, we have

$$\begin{aligned}\hat{f}(n) &= \frac{1}{2\pi}\int_{\mathbb{T}} f(x)e^{-inx}\,dx = \frac{1}{2\pi}\int_{[-\pi,\pi]} f(x)e^{-inx}\,dx\\ &= \frac{1}{2\pi}\left(f(x)\frac{e^{-inx}}{-in}\Big|_{-\pi}^{\pi} - \int_{[-\pi,\pi]} f'(x)\frac{e^{-inx}}{-in}\,dx\right)\end{aligned}$$

We note, however, that because f and e^{-inx} are 2π periodic, □

Exercise 8.3: The impact of regularity on Fourier coefficients

This exercise concerns the absolute decay of Fourier coefficients. It allows you to investigate conditions under which the Fourier series for a function f converges uniformly to f by Theorem 8.22. Do the following:

1. Use induction (on m) and integration by parts to prove the statement: If $f \in C^m(\mathbb{T})$ for $m \in \mathbb{N}$, then, for any non-zero $n \in \mathbb{Z}$,

$$\hat{f}(n) = (in)^{-m}\widehat{f^{(m)}}(n) = \frac{(in)^{-m}}{2\pi}\int_{\mathbb{T}} f^{(m)}(x)e^{-inx}\,dx$$

where $f^{(m)}$ means the mth-derivative of f. Conclude directly that the Fourier series for f converges uniformly to f whenever $f \in C^m(\mathbb{T})$ for $m \geq 2$.

Given a function $f \in R(\mathbb{T})$, we say that f is Hölder continuous of order $\alpha > 0$, for the largest integer $m \in \mathbb{N}$ such that $m < \alpha$, $f \in C^m(\mathbb{T})$ and there exists $C_\alpha > 0$ for which

$$|f^{(m)}(x) - f^{(m)}(y)| \leq C|x - y|^{\alpha-m}$$

for all $x, y \in \mathbb{R}$.

2. Find a function $f \in R(\mathbb{T})$ which is Hölder continuous of order $\alpha = 2$.
3. Use the mean value theorem to prove that, for $f \in C^1(\mathbb{T})$, f is Hölder continuous of order $\alpha = 1$.
4. Using your reasoning from the previous item, what can be said about the Hölder continuity of f for $f \in C^2(\mathbb{T})$?
5. Prove that, if $f \in R(\mathbb{T})$ is Hölder continuous of order $\alpha > 0$, then there is a constant C_α for which

$$|\hat{f}(n)| \leq \frac{C_\alpha}{|n|^\alpha}$$

for nonzero $n \in \mathbb{Z}$. **Hint:** For $n \neq 0$, the 2π-periodicity of f guarantees that

$$\hat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} f\left(x + \frac{\pi}{n}\right)e^{-in(x+\pi/n)}\,dx.$$

Upon noting that $e^{-in(x+\pi/n)} = -e^{-inx}$, you can average the above integrals to obtain a nice way to write $\hat{f}$ as an integral over a difference $f(x) - f(x + \pi/n)$. Now use Hölder continuity.

6. Use your result from the item above and Theorem 8.22 to prove that the Fourier series for f converges uniformly to f whenever f is Hölder continuous of order α for any $\alpha > 1$. Hint: Appeal to the summability of p series.

This concludes our investigation of the uniform convergence of Fourier series. We now move on to the theory of pointwise convergence. In contrast to our study of uniform convergence, the main object in our study of pointwise convergence of Fourier series revolves around a careful (analytical) study of the Dirichlet kernel, which we introduce now.

Proposition 8.24. *Let $f \in R(\mathbb{T})$ and let $(\hat{f}(n))_{n\in\mathbb{Z}}$ be its Fourier coefficients. For each N, let S_N denote the Nth Fourier polynomial of f, i.e.,*

$$S_N(x) = \sum_{n=-N}^{N} \hat{f}(n)e^{inx}.$$

Then, for each $N \in \mathbb{N}$ and $x \in \mathbb{R}$,

$$S_N(x) = \frac{1}{2\pi}\int_{\mathbb{T}} D_N(x-y)f(y)\,dy$$

where

$$D_N(y) := \begin{cases} \frac{\sin((N+1/2)y)}{\sin y/2} & y \neq 2\pi k, k \in \mathbb{Z} \\ 2N+1 & y = 2\pi k, k \in \mathbb{Z} \end{cases} = \sum_{n=-N}^{N} e^{iny};$$

this is called the Dirichlet kernel. For each N,

$$\frac{1}{2\pi}\int_{\mathbb{T}} D_N(y)\,dy = 1$$

and

$$\frac{1}{2\pi}\int_{[-\pi,0]} D_N(y)\,dy = \frac{1}{2\pi}\int_{[0,\pi]} D_N(y)\,dy = \frac{1}{2}.$$

Proof. By virtue of the linearity of the integral, it is evident that

$$\begin{aligned} S_N(x) &= \sum_{n=-N}^{N}\left(\frac{1}{2\pi}\int_{\mathbb{T}} f(y)e^{-iny}\,dy\right)e^{inx} \\ &= \frac{1}{2\pi}\int_{\mathbb{T}}\sum_{n=-N}^{N} f(y)e^{in(x-y)}\,dy \end{aligned}$$

for $N \in \mathbb{N}$ and $x \in \mathbb{R}$. We now show that

$$D_N(y) = \sum_{n=-N}^{N} e^{iny}$$

for $N \in \mathbb{N}$ and $x \in \mathbb{R}$. First, it is clear that, if $y = 2\pi k$ for $k \in \mathbb{Z}$, $D_N(y) = 2N+1 = \sum_{n=-N}^{N} 1 = \sum_{n=-N}^{N} e^{iny}$. Thus, we assume without loss of generality that $y \neq 2\pi k$ for $k \in \mathbb{Z}$ and hence $\sin y/2 \neq 0$. Now,

$$\begin{aligned} D_1(y) &= \frac{\sin((1+1/2)y)}{\sin(y/2)} = \frac{\sin y\cos(y/2)+\sin(y/2)\cos y}{\sin(y/2)} \\ &= \frac{\sin((1/2+1/2)y)\cos(y/2)}{\sin(y/2)} + \cos y = \frac{2\sin(y/2)\cos(y/2)\cos(y/2)}{\sin(y/2)} + \cos y \\ &= 2\cos^2(y/2) + \cos y = 1 + 2\cos y \\ &= e^{i\cdot 0\cdot y} + e^{iy} + e^{iy} = \sum_{n=-1}^{1} e^{iny} \end{aligned}$$

where we have used the half-angle identity $2\cos^2(A/2) = 1+\cos(A)$. Thus, the desired result is true for $N = 1$.

Let's induct on N. Let's assume that the formula holds for $N \geq 1$, we will show it holds for $N+1$. We have

$$\begin{aligned}
D_{N+1}(y) &= \frac{\sin((N+1+1/2)y)}{\sin y/2} \\
&= \frac{\sin y \cos((N+1/2)y) + \cos y \sin((N+1/2)y}{\sin y/2} \\
&= 2\cos(y/2)\cos((N+1/2)) + \cos y D_N(y) \\
&= 2\cos(y/2)(\cos Ny \cos(y/2) - \sin Ny \sin(y/2)) + \cos y D_N(y) \\
&= (1+\cos y)\cos Ny - \sin Ny \sin y + \cos y D_N(y) \\
&= \cos Ny + (\cos y \cos Ny - \sin Ny \sin y) + \cos y D_N(y) \\
&= \cos Ny + \cos((N+1)y) + \cos y D_N(y) \\
&= \frac{e^{iNy}+e^{-iNy}}{2} + \frac{e^{i(N+1)y}+e^{-i(N+1)y}}{2} + \frac{e^{iy}+e^{-iy}}{2} D_N(y) \\
&= \frac{e^{iNy}+e^{-iNy}}{2} + \frac{e^{i(N+1)y}+e^{-i(N+1)y}}{2} + \frac{e^{iy}+e^{-iy}}{2} \sum_{n=-N}^{N} e^{iny} \\
&= \frac{1}{2}\left(e^{iNy}+e^{-iNy}+e^{i(N+1)y}+e^{-i(N+1)y} + \sum_{n=-N}^{N} e^{i(n+1)y} + \sum_{n=-N}^{N} e^{i(n-1)y}\right) \\
&= \frac{1}{2}\left(e^{iNy}+e^{-iNy}+e^{i(N+1)y}+e^{-i(N+1)y} + \sum_{n=-(N-1)}^{N+1} e^{iny} + \sum_{n=-(N+1)}^{N-1} e^{iny}\right) \\
&= \frac{1}{2}\left(\sum_{n=-(N+1)}^{N+1} e^{iny} + \sum_{n=-(N+1)}^{N+1} e^{iny}\right) \\
&= \sum_{n=-(N+1)}^{N+1} e^{iny}
\end{aligned}$$

where we have made use of the induction hypothesis and a tour de force of trigonometric identities.

Now, for any $N \in \mathbb{N}$, due to the periodicity of the functions $x \mapsto e^{inx}$ for $n \neq 0$ and their antiderivatives,

$$\frac{1}{2\pi}\int_{\mathbb{T}} D_N(y)\,dy = \frac{1}{2\pi}\int_{\mathbb{T}} \sum_{n=-N}^{N} e^{iny}\,dy = \frac{1}{2\pi}\int_{\mathbb{T}} e^{i0\cdot y}\,dy = \frac{1}{2\pi}\int_{[-\pi,\pi]} 1\,dy = 1.$$

Finally, by a quick examination, it is clear the $D_N(y)$ is an even function. Consequently

$$1 = 2\left(\frac{1}{2\pi}\int_{[0,\pi]} D_N(y)\,dy\right) = 2\left(\frac{1}{2\pi}\int_{[-\pi,0]} D_N(y)\,dy\right)$$

from which the final result follows. □

The Dirichlet kernels D_5, D_{10} and D_{20} are illustrated in Figure 8.2.

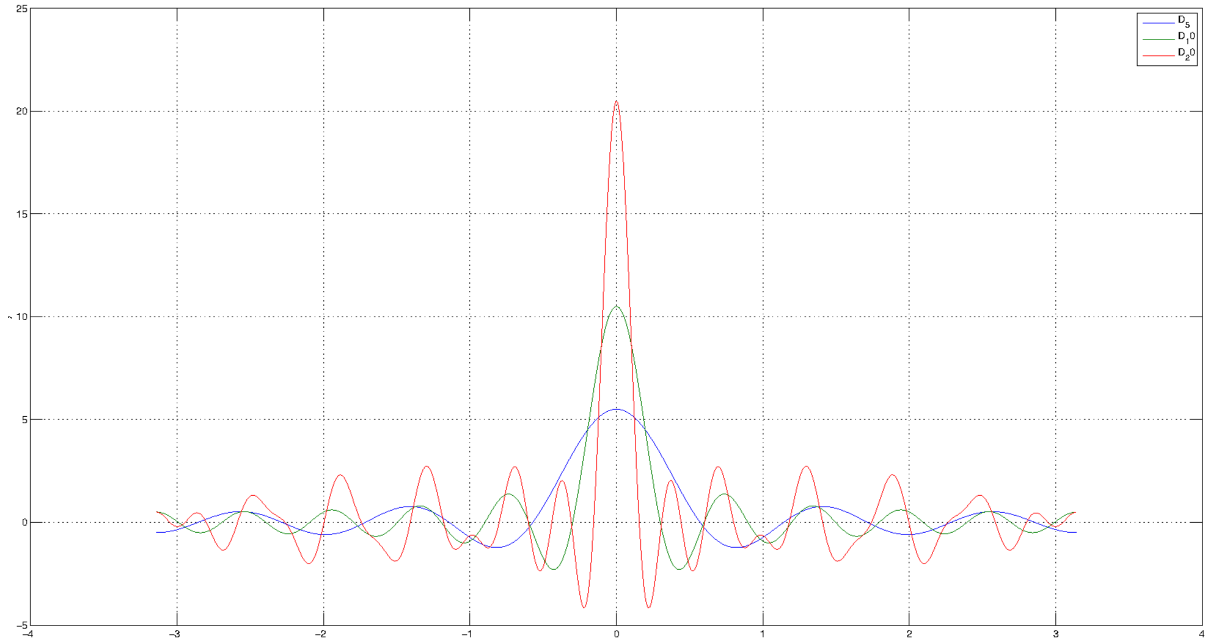

Figure 8.2: The graphs of D_5, D_{10} and $D_2 0$.

8.4.1 Convolutions

Given two functions $f, g \in R(\mathbb{T})$, we define their convolution $f * g : \mathbb{R} \to \mathbb{C}$ by

$$(f * g)(x) = \frac{1}{2\pi} \int_{\mathbb{T}} f(x-y)g(y)\, dy = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-y)g(y)\, dy$$

for $x \in \mathbb{R}$. The convolution operation is a way to combine information about the functions f and g to produce a new function (which will often be nicer) called $f * g$. Let's consider an example.

8.5 All things Fourier

8.5.1 The L^2 theory

In the last subsection, we showed that continuous 2π-periodic functions can be approximated uniformly by trigonometric polynomials. In this subsection, we turn our focus to a larger class of functions, 2π-periodic and Riemann integrable functions, and a completely different mode of approximation, the L^2 approximation. It is my hope that you will see that the L^2 theory, outlined in this subsection, is the cleanest and best-adapted for approximation by trigonometric polynomials. In fact, convergence of Fourier series in L^2 will happen even when other forms of convergence fail.

Let $f : \mathbb{R} \to \mathbb{C}$. We say that f is Riemann integrable on an interval $[a, b]$ if its real and imaginary parts are Riemann integrable on $[a, b]$. In this case, we define the Riemann integral of f (a complex-valued function) on $[a, b]$ by

$$\int_{[a,b]} f = \int_{[a,b]} f(x)\,dx = \int_{[a,b]} \mathrm{Re}(f)(x)\,dx + i\int_{[a,b]} \mathrm{Im}(f)(x)\,dx.$$

Using the properties of the Riemann integral established for real-valued functions, it is easy to check that the Riemann integral, defined here for complex-valued functions, is also linear.

Now, given a 2π-periodic function $f : \mathbb{R} \to \mathbb{C}$, we say that f is Riemann integrable on $\mathbb{T}$ if f is Riemann integrable on $[-\pi, \pi]$ and in this case we write

$$\int_{\mathbb{T}} f = \int_{\mathbb{T}} f(x)\,dx = \int_{[-\pi,\pi]} f(x)\,dx.$$

When the context is clear, we will simply say that f is Riemann integrable. The set of all such functions will henceforth be denoted by $R(\mathbb{T})$. It is clear that $C(\mathbb{T}) \subseteq R(\mathbb{T})$.

Given $f, g \in R(\mathbb{T})$, we define the $L^2 = L^2(\mathbb{T})$ inner product by

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{\mathbb{T}} f(x)\overline{g(x)}\,dx.$$

You should verify that $\langle\cdot,\cdot\rangle$ satisfies all of the properties of an inner product except positive definiteness. In particular, $\langle f, f\rangle \geq 0$ for all $f \in R(\mathbb{T})$ and so it makes sense to define the corresponding L^2 norm by

$$\|f\|_{L^2(\mathbb{T})} = \sqrt{\langle f, f\rangle} = \left(\frac{1}{2\pi}\int_{\mathbb{T}} |f(x)|^2\,dx\right)^{1/2};$$

this is also called the root-mean-square norm. We will often use the shorthand $\|f\|_2$ for $\|f\|_{L^2(\mathbb{T})}$.

As a quick remark, I should point out that $\|\cdot\|_2$ isn't a bona fide norm on the set $R(\mathbb{T})$. It satisfies all of the properties of a norm except positive definiteness. To see that positive definiteness fails, consider the function

$$f(x) = \begin{cases} 1 & \text{if } x \in 2\pi\mathbb{Z} \\ 0 & \text{otherwise.}\end{cases}$$

It is clear that $f \in R(\mathbb{T})$. Further,

$$\|f\|_2^2 = \frac{1}{2\pi}\int_{[-\pi,\pi]} f(x)^2\,dx = 0$$

however f is not the zero function. As it turns out $\|\cdot\|_2$ does become a norm on the Lebesgue space $L^2(\mathbb{T})$ which we do not discuss here. In fact, $L^2(\mathbb{T})$ is a complete (as a metric space) inner product space, such a space is called a Hilbert space. This is standard material in a course on measure theory.

Let us now observe that, for each $n, m \in \mathbb{Z}$,

$$\int_{\mathbb{T}} e^{inx}\overline{e^{imx}}\,dx = \int_{\mathbb{T}} e^{inx-imx}\,dx = \int_{\mathbb{T}} e^{i(n-m)x}\,dx = \int_{[-\pi,\pi]} \cos((n-m)x)\,dx + i\int_{[-\pi,\pi]} \sin((n-m)x)\,dx.$$

Consequently,

$$\langle e^{inx}, e^{imx}\rangle = \frac{1}{2\pi}\int_{\mathbb{T}} e^{inx}\overline{e^{imx}}\,dx = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{if } n \neq m.\end{cases}$$

In view of the preceding calculation, the collection of functions $(e^{inx})_{n\in\mathbb{Z}}$ is called an *orthonormal system.* Let's make a further observation, given $P \in \mathcal{P}(\mathbb{T})$ of the form

$$P(x) = \sum_{n=-N}^{N} c_n e^{inx},$$

for each $-N \leq m \leq N$,

$$\langle P, e^{imx}\rangle = \sum_{n=-N}^{N} c_n \langle e^{inx}, e^{imx}\rangle = c_m$$

and $\langle P, e^{imx}\rangle = 0$ whenever $|m| > N$. In this way, we find a way to find the coefficients of a trigonometric polynomial by simply integrating against the elements of the system (e^{inx}). This is analogous to the way that the coefficients of an analytic function can be computed by taking derivatives. Taking our cues from the above computation, we make a definition.

Definition 8.25. *Let $f \in R(\mathbb{T})$. For each $n \in \mathbb{Z}$, define*

$$\hat{f}(n) = \langle f, e^{inx}\rangle = \frac{1}{2\pi}\int_{\mathbb{T}} f(x)e^{-inx}dx.$$

The collection of complex numbers $(\hat{f}(n))_{n\in\mathbb{Z}}$ are called the Fourier coefficients of f. Considered as a formal series, the series

$$\sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}$$

is called the Fourier series for f and we write

$$f(x) \sim \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}$$

to indicate that the right hand side is the Fourier series for f.

In view of the discussion preceding the definition, we see that computing the Fourier series corresponding to a trigonometric polynomial P returns the polynomial P itself. Do you see an analogue with Taylor series here? Throughout the rest of this section, we begin to analyze the ways in which the Fourier series for a function f converges. As you will see, if the Fourier series is to converge, it will converge back to f. To this end, we will start talking about partial sums.

Let $f \in R(\mathbb{T})$ and $(\hat{f}(n))_{n\in\mathbb{Z}}$ be the Fourier coefficients of f. For each $N \in \mathbb{N}$, the Nth order trigonometric polynomial

$$S_N(x) = \sum_{n=-N}^{N} \hat{f}(n)e^{inx}$$

defined for $x \in \mathbb{R}$ is called the Nth partial sum of the Fourier series $\sum \hat{f}(n)e^{inx}$. Our first main result of the subsection says that, of all Nth order trigonometric polynomials, S_N is the best root-mean-square approximation to f.

Theorem 8.26. *Let $f \in R(\mathbb{T})$ and let $(\hat{f}(n))_{n\in\mathbb{Z}}$ be its Fourier coefficients. Given $N \in \mathbb{N}$ let $S_N(x)$ be the Nth order partial sum of the Fourier series for f and let $P_N \in \mathcal{P}(\mathbb{T})$ be another (possibly different) Nth order trigonometric polynomial of the form*

$$P_N(x) = \sum_{n=-N}^{N} c_n e^{inx}.$$

Then

$$\|f - S_N\|_2 \leq \|f - P_N\|_2$$

where equality holds if and only if $c_n = \hat{f}(n)$ for all n.

Proof. We first recall the basic linearity properties of the inner product: For $u, v, w \in R(\mathbb{T})$, and $a \in \mathbb{C}$,

$$\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle, \qquad \langle u, v+w\rangle = \langle u, v\rangle + \langle u, w\rangle$$

and

$$\langle au, v\rangle = a\langle u, v\rangle, \qquad \langle u, av\rangle = \overline{a}\langle u, v\rangle.$$

By virtue of these properties, we can write

$$\begin{aligned}
\|f-P_N\|_2^2 &= \langle f-P_N, f-P_N\rangle = \langle f, f\rangle - \langle f, P_N\rangle - \langle P_N, f\rangle + \langle P_N, P_N\rangle \\
&= \|f\|_2^2 - \sum_{n=-N}^{N} \overline{c_n}\langle f, e^{inx}\rangle - \sum_{n=-N}^{N} c_n\langle e^{inx}, f\rangle + \sum_{n=-N}^{N}\sum_{m=-N}^{N} c_n\overline{c_m}\langle e^{inx}, e^{imx}\rangle \\
&= \|f\|_2^2 - \sum_{n=-N}^{N} \hat{f}(n)\overline{c_n} - \sum_{n=-N}^{N} c_n\overline{\hat{f}(n)} + \sum_{n=-N}^{N} c_n\overline{c_n} \\
&= \|f\|_2^2 - \sum_{n=-N}^{N} |\hat{f}(n)|^2 + \sum_{n=-N}^{N} |c_n - \hat{f}(n)|^2
\end{aligned}$$

where we have used the fact that $\langle e^{inx}, f\rangle = \overline{\langle f, e^{inx}\rangle} = \overline{\hat{f}(n)}$. Obviously, making the above computation when $c_n = \hat{f}(n)$ yields

$$\|f-S_N\|_2^2 = \|f\|_2^2 - \sum_{n=-N}^{N} |\hat{f}(n)|^2 \tag{8.1}$$

and therefore

$$\|f-P_N\|_2^2 = \|f-S_N\|_2^2 + \sum_{n=-N}^{N} |c_n - \hat{f}(n)|^2$$

from which we observe that

$$\|f-S_N\|_2 \leq \|f-P_N\|_2$$

with equality if and only if $c_n = \hat{f}(n)$ for all n. □

Theorem 8.27 (Bessel's inequality). *Let $f \in R(\mathbb{T})$ and let $(\hat{f}(n))_{n\in\mathbb{Z}}$ be the Fourier coefficients of f. Then*

$$\sum_{n\in\mathbb{Z}} |\hat{f}(n)|^2 \leq \|f\|_2^2 = \frac{1}{2\pi}\int_{\mathbb{T}} |f(x)|^2\,dx.$$

In particular, the series

$$\sum_{n\in\mathbb{Z}} |\hat{f}(n)|^2 = \lim_{N\to\infty}\sum_{n=-N}^{N} |\hat{f}(n)|^2$$

converges.

Proof. In view of (8.1), we have

$$\sum_{n=-N}^{N} |\hat{f}(n)|^2 \leq \|f\|_2^2$$

for all $N \in \mathbb{N}$. The desired result follows by taking the limit of the left hand side as $N \to \infty$ and noting that the partial sums, whose summands are all non-negative, form a non-decreasing sequence of non-negative numbers. □

Corollary 8.28 (The Riemann-Lebesgue lemma). *Let $f \in R(\mathbb{T})$ and let $(\hat{f}(n))_{n\in[Z}$ be the Fourier coefficients of f. Then*

$$\lim_{n\to\infty} \hat{f}(n) = \frac{1}{2\pi}\int_{\mathbb{T}} f(x)e^{-inx}\,dx = 0$$

and

$$\lim_{n\to\infty} \hat{f}(-n) = \frac{1}{2\pi}\int_{\mathbb{T}} f(x)e^{inx}\,dx = 0.$$

Proof. The convergence of the series $\sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2$ implies that the summands for sufficiently large and largely negative n converge to zero. □

Corollary 8.29 (A sharper form of the Riemann-Lebesgue lemma). *Let $[a,b] \subseteq [-\pi,\pi]$ and suppose that f is Riemann integrable on $[a,b]$. Then*

$$\lim_{n\to\infty} \frac{1}{2\pi}\int_{[a,b]} f(x)\cos(nx)\,dx = 0$$

and

$$\lim_{n\to\infty} \frac{1}{2\pi}\int_{[a,b]} f(x)\sin(nx)\,dx = 0.$$

Proof. Given f as above, consider $g \in R(\mathbb{T})$ defined by

$$g(x) = \begin{cases} f(x) & x \in [a,b] \\ 0 & x \in [-\pi,\pi]\setminus[a,b] \end{cases}$$

and extended periodically to $\mathbb{R}$. Then, by the Riemann-Lebesgue lemma applied to g, we have

$$\lim_{n\to\infty} \frac{1}{2\pi}\int_{[a,b]} f(x)e^{-inx}\,dx = \lim_{n\to\infty} \frac{1}{2\pi}\int_{[-\pi,\pi]} g(x)e^{-inx}\,dx = \lim_{n\to\infty} \hat{g}(n) = 0$$

and

$$\lim_{n\to\infty} \frac{1}{2\pi}\int_{[a,b]} f(x)e^{inx}\,dx = \lim_{n\to\infty} \frac{1}{2\pi}\int_{[-\pi,\pi]} g(x)e^{inx}\,dx = \lim_{n\to\infty} \hat{g}(-n) = 0.$$

Consequently,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{2\pi}\int_{[a,b]} f(x)\cos(nx)\,dx &= \lim_{n\to\infty} \frac{1}{2\pi}\int_{[a,b]} f(x)\left(\frac{e^{inx}+e^{-inx}}{2}\right)\,dx \\ &= \frac{1}{2}\left(\lim_{n\to\infty} \frac{1}{2\pi}\int_{[a,b]} f(x)e^{inx}\,dx + \lim_{n\to\infty} \frac{1}{2\pi}\int_{[a,b]} f(x)e^{-inx}\,dx\right) \\ &= 0. \end{aligned}$$

The proof that $\lim_{n\to\infty}(2\pi)^{-1}\int_{[a,b]} f(x)\sin(nx)\,dx = 0$ is similar. □

Our next result shows that the Fourier series for f converges to f with respect to the $L^2(\mathbb{T})$ norm. The result also shows that Bessel's inequality is, in fact, an equality. We first need the following simple lemma that you will prove in your Homework 7.

Lemma 8.30. *Let $f \in \mathbb{R}(\mathbb{T})$. For any $\epsilon > 0$, there exists $g \in C(\mathbb{T})$ such that*

$$\|f-g\|_2 < \epsilon.$$

Theorem 8.31 (Parseval's Theorem)**.** *Let $f \in R(\mathbb{T})$ and let $(\hat{f}(n))_{n\in\mathbb{Z}}$ the Fourier coefficients of f. For each natural number N, denote by S_N the Nth partial sum of the Fourier series for f. Then*

$$\lim_{N\to\infty} \|f - S_N\|_2 = 0;$$

this is the statement that the Fourier series $\sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}$ converges to f with respect to the L^2 norm. Further

$$\|f\|_2^2 = \sum_{n\in\mathbb{Z}} |\hat{f}(n)|^2.$$

Proof. Let $\epsilon > 0$ and by an appeal to the lemma choose $h \in C(\mathbb{T})$ for which $\|f - h\|_2 < \epsilon/2$. Now, in view of Theorem **??** This is the Weierstrass Theorem!, Let $P \in \mathcal{P}(\mathbb{T})$ be a trigonometric polynomial of the form

$$P(x) = \sum_{n=-N}^{N} c_n e^{inx}$$

for which

$$|h(x) - P(x)| < \epsilon/2$$

for all $x \in \mathbb{R}$. From this it follows immediately that

$$\|f - P\|_2 \leq \|f - h\|_2 + \|h - P\|_2 < \epsilon/2 + \left(\frac{1}{2\pi}\int_{[-\pi,\pi]} (\epsilon/2)^2\, dx\right)^{1/2} = \epsilon.$$

Now, for any $M \geq N$, the Mth partial sum of the Fourier series for f, S_M can be compared with the trigonometric polynomial P (which can trivially be though of as of Mth degree by taking its coefficients to be zero for $M \leq |n| > N$. Thus, in view of Theorem 8.26

$$\|f - S_M\|_2 \leq \|f - P\|_2 < \epsilon.$$

Hence, for every $\epsilon > 0$, there exists and $N \in \mathbb{N}$ such that, for every $M \geq N$, $\|f - S_M\|_2 < \epsilon$ and therefore

$$\lim_{N\to\infty} \|f - S_N\|_2 = 0.$$

With this, (8.1) gives

$$\lim_{N\to\infty} \sum_{n=-N}^{N} |\hat{f}(n)|^2 = \|f\|_2^2 - \lim_{N\to\infty} \|f - S_N\|_2^2 = \|f\|_2^2 - 0$$

form which the desired result follows immediately. ☐

Example 8.2:

Consider the so-called sawtooth function defined by

$$f(x) = x \qquad -\pi < x \leq \pi$$

and extended 2π-periodically to $\mathbb{R}$. The graph of f is illustrated in Figure 8.3 (the vertical lines are not part of the graph; they are inserted automatically by Matlab).

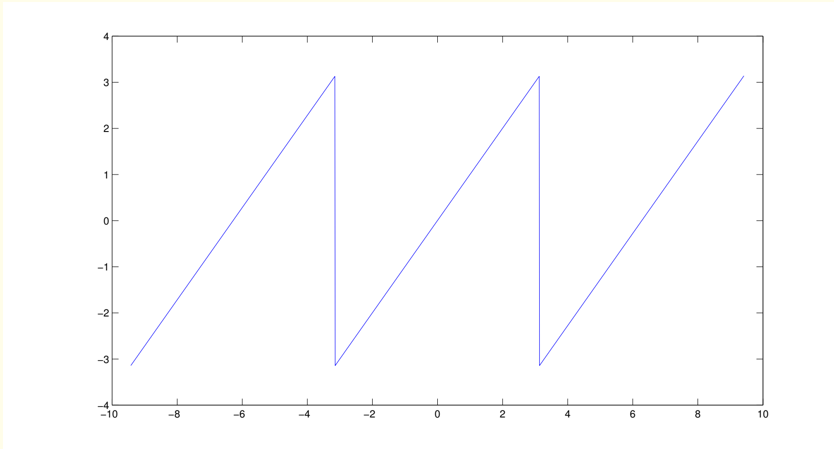

Figure 8.3: The graph of f for $-3\pi < x \leq 3\pi$.

Let's compute the Fourier coefficients of f. For $n = 0$, we have

$$\hat{f}(0) = \frac{1}{2\pi}\int_{\mathbb{T}} f(x)e^{-i\cdot 0\cdot x}\,dx = \frac{1}{2\pi}\int_{\mathbb{T}} x\,dx = \frac{1}{2\pi}\int_{[-\pi,\pi]} x\,dx = 0.$$

For $n \neq 0$,

$$\begin{aligned}
2\pi\hat{f}(n) &= \int_{\mathbb{T}} xe^{-inx}\,dx \\
&= \int_{[-\pi,\pi]} x\cos(-nx)\,dx + i\int_{[-\pi,\pi]} x\sin(-nx)\,dx \\
&= \int_{[-\pi,\pi]} x\cos nx\,dx - i\int_{[-\pi,\pi]} x\sin nx\,dx \\
&= -i\left(\frac{-1}{n}x\cos nx\big|_{-\pi}^{\pi} - \frac{-1}{n}\int_{[-\pi,\pi]}\cos nx\,dx\right) \\
&= \frac{i}{n}(\pi\cos n\pi - (-\pi\cos(-n\pi)) - 0 \\
&= \frac{2\pi i}{n}\cos n\pi = \frac{2\pi i}{n}(-1)^n
\end{aligned}$$

where we have integrated by parts and used (heavily) the periodicity and odd and even properties of sine/cosine. Consequently,

$$\hat{f}(n) = \begin{cases} \frac{i(-1)^n}{n} & n \neq 0 \\ 0 & n = 0. \end{cases}$$

Observe that

$$\sum_{n\in\mathbb{Z}} |\hat{f}(n)|^2 = \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{n^2} = 2\sum_{n=1}^{\infty}\frac{1}{n^2}.$$

Also,

$$\|f\|_2^2 = \frac{1}{2\pi}\int_{\mathbb{T}} |f(x)|^2\,dx = \frac{1}{2\pi}\int_{[-\pi,\pi]} x^2\,dx = \frac{1}{2\pi}\frac{x^3}{3}\big|_{-\pi}^{\pi} = \frac{2\pi^3}{6\pi} = \frac{\pi^2}{3}$$

and so, by an application of Parseval's theorem, we obtain

$$\frac{\pi^2}{3} = \|f\|_2^2 = \sum_{x\in\mathbb{Z}} |\hat{f}(n)|^2 = 2\sum_{n=1}^{\infty} \frac{1}{n^2}.$$

It was long understood that the series $\sum_{n=1}^{\infty} 1/n^2$ converged. The Basel problem, posed by Pietro Mengoli around 1644, asks: What is the exact value of this series? In 1734, a mathematician named Leonhard Euler showed that the value of the series is exactly $\pi^2/6$. The solution made Euler instantly famous. By simply dividing the previous equation by 2 we obtain Euler's result (Theorem 8.32 below). We note that our approach (via Parseval's Theorem) is completely different than that of Euler. This should not be surprising as Fourier series wasn't discovered for nearly one hundred years after Euler presented his result.

Theorem 8.32 (Euler 1734)**.**

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

As our attention will soon turn to the investigation of pointwise convergence of Fourier series, let's continue our Fourier series analysis of the Sawtooth function. As we will see, this example contains a surprising number of the strange phenomena/pathologies found commonplace in the study of Fourier series, including the Gibb's phenomenon. For each N, the Nth Fourier polynomial for f is given by

$$\begin{aligned} S_N(x) &= \sum_{n=-N}^{N} \hat{f}(n)e^{inx} = \hat{f}(0) + \sum_{n=1}^{N} \hat{f}(n)e^{inx} + \sum_{n=1}^{N} \hat{f}(-n)e^{-inx} \\ &= \sum_{n=1}^{N} \frac{i(-1)^n}{n} e^{inx} + \frac{i(-1)^{-n}}{-n} e^{-inx} = \sum_{n=1}^{N} \frac{(-1)^n}{n} i(e^{inx} - e^{-inx}) \\ &= \sum_{n=1}^{N} \frac{(-1)^n}{n} i(2i\sin nx) = \sum_{n=1}^{N} \frac{2(-1)^{n+1}}{n} \sin nx \end{aligned}$$

for $x \in \mathbb{R}$. The Fourier polynomials of degrees $1, 2, 3, 4, 5$ and 40 are illustrated in Figure 8.4 for $-3\pi < x \leq 3\pi$. The reader should note that the partial sums appear to be converging nicely except for the overshoot near the points of discontinuity at $x = \pm 3\pi, \pm\pi$; this is the Gibb's phenomenon.

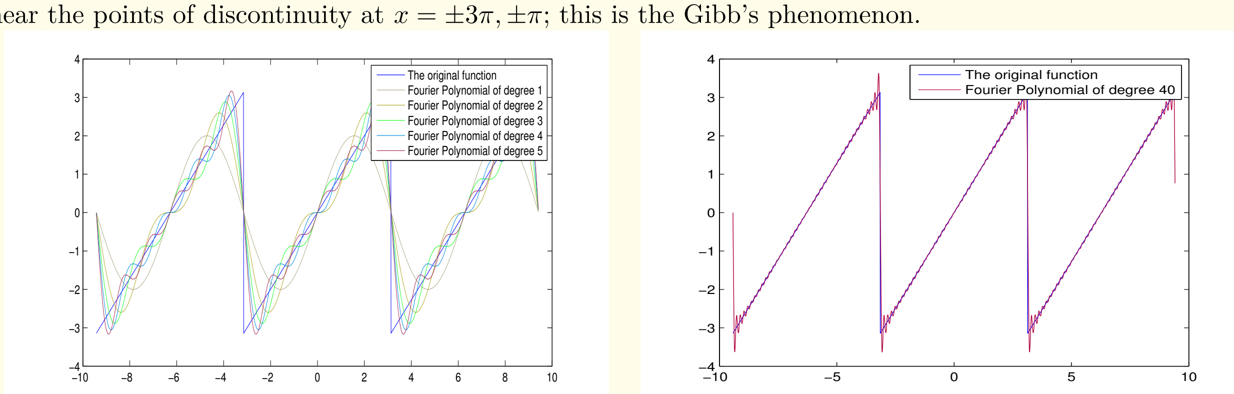

Figure 8.4: The graphs of f and P_n for $n = 1, 2, 3, 4, 5, 40$.

8.5.2 The pointwise and uniform theory theory

Note: I discussed most of these results on Friday, May 8th in the class with optional attendance. Here is a video recording of that day's lecture: Lecture Recording

As we saw in the last subsection, the Fourier series for a function $f \in R(\mathbb{T})$ converges to f with respect to the L^2 norm. In this subsection, we investigate the same question from the perspective of pointwise (and uniform) convergence; as we have seen, both notions are stronger than L^2 convergence. When originally investigating Fourier series in its connection to the theory of heat, Jean-Baptiste Fourier claimed that, given $f \in C(\mathbb{T})$, the Fourier series $\sum_{n\in\mathbb{N}} \hat{f}(n)e^{inx}$ converges to $f(x)$ for all $x \in \mathbb{R}$. As it turns out, this isn't true.

Theorem 8.33 (Du Bois-Reymond, 1873)**.** *There exits $f \in C(\mathbb{T})$ whose Fourier series $\sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}$ diverges at a point $x \in (-\pi, \pi]$. This means specifically that, for some $x \in (-\pi, \pi]$, the limit*

$$\lim_{N\to\infty} \sum_{n=-N}^{N} \hat{f}(n)e^{inx}$$

does not exist.

Following the discovery of this result, a frantic search began in mathematics to find precise conditions on a function f which would guarantee that its Fourier series converged. Throughout this search, it was discovered that the Riemann integral was insufficient for the needed purposes of investigation and this helped lead to a complete revolution in mathematics. Our of the revolution emerged the Lebesgue integral and Lebesgue's theory of integration. Any serious investigation of Fourier series (which ours is unfortunately not) requires one to understand the Lebesgue integral. In 1926 a brilliant young mathematician named Andre Kolmogorov showed that things were much worse than had been previously thought (and had been shown by Du Bois-Reymond).

Theorem 8.34 (Kolmogorov 1926)**.** *There exists a Lebesgue integrable function f (we say $f \in L^1(\mathbb{T})$) whose Fourier series diverges at every point.*

We now begin our investigation of the uniform convergence of Fourier series.

Lemma 8.35. *Let $f, g \in C(\mathbb{T})$, if $\hat{f}(n) = \hat{g}(n)$ for all n, then $f = g$, i.e., $f(x) = g(x)$ for all $x \in \mathbb{R}$.*

Proof. As you show in your homework, $\widehat{(f-g)}(n) = \widehat{f}(n) - \widehat{g}(n)$ for all $n \in \mathbb{Z}$, i.e., the map $f \mapsto \hat{f}$ is linear. Thus, $\widehat{(f-g)}(n) = 0$ for all $n \in \mathbb{Z}$ and so, by an appeal to Parseval's theorem,

$$\|f-g\|_2^2 = \sum_{n\in\mathbb{Z}} |\widehat{(f-g)}(n)|^2 = 0.$$

Consequently,

$$\int_{\mathbb{T}} |f(x) - g(x)|^2\,dx = 2\pi\|f-g\|_2^2 = 0.$$

It now follows, by the result I prove in class during Week 6, that $f(x) - g(x) = 0$ for all $x \in [-\pi, \pi]$. Because f and g are 2π periodic, we conclude that $f(x) = g(x)$ for all $x \in \mathbb{R}$. $\square$

With the properties of the Dirichlet kernel established in the preceding proposition, we have our first (truely) pointwise result.

Theorem 8.36. *Let $f \in R(\mathbb{T})$ and assume that the $f = u+iv$ is piecewise differentiable, i.e., its real and imaginary parts u and v are continuously differentiable on $[-\pi, \pi]$ except possibly at a finite number of points where u, v, u' and v' have (at worse) removable or jump discontinuities. For any $x_0 \in \mathbb{R}$, define*

$$f(x_0^+) = \lim_{x\to x_0; x>x_0} f(x) \qquad \textit{and} \qquad f(x_0^-) = \lim_{x\to x_0 : x<x_0} f(x).$$

Then, at each $x_0 \in \mathbb{R}$, the Fourier series for f converges and

$$\frac{f(x_0^+) + f(x_0^-)}{2} = \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx_0} = \lim_{N\to\infty} \sum_{n=-N}^{N} \hat{f}(n)e^{inx_0}.$$

In particular, if f is continuous at x_0 (or has a removable discontinuity at x_0),

$$f(x_0) = \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx_0};$$

otherwise, the Fourier series for f converges to the average of the left and right limits of f at x_0.

Before proving the theorem, let's illustrate its conclusion by revisiting the sawtooth function and its Fourier series.

Example 8.3: W

e recall the sawtooth function defined by $f(x) = x$ for $-\pi < x \leq \pi$ and extended periodically to $\mathbb{R}$. We previously studied this function in Example 8.5.1 and computed its Fourier series. We found

$$f(x) \sim \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{i(-1)^n}{n} e^{inx} = \sum_{n=1}^{\infty} \frac{2(-1)^{n+1}}{n} \sin nx$$

Now, by a straightforward computation, f is differentiable on the open set $\mathbb{R}\setminus\{(2k+1)\pi : k \in \mathbb{Z}\}$ (consisting of the entire real line except for the *breakpoints* $\pm\pi, \pm3\pi, \pm5\pi, \dots$) and, on this set, $f'(x) = 1$. Consequently, f is piecewise differentiable and so we may apply the theorem.
For $-\pi < x < \pi$, f is continuous and by virtue of the theorem we conclude that the Fourier series for f converges to $f(x) = x$ on the open set $(-\pi, \pi)$. At $x = \pi$, f has a discontinuity. Here we have

$$f(\pi_-) = \lim_{x\to\pi;x<\pi} f(x) = \lim_{x\to\pi;x<\pi} x = \pi$$

and

$$f(\pi_+) = \lim_{x\to\pi;x>\pi} f(x) = \lim_{x\to\pi;x>\pi} (x - 2\pi) = \pi - 2\pi = -\pi.$$

Consequently,

$$\sum_{n\in\mathbb{Z}} \hat{f}(n)e^{in\pi} = \lim_{N\to\infty} S_N(\pi) = \frac{1}{2}(\pi + -\pi) = 0;$$

in fact, this result holds at all the breakpoints $\pm\pi, \pm3\pi, \dots,$. Appealing to the full scope of the theorem (or simply noting that the above conclusion extends by periodicity), we have

$$\sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx} = \lim_{N\to\infty} S_N(x) = \begin{cases} (x - 2\pi k) & (2k-1)\pi < x < (2k+1)\pi, k \in \mathbb{Z} \\ 0 & x = (2k+1)\pi, k \in \mathbb{Z} \end{cases}.$$

In other words, the Fourier series for f converges to f pointwise on the open set $\mathbb{R}\setminus\{(2k+1)\pi : k \in \mathbb{Z}\}$ and it converges to 0 elsewhere. We note that the series does not converge uniformly for, if this were the case, f would be continuous. This convergence is illustrated in the following Figure 8.5.

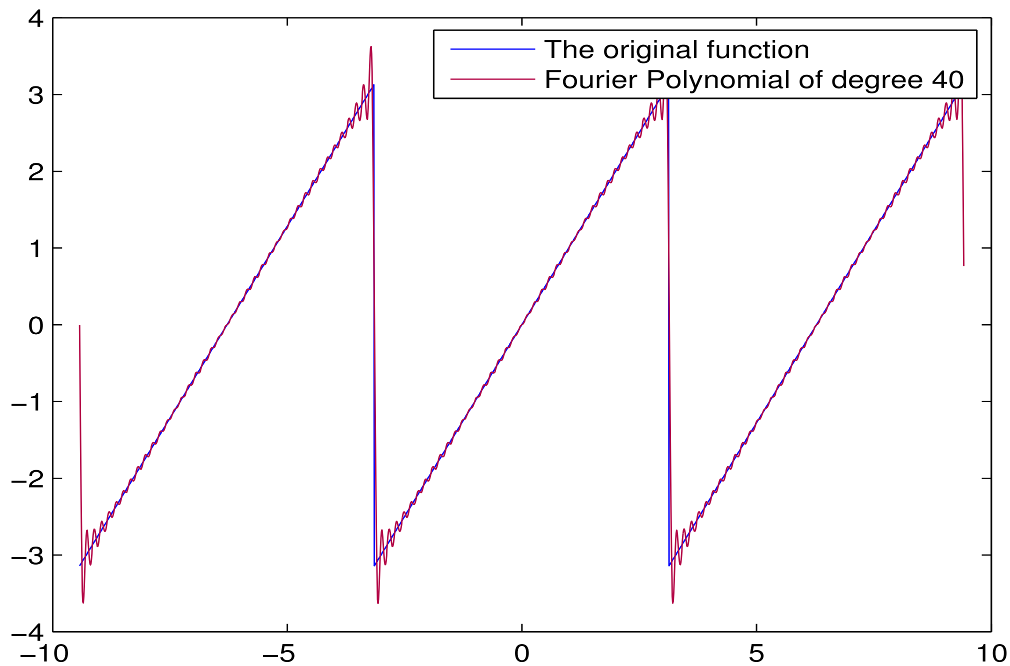

Figure 8.5: f and S_{40}

We now prove the theorem.

Proof of Theorem 8.36. Our aim is to show that

$$\lim_{N\to\infty}\left(S_N(x_0)-\frac{f(x_0^+)+f(x_0^-)}{2}\right)=0.$$

In view of Proposition 8.24, we have

$$\begin{aligned}\left(S_n(x_0)-\frac{f(x_0^+)+f(x_0^-)}{2}\right) &= \frac{1}{2\pi}\int_{\mathbb{T}} f(x_0-y)D_N(y)\,dy-\left(f(x_0^+)\frac{1}{2}+f(x_0^-)\frac{1}{2}\right)\\ &= \frac{1}{2\pi}\int_{[-\pi,0]} f(x_0-y)D_N(y)\,dy+\frac{1}{2\pi}\int_{[0,\pi]} f(x_0-y)D_N(y)\,dy\\ &\qquad -\left(f(x_0^+)\frac{1}{2\pi}\int_{[-\pi,0]}D_N(y)\,dy+f(x_0^-)\frac{1}{2\pi}\int_{[0,\pi]}D_N(y)\,dy\right)\\ &= \frac{1}{2\pi}\int_{[-\pi,0]}(f(x_0-y)-f(x_0^+))D_N(y)\,dy+\frac{1}{2\pi}\int_{[0,\pi]}(f(x_0-y)-f(x_0^-))D_N(y)\,dy\\ &=: I_+(N)+I_-(N).\end{aligned}$$

We consider the integrals I_+ and I_-. By virtue of Proposition 8.24,

$$\begin{aligned} I_+(N) &= \frac{1}{2\pi}\int_{[-\pi,0]}(f(x_0-y)-f(x_0^+))\frac{\sin((N+1/2)y)}{\sin(y/2)}\,dy \\ &= \frac{1}{2\pi}\int_{[0,\pi]}(f(x_0+y)-f(x_0^+))\frac{\sin((N+1/2)y)}{\sin(y/2)}\,dy \\ &= \frac{1}{2\pi}\int_{[0,\pi]}(f(x_0+y)-f(x_0^+))\left(\cos(Ny)+\frac{\cos(y/2)\sin(Ny)}{\sin(y/2)}\right)\,dy \\ &= \frac{1}{2\pi}\int_{[0,\pi]}(f(x_0+y)-f(x_0^+))\cos(Ny)\,dy+\frac{1}{2\pi}\int_{[0,\pi]}\left(\frac{f(x_0+y)-f(x_0^+)}{y}\right)\left(2(y/2)\frac{\cos(y/2)}{\sin(y/2)}\right)\sin(Ny)\,dy \\ &= \frac{1}{2\pi}\int_{[0,\pi]}g(y)\cos(Ny)\,dy+\frac{1}{2\pi}\int_{[0,\pi]}h(y)\sin(Ny)\,dy \end{aligned}$$

where

$$g(y)=f(x_0+y)-f(x_0^+)$$

and

$$h(y)=\left(\frac{f(x_0+y)-f(x_0^+)}{y}\right)\left(2(y/2)\frac{\cos(y/2)}{\sin(y/2)}\right).$$

It is clear that $g\in R(\mathbb{T})$ and thus, by Corollary 8.29,

$$\lim_{N\to\infty}\frac{1}{2\pi}\int_{[0,\pi]}g(y)\cos(Ny)\,dy=0.$$

Making the same conclusion concerning h isn't so straightforward. First, given our hypotheses concerning f, it is clear that h is piecewise continuous (continuous except at a finite number of points) on the interval $(0,\pi]$ and so it is Riemann-integrable on every compact subinterval of $(0,\pi]$. We must examine h near $y=0$ for the only possible impediment for Riemann integrability on $[0,\pi]$ is the behavior of h as $y\to 0$. First,

$$\lim_{y\to 0;y>0}2(y/2)\frac{\cos(y/2)}{\sin(y/2)}=\lim_{y\to 0;y>0}2\cos(y/2)\frac{(y/2)}{\sin(y/2)}=2.$$

Secondly, because f is piecewise differentiable on $[-\pi,\pi]$,

$$\lim_{y\to 0:y>0}\frac{f(x_0+y)-f(x_0^+)}{y}=\lim_{y\to 0:y>0}f'(x_0+y)=f'(x_0^+).$$

We remark that this is really a statement about exchanging limits and derivatives and its validity is far from obvious. A rigorous proof of this limit (which you should attempt), can be seen as an application of Theorem **??**. Putting these two result together shows that

$$\lim_{h\to 0;h>0}h(y)=2f'(x_0^+).$$

In particular, h is bounded (and well-behaved) at 0 and so it follows that h is Riemann integrable on $[0,\pi]$. By an application of Corollary 8.29, we conclude that

$$\lim_{N\to\infty}\frac{1}{2\pi}\int_{[0,\pi]}h(y)\sin(Ny)\,dy=0.$$

Consequently,

$$\lim_{N\to\infty}I_+(N)=0.$$

By making completely analagous reasoning, it follows that

$$\lim_{N\to\infty}I_-(N)=0$$

and therefore

$$\lim_{N\to\infty}\left(S_N(x_0)-\frac{f(x_0^+)+f(x_0^-)}{2}\right)=\lim_{N\to\infty}\left(I_+(N)+I_-(N)\right)=0.$$

$\square$

Exercise 8.4:

We've seen three results so far about the convergence of Fourier series, Theorem 8.22 (and related Corollary 8.23), Theorem 8.31 and Theorem 8.36. Under certain hypotheses, each of these theorems gives an affirmative (in a certain sense) to the question: Can a function be expended in terms of its Fourier series? In this exercise, you are asked to analyze a handful of specific examples through the lens of these three theorems. Specifically, for each of the functions f below, do the following:

a. Determine if the given function f is a member of $R(\mathbb{T})$, $C^m(\mathbb{T})$ for any $m\geq 0$ and/or the set of piecewise differentiable functions on $\mathbb{T}$.

b. Compute the Fourier coefficients of f.

c. Compute and simplify the Fourier series for f.

d. In view of all of the results discussed above, make the strongest statement you can about the convergence of the Fourier series of f. For instance, does it converge pointwise? If so, to what? Does it converge uniformly? If so, to what? Does it converge with respect to the L^2 norm? Please give precise statements.

e. If you concluded that $f\in R(\mathbb{T})$ (the hypothesis of Theorem 8.31), you should expect that

$$\|f\|_2^2=\sum_{n\in\mathbb{Z}}|\hat{f}(n)|^2.$$

Compute both sides of this identity (Parseval's identity). As our example of the sawtooth function gave us the solution to Basel's problem, $\sum_{n=1}^{\infty}1/n^2=\pi^2/6$, what does Parseval's identity give you here?

The following is the list of functions to be analyzed. If the rule of the function f is only given on the interval $(-\pi,\pi]$, the function should be assumed to be 2π-periodic on $\mathbb{R}$ (and so the rule on $(-\pi,\pi]$ determines the function completely).

1.
$$f(x)=|x|\qquad\text{for }\ x\in(-\pi,\pi].$$

2. For $\alpha\in\mathbb{R}$,
$$f(x)=\cos(\alpha x)\qquad\text{for }\ x\in(-\pi,\pi].$$
You should expect a discrepancy between the cases $\alpha\in\mathbb{Z}$ and $\alpha\in\mathbb{R}\setminus\mathbb{Z}$.

3.
$$f(x)=\begin{cases}1 & x\in(0,\pi]\\ 0 & x\in(-\pi,0]\end{cases}.$$

4.
$$f(x)=\begin{cases}x(\pi-x) & x\in(0,\pi]\\ x(\pi+x) & x\in(-\pi,0]\end{cases}.$$

Theorem 8.36 is the strongest result about the convergence of Fourier series we will prove in this course. I hope that you find the result satisfying, it encompasses most of the functions that you know about and can write down. The Nobel prize-winning physicist, Richard Feynman, was quite happy with this results (and ones like it) when made the following statement: "The mathematicians have shown, for a wide class of functions, in fact for all that are of interest to physicists, that if we can do the integrals we will get back $f(t)$." He made this statement in the 1960's in his famous lecture series at Caltech, just before Lennart Carleson completely solved the problem and determined the exact class of functions representable by their Fourier series [?, ?]. Carleson's result, now known as Carleson's theorem, was a long standing conjecture known as Luzin's conjecture. For your cultural benefit, I will state it here; I'll first need to make a definition.

Definition 8.37. *For any interval $I = [a, b] \subseteq \mathbb{R}$, we define*

$$\ell(I) = b - a$$

to be the length of I. Now, for any subset E of $\mathbb{R}$, we say that E is a set of measure zero (or a null set) if, for every $\epsilon > 0$, there is an infinite collection of intervals $\{I_n\}$ such that

$$E \subseteq \bigcup_{n=1}^{\infty} I_n$$

and

$$\sum_{n=1}^{\infty} \ell(I_n) < \epsilon.$$

You should think of a set of measure zero as an extremely small set. For instance, any countable (and hence finite) collection of points is of measure zero. There are, however, uncountable sets of measure zero, an important example of which is the Cantor set.

Exercise 8.5: I

n this exercise, you verify the claim made in the penultimate sentence. Prove the following statement:

> If $E \subseteq \mathbb{R}$ is countable, i.e., there exists a surjection (onto function) $\phi : \mathbb{N} \to E$, then E has measure zero.

Hint: Surround each point $x_k \in E$ by an interval whose length is $\epsilon/2^{k+1}$.

Using this notion of small sets we can state Carleson's theorem in the context of Riemann integrable functions; the general result is formulated using the Lebesgue integral [?]. Here it is:

Theorem 8.38 (Carleson 1966)**.** *For any $f \in R(\mathbb{T})$, there exists a set E of measure zero (which is possibly empty) such that*

$$f(x) = \sum_{n \in \mathbb{Z}} \hat{f}(n) e^{inx} = \lim_{N \to \infty} S_N(x)$$

for all $x \notin E$.

8.5.3 The Gibb's phenomenon

In this final subsection, we will study the Gibb's phenomenon. The Gibb's phenomenon, named after J. Willard Gibb's (yes, the free energy Gibbs), describes the pointwise convergence of Fourier series of a function with a jump discontinuity. Instead of working in the general setting, we will study the Gibb's phenomenon as it occurs when we consider the Fourier series of the sawtooth function. Focusing on this specific case will allow us to very precisely see what's going on. If you are worried about the general case, I'll refer you to a very nice discussion by T. W. Körner in which he describes how to extend the results pertaining to this example to the general class of piecewise differentiable functions in $R(\mathbb{T})$ [?].

So let's return to our favorite example, the sawtooth function f, defined by

$$f(x) = x$$

for all $-\pi < x \leq \pi$ and extended periodically to $\mathbb{R}$. We recall that

$$f(x) \sim \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{(-1)^n}{n} e^{inx} = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(nx) \tag{8.2}$$

Let's again consider the graph of the Fourier polynomial S_{40}; this is illustrated in Figure 8.6.

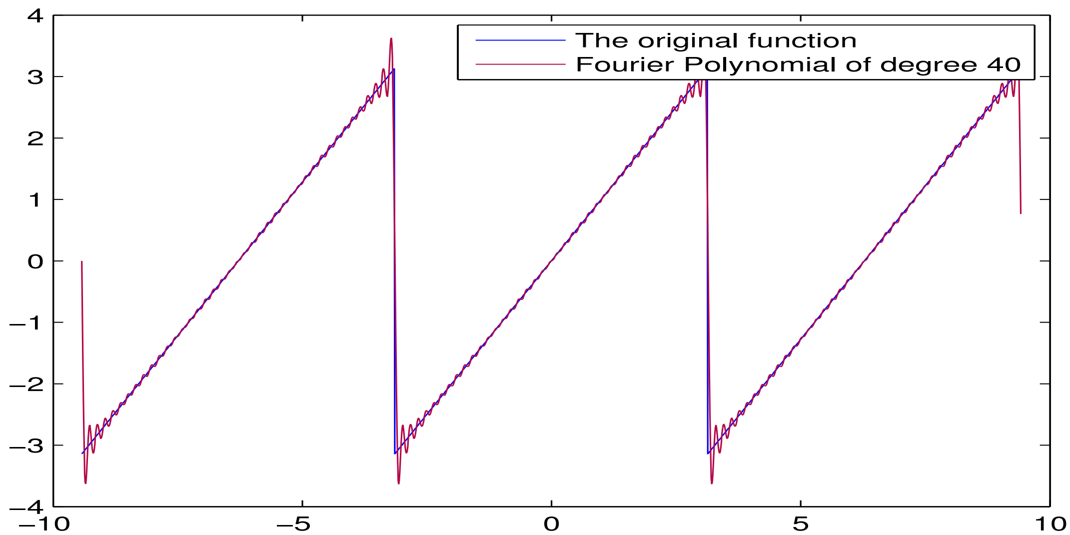

Figure 8.6: f and S_{40}

As we discussed in the last subsection, Theorem 8.36 guarantees that

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

for all $x \neq m\pi$ where $m \in \mathbb{Z}$ is odd. We also showed that, at any $x = m\pi$ where $m \in \mathbb{Z}$ is odd, $S_N(x) \to 0$ as $N \to \infty$. These two things should appear to be somewhat clear by looking at Figure 8.6. There is however one thing that should bother you: near (but not at) the breakpoints, $\{\dots, -3\pi, -\pi, \pi, 3\pi, \dots\}$, the graph of S_{40} seems to "overshoot" (or "undershoot") the graph of f. These are the spikes you see close to the discontinuity in f. This behavior is called the *Gibb's phenomenon*. Let's study this behavior more closely at, say, $x = \pi$; Figure 8.7 shows the graphs of $S_N(x)$ for $N = 25, 26, \dots, 50$.

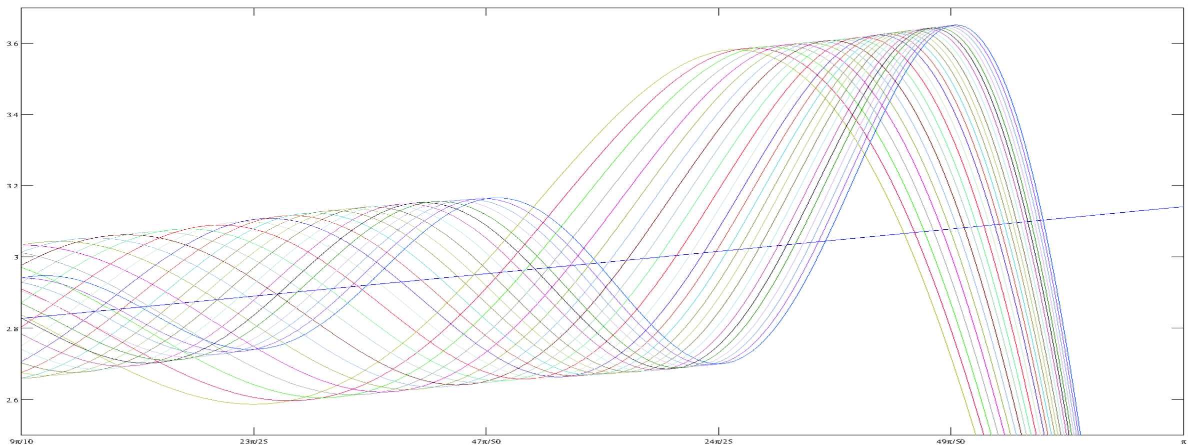

Figure 8.7: The graphs of $S_n(x)$ for $n = 25, 26, \dots 50$ and $f(x)$ for $9\pi/10 \leq x \leq \pi$.

Upon studying Figure 8.7 closely, we see that the overshoot is moving right as N increases. You might say: Theorem 8.36 guarantees that $S_N(x) \to f(x) = x$ for all $-\pi < x < \pi$ but, upon looking at the figure, $S_N(x)$ isn't converging to $f(x)$ for x very close to π. So where did we go wrong? The answer is that we haven't gone wrong at all, the apparent discrepancy can be understood by recognizing that pointwise convergence is weaker than convergence in the graph–this is the difference between pointwise convergence and uniform convergence. Remember, that for pointwise convergence, we first select x and ϵ and find a natural number $M = M(\epsilon, x)$ for which

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

for all $N \geq M$. In the case at hand, we can understand this notion as follows: If I select an $x < \pi$, but as close to π as I want, since the overshoot in the Fourier polynomials are moving to the right, I simply have to wait until they have moved so far right that they've passed x–this will determine N. After this, the Fourier polynomials evaluated at x will get much much closer to $f(x)$. Okay, so now you understand how we still get pointwise convergence. Let's now try to understand the overshoot.

Using the same numbers I've used to make the graphs in Figure 8.7, I can quantify this overshoot. For each $N \in \mathbb{N}$, denote by

$$M_N = \max_{9\pi/10 \leq x \leq \pi} S_N(x),$$

the maximum of the function $S_N(x)$ near π. We also denote by x_N the unique x near π for which

$$f(x_N) = M_N.$$

The following table shows M_N, M_N/π, x_N and $\pi - \pi/N$ to four decimal places for $n = 25, 30, \dots, 50$.

N	M_N	M_N/π	x_N	$\pi - \pi/N$
25	3.5822	1.1403	3.0204	3.0159
30	3.6020	1.1465	3.0404	3.0369
35	3.6172	1.1514	3.0544	3.0518
40	3.6321	1.1561	3.0654	3.0631
45	3.6433	1.1597	3.0734	3.0718
50	3.6516	1.1624	3.0804	3.0788

Upon looking at the table, we see that, as N increases the x_N's are close to $\pi - \pi/N$ and the ratio M_N/π grows toward $1.17\dots$. So, by following the x at which $S_N(x)$ is maximized, the ratio M_N/π approaches some number A as $N \to \infty$ (note that π is half the gap of the discontinuity of f at π); this describes the overshoot. The following theorem formalizes it:

Theorem 8.39. *Let f, S_N be as above. Then*

$$\lim_{N\to\infty} S_N(\pi - \pi/N) = \pi A$$

where

$$A = 1.178979744447216727\dots$$

Thus, the $S_N(\pi-\pi/N)/\pi$ converges to A (called the Gibb's constant) times half of the gap of the jump discontinuity.

Proof. Using our trigonometric identities, we find that

$$\begin{aligned}
S_N(\pi - \pi/N) &= \sum_{n=1}^{N} \frac{2(-1)^{n+1}}{n} \sin(n\pi - n\pi/N) \\
&= \sum_{n=1}^{N} \frac{2(-1)^{n+1}}{n} (\sin(n\pi)\cos(n\pi/N) - \sin(n\pi/N)\cos(n\pi) \\
&= \sum_{n=1}^{N} \frac{2(-1)^{n+1}}{n} (0 - \sin(n\pi/N)\cos(n\pi) \\
&= \sum_{n=1}^{N} \frac{2(-1)^{n+1}}{n} ((-1)^{n+1}\sin(n\pi/N) \\
&= 2\sum_{n=1}^{N} \frac{\sin n\pi/N}{n\pi/N}\frac{\pi}{N}.
\end{aligned}$$

You should recognize that

$$\sum_{n=1}^{N} \frac{\sin n\pi/N}{n\pi/N}\frac{\pi}{N}$$

is a (right) Riemann sum for the integral

$$\int_0^\pi \frac{\sin(x)}{x}\,dx$$

and because $\sin x/x$ is Riemann integrable on $[0,\pi]$, we immediately conclude that

$$\lim_{N\to\infty} S_N(\pi - \pi/N) = \lim_{N\to\infty} 2\sum_{n=1}^{N} \frac{\sin n\pi/N}{n\pi/N}\frac{\pi}{N} = 2\int_0^\pi \frac{\sin x}{x}\,dx = \pi A$$

where

$$A = \frac{2}{\pi}\int_0^\pi \frac{\sin x}{x}\,dx.$$

It remains to compute A. Using the power series representation for $\sin x$ about 0, we have

$$\frac{\sin x}{x} = \frac{1}{x}\left(x - \frac{x^3}{3!} + \frac{x^5}{5!}\cdots\right) = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} + \cdots$$

for $x \in \mathbb{R}$ where we take the left hand side to be 1 when $x = 0$. Using the results of Exercise 7, it is easily verified that the above series converges absolutely and uniformly on $[0,\pi]$. By an application of Corollary 7.10, we conclude that

$$\begin{aligned} A &= \frac{2}{\pi}\int_0^\pi (1-\frac{x^2}{3!}+\frac{x^4}{5!}+\cdots)\,dx = \frac{2}{\pi}\left(\int_0^\pi 1\,dx - \frac{1}{3!}\int_0^\pi x^2\,dx + \frac{1}{5!}\int_0^\pi x^4\,dx\right)\\ &= \frac{2}{\pi}\left(\pi - \frac{\pi^3}{3\cdot 3!}+\frac{\pi^5}{5\cdot 5!}+\cdots\right) = 2-\frac{2\pi^2}{3\cdot 3!}+\frac{2\pi^4}{5\cdot 5!}\cdots\\ &= 1.17897974447216727\dots] \end{aligned}$$

as desired. $\square$

Exercise 8.6: T

here is (at least) one function f in Exercise 16 that has a jump discontinuity in the interval $[-\pi,\pi]$. For this example, plot (in any computing program you want) some Fourier polynomials of f. Print out the results and comment on the appearance (or lack thereof) of the Gibb's phenomenon. If you'd like a source of the Matlab file I used to analyze the sawtooth function, let me know and I'll email it to you.

8.6 The Fourier Transform

Bibliography

[1] T. Apostol. Mathematical Analysis. Pearson; 2nd edition. (1974)

[2] R. Bartle. The Elements of Integration and Lebesgue Measure. (1966)

[3] G. H. Hardy ”Weierstrass’s nondifferentiable function”. *Transactions of the American Mathematical Society.* **17**: 301–325, 1926.

[4] T. Körner. Fourier Analysis.

[5] Rudin, Walter. Principles of Mathematical Analysis. McGraw-Hill Inc. New York. (1976)

[6] Strichartz, Robert S. The way of analysis. Jones & Bartlett Learning. (2000)

[7] W. R. Wade. An Introduction to Analysis. Prentice Hall; 2nd edition. (1999)