MATH 131B, ANALYSIS 2, WINTER 2015

STEVEN HEILMAN

Abstract. These notes are mostly copied from those of T. Tao from 2003, available here

Contents

Date: March 20, 2015.

1. Metric Spaces, Topology, Continuity, Compactness

In this course, we continue our study of analysis from your previous analysis course. Recall that you took an abstract approach to linear algebra by discussing linear transformations and vector spaces. Similarly, in this course, we will begin by presenting an abstract approach to analysis, which generalizes the things we did for analysis on the real line. Specifically, we will first generalize a lot of the arguments from the real line to the setting of metric spaces. We will then apply this general theory in our discussion of analysis on Euclidean spaces of any dimension, power series, and trigonometric functions. We will then discuss Fourier analysis, and the various intricacies of differentiation in Euclidean spaces.

1.1. Metric Spaces.

Recall that a sequence of real numbers $(x_n)_{n=0}^{\infty}$ converges to a real number x if and only if, for every real $\varepsilon > 0$, there exists an integer $N = N(\varepsilon)$ such that, for all $n > N$, we have $|x_n - x| < \varepsilon$. That is, eventually, the sequence $(x_n)_{n=0}^{\infty}$ is within a distance ε of x. And this is true for any $\varepsilon > 0$. So, whenever we have a space of points, and we can define some notion of distance between two points, then we should be able to make a similar definition of convergence of sequences. We are therefore led to consider the following question. What are the crucial properties of the distance function $d\colon \mathbb{R} \times \mathbb{R} \to [0, \infty)$ where $d(x, y) := |x - y|$ that allow us to consider convergence of sequences? The following properties suffice, as we shall see further below.

Definition 1.1 (**Metric Space**)**.** A **metric space** (X, d) is a set X together with a function $d\colon X \times X \to [0, \infty)$ which satisfies the following properties.

- For all $x \in X$, we have $d(x, x) = 0$.
- For all $x, y \in X$ with $x \neq y$, we have $d(x, y) > 0$. (Positivity)
- For all $x, y \in X$, we have $d(x, y) = d(y, x)$. (Symmetry)
- For all $x, y, z \in X$, we have $d(x, z) \leq d(x, y) + d(y, z)$. (Triangle inequality)

Example 1.2 (**The real line**)**.** As mentioned above, define the function $d\colon \mathbb{R}\times\mathbb{R} \to [0, \infty)$ by $d(x, y) := |x - y|$, where $x, y \in \mathbb{R}$. Then $(\mathbb{R}, d)$ is a metric space.

Example 1.3 (**Euclidean space**)**.** Let n be a positive integer. Define $\mathbb{R}^n$ by

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) \colon x_i \in \mathbb{R}\ \forall\, i \in \{1, \ldots, n\}\}.$$

Define the **Euclidean metric** (or ℓ_2 metric) $d\colon \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$ by

$$d_{\ell_2}((x_1, \ldots, x_n), (y_1, \ldots, y_n)) := \left(\sum_{i=1}^{n} (x_i - y_i)^2\right)^{1/2}.$$

Then $(\mathbb{R}^n, d_{\ell_2})$ is a metric space.

There are actually many interesting metrics to consider on $\mathbb{R}^n$. Here is another one.

Example 1.4. Let n be a positive integer. Define the ℓ_1 metric $d_{\ell_1}\colon \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$ by

$$d_{\ell_1}((x_1, \ldots, x_n), (y_1, \ldots, y_n)) := \sum_{i=1}^{n} |x_i - y_i|.$$

Then $(\mathbb{R}^n, d_{\ell_1})$ is a metric space.

The last two examples actually satisfy a few additional properties which you may recall from your linear algebra class.

Definition 1.5 (**Normed linear space**)**.** Let X be a vector space over $\mathbb{R}$. A **normed linear space** $(X, \|\cdot\|)$ is a vector space X over $\mathbb{R}$ together with a norm function $\|\cdot\|: X \to [0,\infty)$ which satisfies the following properties.

- $\|0\| = 0$.
- For all $x \in X$ with $x \neq 0$, we have $\|x\| > 0$. (Positivity)
- For all $x \in X$ and for all $\alpha \in \mathbb{R}$, we $\|\alpha x\| = |\alpha| \, \|x\|$. (Homogeneity)
- For all $x, y \in X$, we have $\|x+y\| \leq \|x\| + \|y\|$. (Triangle inequality)

Exercise 1.6. Let $(X, \|\cdot\|)$ be a normed linear space. Define $d\colon X \times X \to \mathbb{R}$ by $d(x,y) := \|x-y\|$. Show that (X,d) is a metric space.

Example 1.7. Let n be a positive integer. Define the ℓ_2 norm on $\mathbb{R}^n$ by

$$\|(x_1,\ldots,x_n)\|_{\ell_2} := \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}.$$

Then $(\mathbb{R}^n, \|\cdot\|_{\ell_2})$ is a normed linear space. From Exercise 1.6, $(\mathbb{R}^n, d_{\ell_2})$ is a metric space, which we saw in Example 1.3.

Example 1.8. Let n be a positive integer. Define the ℓ_1 norm on $\mathbb{R}^n$ by

$$\|(x_1,\ldots,x_n)\|_{\ell_1} := \sum_{i=1}^{n} |x_i|.$$

Then $(\mathbb{R}^n, \|\cdot\|_{\ell_1})$ is a normed linear space. From Exercise 1.6, $(\mathbb{R}^n, d_{\ell_1})$ is a metric space, which we saw in Example 1.4.

Example 1.9. Let n be a positive integer. Define the ℓ_∞ norm on $\mathbb{R}^n$ by

$$\|(x_1,\ldots,x_n)\|_{\ell_\infty} := \max_{i=1,\ldots,n} |x_i|.$$

Then $(\mathbb{R}^n, \|\cdot\|_{\ell_\infty})$ is a normed linear space. From Exercise 1.6, $(\mathbb{R}^n, d_{\ell_\infty})$ is a metric space, where $d_{\ell_\infty}((x_1,\ldots,x_n),(y_1,\ldots,y_n)) := \max_{i=1}^{n} |x_i - y_i|$.

Example 1.10. Let n be a positive integer and let $1 \leq p < \infty$ be a real number. Define the ℓ_p norm on $\mathbb{R}^n$ by

$$\|(x_1,\ldots,x_n)\|_{\ell_p} := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

Then $(\mathbb{R}^n, \|\cdot\|_{\ell_p})$ is a normed linear space, though the triangle inequality is a bit more difficult to prove. From Exercise 1.6, $(\mathbb{R}^n, d_{\ell_p})$ is a metric space.

Exercise 1.11. Let n be a positive integer and let $x \in \mathbb{R}^n$. Show that $\|x\|_{\ell_\infty} = \lim_{p\to\infty} \|x\|_{\ell_p}$.

Euclidean space is actually even more special than a normed linear space, which we also learned in linear algebra class. Specifically, $\mathbb{R}^n$ is an inner product space.

Definition 1.12 (**Real Inner product space**)**.** Let X be a vector space over $\mathbb{R}$. A **real inner product space** $(X, \langle\cdot,\cdot\rangle)$ is a vector space X over $\mathbb{R}$ together with an inner product function $\langle\cdot,\cdot\rangle\colon X \times X \to \mathbb{R}$ which satisfies the following properties.

- $\langle 0, 0\rangle = 0$.
- For all $x \in X$ with $x \neq 0$, we have $\langle x, x\rangle > 0$.
- For all $x, y \in X$, we have $\langle x, y\rangle = \langle y, x\rangle$. (Symmetry)
- For all $x \in X$ and for all $\alpha \in \mathbb{R}$, we $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$. (Homogeneity)
- For all $x, y, z \in X$, we have $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$. (Linearity)

Exercise 1.13. Let $(X, \langle\cdot,\cdot\rangle)$ be a real inner product space. Define $\|\cdot\| : X \to [0, \infty)$ by $\|x\| := \sqrt{\langle x, x\rangle}$. Show that $(X, \|\cdot\|)$ is a normed linear space. Consequently, from Exercise 1.6, if we define $d\colon X \times X \to [0, \infty)$ by $d(x, y) := \sqrt{\langle (x - y), (x - y)\rangle}$, then (X, d) is a metric space.

In order to prove Exercise 1.13, the following inequality is useful.

Theorem 1.14 (Cauchy-Schwarz Inequality). *Let $(X, \langle\cdot,\cdot\rangle)$ be a real inner product space. Then, for all $x, y \in X$, we have*

$$|\langle x, y\rangle| \leq \langle x, x\rangle^{1/2}\langle y, y\rangle^{1/2}.$$

Proof. It follows from Definition 1.12 that, if $x = 0$, then $\langle x, y\rangle = 0$ for any $y \in X$. (You should have proven this in your linear algebra class.) So, if $x = 0$, then both sides of the Cauchy-Schwarz inequality are zero, and the inequality therefore holds. Similarly, if $y = 0$, then both sides of the inequality are zero. We therefore assume that $x \neq 0$ and $y \neq 0$. For any $x \in X$, define $\|x\| := \sqrt{\langle x, x\rangle}$.

Starting from x, we subtract the projection of x onto y. Define $\delta := -\langle x, y\rangle / \|y\|^2$. We then have

$$0 \leq \|x + \delta y\|^2 = \|x\|^2 + 2\delta\langle x, y\rangle + |\delta|^2 \|y\|^2 = \|x\|^2 - |\langle x, y\rangle|^2 / \|y\|^2 .$$

□

Remark 1.15. For any $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in \mathbb{R}^n$, define

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle := \sum_{i=1}^{n} x_i y_i.$$

Then $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$ is an inner product space. Note that

$$\|(x_1, \ldots, x_n)\|_{\ell_2} = \langle (x_1, \ldots, x_n), (x_1, \ldots, x_n)\rangle^{1/2}.$$

However, there does not exist an inner product $\langle\cdot,\cdot\rangle'$ on $\mathbb{R}^n$ such that $\|(x_1, \ldots, x_n)\|_{\ell_1}$ is equal to $(\langle (x_1, \ldots, x_n), (x_1, \ldots, x_n)\rangle')^{1/2}$. The last statement is more difficult to prove.

In the first part of this course, we will mostly focus on metric spaces. Once we deal with Fourier series, the subject of inner product spaces will reappear. However, we will need to deal with complex inner product spaces, so we now recall their definition. Recall that, for $\alpha, \beta \in \mathbb{R}$ we define

$$\overline{\alpha + \beta\sqrt{-1}} := \alpha - \beta\sqrt{-1}.$$

Definition 1.16 (Complex Inner product space). Let X be a vector space over $\mathbb{C}$. A **complex inner product space** $(X, \langle\cdot,\cdot\rangle)$ is a vector space X together with an inner product function $\langle\cdot,\cdot\rangle\colon X \times X \to \mathbb{C}$ which satisfies the following properties.

- $\langle 0, 0\rangle = 0$.

- For all $x \in X$ with $x \neq 0$, we have $\langle x, x\rangle > 0$.
- For all $x, y \in X$, we have $\langle x, y\rangle = \overline{\langle y, x\rangle}$. (Conjugate Symmetry)
- For all $x \in X$ and for all $\alpha \in \mathbb{C}$, we $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$. (Homogeneity)
- For all $x, y, z \in X$, we have $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$. (Linearity)

Remark 1.17. Let $(X, \langle\cdot,\cdot\rangle)$ be a complex inner product space. Then, for any $x, y, z \in X$ and for any $\alpha \in \mathbb{C}$, it follows from Definition 1.16 that

- $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.
- $\langle x, \alpha y\rangle = \overline{\alpha}\langle x, y\rangle$.

Example 1.18. Let n be a positive integer. Let $(z_1, \ldots, z_n), (w_1, \ldots, x_n) \in \mathbb{C}^n$. The standard inner product on $\mathbb{C}^n$ is defined by

$$\langle (z_1, \ldots, z_n), (w_1, \ldots, w_n)\rangle := \sum_{i=1}^{n} z_i \overline{w_i}.$$

Then $(\mathbb{C}^n, \langle\cdot,\cdot\rangle)$ is a complex inner product space.

Exercise 1.19 (Cauchy-Schwarz Inequality). Let $(X, \langle\cdot,\cdot\rangle)$ be a complex inner product space. Let $x, y \in X$. Modify the proof of Theorem 1.14 to prove the Cauchy-Schwarz inequality for complex inner product spaces:

$$|\langle x, y\rangle| \leq \langle x, x\rangle^{1/2}\langle y, y\rangle^{1/2}.$$

Exercise 1.20. Let $(X, \langle\cdot,\cdot\rangle)$ be a complex inner product space. Let $x, y \in X$. As usual, let $\|x\| := \sqrt{\langle x, x\rangle}$. Prove **Pythagoras's theorem**: if $\langle x, y\rangle = 0$, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$

1.1.1. *Convergence of Sequences.* We now define convergence of sequences in a metric space in a way which imitates the convergence of sequences of real numbers.

Definition 1.21. Let (X, d) be a metric space. Let $(x^{(j)})_{j=k}^{\infty}$ be a sequence of elements of X. Let x be an element of X. We say that the sequence $(x^{(j)})_{j=k}^{\infty}$ **converges to** x **with respect to the metric** d if and only if, for every $\varepsilon > 0$, there exists an integer $J = J(\varepsilon)$ such that, for all $j > J$, we have $d(x^{(j)}, x) < \varepsilon$.

Proposition 1.22. *Let n be a positive integer. Let $x \in \mathbb{R}^n$. Let $(x^{(j)})_{j=k}^{\infty}$ be a sequence of elements of $\mathbb{R}^n$. We write $x^{(j)} = (x_1^{(j)}, \ldots, x_n^{(j)})$, so that for each $1 \leq i \leq n$, we have $x_i^{(j)} \in \mathbb{R}$, that is, $x_i^{(j)}$ is the i^{th} coordinate of $x^{(j)}$. Then the following three statements are equivalent.*

- *$(x^{(j)})_{j=k}^{\infty}$ converges to x with respect to d_{ℓ_1}.*
- *$(x^{(j)})_{j=k}^{\infty}$ converges to x with respect to d_{ℓ_2}.*
- *$(x^{(j)})_{j=k}^{\infty}$ converges to x with respect to d_{ℓ_∞}.*

Exercise 1.23. Prove Proposition 1.22.

Due to Proposition 1.22, we say that the metrics d_{ℓ_1}, d_{ℓ_2} and d_{ℓ_∞} are equivalent on $\mathbb{R}^n$. In fact, for any $p, p' \in \mathbb{R}$ with $1 \leq p, p' \leq \infty$, the metrics d_{ℓ_p} and $d_{\ell_{p'}}$ are equivalent on $\mathbb{R}^n$. In fact, something stronger is true. Let $\|\cdot\|_a, \|\cdot\|_b$ be any two norms on $\mathbb{R}^n$. Define the metrics d_a, d_b so that, for any $x, y \in \mathbb{R}^n$, we have $d_a(x, y) := \|x - y\|_a$ and $d_b(x, y) := \|x - y\|_b$. Then d_a and d_b are equivalent on $\mathbb{R}^n$.

As in the case of the real line, a sequence cannot converge to two distinct points.

Proposition 1.24. *Let (X, d) be a metric space. Let $(x^{(j)})_{j=k}^{\infty}$ be a sequence of elements of X. Let x, x' be elements of X. Assume that the sequence $(x^{(j)})_{j=k}^{\infty}$ converges to x with respect to d. Assume also that the sequence $(x^{(j)})_{j=k}^{\infty}$ converges to x' with respect to d. Then $x = x'$.*

Exercise 1.25. Prove Proposition 1.24.

Due to Proposition 1.24, if $(x^{(j)})_{j=k}^{\infty}$ is a sequence of elements of X which converges to x with respect to d, we then write $\lim_{j\to\infty} x^{(j)} = x$. Although the latter notation has no indication of the metric d, confusion should often not arise as to what metric the convergence is using.

1.2. **Topology of Metric Spaces.** The open intervals such as $(0, 1)$ and closed intervals such as $[1, 2]$ played a central role in analysis on the real line. The open interval is a special case of an open set, and the closed interval is a special case of a closed set. There is a way to generalize the notions of open and closed set to general metric spaces, so we pursue these notions now. We begin by generalizing the notion of an open interval to the notion of a metric ball. The language of topology is used everywhere throughout mathematics, so it is quite useful even just to learn the terminology.

Definition 1.26 (Metric Ball). Let (X, d) be a metric space, let x_0 be a point in X, and let $r > 0$ be a positive real number. We define the **ball** $B_{(X,d)}(x_0, r)$ in X, centered at x_0 with radius r to be the set

$$B_{(X,d)}(x_0, r) := \{x \in X \colon d(x, x_0) < r\}.$$

When the metric space (X, d) is apparent, we abbreviate $B_{(X,d)}(x_0, r)$ as $B(x_0, r)$. Also, if $(X, \|\cdot\|)$ is a normed linear space, we write $B_{(X,\|\cdot\|)}(x_0, r)$ to denote the set $\{x \in X \colon \|x - x_0\| < r\}$.

Example 1.27. Let $x, y \in \mathbb{R}$. In $\mathbb{R}$ with the metric $d(x, y) := |x - y|$, note that $B_{(\mathbb{R},d)}(x_0, r)$ is the open interval $(x_0 - r, x_0 + r)$.

Example 1.28. In $\mathbb{R}^2$ with the metric d_{ℓ_2}, we have

$$B_{(\mathbb{R}^2,d_{\ell_2})}((0, 0), 1) = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 < 1\}.$$

However, with the metric d_{ℓ_1}, we have

$$B_{(\mathbb{R}^2,d_{\ell_1})}((0, 0), 1) = \{(x, y) \in \mathbb{R}^2 \colon |x| + |y| < 1\}.$$

Also, with the metric d_{ℓ_∞}, we have

$$B_{(\mathbb{R}^2,d_{\ell_\infty})}((0, 0), 1) = \{(x, y) \in \mathbb{R}^2 \colon \max\{|x|, |y|\} < 1\}.$$

So, $B_{(\mathbb{R}^2,d_{\ell_1})}((0, 0), 1)$ is a diamond, $B_{(\mathbb{R}^2,d_{\ell_2})}((0, 0), 1)$ is a disc, and $B_{(\mathbb{R}^2,d_{\ell_\infty})}((0, 0), 1)$ is a square. Moreover,

$$B_{(\mathbb{R}^2,d_{\ell_1})}((0, 0), 1) \subseteq B_{(\mathbb{R}^2,d_{\ell_2})}((0, 0), 1) \subseteq B_{(\mathbb{R}^2,d_{\ell_\infty})}((0, 0), 1).$$

Remark 1.29. Note that if (X, d) is a nonempty metric space, and if $x_0 \in X$ with $r > 0$, then $B_{(X,d)}(x_0, r)$ is nonempty, since it contains x_0. Moreover, if $0 < r < r'$, we have the containment $B_{(X,d)}(x_0, r) \subseteq B_{(X,d)}(x_0, r')$.

Definition 1.30. Let (X, d) be a metric space, let E be a subset of X, and let x_0 be a point in X. We say that x_0 is an **interior point** of E if and only if there exists $r > 0$ such that $B(x_0, r) \subseteq E$. We say that x_0 is an **exterior point** of E if and only if there exists $r > 0$ such that $B(x_0, r) \cap E = \emptyset$. We say that x_0 is a **boundary point** of E if and only if x_0 is neither an interior point nor an exterior point of E.

Remark 1.31. The set of all interior points of E is called the **interior** of E, and it is denoted as $\text{int}(E)$. The set of all exterior points of E is called the **exterior** of E, and it is denoted as $\text{ext}(E)$. The set of all boundary points of E is called the **boundary** of E, and it is denoted as ∂E.

Remark 1.32. If x_0 is an interior point of E, then x_0 is an element of E. If x_0 is an exterior point of E, then x_0 is not an element of E. If x_0 is a boundary point of E, then x_0 may or may not be an element of E.

Example 1.33. Consider the real line $\mathbb{R}$ with the usual metric. The open interval $(0, 1)$ has interior $(0, 1)$, it has exterior $(-\infty, 0) \cup (1, \infty)$, and it has boundary $\{0, 1\}$. The closed interval $[0, 1]$ has interior $(0, 1)$, it has exterior $(-\infty, 0) \cup (1, \infty)$, and it has boundary $\{0, 1\}$. The half-open interval $(0, 1]$ has interior $(0, 1)$, it has exterior $(-\infty, 0) \cup (1, \infty)$, and it has boundary $\{0, 1\}$.

Definition 1.34. Let (X, d) be a metric space, let E be a subset of X, and let x_0 be a point in X. We say that x_0 is an **adherent point** of E if and only if for every real $r > 0$, we have $B(x_0, r) \cap E \neq \emptyset$. The set of all adherent points of E is called the **closure** of E and is denoted by $\overline{E}$.

The definitions of interior, exterior, boundary and closure are related by the following proposition.

Proposition 1.35. *Let (X, d) be a metric space, let E be a subset of X, and let x_0 be a point in X. Then the following statements are equivalent.*

- *x_0 is an adherent point of E.*
- *x_0 is either an interior point of E or a boundary point of E.*
- *There exists a sequence $(x^{(j)})_{j=1}^{\infty}$ of elements of E which converges to x_0 with respect to the metric d.*

Exercise 1.36. Prove Proposition 1.35.

We now define open and closed sets in terms of the boundary points of a set. As we will see below, we can equivalently define open and closed sets using only open balls.

Definition 1.37 (Open and Closed Sets). Let (X, d) be a metric space, let E be a subset of X. We say that E is **closed** if and only if E contains all of its boundary points, i.e. when $\partial E \subseteq E$. We say that E is **open** if and only if E contains none of its boundary points, i.e. when $(\partial E) \cap E = \emptyset$. If E contains some of its boundary points but not others, then E is neither open nor closed.

Remark 1.38. If a set E has no boundary, then E is simultaneously open and closed. For example, if $E = X$, then E is both open and closed. Also, if $E = \emptyset$, then E is both open and closed.

Example 1.39. We continue Example 1.33. Consider the real line $\mathbb{R}$ with the usual metric. The open interval $(0,1)$ has boundary $\{0,1\}$, so the open interval is open. The closed interval $[0,1]$ has boundary $\{0,1\}$, so the closed interval is closed. The half-open interval $(0,1]$ has boundary $\{0,1\}$, so the half-open interval is neither open nor closed.

As promised, we now show some equivalent definitions of open and closed sets.

Proposition 1.40 (Properties of Open and Closed Sets). *Let (X,d) be a metric space.*

(i) *Let E be a subset of X. Then E is open if and only if $E = \mathrm{int}(E)$. That is, E is open if and only if, for every $x \in E$, there exists an $r > 0$ such that $B(x,r) \subseteq E$.*
(ii) *Let E be a subset of X. Then E is closed if and only if E contains all of its adherent points, i.e. when $E = \overline{E}$. That is, E is closed if and only if, for every convergent sequence $(x^{(j)})_{j=0}^{\infty}$ consisting of elements of E, the limit $\lim_{n\to\infty} x_n$ of the sequence also lies in E.*
(iii) *For any $x_0 \in X$, for any $r > 0$, the open ball $B(x_0,r)$ is an open set. The set $\{x \in X : d(x,x_0) \leq r\}$ is a closed set. The latter set is sometimes called the **closed ball** of radius r centered at x_0.*
(iv) *Let $x_0 \in X$. Then the singleton set $\{x_0\}$ is closed.*
(v) *If E is a subset of X, then E is open if and only if $X \smallsetminus E$ is closed. Here we have denoted $X \smallsetminus E := \{x \in X : x \notin E\}$ as the complement of E in X.*
(vi) *If $E_1, \ldots, E_n$ is a finite collection of open sets, then $E_1 \cap \cdots \cap E_n$ is an open set. If $F_1, \ldots, F_n$ is a finite collection of closed sets, then $F_1 \cup \cdots \cup F_n$ is a closed set.*
(vii) *If $\{E_\alpha\}_{\alpha \in I}$ is collection of open sets, (where the index set I can be finite, countable, or uncountable), then $\cup_{\alpha \in I} E_\alpha$ is an open set. If $\{F_\alpha\}_{\alpha \in I}$ is collection of closed sets, (where the index set I can be finite, countable, or uncountable), then $\cap_{\alpha \in I} F_\alpha$ is a closed set.*
(viii) *If E is any subset of X, then $\mathrm{int}(E)$ is the largest open set contained in E. That is, $\mathrm{int}(E)$ is open, and if V is any open set such that $V \subseteq E$, then $V \subseteq \mathrm{int}(E)$ also. Similarly, $\overline{E}$ is the smallest closed set containing E. That is, $\overline{E}$ is closed, and if V is any closed set such that $V \supseteq E$, then $V \supseteq \overline{E}$ also.*

Exercise 1.41. Prove Proposition 1.40.

Remark 1.42. Proposition 1.40(vi) does not hold for countable collections of sets, as we can see from the following examples which involve open and closed intervals on the real line.

$$\bigcap_{n \in \mathbb{N}} \left(-\frac{1}{n+1}, 1 + \frac{1}{n+1} \right) = [0,1].$$

$$\bigcup_{n \in \mathbb{N}} \left[\frac{1}{n+1}, 1 - \frac{1}{n+1} \right] = (0,1).$$

As we see from the following example, it is natural to consider the open and closed sets of a subset of a metric space. Such notions are formalized by the relative topology.

Example 1.43. Consider $\mathbb{R}^2$, and define $Y := \{(x,0) : x \in \mathbb{R}\}$, so that Y is a subset of $\mathbb{R}^2$. Let d_{ℓ_2} denote the ℓ_2 metric on $\mathbb{R}^2$. If we restrict d_{ℓ_2} to Y resulting in $d_{\ell_2}|_{Y \times Y}$, then $(Y, d_{\ell_2}|_{Y \times Y})$ is a metric space. In fact, we can identify $(Y, d_{\ell_2}|_{Y \times Y})$ with the real line $\mathbb{R}$ with its usual metric. Now, consider the set

$$E := \{(x, 0) \colon -1 < x < 1\}.$$

Then E is a subset of Y, and E is also a subset of $\mathbb{R}^2$. When we consider E as a subset of Y, then E is an open set, since E is equal to the ball $B_{(Y, d_{\ell_2}|_{Y\times Y})}((0,0),1)$. However, when we consider E as a subset of $\mathbb{R}^2$, then E is no longer an open set. To see this, note that for any $r > 0$, the ball $B_{(\mathbb{R}^2, d_{\ell_2})}((0,0),r)$ is not contained in E. So, by Proposition 1.40(i), E is not open in $(\mathbb{R}^2, d_{\ell_2})$.

To summarize the above example: there is a sensible way to discuss open sets of a subset of a metric space, and it involves restricting the metric.

Definition 1.44 (**Relative Topology**). Let (X, d) be a metric space, let Y be a subset of X, and let E be a subset of Y. We say that E is **relatively open with respect to** Y if and only if E is open in the metric space $(Y, d|_{Y\times Y})$. Similarly, we say that E is **relatively closed with respect to** Y if and only if E is closed in the metric space $(Y, d|_{Y\times Y})$.

The definitions of relatively open and relatively closed sets are consistent with set intersection in the following way.

Proposition 1.45. *Let (X, d) be a metric space, let Y be a subset of X, and let E be a subset of Y.*

- *E is relatively open with respect to Y if and only if $E = V \cap Y$ for some set $V \subseteq X$ which is open in X.*
- *E is relatively closed with respect to Y if and only if $E = V \cap Y$ for some set $V \subseteq X$ which is closed in X.*

Proof. We only prove the first assertion, since the second assertion is proven similarly. First, assume that E is relatively open with respect to Y. Then E is open in the metric space $(Y, d|_{Y\times Y})$. So, by Proposition 1.40(i), for any $e \in E$ there exists $r = r(e) > 0$ such that $B_{(Y, d|_{Y\times Y})}(e, r(e)) \subseteq E$. Note that

$$E = \bigcup_{e\in E} B_{(Y, d|_{Y\times Y})}(e, r(e)). \tag{1}$$

Specifically, every set on the right is contained in E, so the union of the sets on the right is contained in E. And conversely, every element of E appears on the right side, so the right side contains the left side. We therefore deduce the equality (1). With this equality in mind, define

$$V := \bigcup_{e\in E} B_{(X,d)}(e, r(e)). \tag{2}$$

From Proposition 1.40(i), V is open in (X, d). For any $e \in E$, we have

$$Y \cap B_{(X,d)}(e, r(e)) = B_{(Y, d|_{Y\times Y})}(e, r(e)). \tag{3}$$

To see this, note that if $y \in B_{(Y, d|_{Y\times Y})}(e, r(e))$, then $d|_{Y\times Y}(e, y) < r(e)$. Since e and y are in Y, we have $d|_{Y\times Y}(e, y) = d(e, y)$. So, $d(e, y) < r(e)$, so that $y \in B_{(X,d)}(e, r(e))$. Conversely, let $y \in Y \cap B_{(X,d)}(e, r(e))$. then $d(e, y) < r(e)$. Since e and y are in Y, we again have

$d(e,y) = d|_{Y\times Y}(e,y)$. So, $d|_{Y\times Y}(e,y) < r(e)$, so that $y \in B_{(Y,d|_{Y\times Y})}(e,r(e))$. In conclusion, (3) holds. Combining (1), (2) and (3), we conclude that

$$V \cap Y = Y \cap \left(\bigcup_{e\in E} B_{(X,d)}(e,r(e))\right) = \bigcup_{e\in E} \big(Y \cap B_{(X,d)}(e,r(e))\big) = \bigcup_{e\in E} B_{(Y,d|_{Y\times Y})}(e,r(e)) = E.$$

Conversely, assume that there exists $V \subseteq X$ which is open in X, such that $V \cap Y = E$. We will show that E is relatively open in Y. By Proposition 1.40(i), for any $v \in V$ there exists $r = r(v) > 0$ such that $B_{(X,d)}(e,r(e)) \subseteq V$. Note that

$$V = \bigcup_{v\in V} B_{(X,d)}(v,r(v)). \tag{4}$$

Now, consider the set

$$E' := \bigcup_{v\in V\cap Y} B_{(Y,d|_{Y\times Y})}(v,r(v)). \tag{5}$$

From Proposition 1.40(i), E' is open in $(Y,d|_{Y\times Y})$. We will then conclude by showing that $E = E'$. As before, for any $v \in V \cap Y$, we have the equality

$$Y \cap B_{(X,d)}(v,r(v)) = B_{(Y,d|_{Y\times Y})}(v,r(v)). \tag{6}$$

Combining (4), (5) and (6), we get $E = V \cap Y = E'$, completing the proof. □

1.3. **Cauchy sequences and Completeness.** Recall the following definition of a subsequence. Let $(x^{(j)})_{j=k}^{\infty}$ be a sequence of elements of a metric space (X,d). Let $j_1, j_2, \ldots$ be an increasing sequence of integers such that

$$k \le j_1 < j_2 < j_3 < \cdots$$

We then say that $(x^{(j_m)})_{m=1}^{\infty}$ is a **subsequence** of the sequence $(x^{(j)})_{j=k}^{\infty}$.

Lemma 1.46. *Let $(x^{(j)})_{j=k}^{\infty}$ be a sequence of elements of a metric space (X,d) which converges to some limit $x \in X$. Then every subsequence of $(x^{(j)})_{j=k}^{\infty}$ also converges to x.*

Exercise 1.47. Prove Lemma 1.46.

Definition 1.48 (Cauchy sequence). Let $(x^{(j)})_{j=k}^{\infty}$ be a sequence of elements of a metric space (X,d). We say that the sequence $(x^{(j)})_{j=k}^{\infty}$ is a **Cauchy sequence** if and only if, for every $\varepsilon > 0$, there exists an integer $J = J(\varepsilon)$ such that, for all $j,\ell > J$, we have $d(x^{(j)},x^{(\ell)}) < \varepsilon$.

Lemma 1.49. *Let $(x^{(j)})_{j=k}^{\infty}$ be a sequence of elements of a metric space (X,d) which converges to some limit $x \in X$. Then $(x^{(j)})_{j=k}^{\infty}$ is also a Cauchy sequence.*

Exercise 1.50. Prove Lemma 1.46.

The converse is false sometimes, as we learned in the previous real analysis class. If $(x^{(j)})_{j=k}^{\infty}$ is a Cauchy sequence in X, then the sequence $(x^{(j)})_{j=k}^{\infty}$ may not converge to an element of X. For example, we saw that a Cauchy sequence of rationals numbers may converge to a real number which is not itself a rational number.

However, a Cauchy sequence of real numbers does converge to a real number, as we learned in the previous real analysis class.

Theorem 1.51. *Let $(\mathbb{R}, d)$ be the real line with the usual metric, so that for any $x, y \in \mathbb{R}$, we have $d(x, y) := |x - y|$. If $(x^{(j)})_{j=k}^{\infty}$ is a Cauchy sequence of elements of $\mathbb{R}$, then there exists some $x \in \mathbb{R}$ such that $(x^{(j)})_{j=k}^{\infty}$ converges to x with respect to d.*

We cast the latter property into the following definition.

Definition 1.52 (**Completeness**). Let (X, d) be a metric space. We say that (X, d) is **complete** if and only if the following property holds. For any Cauchy sequence $(x^{(j)})_{j=k}^{\infty}$ of elements of X, then there exists some $x \in X$ such that $(x^{(j)})_{j=k}^{\infty}$ converges to x with respect to d.

So, $(\mathbb{R}, d)$ is a complete metric space by Theorem 1.51. However, the metric space $(\mathbb{Q}, d)$ is not complete. For example, we can construct a sequence of rational numbers that converges to $\sqrt{2}$, but $\sqrt{2}$ is not a rational number.

Complete metric spaces are always closed when they are considered as subsets of other metric spaces, as we now show.

Proposition 1.53.

- *Let (X, d) be a metric space, and let Y be a subset of X, so that $(Y, d|_{Y \times Y})$ is a metric space. If $(Y, d|_{Y \times Y})$ is complete, then Y is closed in (X, d).*
- *Conversely, assume that (X, d) is a complete metric space and that Y is a closed subset of X. Then $(Y, d|_{Y \times Y})$ is complete.*

Exercise 1.54. Prove Proposition 1.53.

A metric space (X, d) which is not complete may or may not be closed, when considered as a subset of another metric space. For example, if d is the standard metric on $\mathbb{R}$, then $\mathbb{Q}$ is closed in $(\mathbb{Q}, d|_{\mathbb{Q} \times \mathbb{Q}})$, but $\mathbb{Q}$ is not closed in $(\mathbb{R}, d)$. We briefly mention that, given any metric space (X, d), there is a way to form the **completion** $(\overline{X}, \overline{d})$ of (X, d), so that $(\overline{X}, \overline{d})$ is a complete metric space that contains (X, d). This procedure imitates our construction of the real numbers using Cauchy sequences of rational numbers.

1.4. **Compactness.** We have now arrived at the extremely useful concept of compactness. Compactness expresses the exact properties that are needed to obtain the conclusion of the Bolzano-Weierstrass Theorem. As we recall, the Bolzano-Weierstrass theorem is very useful, and likewise compactness is very useful.

Definition 1.55 (**Boundedness**). A sequence $(x^{(j)})_{j=k}^{\infty}$ in a metric space (X, d) is said to be **bounded** if and only if there exists $x \in X$ and there exists $r > 0$ such that $x^{(j)} \in B(x, r)$ for all $j \geq k$. Similarly, a subset E of a metric space (X, d) is said to be **bounded** if and only if there exists $x \in X$ and there exists $r > 0$ such that $E \subseteq B(x, r)$.

Theorem 1.56 (**Bolzano-Weierstrass**). *Let $(\mathbb{R}, d)$ be the real line with the standard metric. Let $(x^{(j)})_{j=k}^{\infty}$ be a bounded sequence in $\mathbb{R}$. Then there exists a subsequence of $(x^{(j)})_{j=k}^{\infty}$ that converges in $(\mathbb{R}, d)$.*

Corollary 1.57. *Let n be a positive integer. Let $(\mathbb{R}^n, d)$ denote Euclidean space with either of the metrics $d = d_{\ell_2}$ or $d = d_{\ell_1}$. Let $(x^{(j)})_{j=k}^{\infty}$ be a bounded sequence in $\mathbb{R}^n$. Then there exists a subsequence of $(x^{(j)})_{j=k}^{\infty}$ that converges in $(\mathbb{R}^n, d)$.*

This convergent subsequence property is called compactness.

Definition 1.58 (**Compactness**)**.** A metric space (X,d) is said to be **compact** if and only if every sequence in (X,d) has at least one convergent subsequence.

A compact metric space satisfies the following two special properties.

Proposition 1.59. *Let (X,d) be a compact metric space. Then (X,d) is both complete and bounded.*

Exercise 1.60. Prove Proposition 1.59. (Hint: prove each property separately, and use argument by contradiction.)

We often talk about compact sets rather than compact metric spaces, so we make the following definition.

Definition 1.61 (**Compactness of a Set**)**.** Let (X,d) be a metric space, and let Y be a subset of X. We say that Y is **compact** if and only if the metric space $(Y, d|_{Y\times Y})$ is compact.

Corollary 1.62. *Let (X,d) be a metric space, and let Y be a compact subset of X. Then Y is closed and bounded.*

Proof. Apply Proposition 1.59 and then Proposition 1.53. $\square$

In Euclidean space, the converse of Corollary 1.62 is true. The following Theorem therefore gives a useful characterization of compact subsets of Euclidean space.

Theorem 1.63. *Let n be a positive integer. Let $(\mathbb{R}^n, d)$ denote Euclidean space with the metric $d = d_{\ell_2}$ or $d = d_{\ell_1}$. Let E be a subset of $\mathbb{R}^n$. Then E is compact if and only if E is both closed and bounded.*

Exercise 1.64. Prove Theorem 1.63 using Corollary 1.57.

Compact sets of metric spaces can be equivalently characterized using open covers. This open cover property will become useful in our discussion of continuous functions. The property says: any (possibly uncountable) open cover of a compact set has a finite subcover. The following proof is a bit lengthy, so it can be skipped on a first reading.

Theorem 1.65 (**Open Cover Characterization of Compactness**)**.** *Let (X,d) be a metric space and let Y be a compact subset of X. Let I be an index set. Let $\{V_\alpha\}_{\alpha\in I}$ be a collection of open sets in X. Assume that*

$$Y \subseteq \bigcup_{\alpha\in I} V_\alpha.$$

*(That is, the collection $\{V_\alpha\}_{\alpha\in I}$ **covers** Y.) Then, there exists a finite set $A \subseteq I$ such that*

$$Y \subseteq \bigcup_{\alpha\in A} V_\alpha.$$

Proof. Let $y \in Y$. Then there exists $\alpha \in I$ such that $y \in V_\alpha$. Since V_α is open, there exists $r > 0$ such that $B(y,r) \subseteq V_\alpha$. For each $y \in Y$, define $r(y) \in \mathbb{R}$ by

$$r(y) := \sup\{r \in (0,\infty) \colon \exists\, \alpha \in I \,\text{such that}\, B(y,r) \subseteq V_\alpha\}.$$

We showed that for every $y \in Y$, we have $r(y) > 0$. Define $r_0 \in \mathbb{R}$ by

$$r_0 := \inf\{r(y) \colon y \in Y\}.$$

Since $r(y) > 0$ for all $y \in Y$, we have $r_0 \geq 0$. We now consider the cases $r_0 = 0$ and $r_0 > 0$ separately.
Case 1. $r_0 = 0$. In this case, for every positive integer j, there exists $y \in Y$ such that $r(y) < 1/j$. So, for every positive integer j, let $y^{(j)} \in Y$ satisfy $r(y^{(j)}) < 1/j$. (We can do this by the countable axiom of choice.) By the Squeeze Theorem, $\lim_{j\to\infty} r(y^{(j)}) = 0$. Since $(y^{(j)})_{j=1}^{\infty}$ is a sequence in Y, and since Y is compact, there exists a subsequence $(y^{(j_k)})_{k=1}^{\infty}$ that converges to some point $y_0 \in Y$.

Since $y_0 \in Y$, we know as above that there exists $\alpha_0 \in I$ such that $y_0 \in V_{\alpha_0}$. And since V_{α_0} is open, there exists $\varepsilon > 0$ such that $B(y_0, \varepsilon) \subseteq V_{\alpha_0}$. Since $y^{(j_k)}$ converges to y_0 as $k \to \infty$, there exists a positive integer K such that, for all $k > K$, we have $y^{(j_k)} \in B(y_0, \varepsilon/2)$. By the triangle inequality, if $k > K$ and if $z \in B(y^{(j_k)}, \varepsilon/2)$, then $d(z, y_0) \leq d(z, y^{(j_k)}) + d(y^{(j_k)}, y_0) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. That is, if $k > K$, then $B(y^{(j_k)}, \varepsilon/2) \subseteq B(y_0, \varepsilon)$. Since $B(y_0, \varepsilon) \subseteq V_{\alpha_0}$, we conclude that $r(y^{(j_k)}) \geq \varepsilon/2$ for all $k > K$. The last condition implies that $\lim_{k\to\infty} r(y^{(j_k)}) \neq 0$, a contradiction. We conclude that Case 1 does not occur, i.e. we must have $r_0 > 0$.
Case 2. $r_0 > 0$. In this case, for every $y \in Y$, we have $r(y) > r_0/2$. So, for all $y \in Y$, there exists $\alpha = \alpha(y) \in I$ such that $B(y, r_0/2) \in V_\alpha$. We now argue by contradiction. Suppose there does not exist a finite collection $\{V_\alpha\}_{\alpha\in A}$ that covers Y. Let $y^{(1)}$ be any point in Y. We construct a sequence of points in Y recursively. Suppose we are given $y^{(1)}, \ldots, y^{(j)}$ a sequence of points in Y. Given these points, the union $B(y^{(1)}, r_0/2) \cup \cdots \cup B(y^{(j)}, r_0/2)$ is contained in the union $V_{\alpha(y^{(1)})} \cup \cdots \cup V_{\alpha(y^{(j)})}$ for some $\alpha(y^{(1)}), \ldots, \alpha(y^{(j)}) \in I$. By our contradictory assumption, the latter set does not cover Y, so the set $B(y^{(1)}, r_0/2) \cup \cdots \cup B(y^{(j)}, r_0/2)$ does not cover Y. That is, there exists some $y^{(j+1)} \in Y$ such that $y^{(j+1)} \notin B(y^{(i)}, r_0/2)$ for all $1 \leq i \leq j$. That is, $d(y^{(j+1)}, y^{(i)}) \geq r_0/2$ for all $1 \leq i \leq j$. From the latter property, the sequence $(y^{(j)})_{j=1}^{\infty}$ is a sequence that has no convergent subsequence. (If a subsequence $(y^{(j_k)})_{k=1}^{\infty}$ converged to some $z \in Y$, then there would exist a positive integer K such that, for all $k > K$, we would have $d(y^{(j_k)}, z) < r_0/4$, so that $d(y^{(j_{K+2})}, y^{(j_{K+1})}) \leq d(y^{(j_{K+2})}, z) + d(z, y^{(j_{K+1})}) < r_0/2$, a contradiction.) We have therefore contradicted the compactness of Y. Since we have achieved a contradiction, the proof is done. □

Remark 1.66. The converse is also true. If a set Y has the property that every open cover of Y has a finite subcover, then Y is compact.

Theorem 1.65 has the following useful corollary.

Corollary 1.67. *Let (X, d) be a metric space, and let $K_1, K_2, \ldots$ be a sequence of nonempty compact subsets of X such that*

$$K_1 \supseteq K_2 \supseteq K_3 \supseteq \cdots.$$

Then the intersection $\cap_{j=1}^{\infty} K_j$ is nonempty.

Exercise 1.68. Prove Corollary 1.67. (Hint: first, work in the compact metric space $(K_1, d|_{K_1\times K_1})$. Then, consider the sets $K_1 \smallsetminus K_j$ which are open in K_1. Assume for the sake of contradiction that $\cap_{j=1}^{\infty} K_j = \emptyset$. Then apply Theorem 1.65.)

Theorem 1.69. *Let (X, d) be a metric space.*

- (i) *Let Y be a compact subset of X, and let Z be a subset of Y. Then Z is compact if and only if Z is closed.*
- (ii) *Let $Y_1, \ldots, Y_n$ be compact subsets of X. Then $Y_1 \cup \cdots \cup Y_n$ is compact.*

(iii) *Every finite subset of X is compact.*

1.5. **Continuity.** We can readily generalize the notion of continuity of a function $f\colon \mathbb{R} \to \mathbb{R}$ to the setting of a function between metric spaces $f\colon X \to Y$. We just take the usual definition and then we replace the absolute values with the required metric, as follows.

Definition 1.70 (**Continuity**). Let (X, d_X) and let (Y, d_Y) be metric spaces. Let $f\colon X \to Y$ be a function. Let $x_0 \in X$. We say that f is **continuous** at x_0 if and only if, for all $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ such that, if $x \in X$ satisfies $d_X(x, x_0) < \delta$, then $d_Y(f(x), f(x_0)) < \varepsilon$. We say that the function f is **continuous** if and only if it is continuous at every point in X.

Remark 1.71. Suppose $f\colon X \to Y$ is continuous and K is a subset of X. Then the restriction of f to K, $f|_K\colon K \to Y$ is also continuous.

As on the real line, continuous functions maps convergent sequences to convergent sequences.

Theorem 1.72. *Let (X, d_X) and let (Y, d_Y) be metric spaces. Let $f\colon X \to Y$ be a function. Then the following two statements are equivalent.*

- *f is continuous at x_0.*
- *If we have a sequence $(x^{(j)})_{j=1}^{\infty}$ in X which converges to x_0 with respect to d_X, then the sequence $(f(x^{(j)}))_{j=1}^{\infty}$ converges to $f(x_0)$ with respect to the metric d_Y.*

Exercise 1.73. Prove Theorem 1.72.

In fact, there is even a way to characterize continuous functions using the inverse images of open and closed sets.

Theorem 1.74. *Let (X, d_X) and let (Y, d_Y) be metric spaces. Let $f\colon X \to Y$ be a function. Then the following four statements are equivalent.*

- *f is continuous at x_0, for all $x_0 \in X$.*
- *For all $x_0 \in X$, if we have a sequence $(x^{(j)})_{j=1}^{\infty}$ in X which converges to x_0 with respect to d_X, then the sequence $(f(x^{(j)}))_{j=1}^{\infty}$ converges to $f(x_0)$ with respect to the metric d_Y.*
- *For all open sets W in Y, the set $f^{-1}(W) = \{x \in X : f(x) \in W\}$ is an open set in X.*
- *For all closed sets V in Y, the set $f^{-1}(V)$ is a closed set in X.*

Exercise 1.75. Prove Theorem 1.74.

Remark 1.76. For a continuous function, it is not always true that the image of an open set is open, and it is not always true that the image of a closed set is closed.

We can now quickly show that the composition of continuous functions is continuous.

Corollary 1.77. *Let (X, d_X), (Y, d_Y) and (Z, d_Z) be metric spaces. Let $f\colon (X, d_X) \to (Y, d_Y)$ be a continuous function and let $g\colon (Y, d_Y) \to (Z, d_Z)$ be a continuous function. Then $g \circ f\colon (X, d_X) \to (Z, d_Z)$ is a continuous function.*

Exercise 1.78. Prove Corollary 1.77.

1.6. **Continuity and Compactness.**

Remark 1.79. From now on, unless otherwise specified, $\mathbb{R}^n$ refers to Euclidean space $\mathbb{R}^n$ with $n \geq 1$ a positive integer, and where we use the metric d_{ℓ_2} on $\mathbb{R}^n$. In particular, $\mathbb{R}$ refers to the metric space $\mathbb{R}$ equipped with the metric $d(x,y) = |x-y|$.

On the real line, we learned from the Extreme Value Theorem that the continuous image of a closed interval is another closed interval. The appropriate generalization of this statement to metric spaces now follows.

Theorem 1.80. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let $f\colon (X, d_X) \to (Y, d_Y)$ be a continuous function. Suppose $K \subseteq X$ is a compact set. Then $f(K) = \{f(x)\colon x \in K\}$ is also a compact set.*

Exercise 1.81. Prove Theorem 1.80

Combining this Theorem with the characterization of compactness in Euclidean spaces (i.e. Heine-Borel, Theorem 1.63), we get the following statement.

Corollary 1.82. *Let K be a closed and bounded subset of $\mathbb{R}^n$. Let $f\colon K \to \mathbb{R}^m$ be a continuous function. Then the set $f(K)$ is also closed and bounded. In particular, the function f is bounded on K.*

This Corollary allows us to state our generalization of the Extreme Value Theorem, which we now refer to as the Maximum Principle.

Definition 1.83. Let X be a set. Let $f\colon X \to \mathbb{R}$ be a function. We say that f **attains its maximum** at $x_0 \in X$ if and only if $f(x_0) \geq f(x)$ for all $x \in X$. We say that f **attains its minimum** at $x_0 \in X$ if and only if $f(x_0) \leq f(x)$ for all $x \in X$.

Theorem 1.84 (**The Maximum Principle**). *Let K be a closed and bounded subset of $\mathbb{R}^n$, and let $f\colon K \to \mathbb{R}$ be a continuous function. Then there exist points $a, b \in K$ such that f attains its maximum at a and f attains its minimum at b.*

Exercise 1.85. Prove Theorem 1.84. (Hint: use Corollary 1.82 and then consider the numbers $\sup_{x\in K} f(x)$ and $\inf_{x\in K} f(x)$.)

1.7. **Continuity and Connectedness.**

Recall that the Intermediate Value Theorem says that a continuous function on an interval has an interval as its range. The appropriate generalization of this statement to metric spaces involves the concept of connectedness.

Definition 1.86 (**Connectedness**). Let (X, d) be a metric space. We say that X is **disconnected** if and only if there exist disjoint open sets V, W in X such that $V \cup W = X$. (Equivalently, X is disconnected if and only if X contains a proper non-empty subset which is both open and closed.) We say that X is **connected** if and only if X is not disconnected.

Example 1.87. The set $X = [0,1] \cup [2,3]$ with the metric $d(x,y) = |x-y|$ is disconnected, since the sets $[0,1]$ and $[2,3]$ are both open in X.

Definition 1.88. Let (X, d) be a metric space and let Y be a subset of X. We say that Y is **connected** if and only if the metric space $(Y, d|_{Y\times Y})$ is connected. We say that Y is **disconnected** if and only if the metric space $(Y, d|_{Y\times Y})$ is disconnected.

For the sake of examples, we now identify the connected subsets of the real line.

Theorem 1.89. *Let X be a subset of the real line $\mathbb{R}$. Then the following statements are equivalent.*

- *X is connected.*
- *For any $x, y \in X$ with $x < y$, the closed interval $[x, y]$ is also contained in X.*

Proof. We first show the forward implication. Suppose X is connected. We argue by contradiction. Let $x, y \in X$ with $x < y$ such that $[x, y]$ is not contained in X. Then there exists $x < z < y$ such that $z \notin X$. Then the sets $(-\infty, z) \cap X$ and $(z, \infty) \cap X$ are both disjoint, nonempty, relatively open sets whose union in X. Therefore, X is disconnected, a contradiction. We conclude that the forward implication holds.

We now prove the more involved reverse implication. Suppose for any $x, y \in X$ with $x < y$, the closed interval $[x, y]$ is also contained in X. We need to show that X is connected. We argue by contradiction. Suppose that X is disconnected. Then there exist two disjoint, nonempty, relatively open sets V, W such that $V \cup W = X$. Since V, W are nonempty, let $v \in V$ and let $w \in W$. Without loss of generality, $v < w$. By assumption, the closed interval $[v, w]$ is contained in X. Consider the real number

$$x = \sup([v, w] \cap V).$$

By the definition of x, we have $x \in [v, w]$. We will derive a contradiction by trying to determine whether or not $x \in V$.

Suppose $x \in V$. Since $w \notin V$, we have $x \neq w$, so $x \in [v, w)$. Since V is relatively open in X, there exists $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \cap X \subseteq V$. Since $x \in [v, w) \subseteq X$ as well, there exists $\delta > 0$ such that $[x, x + \delta) \subseteq V$. But then x is not the least upper bound of $[v, w] \cap V$, a contradiction.

We must therefore have $x \notin V$. Since $x \in [v, w] \subseteq X$, and since V, W are disjoint, we must have $x \in W$. Since $v \in V$, we have $x \neq v$, so $x \in (v, w]$. Since W is relatively open in X, there exists $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \cap X \subseteq W$. Since $x \in (v, w] \subseteq X$ as well, there exists $\delta > 0$ such that $(x - \delta, x] \subseteq W$. Since V and W are disjoint, we once again conclude that x is not the least upper bound of $[v, w] \cap V$. In any case, we have achieved a contradiction. We finally conclude that X is connected, as desired. □

Remark 1.90. So, $\mathbb{R}$ is connected, and so are the intervals $(a, b]$, $[a, b)$, (a, b), $[a, b]$, $(-\infty, b)$, $(-\infty, b]$, (a, ∞), $[a, \infty)$. Additionally, the empty set $\emptyset$ and singleton sets $\{a\}$ are connected. We therefore have a complete list of connected subsets of $\mathbb{R}$.

It turns out that connected sets are mapped to connected sets by continuous functions. This fact particularly implies the Intermediate Value Theorem.

Theorem 1.91. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let $f \colon X \to Y$ be a continuous function. Let E be a connected subset of X. Then $f(E)$ is connected.*

Exercise 1.92. Prove Theorem 1.91.

Theorem 1.93 (**Intermediate Value Theorem**)**.** *Let (X, d) be a metric space. Let $f \colon X \to \mathbb{R}$ be a continuous function. Let E be a connected subset of X and let a, b be any two elements of E. Let y be a real number between $f(a)$ and $f(b)$, so that either $f(a) \leq y \leq f(b)$ or $f(b) \leq y \leq f(a)$. Then there exists $c \in E$ such that $f(c) = y$.*

Exercise 1.94. Prove Theorem 1.93 using Theorem 1.91.

2. Sequences and Series of Functions, Convergence

Remark 2.1. From now on, unless otherwise specified, $\mathbb{R}^n$ refers to Euclidean space $\mathbb{R}^n$ with $n \geq 1$ a positive integer, and where we use the metric d_{ℓ_2} on $\mathbb{R}^n$. In particular, $\mathbb{R}$ refers to the metric space $\mathbb{R}$ equipped with the metric $d(x,y) = |x-y|$.

Proposition 2.2. *Let $a < b$ be real numbers, and let $f\colon [a,b] \to \mathbb{R}$ be a function which is both continuous and strictly monotone increasing. Then f is a bijection from $[a,b]$ to $[f(a), f(b)]$, and the inverse function $f^{-1}\colon [f(a), f(b)] \to [a,b]$ is also continuous and strictly monotone increasing.*

Theorem 2.3 (Inverse Function Theorem). *Let X, Y be subsets of $\mathbb{R}$. Let $f\colon X \to Y$ be bijection, so that $f^{-1}\colon Y \to X$ is a function. Let $x_0 \in X$ and $y_0 \in Y$ such that $f(x_0) = y_0$. If f is differentiable at x_0, if f^{-1} is continuous at y_0, and if $f'(x_0) \neq 0$, then f^{-1} is differentiable at y_0 with*

$$(f^{-1})'(y_0) = \frac{1}{f'(x_0)}.$$

2.1. **Sequences of Functions.**

As we have seen in analysis, it is often desirable to discuss sequences of points that converge. Below, we will see that it is similarly desirable to discuss sequences of functions that converge in various senses. There are many distinct ways of discussing the convergence of sequences of functions. We will only discuss two such modes of convergence, namely pointwise and uniform convergence. Before beginning this discussion, we discuss the limiting values of functions between metric spaces, which should generalize our notion of limiting values of functions on the real line.

2.1.1. *Limiting Values of Functions.*

Definition 2.4. Let (X, d_X) and (Y, d_Y) be metric spaces, let E be a subset of X, let $f\colon X \to Y$ be a function, let $x_0 \in X$ be an adherent point of E, and let $L \in Y$. We say that $f(x)$ **converges to** L **in** Y **as** x **converges to** x_0 **in** E, and we write $\lim_{x\to x_0; x\in E} f(x) = L$, if and only if, for every $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ such that, if $x \in E$ satisfies $d_X(x, x_0) < \delta$, then $d_Y(f(x), L) < \varepsilon$.

Remark 2.5. So, f is continuous at x_0 if and only if

$$\lim_{x\to x_0; x\in X} f(x) = f(x_0). \qquad (*)$$

And f is continuous on X if and only if, for all $x_0 \in X$, $(*)$ holds.

Remark 2.6. When the domain of x of the limit $\lim_{x\to x_0; x\in X} f(x)$ is clear, we will often instead write $\lim_{x\to x_0} f(x)$.

The following equivalence is generalized from its analogue on the real line.

Proposition 2.7. *Let (X, d_X) and (Y, d_Y) be metric spaces, let E be a subset of X, let $f\colon X \to Y$ be a function, let $x_0 \in X$ be an adherent point of E, and let $L \in Y$. Then the following statements are equivalent.*

- $\lim_{x\to x_0; x\in E} f(x) = L$.
- *For any sequence $(x^{(j)})_{j=1}^{\infty}$ in E which converges to x_0 with respect to the metric d_X, the sequence $(f(x^{(j)}))_{j=1}^{\infty}$ converges to L with respect to the metric d_Y.*

Exercise 2.8. Prove Proposition 2.7.

Remark 2.9. From Propositions 2.7 and 1.24, the function f can converge to at most one limit L as x converges to x_0.

Remark 2.10. The notation $\lim_{x\to x_0;x\in E} f(x)$ implicitly refers to a convergence of the function values $f(x)$ in the metric space (Y, d_Y). Strictly speaking, it would be better to write d_Y somewhere next to the notation $\lim_{x\to x_0;x\in E} f(x)$. However, this omission of notation should not cause confusion.

2.1.2. *Pointwise Convergence and Uniform Convergence.*

Definition 2.11 (**Pointwise Convergence**)**.** Let (X, d_X) and (Y, d_Y) be metric spaces. Let $(f_j)_{j=1}^{\infty}$ be a sequence of functions from X to Y. Let $f\colon X \to Y$ be another function. We say that $(f_j)_{j=1}^{\infty}$ **converges pointwise** to f on X if and only if, for every $x \in X$, we have

$$\lim_{j\to\infty} f_j(x) = f(x).$$

That is, for all $x \in X$, we have

$$\lim_{j\to\infty} d_Y(f_j(x), f(x)) = 0.$$

That is, for every $x \in X$ and for every $\varepsilon > 0$, there exists $J > 0$ such that, for all $j > J$, we have $d_Y(f_j(x), f(x)) < \varepsilon$.

Remark 2.12. Note that, if we change the point x, then the limiting behavior of $f_j(x)$ can change quite a bit. For example, let j be a positive integer, and consider the functions $f_j\colon [0,1] \to \mathbb{R}$ where $f_j(x) = j$ for all $x \in (0, 1/j)$, and $f_j(x) = 0$ otherwise. Let $f\colon [0,1] \to \mathbb{R}$ be the zero function. Then f_j converges pointwise to zero, since for any $x \in (0,1]$, we have $f_j(x) = 0$ for all $j > 1/x$. (And $f_j(0) = 0$ for all positive integers j.) However, given any fixed positive integer j, there exists an x such that $f_j(x) = j$. Moreover, $\int_0^1 f_j = 1$ for all positive integers j, but $\int_0^1 f = 0$. So, we see that pointwise convergence does not preserve the integral of a function.

Remark 2.13. Pointwise convergence also does not preserve continuity. For example, consider $f_j\colon [0,1] \to \mathbb{R}$ defined by $f_j(x) = x^j$, where $j \in \mathbb{N}$ and $x \in [0,1]$. Define $f\colon [0,1] \to \mathbb{R}$ so that $f(1) = 1$ and so that $f(x) = 0$ for $x \in [0,1)$. Then f_j converges pointwise to f as $j \to \infty$, and each f_j is continuous, but f is not continuous.

In summary, pointwise convergence doesn't really preserve any useful analytic quantities. The above remarks show that some points are changing at much different rates than other points as $j \to \infty$. A stronger notion of convergence will then fix these issues, where all points in the domain are controlled simultaneously.

Definition 2.14 (**Uniform Convergence**)**.** Let (X, d_X) and (Y, d_Y) be metric spaces. Let $(f_j)_{j=1}^{\infty}$ be a sequence of functions from X to Y. Let $f\colon X \to Y$ be another function. We say that $(f_j)_{j=1}^{\infty}$ **converges uniformly** to f on X if and only if, for every $\varepsilon > 0$, there exists $J > 0$ such that, for all $j > J$ and for all $x \in X$ we have $d_Y(f_j(x), f(x)) < \varepsilon$.

Remark 2.15. Note that the difference between uniform and pointwise convergence is that we simply moved the quantifier "for all $x \in X$" within the statement. This change means that the integer J does not depend on x in the case of uniform convergence.

Remark 2.16. The sequences of functions from Remarks 2.12 and 2.13 do not converge uniformly. So, pointwise convergence does not imply uniform convergence. However, uniform convergence does imply pointwise convergence.

2.2. **Uniform Convergence and Continuity.** We saw that pointwise convergence does not preserve continuity. However, uniform convergence does preserve continuity.

Theorem 2.17. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let $(f_j)_{j=1}^{\infty}$ be a sequence of functions from X to Y. Let $f\colon X \to Y$ be another function. Let $x_0 \in X$. Suppose f_j converges uniformly to f on X. Suppose that, for each $j \geq 1$, we know that f_j is continuous at x_0. Then f is also continuous at x_0.*

Exercise 2.18. Prove Theorem 2.17. Hint: it is probably easiest to use the $\varepsilon - \delta$ definition of continuity. Once you do this, you may require the triangle inequality in the form

$$d_Y(f(x), f(x_0)) \leq d_Y(f(x), f_j(x)) + d_Y(f_j(x), f_j(x_0)) + d_Y(f_j(x_0), f(x_0)).$$

Corollary 2.19. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let $(f_j)_{j=1}^{\infty}$ be a sequence of functions from X to Y. Let $f\colon X \to Y$ be another function. Suppose $(f_j)_{j=1}^{\infty}$ converges uniformly to f on X. Suppose that, for each $j \geq 1$, we know that f_j is continuous on X. Then f is also continuous on X.*

Uniform limits of bounded functions are also bounded. Recall that a function $f\colon X \to Y$ between metric spaces (X, d_X) and (Y, d_Y) is **bounded** if and only if there exists a radius $R > 0$ and a point $y_0 \in Y$ such that $f(x) \in B_{(Y,d_Y)}(y_0, R)$ for all $x \in X$.

Proposition 2.20. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let $(f_j)_{j=1}^{\infty}$ be a sequence of functions from X to Y. Let $f\colon X \to Y$ be another function. Suppose $(f_j)_{j=1}^{\infty}$ converges uniformly to f on X. Suppose also that, for each $j \geq 1$, we know that f_j is bounded. Then f is also bounded.*

Exercise 2.21. Prove Proposition 2.20.

2.2.1. *The Metric of Uniform Convergence.* We will now see one advantage to our abstract approach to analysis on metric spaces. We can in fact talk about uniform convergence in terms of a metric on a space of functions, as follows.

Definition 2.22. Let (X, d_X) and (Y, d_Y) be metric spaces. Let $B(X;Y)$ denote the set of functions $f\colon X \to Y$ that are bounded. Let $f, g \in B(X;Y)$. We define the metric $d_\infty\colon B(X;Y) \times B(X;Y) \to [0, \infty)$ by

$$d_\infty(f, g) := \sup_{x \in X} d_Y(f(x), g(x)).$$

This metric is known as the **sup norm metric** or the L_∞ **metric**. We also use $d_{B(X;Y)}$ as a synonym for d_∞. Note that $d_\infty(f, g) < \infty$ since f, g are assumed to be bounded.

Exercise 2.23. Show that the space $(B(X;Y), d_\infty)$ is a metric space.

Example 2.24. Let $X = [0, 1]$ and let $Y = \mathbb{R}$. Consider the functions $f(x) = x$ and $g(x) = 2x$ where $x \in [0, 1]$. Then f, g are bounded, and

$$d_\infty(f, g) = \sup_{x \in [0,1]} |x - 2x| = \sup_{x \in [0,1]} |x| = 1.$$

Here is our promised characterization of uniform convergence in terms of the metric d_∞.

Proposition 2.25. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let $(f_j)_{j=1}^\infty$ be a sequence of functions in $B(X;Y)$. Let $f \in B(X;Y)$. Then $(f_j)_{j=1}^\infty$ converges uniformly to f on X if and only if $(f_j)_{j=1}^\infty$ converges to f in the metric $d_{B(X;Y)}$.*

Exercise 2.26. Prove Proposition 2.25.

Definition 2.27. Let (X, d_X) and (Y, d_Y) be metric spaces. Define the set of bounded continuous functions from X to Y as

$$C(X;Y) := \{f \in B(X;Y) \colon f \text{ is continuous}\}.$$

Note that $C(X;Y) \subseteq B(X;Y)$ by the definition of $C(X;Y)$. Also, by Corollary 2.19, $C(X;Y)$ is closed in $B(X;Y)$ with respect to the metric d_∞. In fact, more is true.

Theorem 2.28. *Let (X, d_X) be a metric space, and let (Y, d_Y) be a complete metric space. Then the space $(C(X;Y), d_{B(X;Y)}|_{C(X;Y)\times C(X;Y)})$ is a complete subspace of $B(X;Y)$. That is, every Cauchy sequence of functions in $C(X;Y)$ converges to a function in $C(X;Y)$.*

Exercise 2.29. Prove Theorem 2.28

2.3. **Series of Functions and the Weierstrass M-test.** For each positive integer j, let $f_j \colon X \to \mathbb{R}$ be a function. We will now consider infinite series of the form $\sum_{j=1}^\infty f_j$. The most natural thing to do now is to determine in what sense the series $\sum_{j=1}^\infty f_j$ is a function, and if it is a function, determine if it is continuous. Note that we have restricted the range to be $\mathbb{R}$ since it does not make sense to add elements in a general metric space. Power series and Fourier series perhaps give the most studied examples of series of functions. If $x \in [0,1]$ and if a_j are real numbers for all $j \geq 1$, we want to make sense of the series $\sum_{j=1}^\infty a_j \cos(2\pi j x)$. We want to know in what sense this infinite series is a function, and if it is a function, do the partial sums converge in any reasonable manner? We will return to these issues later on.

Definition 2.30. Let (X, d_X) be a metric space. For each positive integer j, let $f_j \colon X \to \mathbb{R}$ be a function, and let $f \colon X \to \mathbb{R}$ be another function. If the partial sums $\sum_{j=1}^J f_j$ converge pointwise to f as $J \to \infty$, then we say that the infinite series $\sum_{j=1}^\infty f_j$ **converge pointwise** to f, and we write $f = \sum_{j=1}^\infty f_j$. If the partial sums $\sum_{j=1}^J f_j$ converge uniformly to f as $J \to \infty$, then we say that the infinite series $\sum_{j=1}^\infty f_j$ **converge uniformly** to f, and we write $f = \sum_{j=1}^\infty f_j$. (In particular, the notation $f = \sum_{j=1}^\infty f_j$ is ambiguous, since the nature of the convergence of the series is not specified.)

Remark 2.31. If a series converges uniformly then it converges pointwise. However, the converse is false in general.

Exercise 2.32. Let $x \in (-1,1)$. For each integer $j \geq 1$, define $f_j(x) := x^j$. Show that the series $\sum_{j=1}^\infty f_j$ converges pointwise, but not uniformly, on $(-1,1)$ to the function $f(x) = x/(1-x)$. Also, for any $0 < t < 1$, show that the series $\sum_{j=1}^\infty f_j$ converges uniformly to f on $[-t,t]$.

Definition 2.33. Let $f \colon X \to \mathbb{R}$ be a bounded real-valued function. We define the **sup-norm** $\|f\|_\infty$ of f to be the real number

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

Exercise 2.34. Let X be a set. Show that $\|\cdot\|_\infty$ is a norm on the space $B(X;\mathbb{R})$.

Theorem 2.35 (Weierstrass M-test). *Let (X,d) be a metric space and let $(f_j)_{j=1}^\infty$ be a sequence of bounded real-valued continuous functions on X such that the series (of real numbers) $\sum_{j=1}^\infty \|f_j\|_\infty$ is absolutely convergent. Then the series $\sum_{j=1}^\infty f_j$ converges uniformly to some continuous function $f\colon X\to\mathbb{R}$.*

Exercise 2.36. Prove Theorem 2.35. (Hint: first, show that the partial sums $\sum_{j=1}^J f_j$ form a Cauchy sequence in $C(X;\mathbb{R})$. Then, use Theorem 2.28 and the completeness of the real line $\mathbb{R}$.)

Remark 2.37. The Weierstrass M-test will be useful in our investigation of power series.

2.4. **Uniform Convergence and Integration.**

Theorem 2.38. *Let $a<b$ be real numbers. For each integer $j\geq 1$, let $f_j\colon[a,b]\to\mathbb{R}$ be a Riemann integrable function on $[a,b]$. Suppose f_j converges uniformly on $[a,b]$ to a function $f\colon[a,b]\to\mathbb{R}$, as $j\to\infty$. Then f is also Riemann integrable, and*

$$\lim_{j\to\infty}\int_a^b f_j=\int_a^b f.$$

Remark 2.39. Before we begin, recall that we require any Riemann integrable function g to be bounded. Also, for a Riemann integrable function g, we denote $\underline{\int_a^b} g$ as the supremum of all lower Riemann sums of g over all partitions of $[a,b]$. And we denote $\overline{\int_a^b} g$ as the infimum of all upper Riemann sums of g over all partitions of $[a,b]$. Recall also that a function g is defined to be Riemann integrable if and only if $\underline{\int_a^b} g=\overline{\int_a^b} g$.

Proof. We first show that f is Riemann integrable. First, note that f_j is bounded for all $j\geq 1$, since that is part of the definition of being Riemann integrable. So, f is bounded by Proposition 2.20. Now, let $\varepsilon>0$. Since f_j converges uniformly to f on $[a,b]$, there exists $J>0$ such that, for all $j>J$, we have

$$f_j(x)-\varepsilon\leq f(x)\leq f_j(x)+\varepsilon,\qquad\forall\,x\in[a,b].$$

Integrating this inequality on $[a,b]$, we have

$$\underline{\int_a^b}(f_j(x)-\varepsilon)\leq\underline{\int_a^b} f\leq\overline{\int_a^b} f\leq\overline{\int_a^b}(f_j(x)+\varepsilon).$$

Since f_j is Riemann integrable for all $j\geq 1$, we therefore have

$$-(b-a)\varepsilon+\int_a^b f_j\leq\underline{\int_a^b} f\leq\overline{\int_a^b} f\leq(b-a)\varepsilon+\int_a^b f_j.\qquad(*)$$

In particular, we get

$$0\leq\overline{\int_a^b} f-\underline{\int_a^b} f\leq 2(b-a)\varepsilon.$$

Since $\varepsilon>0$ is arbitrary, we conclude that $\overline{\int_a^b} f=\underline{\int_a^b} f$, so f is Riemann integrable.

Now, from $(*)$, we have: for any $\varepsilon > 0$, there exists J such that, for all $j > J$, we have

$$\left|\int_a^b f - \int_a^b f_j\right| \leq (b-a)\varepsilon.$$

Since this holds for any $\varepsilon > 0$, we conclude that $\lim_{j\to\infty} \int_a^b f_j = \int_a^b f$, as desired. □

Remark 2.40. In summary, if a sequence of Riemann integrable functions $(f_j)_{j=1}^\infty$ converges to f uniformly, then we can interchange limits and integrals

$$\lim_{j\to\infty} \int f_j = \int \lim_{j\to\infty} f_j.$$

Recall that this equality does not hold if we only assume that the functions converge pointwise.

An analogous statement holds for series.

Theorem 2.41. *Let $a < b$ be real numbers. For each integer $j \geq 1$, let $f_j\colon [a,b] \to \mathbb{R}$ be a Riemann integrable function on $[a,b]$. Suppose $\sum_{j=1}^\infty f_j$ converges uniformly on $[a,b]$. Then $\sum_{j=1}^\infty f_j$ is also Riemann integrable, and*

$$\sum_{j=1}^\infty \int_a^b f_j = \int_a^b \sum_{j=1}^\infty f_j.$$

Exercise 2.42. Prove Theorem 2.41.

Example 2.43. Let $x \in (-1,1)$. We know that $\sum_{j=1}^\infty x^j = x/(1-x)$, and the convergence is uniform on $[-r,r]$ for any $0 < r < 1$. Adding 1 to both sides, we get

$$\sum_{j=0}^\infty x^j = \frac{1}{1-x}.$$

And this sum also converges uniformly on $[-r,r]$ for any $0 < r < 1$. Applying Theorem 2.41 and integrating on $[0,r]$, we get

$$\sum_{j=0}^\infty \frac{r^{j+1}}{j+1} = \sum_{j=0}^\infty \int_0^r x^j = \int_0^r \frac{1}{1-x}.$$

The last function is equal to $-\log(1-r)$, though we technically have not defined the logarithm function yet. We will define the logarithm further below.

2.5. **Uniform Convergence and Differentiation.** We now investigate the relation between uniform convergence and differentiation.

Remark 2.44. Suppose a sequence of differentiable functions $(f_j)_{j=1}^\infty$ converges uniformly to a function f. We first show that f need not be differentiable. Consider the functions $f_j(x) := \sqrt{x^2 + 1/j}$, where $x \in [-1,1]$. Let $f(x) = |x|$. Note that

$$|x| \leq \sqrt{x^2 + 1/j} \leq |x| + 1/\sqrt{j}.$$

These inequalities follow by taking the square root of $x^2 \leq x^2 + 1/j \leq x^2 + 1/j + 2|x|/\sqrt{j}$. So, by the Squeeze Theorem, $(f_j)_{j=1}^\infty$ converges uniformly to f on $[-1,1]$. However, f is not differentiable at 0. In conclusion, uniform convergence does not preserve differentiability.

Remark 2.45. Suppose a sequence of differentiable functions $(f_j)_{j=1}^{\infty}$ converge uniformly to a function f. Even if f is assumed to be differentiable, we show that $(f_j)'$ may not converge to f'. Consider the functions $f_j(x) := j^{-1/2}\sin(j\pi x)$, where $x \in [-1,1]$. (We will assume some basic properties of trigonometric functions which we will prove later on. Since we are only providing a motivating example, we will not introduce any circular reasoning.) Let f be the zero function. Since $|\sin(j\pi x)| \leq 1$, we have $d_\infty(f_j, f) \leq j^{-1/2}$, so $(f_j)_{j=1}^{\infty}$ converges uniformly on $[-1,1]$. However, $f_j'(x) = j^{1/2}\pi\cos(j\pi x)$. So, $f_j'(0) = j^{1/2}\pi$. That is, $(f_j')_{j=1}^{\infty}$ does not converge pointwise to f. So, $(f_j')_{j=1}^{\infty}$ does not converge uniformly to $f' = 0$. In conclusion, uniform convergence does not imply uniform convergence of derivatives.

However, the converse statement is true, as long as the sequence of functions converges at one point.

Theorem 2.46. *Let $a < b$. For every integer $j \geq 1$, let $f_j \colon [a,b] \to \mathbb{R}$ be a differentiable function whose derivative $(f_j)' \colon [a,b] \to \mathbb{R}$ is continuous. Assume that the derivatives $(f_j)'$ converge uniformly to a function $g \colon [a,b] \to \mathbb{R}$ as $j \to \infty$. Assume also that there exists a point $x_0 \in [a,b]$ such that $\lim_{j\to\infty} f_j(x_0)$ exists. Then the functions f_j converge uniformly to a differentiable function f as $j \to \infty$, and $f' = g$.*

Proof. Let $x \in [a,b]$. From the Fundamental Theorem of Calculus, for each $j \geq 1$,

$$f_j(x) - f_j(x_0) = \int_{x_0}^{x} f_j'. \qquad (*)$$

By assumption, $L := \lim_{j\to\infty} f_j(x_0)$ exists. From Theorem 2.17, g is continuous, and in particular, g is Riemann integrable on $[a,b]$. Also, by Theorem 2.38, $\lim_{j\to\infty}\int_{x_0}^{x} f_j'$ exists and is equal to $\int_{x_0}^{x} g$. We conclude by $(*)$ that $\lim_{j\to\infty} f_j(x)$ exists, and

$$\lim_{j\to\infty} f_j(x) = L + \int_{x_0}^{x} g.$$

Define the function f on $[a,b]$ so that

$$f(x) = L + \int_{x_0}^{x} g.$$

We know so far that $(f_j)_{j=1}^{\infty}$ converges pointwise to f. We now need to show that this convergence is in fact uniform. We defer this part to the exercises. $\square$

Exercise 2.47. Complete the proof of Theorem 2.46.

Corollary 2.48. *Let $a < b$. For every integer $j \geq 1$, let $f_j \colon [a,b] \to \mathbb{R}$ be a differentiable function whose derivative $f_j' \colon [a,b] \to \mathbb{R}$ is continuous. Assume that the series of real numbers $\sum_{j=1}^{\infty} \|f_j'\|_\infty$ is absolutely convergent. Assume also that there exists $x_0 \in [a,b]$ such that the series of real numbers $\sum_{j=1}^{\infty} f_j(x_0)$ converges. Then the series $\sum_{j=1}^{\infty} f_j$ converges uniformly on $[a,b]$ to a differentiable function. Moreover, for all $x \in [a,b]$,*

$$\frac{d}{dx}\sum_{j=1}^{\infty} f_j(x) = \sum_{j=1}^{\infty}\frac{d}{dx} f_j(x)$$

Exercise 2.49. Prove Corollary 2.48.

The following exercise is a nice counterexample to keep in mind, and it also shows the necessity of the assumptions of Corollary 2.48.

Exercise 2.50. (For this exercise, you can freely use facts about trigonometry that you learned in your previous courses.) Let $x \in \mathbb{R}$ and let $f\colon \mathbb{R} \to \mathbb{R}$ be the function $f(x) := \sum_{j=1}^{\infty} 4^{-j} \cos(32^j \pi x)$. Note that this series is uniformly convergent by the Weierstrass M-test (Theorem 2.35). So, f is a continuous function. However, at every point $x \in \mathbb{R}$, f is not differentiable, as we now discuss.

- Show that, for all positive integers j, m, we have
$$|f((j+1)/32^m) - f(j/32^m)| \geq 4^{-m}.$$
(Hint: for certain sequences of numbers $(a_j)_{j=1}^{\infty}$, use the identity
$$\sum_{j=1}^{\infty} a_j = (\sum_{j=1}^{m-1} a_j) + a_m + \sum_{j=m+1}^{\infty} a_j.$$
Also, use the fact that the cosine function is periodic with period 2π, and the summation $\sum_{j=0}^{\infty} r^j = 1/(1-r)$ for all $-1 < r < 1$. Finally, you should require the inequality: for all real numbers x, y, we have $|\cos(x) - \cos(y)| \leq |x - y|$. This inequality follows from the Mean Value Theorem or the Fundamental Theorem of Calculus.)
- Using the previous result, show that, for ever $x \in \mathbb{R}$, f is not differentiable at x. (Hint: for every $x \in \mathbb{R}$ and for every positive integer m, there exists an integer j such that $j \leq 32^m x \leq j + 1$.)
- Explain briefly why this result does not contradict Corollary 2.48.

2.6. **Uniform Approximation by Polynomials.**

Definition 2.51 (Polynomial). Let $a < b$ be real numbers and let $x \in [a, b]$. A **polynomial** on $[a, b]$ is a function $f\colon [a, b] \to \mathbb{R}$ of the form $f(x) = \sum_{j=0}^{k} a_j x^j$, where k is a natural number and $a_0, \ldots, a_k$ are real numbers. If $a_k \neq 0$, then k is called the **degree of** f.

From the previous exercise, we have seen that general continuous functions can behave rather poorly, in that they may never be differentiable. Polynomials on the other hand are infinitely differentiable. And it is often beneficial to deal with polynomials instead of general functions. So, we mention below a result of Weierstrass which says: any continuous function on an interval $[a, b]$ can be uniformly approximated by polynomials.

This fact seems to be related to power series, but it is something much different. It may seem possible to take a general (infinitely differentiable) function, take a high degree Taylor polynomial of this function, and then claim that this polynomial approximates our original function well. There are two problems with this approach. First of all, the continuous function that we start with may not even be differentiable. Second of all, even if we have an infinitely differentiable function, its power series may not actually approximate that function well. Recall that the function $f(x) = e^{-1/x^2}$ (where $f(0) := 0$) is infinitely differentiable, but its Taylor polynomial is identically zero at $x = 0$. In conclusion, we need to use something other than Taylor series to approximate a general continuous function by polynomials.

The proof of the Weierstrass approximation theorem introduces several useful ideas, but it is typically only proven in the honors class. However, later on, we will prove a version of this theorem for trigonometric polynomials, and this proof will be analogous to the proof of the current theorem.

Theorem 2.52 (**Weierstrass approximation**). *Let $a < b$ be real numbers. Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function, and let $\varepsilon > 0$. Then there exists a polynomial P on $[a,b]$ such that $d_\infty(P,f) < \varepsilon$. (That is, $|f(x) - P(x)| < \varepsilon$ for all $x \in [a,b]$.)*

Remark 2.53. We can also state this Theorem using metric space terminology. Recall that $C([a,b];\mathbb{R})$ is the space of continuous functions from $[a,b]$ to $\mathbb{R}$, equipped with the sup-norm metric d_∞. Let $P([a,b];\mathbb{R})$ be the space of all polynomials on $[a,b]$, so that $P([a,b];\mathbb{R})$ is a subspace of $C([a,b];\mathbb{R})$, since polynomials are continuous. Then the Weierstrass approximation theorem says that every continuous function is an adherent point of $P([a,b];\mathbb{R})$. Put another way, the closure of $P([a,b];\mathbb{R})$ is $C([a,b];\mathbb{R})$.

$$\overline{P([a,b];\mathbb{R})} = C([a,b];\mathbb{R}).$$

Put another way, every continuous function on $[a,b]$ is the uniform limit of polynomials.

2.7. **Power Series.** We now focus our discussion of series to power series.

Definition 2.54 (**Power Series**). Let a be a real number, let $(a_j)_{j=0}^\infty$ be a sequence of real numbers, and let $x \in \mathbb{R}$. A **formal power series centered at** a is a series of the form

$$\sum_{j=0}^\infty a_j (x-a)^j,$$

For a natural number j, we refer to a_j as the j^{th} **coefficient** of the power series.

Remark 2.55. We refer to these power series as formal since their convergence is not guaranteed. Note however that any formal power series centered at a converges at $x = a$. It turns out that we can precisely identify where a formal power series converges just from the asymptotic behavior of the coefficients.

Definition 2.56 (**Radius of Convergence**). Let $\sum_{j=0}^\infty a_j (x-a)^j$ be a formal power series. The **radius of convergence** $R \geq 0$ of this series is defined to be

$$R := \frac{1}{\limsup_{j\to\infty} |a_j|^{1/j}}.$$

In the definition of R, we use the convention that $1/0 = +\infty$ and $1/(+\infty) = 0$. Note that it is possible for R to then take any value between and including 0 and $+\infty$. Note also that R always exists as a nonnegative real number, or as $+\infty$, since the limit superior of a positive sequence always exists as a nonegative number, or $+\infty$.

Example 2.57. The radius of convergence of the series $\sum_{j=0}^\infty j(-2)^j (x-3)^j$ is

$$\frac{1}{\limsup_{j\to\infty} |j(-2)^j|^{1/j}} = \frac{1}{\limsup_{j\to\infty} 2j^{1/j}} = \frac{1}{2}.$$

The radius of convergence of the series $\sum_{j=0}^\infty 2^{j^2} (x+2)^j$ is

$$\frac{1}{\limsup_{j\to\infty} 2^j} = \frac{1}{+\infty} = 0.$$

The radius of convergence of the series $\sum_{j=0}^{\infty} 2^{-j^2}(x+2)^j$ is

$$\frac{1}{\limsup_{j\to\infty} 2^{-j}} = \frac{1}{0} = +\infty.$$

As we now show, the radius of convergence tells us exactly where the power series converges.

Theorem 2.58. *Let $\sum_{j=0}^{\infty} a_j(x-a)^j$ be a formal power series, and let R be its radius of convergence.*

(a) *(Divergence outside of the radius of convergence) If $x \in \mathbb{R}$ satisfies $|x-a| > R$, then the series $\sum_{j=0}^{\infty} a_j(x-a)^j$ is divergent at x.*

(b) *(Convergence inside the radius of convergence) If $x \in \mathbb{R}$ satisfies $|x-a| < R$, then the series $\sum_{j=0}^{\infty} a_j(x-a)^j$ is convergent at x.*

- *For the following items (c),(d) and (e), we assume that $R > 0$. Then, let $f\colon (a-R, a+R)$ be the function $f(x) = \sum_{j=0}^{\infty} a_j(x-a)^j$, which exists by part (b).*

(c) *(Uniform convergence on compact intervals) For any $0 < r < R$, we know that the series $\sum_{j=0}^{\infty} a_j(x-a)^j$ converges uniformly to f on $[a-r, a+r]$. In particular, f is continuous on $(a-R, a+R)$ (by Theorem 2.17.)*

(d) *(Differentiation of power series) The function f is differentiable on $(a-R, a+R)$. For any $0 < r < R$, the series $\sum_{j=0}^{\infty} ja_j(x-a)^{j-1}$ converges uniformly to f' on the interval $[a-r, a+r]$.*

(e) *(Integration of power series) For any closed interval $[y, z]$ contained in $(a-R, a+R)$, we have*

$$\int_y^z f = \sum_{j=0}^{\infty} a_j \frac{(z-a)^{j+1} - (y-a)^{j+1}}{j+1}.$$

Exercise 2.59. Prove Theorem 2.58. (Hints: for parts (a),(b), use the root test. For part (c), use the Weierstrass M-test. For part (d), use Theorem 2.46. For part (e), use Theorem 2.41.)

Remark 2.60. A power series may converge or diverge when $|x-a| = R$.

Exercise 2.61. Give examples of formal power series centered at 0 with radius of convergence $R = 1$ such that

- The series diverges at $x = 1$ and at $x = -1$.
- The series diverges at $x = 1$ and converges at $x = -1$.
- The series converges at $x = 1$ and diverges at $x = 1$.
- The series converges at $x = 1$ and at $x = -1$.

We now discuss functions that are equal to convergent power series.

Definition 2.62. Let $a \in \mathbb{R}$ and let $r > 0$. Let E be a subset of $\mathbb{R}$ such that $(a-r, a+r) \subseteq E$. Let $f\colon E \to \mathbb{R}$. We say that the function f is **real analytic on** $(a-r, a+r)$ if and only if there exists a power series $\sum_{j=0}^{\infty} a_j(x-a)^j$ centered at a with radius of convergence R such that $R \geq r$ and such that this power series converges to f on $(a-r, a+r)$.

Example 2.63. The function $f\colon (0,2) \to \mathbb{R}$ defined by $f(x) = \sum_{j=0}^{\infty} j(x-1)^j$ is real analytic on $(0, 2)$.

From Theorem 2.58, if a function f is real analytic on $(a-r, a+r)$, then f is continuous and differentiable. In fact, f is can be differentiated any number of times, as we now show.

Definition 2.64. Let E be a subset of $\mathbb{R}$. We say that a function $f\colon E \to \mathbb{R}$ is **once differentiable on** E if and only if f is differentiable on E. More generally, for any integer $k \geq 2$, we say that $f\colon E \to \mathbb{R}$ is k **times differentiable on** E, or just k **times differentiable**, if and only if f is differentiable and f' is $k-1$ times differentiable. If f is k times differentiable, we define the k^{th} derivative $f^{(k)}\colon E \to \mathbb{R}$ by the recursive rule $f^{(1)} := f'$ and $f^{(k)} := (f^{(k-1)})'$, for all $k \geq 2$. We also define $f^{(0)} := f$. A function is said to be **infinitely differentiable** if and only if f is k times differentiable for every $k \geq 0$.

Example 2.65. The function $f(x) = |x|^3$ is twice differentiable on $\mathbb{R}$, but not three times differentiable on $\mathbb{R}$. Note that $f''(x) = 6\,|x|$, which is not differentiable at $x = 0$.

Proposition 2.66. *Let $a \in \mathbb{R}$ and let $r > 0$. Let f be a function that is real analytic on $(a-r, a+r)$, with the power series expansion*

$$f(x) = \sum_{j=0}^{\infty} a_j (x-a)^j, \quad \forall\, x \in (a-r, a+r).$$

Then, for any integer $k \geq 0$, the function f is k times differentiable on $(a-r, a+r)$, and the k^{th} derivative is given by

$$f^{(k)}(x) = \sum_{j=0}^{\infty} a_{j+k}(j+1)(j+2)\cdots(j+k)(x-a)^j, \quad \forall\, x \in (a-r, a+r).$$

Exercise 2.67. Prove Proposition 2.66.

Corollary 2.68 (Taylor's formula). *Let $a \in \mathbb{R}$ and let $r > 0$. Let f be a function that is real analytic on $(a-r, a+r)$, with the power series expansion*

$$f(x) = \sum_{j=0}^{\infty} a_j (x-a)^j, \quad \forall\, x \in (a-r, a+r).$$

Then, for any integer $k \geq 0$, we have

$$f^{(k)}(a) = k! a_k,$$

where $k! = 1 \times 2 \times \cdots \times k$, and we denote $0! := 1$. In particular, we have ***Taylor's formula***

$$f(x) = \sum_{j=0}^{\infty} \frac{f^{(j)}(a)}{j!}(x-a)^j, \quad \forall\, x \in (a-r, a+r).$$

Exercise 2.69. Prove Corollary 2.68 using Proposition 2.66.

Remark 2.70. The series $\sum_{j=0}^{\infty} \frac{f^{(j)}(a)}{j!}(x-a)^j$ is sometimes called the **Taylor series** of f around a. Taylor's formula says that if f is real analytic, then f is equal to its Taylor series. In the following exercise, we see that even if f is infinitely differentiable, it may not be equal to its Taylor series.

Exercise 2.71. Define a function $f\colon \mathbb{R} \to \mathbb{R}$ by $f(0) := 0$ and $f(x) := e^{-1/x^2}$ for $x \neq 0$. Show that f is infinitely differentiable, but $f^{(k)}(0) = 0$ for all $k \geq 0$. So, being infinitely differentiable does not imply that f is equal to its Taylor series. (You may freely use properties of the exponential function that you have learned before.)

Corollary 2.72 (**Uniqueness of power series**)**.** *Let $a \in \mathbb{R}$ and let $r > 0$. Let f be a function that is real analytic on $(a-r, a+r)$, with two power series expansions*

$$f(x) = \sum_{j=0}^{\infty} a_j (x-a)^j, \quad \forall\, x \in (a-r, a+r).$$

$$f(x) = \sum_{j=0}^{\infty} b_j (x-a)^j, \quad \forall\, x \in (a-r, a+r).$$

Then $a_j = b_j$ for all $j \geq 0$.

Proof. By Corollary 2.68, we have $k!a_k = f^{(k)}(a) = k!b_k$ for all $k \geq 0$. Since $k! \neq 0$ for all $k \geq 0$, we divide by $k!$ to get $a_k = b_k$ for all $k \geq 0$. □

Remark 2.73. Note however that a power series can have very different expansions if we change the center of the expansion. For example, the function $f(x) = 1/(1-x)$ satisfies

$$f(x) = \sum_{j=0}^{\infty} x^j, \quad \forall\, x \in (-1, 1).$$

However, at the point 1/2, we have the different expansion

$$f(x) = \frac{1}{1-x} = \frac{2}{1-2(x-1/2)} = \sum_{j=0}^{\infty} 2(2(x-1/2))^j = \sum_{j=0}^{\infty} 2^{j+1}(x-1/2)^j, \quad \forall x \in (0, 1).$$

Note also that the first series has radius of convergence 1 and the second series has radius of convergence 1/2.

2.7.1. *Multiplication of Power Series.*

Lemma 2.74 (**Fubini's Theorem for Series**)**.** *Let $f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ be a function such that $\sum_{(j,k)\in\mathbb{N}\times\mathbb{N}} f(j,k)$ is absolutely convergent. (That is, for any bijection $g\colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$, the sum $\sum_{\ell=0}^{\infty} f(g(\ell))$ is absolutely convergent.) Then*

$$\sum_{j=1}^{\infty} (\sum_{k=1}^{\infty} f(j,k)) = \sum_{(j,k)\in\mathbb{N}\times\mathbb{N}} f(j,k) = \sum_{k=1}^{\infty} (\sum_{j=1}^{\infty} f(j,k)).$$

Proof Sketch. We only consider the case $f(j,k) \geq 0$ for all $(j,k) \in \mathbb{N}$. The general case then follows by writing $f = \max(f,0) - \min(f,0)$, and applying this special case to $\max(f,0)$ and $\min(f,0)$, separately.

Let $L := \sum_{(j,k)\in\mathbb{N}\times\mathbb{N}} f(j,k)$. For any $J, K > 0$, we have $\sum_{j=1}^{J} \sum_{k=1}^{K} f(j,k) \leq L$. Letting $J, K \to \infty$, we conclude that $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} f(j,k) \leq L$. Let $\varepsilon > 0$. It remains to find J, K such that $\sum_{j=1}^{J} \sum_{k=1}^{K} > L - \varepsilon$. Since $\sum_{(j,k)\in\mathbb{N}\times\mathbb{N}} f(j,k)$ converges absolutely, there exists a finite set $X \subseteq \mathbb{N} \times \mathbb{N}$ such that $\sum_{(j,k)\in X} f(j,k) > L - \varepsilon$. But then we can choose J, K sufficiently large such that $\{(j,k) \in X\} \subseteq \{(j,k)\colon 1 \leq j \leq J, 1 \leq k \leq K\}$. Therefore, $\sum_{j=1}^{J} \sum_{k=1}^{K} f(j,k) \geq \sum_{(j,k)\in X} f(j,k) > L - \varepsilon$, as desired. □

Theorem 2.75. *Let $a \in \mathbb{R}$ and let $r > 0$. Let f and g be functions that are real analytic on $(a-r, a+r)$, with power series expansions*

$$f(x) = \sum_{j=0}^{\infty} a_j (x-a)^j, \quad \forall\, x \in (a-r, a+r).$$

$$g(x) = \sum_{j=0}^{\infty} b_j (x-a)^j, \quad \forall\, x \in (a-r, a+r).$$

Then the function fg is also real analytic on $(a-r, a+r)$. For each $j \geq 0$, define $c_j := \sum_{k=0}^{j} a_k b_{j-k}$. Then fg has the power series expansion

$$f(x)g(x) = \sum_{j=0}^{\infty} c_j (x-a)^j, \quad \forall\, x \in (a-r, a+r).$$

Proof. Fix $x \in (a-r, a+r)$. By Theorem 2.58, both f and g have radius of convergence $R \geq r$. So, both $\sum_{j=0}^{\infty} a_j (x-a)^j$ and $\sum_{j=0}^{\infty} b_j (x-a)^j$ are absolutely convergent. Define

$$C := \sum_{j=0}^{\infty} \left| a_j (x-a)^j \right|, \qquad D := \sum_{j=0}^{\infty} \left| b_j (x-a)^j \right|.$$

Then both C, D are finite.

For any $N \geq 0$, consider the partial sum

$$\sum_{j=0}^{N} \sum_{k=0}^{N} \left| a_j (x-a)^j b_k (x-a)^k \right|.$$

We can re-write this sum as

$$\sum_{j=0}^{N} \left| a_j (x-a)^j \right| \sum_{k=0}^{N} \left| b_k (x-a)^k \right| \leq \sum_{j=0}^{N} \left| a_j (x-a)^j \right| D \leq CD.$$

Since this inequality holds for all $N \geq 0$, the series

$$\sum_{(j,k) \in \mathbb{N} \times \mathbb{N}} \left| a_j (x-a)^j b_k (x-a)^k \right|$$

is convergent. That is, the following series is absolutely convergent.

$$\sum_{(j,k) \in \mathbb{N} \times \mathbb{N}} a_j (x-a)^j b_k (x-a)^k.$$

Now, using Lemma 2.74,

$$\sum_{(j,k) \in \mathbb{N} \times \mathbb{N}} a_j (x-a)^j b_k (x-a)^k = \sum_{j=0}^{\infty} a_j (x-a)^j \sum_{k=0}^{\infty} b_k (x-a)^k = \sum_{j=0}^{\infty} a_j (x-a)^j g(x) = f(x)g(x).$$

Rewriting this equality,

$$f(x)g(x) = \sum_{(j,k) \in \mathbb{N} \times \mathbb{N}} a_j (x-a)^j b_k (x-a)^k = \sum_{(j,k) \in \mathbb{N} \times \mathbb{N}} a_j b_k (x-a)^{j+k}.$$

Since the sum is absolutely convergent, we can rearrange the order of summation. For any fixed positive integer ℓ, we sum over all positive integers j, k such that $j + k = \ell$. That is, we have

$$f(x)g(x) = \sum_{\ell=0}^{\infty} \sum_{(j,k)\in\mathbb{N}\times\mathbb{N}\colon\, j+k=\ell} a_j b_k (x-a)^{\ell} = \sum_{\ell=0}^{\infty} (x-a)^{\ell} \sum_{s=0}^{\ell} a_s b_{s-\ell}.$$

□

2.8. **The Exponential and Logarithm.** We can now use the material from the previous sections to define and investigate various special functions.

Definition 2.76. For every real number x, we define the **exponential function** $\exp(x)$ to be the real number

$$\exp(x) := \sum_{j=0}^{\infty} \frac{x^j}{j!}.$$

Theorem 2.77 (**Properties of the Exponential Function**).

(a) *For every real number x, the series $\sum_{j=0}^{\infty} \frac{x^j}{j!}$ is absolutely convergent. So,* $\exp(x)$ *exists and is a real number for every $x \in \mathbb{R}$, the power series $\sum_{j=0}^{\infty} \frac{x^j}{j!}$ has radius of convergence $R = +\infty$, and* exp *is an analytic function on $(-\infty, +\infty)$.*

(b) exp *is differentiable on $\mathbb{R}$, and for every $x \in \mathbb{R}$, we have* $\exp'(x) = \exp(x)$.

(c) exp *is continuous on $\mathbb{R}$, and for all real numbers $a < b$, we have* $\int_a^b \exp = \exp(b) - \exp(a)$.

(d) *For every $x, y \in \mathbb{R}$, we have* $\exp(x + y) = \exp(x)\exp(y)$.

(e) $\exp(0) = 1$. *Also, for every $x \in \mathbb{R}$, we have* $\exp(x) > 0$, *and* $\exp(-x) = 1/\exp(x)$.

(f) exp *is strictly monotone increasing. That is, whenever x, y are real numbers with $x < y$, we have* $\exp(x) < \exp(y)$.

Exercise 2.78. Prove Theorem 2.77. (Hints: for part (a), use the ratio test. For parts (b) and (c), use Theorem 2.58. For part (d), you may need the binomial formula $(x + y)^k = \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} x^j y^{k-j}$. For part (e), use part (d). For part (f), use part (d) and show that $\exp(x) > 1$ for all $x > 0$.)

Definition 2.79. We define the real number e by

$$e := \exp(1) = \sum_{j=0}^{\infty} \frac{1}{j!}$$

Proposition 2.80. *For every real number x, we have*

$$\exp(x) = e^x.$$

Exercise 2.81. Prove Proposition 2.80. (Hint: first prove the proposition for natural numbers x. Then, prove the proposition for integers. Then, prove the proposition for rational numbers. Finally, use the density of the rationals to prove the proposition for real numbers. You should find useful identifies for exponentiation by rational numbers.)

From now on, we use $\exp(x)$ and e^x interchangeably.

Remark 2.82. Since $e > 1$ by the definition of e, we have $e^x \to +\infty$ as $x \to +\infty$ and $e^x \to 0$ as $x \to -\infty$. So, from the Intermediate Value Theorem, the range of exp is $(0, \infty)$. Since exp is strictly increasing on $\mathbb{R}$, exp is therefore injective on $\mathbb{R}$, so exp is a bijection from $\mathbb{R}$ to $(0, \infty)$. Therefore, exp has an inverse function from $(0, \infty)$ to $\mathbb{R}$.

Definition 2.83. We define the **natural logarithm function** $\log\colon (0, \infty) \to \mathbb{R}$ (which is also called ln) to be the inverse of the exponential function. So, $\exp(\log(x)) = x$ for every $x \in (0, \infty)$, and $\log(\exp(x)) = x$ for every $x \in \mathbb{R}$.

Remark 2.84. Since exp is continuous and strictly monotone increasing, log is also continuous and strictly monotone increasing by Proposition 2.2. Since exp is differentiable and its derivative is never zero, the Inverse Function Theorem (Theorem 2.3) implies that log is also differentiable.

Theorem 2.85.

(a) *For every* $x \in (0, \infty)$, *we have* $\log'(x) = 1/x$. *So, by the Fundamental Theorem of Calculus, for any* $0 < a < b$, *we have* $\int_a^b (1/t)dt = \log(b) - \log(a)$.
(b) *For all* $x, y \in (0, \infty)$, *we have* $\log(x) + \log(y) = \log(xy)$.
(c) *For all* $x \in (0, \infty)$, *we have* $\log(1/x) = -\log x$. *In particular,* $\log(1) = 0$.
(d) *For any* $x \in (0, \infty)$ *and* $y \in \mathbb{R}$, *we have* $\log(x^y) = y \log x$.
(e) *For any* $x \in (-1, 1)$, *we have*

$$-\log(1 - x) = \sum_{j=1}^{\infty} \frac{x^j}{j}.$$

In particular, log *is analytic on* $(0, 2)$ *with the power series expansion*

$$\log(x) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j}(x - 1)^j, \quad \forall\, x \in (0, 2).$$

Exercise 2.86. Prove Theorem 2.85. (Hints: for part (a), use the Inverse Function Theorem or Chain Rule. For parts (b),(c) and (d), use Theorem 2.77 and the laws of exponentiation. For part (e), let $x \in (-1, 1)$, use the geometric series formula $1/(1 - x) = \sum_{j=0}^{\infty} x^j$ and integrate using Theorem 2.58.)

2.8.1. *A Digression concerning Complex Numbers.* Our investigation of trigonometric functions below is significantly improved by the introduction of the complex number system. We will also use the complex exponential in our discussion of Fourier series.

Definition 2.87 (**Complex Numbers**)**.** A **complex number** is any expression of the form $a + bi$ where a, b are real numbers. Formally, the symbol i is a placeholder with no intrinsic meaning. Two complex numbers $a + bi$ and $c + di$ are said to be equal if and only if $a = c$ and $b = d$. Every real number x is considered a complex number, with the identification $x = x + 0i$. The sum of two complex numbers is defined by $(a+bi)+(c+di) := (a+c)+(b+d)i$. The difference of two complex numbers is defined by $(a+bi)-(c+di) := (a-c)+(b-d)i$. The product of two complex numbers is defined by $(a+bi)(c+di) := (ac-bd)+(ad+bc)i$. If $c+di \neq 0$, the quotient of two complex numbers is defined by $(a+bi)/(c+di) := (a+bi)(\frac{c}{c^2+d^2} - \frac{d}{c^2+d^2}i)$. The **complex conjugate** of a complex number $a + bi$ is defined by $\overline{a + bi} := a - bi$. The absolute value of a complex number $a + bi$ is defined by $|a + bi| := \sqrt{a^2 + b^2}$. The space of all complex numbers is called $\mathbb{C}$.

Remark 2.88. We write i as shorthand for $0 + i$. Note that $i^2 = -1$.

Remark 2.89. The complex numbers obey all of the usual rules of algebra. For example, if v, w, z are complex numbers, then $v(w+z) = vw+vz$, $v(wz) = (vw)z$, and so on. Specifically, the complex numbers $\mathbb{C}$ form a **field**. Also, the rules of complex arithmetic are consistent with the rules of real arithmetic. That is, $3 + 5 = 8$ whether or not we use addition in $\mathbb{R}$ or addition in $\mathbb{C}$.

The operation of complex conjugation preserves all of the arithmetic operations. If w, z are complex numbers, then $\overline{w+z} = \overline{w} + \overline{z}$, $\overline{w-z} = \overline{w} - \overline{z}$, $\overline{w \cdot z} = \overline{w} \cdot \overline{z}$, and $\overline{w/z} = \overline{w}/\overline{z}$ for $z \neq 0$. The complex conjugate and absolute value satisfy $|z|^2 = z\overline{z}$.

Remark 2.90. If $z \in \mathbb{C}$, then $|z| = 0$ if and only if $z = 0$. If $z, w \in \mathbb{C}$, then it can be shown that $|zw| = |z|\,|w|$, and if $w \neq 0$, then $|z/w| = |z| / |w|$. Also, the triangle inequality holds: $|z + w| \leq |z| + |w|$. So, $\mathbb{C}$ is a metric space if we use the metric $d(z, w) := |z - w|$. Moreover, $\mathbb{C}$ is a complete metric space.

The theory we have developed to deal with series of real functions also covers complex-valued functions, with almost no change to the proofs. For example, we can define the exponential function of a complex number z by

$$\exp(z) := \sum_{j=0}^{\infty} \frac{z^j}{j!}.$$

The ratio test then can be proven in exactly the same manner for complex series, and it follows that $\exp(z)$ converges for every $z \in \mathbb{C}$. Many of the properties of Theorem 2.77 still hold, though we cannot deal with all of these properties in this class. However, the following identity is proven in the exact same way as in the setting of real numbers: for any $z, w \in \mathbb{C}$, we have

$$\exp(z + w) = \exp(z)\exp(w).$$

Also, we should note that $\overline{\exp(z)} = \exp(\overline{z})$, which follows by conjugating the partial sums $\sum_{j=0}^{J} z^j/j!$, and then letting $J \to \infty$.

We briefly mention that the complex logarithm is more difficult to define, mainly because the exponential function is not invertible on $\mathbb{C}$. This topic is deferred to the complex analysis class.

2.9. **Trigonometric Functions.** Besides the exponential and logarithmic functions, there are many different kinds of special functions. Here, we will only mention the sine and cosine functions. One's first encounter with the sine and cosine functions probably involved their definition in terms of the edge lengths of right triangles. However, we will show below an analytic definition of these functions, which will also facilitate the investigation of the properties that they possess. The complex exponential plays a crucial role in this development.

Definition 2.91. Let x be a real number. We then define

$$\cos(x) := \frac{e^{ix} + e^{-ix}}{2}.$$

$$\sin(x) := \frac{e^{ix} - e^{-ix}}{2i}.$$

We refer to cos as the **cosine** function, and we refer to sin as the **sine** function.

Remark 2.92. Using the power series expansion for the exponential, we can then derive power series expansions for sine and cosine as follows. Let $x \in \mathbb{R}$. Then

$$e^{ix} = 1 + ix - x^2/2! - ix^3/3! + x^4/4! + \cdots$$

$$e^{-ix} = 1 - ix - x^2/2! + ix^3/3! + x^4/4! - \cdots$$

Therefore, using the definitions of sine and cosine,

$$\cos(x) = 1 - x^2/2! + x^4/4! - \cdots = \sum_{j=0}^{\infty} \frac{(-1)^j x^{2j}}{(2j)!}.$$

$$\sin(x) = x - x^3/3! + x^5/5! - \cdots = \sum_{j=0}^{\infty} \frac{(-1)^j x^{2j+1}}{(2j+1)!}.$$

So, if $x \in \mathbb{R}$ then $\cos(x) \in \mathbb{R}$ and $\sin(x) \in \mathbb{R}$. Also, sine and cosine real analytic on $(-\infty, \infty)$, e.g. since their power series converge on $(-\infty, \infty)$ by the ratio test. In particular, the sine and cosine functions are continuous and infinitely differentiable.

Theorem 2.93 (Properties of Sine and Cosine).

(a) *For any real number x we have $\cos(x)^2 + \sin(x)^2 = 1$. In particular, $\sin(x) \in [-1, 1]$ and $\cos(x) \in [-1, 1]$ for all real numbers x.*
(b) *For any real number x, we have $\sin'(x) = \cos(x)$, and $\cos'(x) = -\sin(x)$.*
(c) *For any real number x, we have $\sin(-x) = -\sin(x)$ and $\cos(-x) = \cos(x)$.*
(d) *For any real numbers x, y we have $\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$ and $\sin(x + y) = \sin(x)\cos(y) + \cos(x)\sin(y)$.*
(e) *$\sin(0) = 0$ and $\cos(0) = 1$.*
(f) *For every real number x, we have $e^{ix} = \cos(x) + i\sin(x)$ and $e^{-ix} = \cos(x) - i\sin(x)$.*

Exercise 2.94. Prove Theorem 2.93. (Hints: whenever possible, write everything in terms of exponentials.)

Lemma 2.95. *There exists a positive real number x such that $\sin(x) = 0$.*

Proof. We argue by contradiction. Suppose $\sin(x) \neq 0$ for all $x > 0$. We conclude that $\cos(x) \neq 0$ for all $x > 0$, since $\cos(x) = 0$ implies that $\sin(2x) = 0$, by Theorem 2.93(d). Since $\cos(0) = 1$, we conclude that $\cos(x) > 0$ for all $x > 0$ by the Intermediate Value Theorem. Since $\sin(0) = 0$ and $\sin'(0) = 1 > 0$, we know that sin is positive for small positive x. Therefore, $\sin(x) > 0$ for all $x > 0$ by the Intermediate Value Theorem.

Define $\cot(x) := \cos(x)/\sin(x)$. Then cot is positive on $(0, \infty)$, and cot is differentiable for $x > 0$. From the quotient rule and Theorem 2.93(a), we have $\cot'(x) = -1/\sin^2(x)$. So, $\cot'(x) \leq -1$ for all $x > 0$. Then, by the Fundamental Theorem of Calculus, for all $x, s > 0$, we have $\cot(x + s) \leq \cot(x) - s$. Letting $s \to \infty$ shows that cot eventually becomes negative on $(0, \infty)$, a contradiction. □

Let E be the set $E := \{x \in (0, \infty) \colon \sin(x) = 0\}$, so that E is the set of zeros of the sine function. By Lemma 2.95, E is nonempty. Also, since sin is continuous, E is a closed set. (Note that $E = \sin^{-1}(0)$.) In particular, E contains all of its adherent points, so E contains $\inf(E)$.

Definition 2.96. We define π to be the number

$$\pi := \inf\{x \in (0,\infty) \colon \sin(x) = 0\}.$$

Then $\pi > 0$ and $\sin(\pi) = 0$. Since sin is nonzero on $(0,\pi)$ and $\sin'(0) = 1 > 0$, we conclude that sin is positive on $(0,\pi)$. Since $\cos'(x) = -\sin(x)$, we see that cos is decreasing on $(0,\pi)$. Since $\cos(0) = 1$, we therefore have $\cos(\pi) < 1$. Since $\sin^2(\pi) + \cos^2(\pi) = 1$ and $\sin(\pi) = 0$, we conclude that $\cos(\pi) = -1$.

We therefore deduce Euler's famous formula

$$e^{i\pi} = \cos(\pi) + i\sin(\pi) = -1.$$

Here are some more properties of sine and cosine.

Theorem 2.97.

(a) *For any real x we have* $\cos(x+\pi) = -\cos(x)$ *and* $\sin(x+\pi) = -\sin(x)$. *In particular, we have* $\cos(x+2\pi) = \cos(x)$ *and* $\sin(x+2\pi) = \sin(x)$, *so that* sin *and* cos *are* 2π-*periodic.*

(b) *If x is real, then* $\sin(x) = 0$ *if and only if x/π is an integer.*

(c) *If x is real, then* $\cos(x) = 0$ *if and only if x/π is an integer plus* $1/2$.

Exercise 2.98. Prove Theorem 2.97.

3. Fourier series

A general problem in analysis is to approximate a general function by a series that is relatively easy to describe. With the Weierstrass Approximation theorem, we saw that it is possible to achieve this goal by approximating compactly supported continuous functions by polynomials. The notion of approximation here uses the sup norm. After our discussion of general series of functions, we focused on power series. In this case, real analytic functions can be written exactly in terms of their power series expansions. However, power series do not provide the best approximations for general functions.

There is a different notion of approximation of general functions which we will now discuss. We will focus on periodic functions, and we will try to approximate these functions by trigonometric polynomials. As before, there are many choices of metrics in which we can say how close the approximating function is to the original function. These issues will be dealt with below, in our discussion of Fourier series. The topic of Fourier analysis can occupy more than one course, so we only select the introductory parts in this course.

3.1. Periodic Functions.

Fourier series begins with the analysis of complex-valued, periodic functions.

Definition 3.1. Let $L > 0$ be a real number. A function $f \colon \mathbb{R} \to \mathbb{C}$ is **periodic with period** L, or $L-$**periodic**, if and only if $f(x+L) = f(x)$ for every real number x.

Example 3.2. The functions $\sin(x)$, $\cos(x)$ and e^{ix} are all 2π-periodic. They are also 4π-periodic, 6π-periodic, and so on. The function $f(x) = x$ is not periodic. The constant function $f(x) = 1$ is L-periodic for every $L > 0$.

Remark 3.3. If a function $f \colon \mathbb{R} \to \mathbb{C}$ is L-periodic, then $f(x+kL) = f(x)$ for every integer k. In particular, if f is 1-periodic, then $f(x+k) = f(x)$ for every integer k. So, 1-periodic functions are sometimes called $\mathbb{Z}$-**periodic functions** (and L-periodic functions are sometimes called $L\mathbb{Z}$-periodic functions.)

Example 3.4. For any integer n, the functions $\sin(2\pi nx)$, $\cos(2\pi nx)$ and $e^{2\pi inx}$ are $\mathbb{Z}$-periodic. For another example, consider the function where $f(x) = 1$ when $x \in [n, n + 1/2)$ for any integer n, and $f(x) = -1$ when $x \in [n + 1/2, n + 1)$ for any integer n. This function is an example of a **square wave**.

Remark 3.5. For simplicity, we will only deal with $\mathbb{Z}$-periodic functions below. The theory of general L-periodic functions follows relatively easily once the $\mathbb{Z}$-periodic theory has been developed. Note that a $\mathbb{Z}$-periodic function f is entirely determined by its values on the interval $[0, 1)$, since any $x \in \mathbb{R}$ can be written as $x = k + y$ where $k \in \mathbb{Z}$ and $y \in [0, 1)$, so that $f(x) = f(k + y) = f(y)$. Consequently, we sometimes describe a $\mathbb{Z}$-periodic function f by defining the function f on $[0, 1)$, and we then say that f is **extended periodically** by setting $f(k + y) = f(x) := f(y)$. (As before, we write $x \in \mathbb{R}$ as $x = k + y$ where $k \in \mathbb{Z}$ and $y \in [0, 1)$.)

The space of continuous complex-valued $\mathbb{Z}$-periodic functions is denoted by $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$. The notation $\mathbb{R}/\mathbb{Z}$ comes from algebra, where we consider the quotient of the additive group $\mathbb{R}$ by the additive group $\mathbb{Z}$. When we say that f is continuous and $\mathbb{Z}$-periodic, we mean that f is continuous on all of $\mathbb{R}$. If f is only continuous on the interval $[0, 1]$, then f may have a discontinuity at 0, so f may not be in $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$. For any integer n, the functions $\sin(2\pi nx)$, $\cos(2\pi nx)$ and $e^{2\pi inx}$ are in $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$. However, the square wave is not in $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$. Also, $\sin(x)$ is not in $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ since it is not $\mathbb{Z}$-periodic.

Lemma 3.6.

(a) *(Continuous periodic functions are bounded.) If $f \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$, then f is bounded. (That is, given $f \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$, there exists $M > 0$ such that $|f(x)| \leq M$ for all $x \in \mathbb{R}$.)*

(b) *(Continuous periodic functions form a vector space and an algebra.) Let $f, g \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$. Then $f + g$, $f - g$ and fg are all in $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$. Also, if $c \in \mathbb{C}$, then $cf \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$.*

(c) *(Uniform limits of continuous periodic functions are continuous periodic.) Let $(f_j)_{j=1}^{\infty}$ be a sequence of functions in $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ which converges uniformly to a function $f: \mathbb{R} \to \mathbb{C}$. Then $f \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$.*

Exercise 3.7. Prove Lemma 3.6. (Hint: for (i), first show that f is bounded on $[0, 1]$.)

Remark 3.8. $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ becomes a metric space by re-introducing the sup-norm metric. Let $f, g \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$, and define

$$d_\infty(f, g) := \sup_{x \in \mathbb{R}} |f(x) - g(x)| = \sup_{x \in [0,1)} |f(x) - g(x)| .$$

In fact, $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ is a normed linear space with the norm

$$\|f\|_\infty := \sup_{x \in \mathbb{R}} |f(x)| = \sup_{x \in [0,1)} |f(x)| .$$

One can also show that $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ is a complete metric space.

3.2. **Inner Products on Periodic Functions.** We just discussed how to make $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ a normed linear space. We can also realize $\mathbb{C}(\mathbb{R}/\mathbb{Z};\mathbb{C})$ as a complex inner product space. However, the norm that is induced by this inner product will be *different* then the sup-norm. As we have mentioned above, there are many different norms in which we deal with functions. In this particular case, the most natural norm will not be the sup-norm. Instead, we will see that the norm that comes from the inner product will be more natural. We will discuss this issue further below, but for now we begin by defining the inner product on $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$.

Definition 3.9. Let $f,g\in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$. We define the (complex) **inner product** $\langle f,g\rangle$ to be the quantity

$$\langle f,g\rangle := \int_0^1 f(x)\overline{g(x)}dx.$$

Exercise 3.10. Verify that this inner product on $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ satisfies the axioms of a complex inner product space.

Remark 3.11. In order to integrate a general complex-valued function of the form $f(x) = g(x)+ih(x)$ where $h(x),g(x)\in\mathbb{R}$ for all $x\in[a,b]$, we define $\int_a^b f := (\int_a^b g) + i(\int_a^b h)$. For example,

$$\int_0^1 (1+ix) = 1+i(\int_0^1 xdx) = 1+i/2.$$

One can verify that all standard rules of calculus (integration by parts, the fundamental theorem of calculus, substitution, etc.) still hold when the function is complex-valued instead of real-valued.

Example 3.12. Let $f,g\in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ be the functions $f(x)=1$ and $g(x)=e^{2\pi ix}$, for all $x\in\mathbb{R}$. Then

$$\langle f,g\rangle = \int_0^1 \overline{e^{2\pi ix}}dx = \int_0^1 e^{-2\pi ix}dx = \frac{e^{-2\pi ix}}{-2\pi i}|_{x=0}^{x=1} = \frac{e^{-2\pi i}-1}{-2\pi i} = 0.$$

Remark 3.13. Let $f,g\in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$. In general, $\langle f,g\rangle$ will be a complex number. Note also that since f,g are bounded and continuous, the function $f\overline{g}$ is Riemann integrable.

Definition 3.14. Let $f,g\in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$. From Exercise 3.10, we see that $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ is a complex inner product space when equipped with the inner product $\langle f,g\rangle = \int_0^1 f\overline{g}$. So, from Exercise 1.13, we recall that this inner product makes $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ a normed linear space. We refer to this norm $\|f\|_2$ as the L_2**-norm** of f:

$$\|f\|_2 := \sqrt{\langle f,f\rangle} = \left(\int_0^1 f(x)\overline{f(x)}dx\right)^{1/2} = \left(\int_0^1 |f(x)|^2\,dx\right)^{1/2}.$$

The norm $\|f\|_2$ is sometimes called the **root mean square** of f.

Example 3.15. Let $f(x):=e^{2\pi ix}$. Then

$$\|f\|_2 = (\int_0^1 e^{2\pi ix}e^{-2\pi ix}dx)^{1/2} = (1)^{1/2} = 1.$$

Exercise 3.16. Let $M > 0$ be any positive real number. Find a function $f \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ such that $\|f\|_2 \leq 1$ but such that $\|f\|_\infty > M$. On the other hand, if $g \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$, show that $\|g\|_2 \leq \|g\|_\infty$. So, the L_2 and sup-norms on $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ are related, but they can also be very different.

Definition 3.17. Due to Pythagoras's Theorem (Exercise 1.20), if $f, g \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ satisfy $\langle f, g\rangle = 0$, we sometimes say that f, g are **orthogonal**.

Definition 3.18. Let $f, g \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$. From Exercise 1.13, we recall that the inner product on $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ also gives a metric on $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$. We refer to this metric d_{L_2} as the L_2-metric:

$$d_{L_2}(f,g) := \sqrt{\langle (f-g), (f-g)\rangle} = \|f-g\|_2 = \left(\int_0^1 |f(x)-g(x)|^2\,dx\right)^{1/2}.$$

The L_2 metric shares many characteristics with the ℓ_2-metric on $\mathbb{R}^n$.

Remark 3.19. A sequence of functions $(f_j)_{j=1}^\infty$ in $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ will **converge in the L_2 metric** to $f \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ if and only if $d_{L_2}(f_j, f) \to 0$ as $j \to \infty$. Equivalently,

$$\lim_{j\to\infty} \int_0^1 |f_j(x) - f(x)|^2\,dx = 0.$$

As we now show, convergence in the L_2 metric is different than both uniform convergence and pointwise convergence.

Exercise 3.20. Let $(f_j)_{j=1}^\infty$ be a sequence of functions in $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$, and let f be another function in $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$.

- Show that if $(f_j)_{j=1}^\infty$ converges uniformly to f, then $(f_j)_{j=1}^\infty$ also converges to f in the L_2 metric.
- Find a sequence $(f_j)_{j=1}^\infty$ in $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ which converges to some $f \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ in the L_2 metric, so that $(f_j)_{j=1}^\infty$ does not converge to f uniformly. (Hint: consider $f = 0$ and use Exercise 3.16.)
- Find a sequence $(f_j)_{j=1}^\infty$ in $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ which converges to some $f \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ in the L_2 metric, so that $(f_j)_{j=1}^\infty$ does not converge pointwise to f. (Hint: consider $f = 0$ and try to make the functions f_j large at one point.)
- Find a sequence $(f_j)_{j=1}^\infty$ in $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ which converges to some $f \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ pointwise, so that $(f_j)_{j=1}^\infty$ does not converge to f in the L_2 metric. (Hint: consider $f = 0$ and try to make the functions f_j large in L_2 norm.)

Remark 3.21. Even though $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ is complete with respect to the sup-norm metric, it turns out that $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ is not complete with respect to the L_2 metric. For example, try to find a sequence of continuous functions that converges to the (discontinuous) square wave function.

3.3. **Trigonometric Polynomials.** In the theory of power series, we approximated functions by linear combinations of the monomials x^n where n is a positive integer. Now, in our discussion of Fourier series, we will approximate functions by linear combinations of the functions $e^{2\pi i n x}$ where $n \in \mathbb{Z}$. The functions $e^{2\pi i n x}$ are sometimes called **characters**.

To keep some simplicity in our notation, we make the following definition.

Definition 3.22. For every integer n, let $e_n \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ denote the function

$$e_n(x) := e^{2\pi i n x}, \quad x \in \mathbb{R}.$$

We sometimes refer to e_n as the **character with frequency** n.

Definition 3.23. A function $f \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ is said to be a **trigonometric polynomial** if and only if there exists an integer $N \geq 0$ and there exists a sequence of complex numbers $(c_n)_{n=-N}^{N}$ such that

$$f = \sum_{n=-N}^{N} c_n e_n.$$

Example 3.24. The function $f = 2e_{-2} + 1 + 3e_1$ is a trigonometric polynomial. More explicitly, for all $x \in \mathbb{R}$, we have

$$f(x) = 2e^{-4\pi i x} + 1 + 3e^{2\pi i x}.$$

Example 3.25. For any integer n, the function $\cos(2\pi n x)$ is a trigonometric polynomial, since $\cos(2\pi n x) = (1/2)(e^{2\pi i n x} + e^{-2\pi i n x})$. Similarly, for any integer n, the function $\sin(2\pi n x)$ is a trigonometric polynomial, since $\sin(2\pi n x) = (1/(2i))(e_n(x) - e_{-n}(x))$. In particular, any linear combination of sines and cosines of this form is a trigonometric polynomial. For example, $\sin(4\pi x) + 3i\cos(2\pi x)$ is a trigonometric polynomial.

Remark 3.26. It turns out that *any* function in $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ can be written as an infinite sum of characters. That is, any function in $C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ can be written as its Fourier series. The analogous statement for power series is that any real analytic function is equal to its power series.

The key fact used in proving this statement is given by the following computation.

Lemma 3.27 (**Characters are an Orthonormal System**). *Let n, m be integers. If $n = m$, then $\langle e_n, e_m \rangle = 1$. If $n \neq m$, then $\langle e_n, e_m \rangle = 0$. Also, $\|e_n\|_2 = 1$.*

Exercise 3.28. Prove Lemma 3.27.

Consequently, there is a nice formula to find the coefficients of a trigonometric polynomial.

Corollary 3.29. *Let $f = \sum_{n=-N}^{N} c_n e_n$ be a trigonometric polynomial. Then, for all integers $-N \leq n \leq N$, we have*

$$c_n = \langle f, e_n \rangle.$$

Also, for any integer n with $|n| > N$, we have $\langle f, e_n \rangle = 0$. And, we have the identity

$$\|f\|_2^2 = \sum_{n=-N}^{N} |c_n|^2.$$

Exercise 3.30. Prove Corollary 3.29. (Hint: for the final identity, use either the Pythagorean Theorem and induction, or substitute $f = \sum_{n=-N}^{N} c_n e_n$ into $\|f\|_2^2$ and expand out all of the terms.)

Definition 3.31. Let $f \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$, and let $n \in \mathbb{Z}$. We define the n^{th} **Fourier coefficient of** f, denoted $\widehat{f}(n)$, to be the complex number

$$\widehat{f}(n) := \langle f, e_n \rangle = \int_0^1 f(x)e^{-2\pi i n x}dx.$$

The function $\widehat{f}\colon \mathbb{Z} \to \mathbb{C}$ is called the **Fourier transform** of f.

We now restate Corollary 3.29. From Corollary 3.29, whenever $f = \sum_{n=-N}^{N} c_n e_n$ is a trigonometric polynomial, we have

$$f = \sum_{n=-N}^{N} \langle f, e_n \rangle e_n = \sum_{n=-\infty}^{\infty} \langle f, e_n \rangle e_n.$$

That is, we have the **Fourier inversion formula**

$$f = \sum_{n=-\infty}^{\infty} \widehat{f}(n) e_n.$$

Put another way, for all $x \in \mathbb{R}$,

$$f(x) = \sum_{n=-\infty}^{\infty} \widehat{f}(n) e^{2\pi i n x}.$$

Also, from the second part of Corollary 3.29, we have **Plancherel's formula**

$$\|f\|_2^2 = \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2.$$

Note that even though we have written these sums as infinite sums, they are actually finite sums, so there is no issue talking about their convergence. Below, we will extend the Fourier inversion and Plancherel formulas to general functions in $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$. These formulas hold on even larger classes of functions, but we may not have time to elaborate on this point. To prove the Fourier inversion formula, we will need a version of the Weierstrass approximation theorem for trigonometric polynomials. This proof will be analogous to the proof of the previous case of Weierstrass approximation (which we omitted).

3.4. **Periodic Convolutions.** In this section we will prove the following theorem.

Theorem 3.32 (Weierstrass approximation theorem for trigonometric polynomials). *Let $f \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ and let $\varepsilon > 0$. Then there exists a trigonometric polynomial P such that $\|f - P\|_\infty < \varepsilon$.*

In other words, any continuous periodic function can be uniformly approximated by trigonometric polynomials. In other words, if we let $P(\mathbb{R}/\mathbb{Z};\mathbb{C})$ denote the space of all trigonometric polynomials, then the closure of $P(\mathbb{R}/\mathbb{Z};\mathbb{C})$ in the L_∞ metric is $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$.

The main tool in proving the Weierstrass approximation theorem is convolution.

Definition 3.33 (Convolution). Let $f, g \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$. Then we define the **periodic convolution** $f * g\colon \mathbb{R} \to \mathbb{C}$ of f and g by the formula

$$f * g(x) := \int_0^1 f(y)g(x-y)dy, \qquad x \in \mathbb{R}.$$

For a fixed $x \in \mathbb{R}$, we can consider $f * g(x)$ to be a shifted average of the values of f and g. Below, we will see that the convolution can also be understood by looking at the Fourier transform of $f * g$.

Lemma 3.34 (**Properties of Convolution**). *Let $f, g, h \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$.*

(a) *The convolution $f * g$ is continuous and $\mathbb{Z}$-periodic. That is, $f * g \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$.*
(b) *$f * g = g * f$.*
(c) *$f * (g + h) = f * g + f * h$ and $(f + g) * h = f * g + g * h$. For any complex number c, we have $c(f * g) = (cf) * g = f * (cg)$.*

Exercise 3.35. Prove Lemma 3.34. (Hints: to prove (a), you may need to use the uniform continuity of f and the boundedness of g, or vice versa. To prove $f * g = g * f$, you may need to use periodicity to "cut and paste" the interval $[0, 1]$.)

Let $f \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ and let $n \in \mathbb{Z}$. Then

$$f * e_n = \widehat{f}(n) e_n.$$

Indeed, note that

$$\begin{aligned} f * e_n &= \int_0^1 f(y) e^{2\pi i n(x-y)} dy = e^{2\pi i n x} \int_0^1 f(y) e^{-2\pi i n y} dy \\ &= e^{2\pi i n x} \widehat{f}(n) = \widehat{f}(n) e_n. \end{aligned}$$

More generally, from Lemma 3.34(c), if $P = \sum_{n=-N}^{N} c_n e_n$ is any trigonometric polynomial, then

$$f * P = \sum_{n=-N}^{N} c_n (f * e_n) = \sum_{n=-N}^{N} \widehat{f}(n) c_n e_n \qquad (\ddagger)$$

In particular, we have the following

Lemma 3.36. *Let $f \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$ and let P be a trigonometric polynomial. Then $f * P$ is also a trigonometric polynomial.*

We can actually rewrite the equation $(\ddagger)$ as

$$\widehat{f * P}(n) = \widehat{f}(n) c_n = \widehat{f}(n) \widehat{P}(n).$$

In fact, even more generally, for any $f, g \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$, we have

$$\widehat{f * g}(n) = \widehat{f}(n) \widehat{g}(n). \qquad (\dagger)$$

In particular, even though the convolution can look a bit strange, we can instead interpret $f * g$ as simply being a function whose Fourier transform multiplies the Fourier transforms of f and g. The identity $(\dagger)$ is very important, though we will not use it this course.

Our strategy for proving the Weierstrass approximation theorem is the following. We will find a trigonometric polynomial P such that $f * P$ is close to f. From Lemma 3.36, we know that $f * P$ is a trigonometric polynomial. So, the strategy reduces to finding a trigonometric polynomial P such that $f * P$ is close to f. Since $f * P$ can be considered an average of the values of f and P, it turns out that we want to choose the polynomial P to be a positive function whose integral on $[0, 1]$ is mostly concentrated at a single point in $[0, 1]$. So, if P is very concentrated, then $f * P(x)$ will be mostly an average of the values of f near x. Then, the (uniform) continuity of f will guarantee that this average will be close to $f(x)$.

With this strategy in mind, we therefore look for a polynomial P that is positive and mostly concentrated at a single point. We call such a function an approximation of the identity.

Definition 3.37. Let $\varepsilon > 0$ and let $0 < \delta < 1/2$. A function $f \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ is said to be an (ε, δ) **approximation of the identity** if and only if the following properties hold:

- $f(x) \geq 0$ for all $x \in \mathbb{R}$, and $\int_0^1 f = 1$.
- $f(x) < \varepsilon$ for all x with $\delta \leq |x| \leq 1 - \delta$.

Lemma 3.38. *For every $\varepsilon > 0$ and $0 < \delta < 1/2$, there exists a trigonometric polynomial P such that P is an (ε, δ) approximation of the identity.*

Proof. Let $N \geq 1$. We define the **Fejér kernel** F_N to be the function

$$F_N := \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) e_n.$$

Note that F_N is a trigonometric polynomial. Also,

$$\begin{aligned} F_N &= \frac{1}{N} \sum_{\ell=-N}^{N} (N - |\ell|) e_\ell = \frac{1}{N} \sum_{\ell=-N}^{N} \left(\sum_{-N+1 \leq j \leq 0 \leq k \leq N-1\colon\, j+k=\ell} e_\ell \right) \\ &= \frac{1}{N} \sum_{\ell=-N}^{N} \left(\sum_{-N+1 \leq j \leq 0 \leq k \leq N-1\colon\, j+k=\ell} e_j e_k \right) = \frac{1}{N} \sum_{k=0}^{N-1} e_k \overline{\big(\sum_{j=0}^{N-1} e_j\big)} = \frac{1}{N} \left| \sum_{n=0}^{N-1} e_n \right|^2. \qquad (*) \end{aligned}$$

And from the geometric series formula, for any real x that is not an integer,

$$\sum_{n=0}^{N-1} e_n(x) = \frac{e^{2\pi i N x} - 1}{e^{2\pi i x} - 1} = \frac{e^{\pi i N x}}{e^{\pi i x}} \cdot \frac{e^{\pi i N x} - e^{-\pi i N x}}{e^{\pi i x} - e^{-\pi i x}} = \frac{e^{\pi i (N-1) x} \sin(N \pi x)}{\sin(\pi x)}.$$

Combining this equation with $(*)$, we have for any real x that is not an integer,

$$F_N(x) = \frac{\sin^2(N\pi x)}{N \sin^2(\pi x)}. \qquad (**)$$

Also, when x is an integer, we see directly from $(*)$ that $F_N(x) = N$. So, $F_N(x) \geq 0$ for all real x. And by the definition of F_N and Lemma 3.27,

$$\int_0^1 F_N = \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) \int_0^1 e_n = (1 - 0/N)1 = 1.$$

Finally, since $|\sin(\pi N x)| \leq 1$, we conclude by $(**)$ that, if $\delta \leq |x| \leq 1 - \delta$, then

$$F_N(x) \leq \frac{1}{N \sin^2(\pi x)} \leq \frac{1}{N \sin^2(\pi \delta)}.$$

The last inequality follows since sin is increasing on $[0, \pi/2]$, it is decreasing on $[\pi/2, \pi]$, and $\sin(\pi\delta) = \sin(-\pi\delta) = \sin(\pi(1-\delta))$, which uses that sin is odd and $\mathbb{Z}$-periodic. So, by choosing N large enough, we have $|F_N(x)| < \varepsilon$ for all x with $\delta \leq |x| \leq 1 - \delta$. $\square$

We can now prove the Weierstrass approximation theorem.

Proof of Theorem 3.32. Let $f \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$. Then f is bounded, so there exists $M > 0$ such that $|f(x)| \leq M$ for all $x \in \mathbb{R}$.

Let $\varepsilon > 0$. Since f is uniformly continuous, there exists $\delta > 0$ such that, if $|x - y| < \delta$, then $|f(x) - f(y)| < \varepsilon$. By Lemma 3.38, let P be a trigonometric polynomial that is also an (ε, δ) approximation of the identity. Then $f * P$ is also a trigonometric polynomial by Lemma 3.36. So, it remains to show that $\|f - f * P\|_\infty < \varepsilon(2M + 2)$. Let $x \in \mathbb{R}$. Then

$$\begin{aligned}
|f(x) - f * P(x)| &= \left| f(x) - \int_0^1 P(y) f(x-y) dy \right| \\
&= \left| \int_0^1 P(y) f(x) dy - \int_0^1 P(y) f(x-y) dy \right| \qquad \text{, since } \int_0^1 P(y) dy = 1 \\
&= \left| \int_0^1 P(y) [f(x) - f(x-y)] dy \right| \\
&\leq \int_0^1 P(y) \left| f(x) - f(x-y) \right| dy \qquad \text{, using } P(y) \geq 0 \\
&= \int_0^\delta P(y) \left| f(x) - f(x-y) \right| dy + \int_\delta^{1-\delta} P(y) \left| f(x) - f(x-y) \right| dy \\
&\qquad + \int_{1-\delta}^1 P(y) \left| f(x) - f(x-y) \right| dy \\
&\leq \int_0^\delta P(y) \varepsilon dy + \int_\delta^{1-\delta} P(y) 2M dy + \int_{1-\delta}^1 P(y) \left| f(x-1) - f(x-y) \right| dy \\
&\leq \varepsilon + 2M\varepsilon + \int_{1-\delta}^1 P(y) \varepsilon dy \leq \varepsilon(2M + 2).
\end{aligned}$$

In conclusion, $\|f - f * P\|_\infty \leq \varepsilon(2M + 2)$. Since $\varepsilon > 0$ is arbitrary, we can find $f * P$ arbitrarily close to f in the sup norm, as desired. ☐

3.5. **Fourier Inversion and Plancherel Theorems.** Using the Weierstrass approximation (Theorem 3.32), we can now prove the Fourier inversion and Plancherel theorems for arbitrary functions in $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$. The general theme here is that, Fourier inversion holds for any trigonometric polynomial, but trigonometric polynomials approximate functions in $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$ arbitrarily well, so Fourier inversion also holds for functions in $C(\mathbb{R}/\mathbb{Z};\mathbb{C})$. Analogously, we know that a continuous function $F \colon \mathbb{R} \to \mathbb{R}$ which is zero on rational numbers is actually zero on all of $\mathbb{R}$.

Theorem 3.39 (**Fourier inversion/ Best approximation**)**.** *For any $f \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$, the series $\sum_{n=-N}^{N} \widehat{f}(n) e_n$ converges to f in the L_2 metric. That is,*

$$\lim_{N \to \infty} \|f - \sum_{n=-N}^{N} \widehat{f}(n) e_n\|_2 = 0.$$

Proof. Let $\varepsilon > 0$. We need N_0 such that, for all $N > N_0$, we have $\|f - \sum_{n=-N}^{N} \widehat{f}(n) e_n\|_2 < \varepsilon$.

From the Weierstrass approximation theorem, there exists a natural number N_0 and there exists a trigonometric polynomial $P = \sum_{n=-N_0}^{N_0} c_n e_n$ such that $\|f - P\|_\infty < \varepsilon$. So, $\|f - P\|_2 \leq \|f - P\|_\infty < \varepsilon$.

Let $N > N_0$, and let $f_N := \sum_{n=-N}^{N} \widehat{f}(n) e_n$. We will conclude by showing that $\|f - f_N\|_2 < \varepsilon$. First, note that, for any $m \in \mathbb{Z}$ with $|m| \leq N$, we have by Lemma 3.27

$$\langle f - f_N, e_m \rangle = \langle f, e_m \rangle - \sum_{n=-N}^{N} \widehat{f}(n) \langle e_n, e_m \rangle = \widehat{f}(m) - \widehat{f}(m) = 0.$$

In particular, since $f_N - P$ is a linear combination of e_m where $|m| \leq N$, we have

$$\langle f - f_N, f_N - P \rangle = 0.$$

By the Pythagorean Theorem (Exercise 1.20), we therefore have

$$\|f - P\|_2^2 = \|f - f_N\|_2^2 + \|f_N - P\|_2^2 .$$

Consequently,

$$\|f - f_N\|_2 \leq \|f - P\|_2 < \varepsilon.$$

□

Remark 3.40. Note that we have only proven convergence in the L_2 metric, and it is natural to look for other kinds of convergence. However, in general, f_N does not converge to f pointwise, and f_N does not converge to f uniformly. On the other hand, if we assume more about the function f, then we can get better convergence results. For example, if f is continuously differentiable, then f_N converges to f pointwise. And if f is twice continuously differentiable, then f_N converges to f uniformly. We will not cover these results here, and we instead defer them to the Fourier analysis course. Below, we only mention one theorem concerning uniform convergence.

Theorem 3.41. *Let $f \in C(\mathbb{R}/\mathbb{Z}; \mathbb{C})$, and assume that the series (of numbers) $\sum_{n=-\infty}^{\infty} |\widehat{f}(n)|$ is absolutely convergent. Then the series $\sum_{n=-\infty}^{\infty} \widehat{f}(n) e_n$ converges uniformly to f. That is,*

$$\lim_{N \to \infty} \|f - \sum_{n=-N}^{N} \widehat{f}(n) e_n\|_\infty = 0.$$

Proof. By the Weierstrass M-test (Theorem 2.35), we know that $\sum_{n=-N}^{N} \widehat{f}(n) e_n$ converges uniformly to *some* function g. (Strictly speaking, the Weierstrass M-test applies to functions summed from $n = 0$ to $n = -\infty$, but this result applies to the situation at hand by splitting the double sum into two separate infinite sums.) By Lemma 3.6(c), g is continuous and periodic. So,

$$\lim_{N \to \infty} \|g - \sum_{n=-N}^{N} \widehat{f}(n) e_n\|_\infty = 0 \qquad (*).$$

Since the L_2 norm is bounded by the L_∞ norm, we conclude that

$$\lim_{N \to \infty} \|g - \sum_{n=-N}^{N} \widehat{f}(n) e_n\|_2 = 0.$$

By Theorem 3.39, we already know that

$$\lim_{N\to\infty} \|f - \sum_{n=-N}^{N} \widehat{f}(n)e_n\|_2 = 0.$$

By the uniqueness of limits in metric spaces (in this case, uniqueness of limits with respect to the L_2 metric), we conclude that $f = g$. That is, $(*)$ concludes the proof. □

As a Corollary of Fourier inversion, we obtain the following theorem.

Theorem 3.42 (**Plancherel theorem**). *Let $f \in C(\mathbb{R}/\mathbb{Z};\mathbb{C})$. Then the series (of numbers) $\sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2$ is absolutely convergent. Also,*

$$\|f\|_2^2 = \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2.$$

Proof. Let $\varepsilon > 0$. By the Fourier inversion theorem (Theorem 3.39), there exists N_0 such that, for all $N > N_0$, we have

$$\|f - \sum_{n=-N}^{N} \widehat{f}(n)e_n\|_2 < \varepsilon.$$

So, by the triangle inequality,

$$\|f\|_2 - \varepsilon < \|\sum_{n=-N}^{N} \widehat{f}(n)e_n\|_2 \leq \|f\|_2 + \varepsilon.$$

Moreover, by Corollary 3.29, we have

$$\|\sum_{n=-N}^{N} \widehat{f}(n)e_n\|_2 = (\sum_{n=-N}^{N} |\widehat{f}(n)|^2)^{1/2}.$$

Therefore,

$$(\|f\|_2 - \varepsilon)^2 < \sum_{n=-N}^{N} |\widehat{f}(n)|^2 < (\|f\|_2 + \varepsilon)^2.$$

Taking the limit superior and limit inferior, we have

$$(\|f\|_2 - \varepsilon)^2 \leq \liminf_{N\to\infty} \sum_{n=-N}^{N} |\widehat{f}(n)|^2 \leq \limsup_{N\to\infty} \sum_{n=-N}^{N} |\widehat{f}(n)|^2 \leq (\|f\|_2 + \varepsilon)^2.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $\lim_{N\to\infty} \sum_{n=-N}^{N} |\widehat{f}(n)|^2 = \|f\|_2^2$, as desired. □

4. Differentiation in several variables

Definition 4.1 (**Derivative on the real line**). Let E be a subset of $\mathbb{R}$, and let x_0 be a limit point of E, and let $f\colon E \to \mathbb{R}$. If the limit

$$\lim_{x\to x_0; x\in E\setminus\{x_0\}} \frac{f(x) - f(x_0)}{x - x_0}.$$

exists and converges to a real number $L \in \mathbb{R}$, then we write $f'(x_0) = L$ and we say that f is **differentiable** at x_0. If this limit does not exist, then we say that f is **not differentiable** at x_0.

Lemma 4.2. *Let E be a subset of $\mathbb{R}$, let $f\colon E \to \mathbb{R}$, let $x_0 \in E$, and let $L \in \mathbb{R}$. Then the following two statements are equivalent.*

- *f is differentiable at x_0 and $f'(x_0) = L$.*
- *We have $\lim_{x \to x_0; x \in E \setminus \{x_0\}} \frac{|f(x) - (f(x_0) + L(x - x_0))|}{|x - x_0|} = 0$.*

Definition 4.3. Let n, m be positive integers. A **linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$** is a function $L\colon \mathbb{R}^n \to \mathbb{R}^m$ which satisfies the following properties.

- For all $x, y \in \mathbb{R}^n$, we have $L(x + y) = L(x) + L(y)$.
- For all $x \in \mathbb{R}^n$ and for all $\alpha \in \mathbb{R}$, we have $L(\alpha x) = \alpha L(x)$.

Remark 4.4. Given a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$, there exists an $m \times n$ matrix A (that is, a matrix A with m rows and n columns) such that

$$L(x) = Ax, \quad \forall\, x \in \mathbb{R}^n.$$

Conversely, given a matrix A, the function $L\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by $L(x) := Ax$ for all $x \in \mathbb{R}^n$, is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. So, on Euclidean spaces, the notions of matrices and linear transformations are interchangeable.

Our final topic in this course will be differentiation in several variables. Here the theory somewhat resembles the theory of differentiation in one variable, however there are many key differences. The first obstacle we need to overcome is to simply define the derivative in the higher dimensional setting. We therefore begin with this task.

4.1. **Differentiation in multiple variables.** Let n, m be positive integers. Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$. In order to define the derivative of f, we cannot simply copy and paste Definition 4.1, since we would need to let $x \in \mathbb{R}^n$ and then divide by x, which is meaningless unless $n = 1$. We instead use the equivalent definition within Lemma 4.2. In this case, we can successfully define differentiation by replacing the absolute values by the appropriate norm, and by replacing L by a linear map.

Definition 4.5 (**Derivatives in multiple variables**)**.** Let E be a subset of $\mathbb{R}^n$, let $f\colon E \to \mathbb{R}^m$ be a function, let $x_0 \in E$, and let $L\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. We say that f is **differentiable at x_0 with derivative** L if and only if we have

$$\lim_{x \to x_0; x \in E} \frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} = 0.$$

Example 4.6. Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $f(x_1, x_2) = (x_1^2, x_2^2)$. Define the linear transformation $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ by $L(x_1, x_2) := (2x_1, 4x_2)$. We will show that L is the derivative of f at the point $x_0 = (1, 2)$. We want to show that

$$\lim_{x \to (1,2); x \neq (1,2)} \frac{\|f(x) - (f(1,2) + L(x - (1,2)))\|}{\|x - (1,2)\|} = 0.$$

Now, note that

$$\begin{aligned} f(x) - (f(1,2) + L(x-(1,2))) &= (x_1^2, x_2^2) - ((1,4) + (2x_1, 4x_2) - (2,8)) \\ &= (x_1^2, x_2^2) - (2x_1 - 1, 4x_2 - 4) \\ &= ((x_1-1)^2, (x_2-2)^2). \end{aligned}$$

So, using the triangle inequality,

$$\|f(x) - (f(1,2) + L(x-(1,2)))\| \leq \|((x_1-1)^2, 0)\| + \|(0, (x_2-2)^2)\| = (x_1-1)^2 + (x_2-2)^2.$$

In conclusion,

$$0 \leq \lim_{x \to (1,2); x \neq (1,2)} \frac{(x_1-1)^2 + (x_2-2)^2}{\sqrt{(x_1-1)^2 + (x_2-2)^2}} = \lim_{x \to (1,2); x \neq (1,2)} \sqrt{(x_1-1)^2 + (x_2-2)^2} = 0.$$

So, we have proven our desired statement.

The following lemma shows that a function can have at most one derivative at an interior point of E.

Lemma 4.7. *Let E be a subset of $\mathbb{R}^n$, let $f\colon E \to \mathbb{R}^m$ be a function, and let x_0 be an interior point of E. Let $L_a\colon \mathbb{R}^n \to \mathbb{R}^m$ and let $L_b\colon \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. Suppose f is differentiable at x_0 with derivative L_a, and f is differentiable at x_0 with derivative L_b. Then $L_a = L_b$.*

Exercise 4.8. Prove Lemma 4.7. (Hint: argue by contradiction. Assume that $L_a \neq L_b$. Then there exists a nonzero vector $v \in \mathbb{R}^n$ such that $L_a v \neq L_b v$. Then, apply the definition of the derivative, and try to specialize to the case where $x = x_0 + tv$ for some scalar t, in order to obtain a contradiction.)

Using Lemma 4.7, we can now talk about the derivative of f at interior points x_0, and we will label this derivative as $f'(x_0)$. That is, if x_0 is an interior point of E, then $f'(x_0)$ is the unique linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ such that

$$\lim_{x \to x_0; x \in E} \frac{\|f(x) - (f(x_0) + f'(x_0)(x - x_0))\|}{\|x - x_0\|} = 0.$$

Informally, we therefore have **Newton's approximation**:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$

Remark 4.9. We sometimes refer to $f'(x_0)$ as the **total derivative** of f, to distinguish $f'(x_0)$ from the related directional and partial derivatives.

4.2. **Partial and Directional Derivatives.** We now relate the total derivative to the partial and directional derivatives. Let n, m be positive integers.

Definition 4.10. Let E be a subset of $\mathbb{R}^n$, let $f\colon E \to \mathbb{R}^m$ be a function, let x_0 be an interior point of E, let $v \in \mathbb{R}^n$, and let t be a real number. If the limit

$$\lim_{t \to 0; t \neq 0, x_0 + tv \in E} \frac{f(x_0 + tv) - f(x_0)}{t}.$$

exists, we say that f is **differentiable in the direction** v **at** x_0, and we denote this limit by $D_v f(x_0)$.

$$D_v f(x_0) := \lim_{t\to 0; t\neq 0, x_0+tv\in E} \frac{f(x_0+tv)-f(x_0)}{t}.$$

Equivalently, we have

$$D_v f(x_0) := \frac{d}{dt} f(x_0+tv)|_{t=0}.$$

Note that in this definition we are dividing by the scalar t, so this division is okay, and $D_v f(x_0) \in \mathbb{R}^m$.

Example 4.11. Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $f(x_1,x_2) = (x_1^2, x_2^2)$. Let $x_0 := (1,2)$ and let $v := (3,4)$. We then compute

$$\frac{((1+3t)^2,(2+4t)^2)-(1,4)}{t} = \frac{(1+6t+9t^2, 4+16t+16t^2)-(1,4)}{t} = (6+9t, 16+16t).$$

Therefore,

$$D_v f(x_0) = \lim_{t\to 0; t\neq 0} (6+9t, 16+16t) = (6,16).$$

If v is a standard basis vector, then we write $\frac{\partial f}{\partial x_j}(x_0)$ or $\frac{\partial}{\partial x_j} f(x_0)$ for $D_{e_j} f(x_0)$. We refer to $\frac{\partial f}{\partial x_j}(x_0)$ as the **partial derivative of** f **with respect to** x_j. So,

$$\frac{\partial f}{\partial x_j}(x_0) := \lim_{t\to 0; t\neq 0, x_0+te_j\in E} \frac{f(x_0+te_j)-f(x_0)}{t} = \frac{d}{dt} f(x_0+te_j)|_{t=0}.$$

Note that if $f\colon E \to \mathbb{R}^m$, then $\frac{\partial f}{\partial x_j} \in \mathbb{R}^m$. And if we write f in its components as $f = (f_1,\ldots,f_m)$, then

$$\frac{\partial f}{\partial x_j}(x_0) = \left(\frac{\partial f_1}{\partial x_j}(x_0), \ldots, \frac{\partial f_1}{\partial x_j}(x_0)\right).$$

The total derivative and directional derivative are related in the following way.

Lemma 4.12. *Let E be a subset of $\mathbb{R}^n$, let $f\colon E \to \mathbb{R}^m$ be a function, let x_0 be an interior point of E, and let $v \in \mathbb{R}^n$. If f is differentiable at x_0, then f is also differentiable in the direction v at x_0, and*

$$D_v f(x_0) = f'(x_0)v.$$

Exercise 4.13. Prove Lemma 4.12.

From Lemma 4.12, total differentiability implies directional differentiability. Unfortunately, the converse is false.

Exercise 4.14. Define $f\colon \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) := x^3/(x^2+y^2)$ when $(x,y) \neq (0,0)$, and $f(0,0) := 0$. Show that for any $v \in \mathbb{R}^2$, f is differentiable at $(0,0)$ in the direction v. However, show that f is not differentiable at $(0,0)$.

Remark 4.15. From Lemma 4.12, if $E \subseteq \mathbb{R}^n$ and if $f\colon E \to \mathbb{R}^m$ is differentiable at $x_0 \in E$, then all partial derivatives $\frac{\partial f}{\partial x_j}$ exist at x_0, for all $j \in \{1,\ldots,n\}$, and

$$\frac{\partial f}{\partial x_j} = f'(x_0)e_j, \quad \forall j \in \{1,\ldots,n\}.$$

Also, given $v = (v_1, \dots, v_n) = \sum_{j=1}^{n} v_j e_j \in \mathbb{R}^n$, we have

$$D_v f(x_0) = f'(x_0) \sum_{j=1}^{n} v_j e_j = \sum_{j=1}^{n} v_j f'(x_0) e_j = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0). \qquad (*)$$

From Exercise 4.14, partial differentiability does not imply differentiability. However, if the partial derivatives of a function are continuous, then partial differentiability does imply differentiability. We will use equation $(*)$ to prove this assertion.

Theorem 4.16. *Let E be a subset of $\mathbb{R}^n$, let $f \colon E \to \mathbb{R}^m$ be a function, let F be a subset of E, and let x_0 be an interior point of F. If the partial derivatives $\frac{\partial f}{\partial x_j}$ exist on F and are continuous at x_0 for all $j \in \{1, \dots, n\}$, then f is differentiable at x_0. Moreover, $f'(x_0) \colon \mathbb{R}^n \to \mathbb{R}^m$ is defined by*

$$f'(x_0)(v_1, \dots, v_n) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$

Proof. Define a linear transformation $L \colon \mathbb{R}^n \to \mathbb{R}^m$ by

$$L(v_1, \dots, v_n) := \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$

We need to show that

$$\lim_{x \to x_0; x \in E \setminus \{x_0\}} \frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} = 0.$$

Let $\varepsilon > 0$. We will find $\delta > 0$ such that, if x satisfies $0 < \|x - x_0\| < \delta$, then

$$\frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} < \varepsilon.$$

That is, we will show, if x satisfies $0 < \|x - x_0\| < \delta$, then

$$\|f(x) - (f(x_0) + L(x - x_0))\| < \varepsilon \|x - x_0\|.$$

Since x_0 is an interior point of F, there exists $r > 0$ such that $B(x_0, r) \subseteq F$. Since the partial derivative $\frac{\partial f}{\partial x_j}$ is continuous on F for each $j \in \{1, \dots, n\}$, there exists $0 < \delta_j < r$ such that $\|\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0)\| < \varepsilon/(nm)$, for every $x \in B(x_0, \delta_j)$, for every $j \in \{1, \dots, n\}$. Define $\delta := \min_{j=1,\dots,n} \delta_j$. Then $\|\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0)\| < \varepsilon/(nm)$, for every $x \in B(x_0, \delta)$, for every $j \in \{1, \dots, n\}$.

Let $x \in B(x_0, \delta)$, and write $x = x_0 + v_1 e_1 + \cdots + v_n e_n$ for some scalars $v_1, \dots, v_n$. Note that

$$\|x - x_0\| = \sqrt{v_1^2 + \cdots + v_n^2}.$$

In particular, we have $|v_j| \leq \|x - x_0\|$ for all $j \in \{1, \dots, n\}$. Recall that we need to show

$$\|f(x_0 + v_1 e_1 + \cdots + v_n e_n) - f(x_0) - \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0)\| < \varepsilon \|x - x_0\|.$$

Write f in its components as $f = (f_1, \ldots, f_m)$, so that $f_i \colon E \to \mathbb{R}$ for all $i \in \{1, \ldots, m\}$. Applying the Mean Value Theorem in the first variable, there exists a real number t_i between 0 and v_1 such that

$$f_i(x_0 + v_1 e_1) - f_i(x_0) = \frac{\partial f_i}{\partial x_1}(x_0 + t_i e_1) v_1.$$

Note that, for all $i \in \{1, \ldots, m\}$, for all $j \in \{1, \ldots, n\}$, we have

$$|\frac{\partial f_i}{\partial x_1}(x_0 + t_i e_1) - \frac{\partial f_i}{\partial x_1}(x_0)| \leq \|\frac{\partial f}{\partial x_1}(x_0 + t_i e_1) - \frac{\partial f}{\partial x_1}(x_0)\| \leq \varepsilon/(nm).$$

Therefore,

$$|f_i(x_0 + v_1 e_1) - f_i(x_0) - \frac{\partial f_i}{\partial x_1}(x_0) v_1| \leq \varepsilon |v_1| /(nm).$$

Summing this inequality over $i \in \{1, \ldots, m\}$ and using $\|(y_1, \ldots, y_m)\| \leq |y_1| + \cdots + |y_m|$, we have

$$\|f(x_0 + v_1 e_1) - f(x_0) - \frac{\partial f}{\partial x_1}(x_0) v_1\| \leq \varepsilon |v_1| /n \leq \varepsilon \|x - x_0\| /n.$$

In the last inequality, we used $|v_1| \leq \|x - x_0\|$.

Using a similar argument, we conclude that

$$\|f(x_0 + v_1 e_1 + v_2 e_2) - f(x_0 + v_1 e_1) - \frac{\partial f}{\partial x_2}(x_0) v_2\| \leq \varepsilon \|x - x_0\| /n.$$

And so on, until we get

$$\|f(x_0 + v_1 e_1 + \cdots + v_n e_n) - f(x_0 + v_1 e_1 + \cdots + v_{n-1} e_{n-1}) - \frac{\partial f}{\partial x_n}(x_0) v_n\| \leq \varepsilon \|x - x_0\| /n.$$

Summing these n inequalities and using the triangle inequality $\|x + y\| \leq \|x\| + \|y\|$, we get a telescoping sum which finally gives

$$\|f(x_0 + v_1 e_1 + \cdots + v_n e_n) - f(x_0) - \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0)\| < \varepsilon \|x - x_0\|.$$

□

From Theorem 4.16 and Lemma 4.12, if the partial derivatives of a function $f \colon E \to \mathbb{R}^m$ exist and are continuous on a set F, then all directional derivatives of f exist at every interior point x_0 of F, and

$$D_{(v_1, \ldots, v_n)} f(x_0) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$

In particular, if $f \colon E \to \mathbb{R}$ is a real-valued function, and if we define the **gradient** $\nabla f(x_0)$ of f at x_0 to be the n-dimensional row vector

$$\nabla f(x_0) := (\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_n}(x_0)),$$

then we have the formula

$$D_v f(x_0) = \langle \nabla f(x_0), v \rangle.$$

More generally, if $f\colon E \to \mathbb{R}^m$ is a function with $f = (f_1, \ldots, f_m)$, and x_0 is in the interior of the region where the partial derivatives of f exist and are continuous, then Theorem 4.16 says

$$f'(x_0)(v_1, \ldots, v_n) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0) = \left(\sum_{j=1}^{n} v_j \frac{\partial f_i}{\partial x_j}(x_0) \right)_{i=1}^{m}.$$

So, if we define the matrix

$$Df(x_0) = \left(\frac{\partial f_i}{\partial x_j}(x_0) \right)_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) & \cdots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \frac{\partial f_2}{\partial x_1}(x_0) & \frac{\partial f_2}{\partial x_2}(x_0) & \cdots & \frac{\partial f_2}{\partial x_n}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \frac{\partial f_m}{\partial x_2}(x_0) & \cdots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix},$$

then we have

$$D_v f(x_0) = f'(x_0)v = Df(x_0)v.$$

The matrix $Df(x_0)$ is sometimes called the **derivative** or the **differential** of f at x_0. We still wish to distinguish the matrix $Df(x_0)$ from the linear transformation $f'(x_0)$, since the latter is defined in a way which does not depend on the chosen basis of Euclidean space.

4.3. **The Chain Rule in Several Variables.** Let n, m, p be positive integers. Recall that if $f\colon X \to Y$ and $g\colon Y \to Z$ are functions, then the composition $g \circ f\colon X \to Z$ is defined by $g \circ f(x) := g(f(x))$, for all $x \in X$.

Theorem 4.17 (**The Chain Rule in Multiple Variables**). *Let E be a subset of $\mathbb{R}^n$, let F be a subset of $\mathbb{R}^m$, let $f\colon E \to F$ be a function, and let $g\colon F \to \mathbb{R}^p$. Let x_0 be a point in the interior of E. Assume that f is differentiable at x_0 and that $f(x_0)$ is in the interior of F. Assume also that g is differentiable at $f(x_0)$. Then $g \circ f\colon E \to \mathbb{R}^p$ is also differentiable at x_0, and*

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

Remark 4.18. We can intuitively think of the chain rule as follows. From Newton's approximation, we have

$$f(x) - f(x_0) \approx f'(x_0)(x - x_0).$$

Also, using Newton's approximation again,

$$g(f(x)) - g(f(x_0)) \approx g'(f(x_0))(f(x) - f(x_0)).$$

So, combining these two approximations, we have

$$g(f(x)) - g(f(x_0)) \approx g'(f(x_0))f'(x_0)(x - x_0).$$

That is, $(g \circ f)'(x_0) = g'(f(x_0))f'(x_0)$. The rigorous version of this proof irons out the details inherent in Newton's approximation.

Exercise 4.19.

- Let $L\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Show that there exists a real number $M > 0$ such that $\|Lx\| \leq M \|x\|$, for all $x \in \mathbb{R}^n$. (Hint: first, using Remark 4.4, write L in terms of a matrix A. Then, set M to be equal to the sum of the absolute values of the entries of A. Use the triangle inequality a lot. There are many different ways to do this exercise, some of which use a different value of M. For example, you could try using the Cauchy-Schwarz inequality.) In particular, conclude that any linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is continuous.

- Let E be a subset of $\mathbb{R}^n$. Assume that $f\colon E \to \mathbb{R}^m$ is differentiable at an interior point x_0 of E. Then f is also continuous at x_0.
- Prove Theorem 4.17. (Hint: it may be helpful to review the proof of the single variable chain rule. It is probably easiest to use the sequence definition of a limit.)

Example 4.20. Suppose $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is a differentiable function, and $x_j\colon \mathbb{R} \to \mathbb{R}$ are differentiable functions for all $j \in \{1, \ldots, n\}$. Then

$$\frac{d}{dt} f(x_1(t), \ldots, x_n(t)) = \sum_{j=1}^{n} x_j'(t) \frac{\partial f}{\partial x_j}(x_1(t), \ldots, x_n(t)).$$

This follows from the chain rule.

4.4. **Iterated Derivatives and Clairaut's Theorem.** We now investigate what happens when we differentiate a function twice, in two different directions.

Definition 4.21. Let E be a subset of $\mathbb{R}^n$, and let $f\colon E \to \mathbb{R}^m$ be a function. We say that f is **continuously differentiable** if and only if the partial derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ exist and are continuous on E. We say that f is **twice continuously differentiable** if and only if it is continuously differentiable, and the partial derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ are themselves continuously differentiable.

Continuously differentiable functions are sometimes called C^1 functions. Twice continuously differentiable functions are sometimes called C^2 functions. One can also define C^3 functions, C^4 functions, etc., but we will not do so here.

Let $f\colon \mathbb{R}^2 \to \mathbb{R}$. As you may have learned, it is often true that $\frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2} f = \frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1} f$. Unfortunately, this equality does not always hold.

Exercise 4.22. Define $f\colon \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) := (x^3 y)/(x^2 + y^2)$ when $(x,y) \neq (0,0)$, and $f(0,0) := 0$. Show that f is continuously differentiable, and the double derivatives $\frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2} f$ and $\frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1}$ exist, but these derivatives are not equal at $(0,0)$.

Thankfully, if f is twice continuously differentiable, then the order of differentiation does not matter.

Theorem 4.23 (**Clairaut's Theorem**). *Let E be an open subset of $\mathbb{R}^n$, and let $f\colon E \to \mathbb{R}^m$ be a twice continuously differentiable function. Then, for all $1 \leq i, j \leq n$ and for all interior points x_0 of E, we have $\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f(x_0) = \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} f(x_0)$.*

Proof. The claim is certainly true for $i = j$ so assume that $i \neq j$. By replacing f by $f(x - x_0)$ as necessary, we may assume that $x_0 = 0$.

Define $a := \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f(x_0)$ and define $a' := \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} f(x_0)$. We need to show that $a = a'$.

Let $\varepsilon > 0$. Since f is twice continuously differentiable, there exists $\delta > 0$ such that, for all x with $\|x\| < 2\delta$, we have

$$\left| \frac{\partial}{x_i}\frac{\partial}{x_j} f(x) - a \right| < \varepsilon, \qquad \left| \frac{\partial}{x_j}\frac{\partial}{x_i} f(x) - a' \right| < \varepsilon.$$

Define

$$M := f(\delta e_i + \delta e_j) - f(\delta e_i) - f(\delta e_j) + f(0).$$

Applying the Fundamental Theorem of Calculus to the e_i variable, we have

$$f(\delta e_i + \delta e_j) - f(\delta e_j) = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) dx_i.$$

And

$$f(\delta e_i) - f(0) = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i) dx_i.$$

Therefore,

$$M = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) dx_i$$

For each $x_i \in (0, \delta)$, there exists $x_j \in [0, \delta]$ such that, by the Mean Value Theorem, we have

$$\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) = \delta \frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i}(x_i e_i + x_j e_j).$$

By our choice of δ (noting that $\|x_i e_i + x_j e_j\| < 2\delta$), we therefore have

$$\left| \frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) - \delta a' \right| < \varepsilon \delta.$$

So, integrating this inequality over $x_i \in [0, \delta]$, we get

$$\left| M - \delta^2 a' \right| < \varepsilon \delta^2.$$

We can run this same argument with the roles of i and j reversed (noting that M is symmetric in i, j) to get

$$\left| M - \delta^2 a \right| < \varepsilon \delta^2.$$

So, from the triangle inequality, we conclude that

$$|a - a'| < 2\varepsilon.$$

Since this inequality holds for all $\varepsilon > 0$, we conclude that $a = a'$, as desired. □

4.5. **Appendix: Notation.** Let A, B be sets in a space X. Let m, n be a nonnegative integers.

$$
\begin{aligned}
\mathbb{Z} &:= \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}, \text{ the integers} \\
\mathbb{N} &:= \{0, 1, 2, 3, 4, 5, \ldots\}, \text{ the natural numbers} \\
\mathbb{Z}_+ &:= \{1, 2, 3, 4, \ldots\}, \text{ the positive integers} \\
\mathbb{Q} &:= \{m/n \colon m, n \in \mathbb{Z}, n \neq 0\}, \text{ the rationals} \\
\mathbb{R} & \text{ denotes the set of real numbers} \\
\mathbb{R}^* &= \mathbb{R} \cup \{-\infty\} \cup \{+\infty\} \text{ denotes the set of extended real numbers} \\
\mathbb{C} &:= \{x + y\sqrt{-1} : x, y \in \mathbb{R}\}, \text{ the complex numbers} \\
\emptyset & \text{ denotes the empty set, the set consisting of zero elements} \\
\in & \text{ means “is an element of.” For example, } 2 \in \mathbb{Z} \text{ is read as “2 is an element of } \mathbb{Z}.\text{”} \\
\forall & \text{ means “for all”} \\
\exists & \text{ means “there exists”} \\
\mathbb{R}^n &:= \{(x_1, \ldots, x_n) \colon x_i \in \mathbb{R}, \forall\, i \in \{1, \ldots, n\}\} \\
A \subseteq B & \text{ means } \forall\, a \in A, \text{ we have } a \in B, \text{ so } A \text{ is contained in } B \\
A \smallsetminus B &:= \{x \in A \colon x \notin B\} \\
A^c &:= X \smallsetminus A, \text{ the complement of } A \\
A \cap B & \text{ denotes the intersection of } A \text{ and } B \\
A \cup B & \text{ denotes the union of } A \text{ and } B
\end{aligned}
$$

Let (X, d) be a metric space, let $x_0 \in X$, let $r > 0$ be a real number, and let E be a subset of X. Let $(x_1, \ldots, x_n)$ be an element of $\mathbb{R}^n$, and let $p \geq 1$ be a real number.

$$
\begin{aligned}
B_{(X,d)}(x_0, r) = B(x_0, r) &:= \{x \in X \colon d(x, x_0) < r\}. \\
\overline{E} & \text{ denotes the closure of } E \\
\mathrm{int}(E) & \text{ denotes the interior of } E \\
\partial E & \text{ denotes the boundary of } E
\end{aligned}
$$

$$
\begin{aligned}
\|(x_1, \ldots, x_n)\|_{\ell_p} &:= (\sum_{i=1}^{n} |x_i|^p)^{1/p} \\
\|(x_1, \ldots, x_n)\|_{\ell_\infty} &:= \max_{i=1,\ldots,n} |x_i|
\end{aligned}
$$

Let $f, g\colon (X, d_X) \to (Y, d_Y)$ be maps between metric spaces. Let $V \subseteq X$, and let $W \subseteq Y$.

$$f(V) := \{f(v) \in Y : v \in V\}.$$
$$f^{-1}(W) := \{x \in X : f(x) \in W\}.$$
$$d_\infty(f, g) := \sup_{x \in X} d_Y(f(x), g(x)).$$

$B(X; Y)$ denotes the set of functions $f\colon X \to Y$ that are bounded.

$$C(X; Y) := \{f \in B(X; Y) : f \text{ is continuous}\}.$$

Let $f, g\colon \mathbb{R} \to \mathbb{C}$ be $\mathbb{Z}$-periodic functions.

$$\|f\|_\infty := \sup_{x \in [0,1]} |f(x)|.$$
$$\langle f, g \rangle := (\int_0^1 f(x)\overline{g(x)}dx)^{1/2}.$$
$$\|f\|_2 := \sqrt{\langle f, f \rangle} = (\int_0^1 |f(x)|^2 \, dx)^{1/2}$$
$$d_{L_2}(f, g) := \|f - g\|_2 = (\int_0^1 |f(x) - g(x)|^2 \, dx)^{1/2}.$$

Let n, m be positive integers, let $(e_1, \ldots, e_n)$ denote the standard basis of $\mathbb{R}^n$, let E be a subset of $\mathbb{R}^n$, let $f\colon E \to \mathbb{R}^m$ be a function, let $x_0 \in E$ be an interior point of E, let $v \in \mathbb{R}^n$, and let $j \in \{1, \ldots, n\}$.

$f'(x_0)$ denotes the total derivative of f.

$D_v f(x_0)$ denotes the derivative of f in the direction v.

$$\frac{\partial f}{\partial x_j}(x_0) = \frac{\partial}{\partial x_j} f(x_0) = D_{e_j} f(x_0).$$

Let E be a subset of $\mathbb{R}^n$, let $f\colon E \to \mathbb{R}$ be a function, and let x_0 be an interior point of E.

$$\nabla f(x_0) = (\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_n}(x_0)).$$

4.5.1. *Set Theory.* Let X, Y be sets, and let $f\colon X \to Y$ be a function. The function $f\colon X \to Y$ is said to be **injective** (or **one-to-one**) if and only if: for every $x, x' \in V$, if $f(x) = f(x')$, then $x = x'$.

The function $f\colon X \to Y$ is said to be **surjective** (or **onto**) if and only if: for every $y \in Y$, there exists $x \in X$ such that $f(x) = y$.

The function $f\colon X \to Y$ is said to be **bijective** (or a **one-to-one correspondence**) if and only if: for every $y \in Y$, there exists exactly one $x \in X$ such that $f(x) = y$. A function $f\colon X \to Y$ is bijective if and only if it is both injective and surjective.

Two sets X, Y are said to have the same **cardinality** if and only if there exists a bijection from X onto Y.

UCLA Department of Mathematics, Los Angeles, CA 90095-1555
E-mail address: heilman@math.ucla.edu