Math Review Notes—Real Analysis

Gregory Faletto

Contents

Last updated November 22, 2020

Chapter 1

Real Analysis

These are my notes from Math 4650: Analysis I at Cal State LA, Math 425B: Fundamental Concepts of Analysis at USC taught by Andrew Manion and the textbook [Pugh, 2015], the textbook [Rudin, 1976], as well as Prof. Steven Heilman's notes from Math 541A at USC.

1.1   Midterm 1

1.1.1   Homework 1

Definition 1.1.1. Let $S \subseteq \mathbb{R}$. We say that S is **bounded from above** if $\exists\, b \in \mathbb{R}$ where

$$s \leq b \ \forall\, s \in S$$

If this is the case, we call b an **upper bound** of S.

If $b \leq c$ for all upper bounds c of S, we call b the **supremum** of S: $b = \sup(S)$.

Definition 1.1.2. We say that S is **bounded from below** if $\exists\, a \in \mathbb{R}$ where

$$s \geq a \ \forall\, s \in S$$

If this is the case, we call a a **lower bound** of S.

If $a \geq d$ for all lower bounds d of S, we call a the **infimum** of S: $a = \inf(S)$.

Proposition 1. Useful Sup/Inf Fact: Let $S \in \mathbb{R}$, $S \neq \emptyset$.

(1) Suppose S is bounded from above by an element b. Then $b = \sup(S) \iff \forall\, \epsilon > 0 \ \exists\, x \in S$ with

$$b - \epsilon < x \leq b$$

(2) Suppose S is bounded from below by an element a. Then $a = \inf(S) \iff \forall\, \epsilon > 0 \ \exists\, x \in S$ with

$$a \leq x < a + \epsilon$$

Completeness Axiom: Let S be a nonempty subset of $\mathbb{R}$. If S is bounded from above, then $\sup(S)$ exists. If S is bounded from below, then $\inf(S)$ exists.

Facts about absolute value:

-

  Proposition 2. $|x - y| < \epsilon \iff y - \epsilon < x < y + \epsilon$.

  Proof. In notes 08/23. $\square$

-

  Proposition 3. $|ab| = |a||b|$.

  Proof.
  $$|ab| = \begin{cases} ab & ab \geq 0 \\ -ab & ab < 0 \end{cases} = \begin{cases} ab & a \geq 0, b \geq 0 \\ -ab & a \geq 0, b < 0 \\ -ab & a < 0, b \geq 0 \\ ab & a < 0, b < 0 \end{cases} = \begin{cases} ab & a \geq 0, b \geq 0 \\ a(-b) & a \geq 0, b < 0 \\ (-a)b & a < 0, b \geq 0 \\ (-a)(-b) & a < 0, b < 0 \end{cases}$$

  $$= \begin{cases} |a||b| & a \geq 0, b \geq 0 \\ |a||b| & a \geq 0, b < 0 \\ |a||b| & a < 0, b \geq 0 \\ |a||b| & a < 0, b < 0 \end{cases} \implies |ab| = |a||b|$$

  $\square$

-

  Proposition 4. Let $\epsilon > 0$. Then $|a| < \epsilon \iff -\epsilon < a < \epsilon$.

  Proof. Follows from Proposition 2 if $x = a$, $y = 0$. $\square$

-

  Proposition 5. $-|a| \leq a \leq |a|$

  Proof. Follows from Proposition 2 if $x = a$, $y = 0$, $\epsilon = |a|$. $\square$

-

  Theorem 6. Triangle Inequality: $|a + b| \leq |a| + |b|$.

  Proof. In notes 08/23. $\square$

  Corollary 6.1. Triangle Inequality: $|a - b| \leq |a| + |b|$.

  Proof. Follows from Theorem 6, let $b = -b$. $\square$

  Remark 1. See also Theorem **??**.

•

Proposition 7. $\big|\, |a| - |b| \,\big| \leq |a - b|$.

Proof. By Proposition 2, $\big|\, |a| - |b| \,\big| \leq |a - b|$ if and only if

$$|b| - |a - b| \leq |a| \leq |b| + |a - b| \tag{1.1}$$

The left half of (1.1) is true by the Triangle Inequality (Theorem 6):

$$|b| = |a - (a - b)| \leq |a| + |a - b| \iff |b| \leq |a| + |a - b| \iff |b| - |a - b| \leq |a|$$

The right half of (1.1) is also true by the Triangle Inequality (Theorem 6):

$$|a| = |b + a - b| \leq |b| + |a - b|$$

Therefore

$$\big|\, |a| - |b| \,\big| \leq |a - b|.$$

$\square$

Proof. (Alternative proof.) Note that by the Triangle Inequality (Theorem 6),

$$|a| = |a - b + b| \leq |a - b| + |b| \implies |a| - |b| \leq |a - b|$$

Also,

$$|b| = |b - a + a| \leq |b - a| + |a| \implies -|b - a| \leq |a| - |b| \implies -|a - b| \leq |a| - |b|$$

where the last step follows from Proposition 9. Therefore

$$-|a - b| \leq |a| - |b| \leq |a - b|$$

and by Proposition 2,

$$\big|\, |a| - |b| \,\big| \leq |a - b|.$$

$\square$

•

Proposition 8. If $a < x < b$ and $a < y < b$ then $|x - y| < b - a$.

Proof.
$$y > a \implies -y < -a \implies b - y < b - a$$

$$b > y \implies b - y = |b - y| \implies \boxed{|b - y| < b - a}$$

By the Triangle Inequality (Theorem 6),

$$|x - y| = |x - b + b - y| \leq |x - b| + |b - y|$$

Since $b < x$, $|x - b| > 0$. Therefore $\boxed{|x - y| < |b - y|}$.

$$\implies |x - y| < |b - y| < b - a$$

$$\implies |x - y| < b - a$$

$\square$

Proof. (Alternative proof.) Break into two cases.

- **Case 1:** $x \geq y$. Then $|x - y| = x - y$. We know $a < x < b \implies 0 < x - a < b - a$.

$$a < y \implies -a > -y \implies x - a > x - y \implies x - y < x - a < b - a$$

$$\implies \boxed{|x - y| < b - a}$$

- **Case 2:** $x < y$. Then $|x - y| = y - x$. We know $a < y < b \implies 0 < y - a < b - a$.

$$a < x \implies -a > -x \implies y - a > y - x \implies y - x < y - a < b - a$$

$$\implies \boxed{|x - y| < b - a}$$

$\square$

-

Proposition 9. $|a - b| = |b - a|$

Proof. $|a - b| = |(-1)(b - a)| = |-1||b - a| = |b - a|$, where the second-to-last step follows from Proposition 2.

$\square$

1.1.2 Homework 2

Definition 1.1.3. A sequence (a_n) of real numbers is said to **converge** to a **limit** $L \in \mathbb{R}$ if $\forall\, \epsilon > 0\, \exists\, N > 0$ where

$$n \geq N \implies |a_n - L| < \epsilon$$

We say that (a_n) **diverges** if it does not converge.

Definition 1.1.4. A sequence (a_n) of real numbers is **bounded** if $\exists\, M > 0$ where $\forall\, n \in \mathbb{N}$

$$|a_n| \leq M.$$

Theorem 10. If (a_n) converges then (a_n) is bounded.

Definition 1.1.5. Let (a_n) be a sequence of real numbers. We say that (a_n) is a **Cauchy sequence** if $\forall\, \epsilon > 0\ \exists\, N$ where

$$n, m \geq N \implies |a_n - a_m| < \epsilon$$

Theorem 11. (a_n) is Cauchy if and only if (a_n) converges.

Corollary 11.1. If (a_n) is Cauchy then (a_n) is bounded.

Proof. Let $\epsilon = 1$. Since (a_n) is Cauchy, $\exists\, N > 0 \mid n, m \geq N \implies$

$$|a_n - a_m| < 1$$

So, $n \geq N \implies$

$$|a_n - a_N| < 1 \iff a_N - 1 < a_n < a_N + 1 \implies |a_n| < |a_N + 1| \leq |a_N| + 1$$

Let $M = \max\{|a_1|, |a_2|, \ldots, |a_{N-1}|, |a_N| + 1\}$. Then $|a_n| \leq M\ \forall\, n \geq 1$. Therefore (a_n) is bounded.

$\square$

Theorem 12. (Squeeze theorem.) Suppose that $\{a_n\}, \{b_n\}$, and $\{c_n\}$ are sequences of real numbers such that $a_n \leq b_n \leq c_n$ for all n. If both $\{a_n\}$ and $\{c_n\}$ converge to L, then $\{b_n\}$ converges to L.

Proof. Let $\epsilon > 0$. (a_n) converges to $L \implies$

$$\forall\, \epsilon > 0\ \exists\, N_A \mid n \geq N_A \implies |a_n - L| < \epsilon$$

(c_n) converges to $L \implies$

$$\forall\, \epsilon > 0\ \exists\, N_C \mid n \geq N_C \implies |c_n - L| < \epsilon$$

Let $N = \max\{N_A, N_C\}$. Then by one of our absolute values rules, $n \geq N \implies$

$$|a_n - L| < \epsilon \iff L - \epsilon < a_n < L + \epsilon$$

$$|c_n - L| < \epsilon \iff L - \epsilon < c_n < L + \epsilon$$

Therefore since $a_n \leq b_n \leq c_n$,

$$L - \epsilon < a_n \leq b_n \leq c_n < L + \epsilon \implies L - \epsilon < b_n < L + \epsilon \iff |b_n - L| < \epsilon$$

Therefore (b_n) converges to L.

$\square$

Proposition 13. Limits of sequences in $\mathbb{R}$ are unique: if $a_n : \mathbb{N} \to \mathbb{R}$ is a sequence, $\lim a_n = L$, and $\lim a_n = L'$, then $L = L'$.

Proof. Let $L, L' \in \mathbb{R}$, with $\lim_{n\to\infty} a_n = L$ and $\lim_{n\to\infty} a_n = L'$. Then for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have $|a_n - L| < \epsilon$, and there exists $N' \in \mathbb{N}$ such that for all $n \geq N'$ we have $|a_n - L'| < \epsilon$. Suppose $L \neq L'$, and let $\epsilon := \frac{1}{2}|L - L'| > 0$. But then for all $n \geq \max\{N, N'\}$,

$$2\epsilon = |L - L'| = |L - a_n + a_n - L'| \leq |a_n - L| + |a_n - L'| < \epsilon + \epsilon = 2\epsilon,$$

contradiction. Therefore $L = L'$.

$\square$

Theorem 14. Suppose that $\{a_n\}$ and $\{b_n\}$ are sequences of real numbers such that $a_n \leq b_n$ for all n. If $\{a_n\}$ and $\{b_n\}$ converge to A and B respectively, then $A \leq B$.

Proof. Suppose $A > B$. Then let $\epsilon = \frac{A-B}{4} > 0$. (a_n) converges to $A \implies$

$$\exists \, N_A \mid n \geq N_A \implies |a_n - A| < \epsilon \iff A - \epsilon < a_n < A + \epsilon$$

(b_n) converges to $B \implies$

$$\exists \, N_B \mid n \geq N_B \implies |b_n - B| < \epsilon \iff B - \epsilon < b_n < B + \epsilon$$

Then if $n > \max\{N_A, N_B\}$,

$$A - \epsilon < a_n < A + \epsilon \iff A - \frac{A-B}{4} < a_n < A + \frac{A-B}{4} \iff \frac{3A}{4} + \frac{B}{4} < a_n < \frac{5A}{4} - \frac{B}{4}$$

$$B - \epsilon < b_n < B + \epsilon \iff B - \frac{A-B}{4} < b_n < B + \frac{A-B}{4} \iff \frac{5B}{4} - \frac{A}{4} < b_n < \frac{3B}{4} + \frac{A}{4}$$

This implies

$$b_n < \frac{3B}{4} + \frac{A}{4} = \frac{B}{4} + \frac{A}{4} + \frac{2B}{4} < \frac{B}{4} + \frac{A}{4} + \frac{2A}{4} = \frac{3A}{4} + \frac{B}{4} < a_n$$

Contradiction, since it is given that $a_n \leq b_n \; \forall \, n$. Therefore $A \leq B$. $\square$

1.2   Midterm 2

1.2.1   Homework 3

Definition 1.2.1. (Limits of functions at infinity.) Let f be a real-valued function defined on some set D where D contains an interval of the form (a, ∞). Let $L \in \mathbb{R}$. We say

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

if $\forall\, \epsilon > 0\ \exists\, N \in \mathbb{R}$ where

$$x \geq N \implies |f(x) - L| < \epsilon.$$

Definition 1.2.2. Let $D \subseteq \mathbb{R}$. Let $a \in \mathbb{R}$. We say that a is a **limit point** (or "cluster point," or "accumulation point") of D if $\forall\, \delta > 0\ \exists\, x \in D$ where

$$x \neq a \text{ and } |x - a| < \delta$$

(Note that a may or may not be contained in D.)

Definition 1.2.3. (Limit of a function at a.): Let $D \subseteq \mathbb{R}$ and $f : d \to \mathbb{R}$. Let a be a limit point of D. Let $x \in D$. We say that f has a *limit as x tends to a* if $\exists\, L \in \mathbb{R}$ where $\forall\, \epsilon > 0\ \exists\, \delta > 0$ such that

$$0 < |x - a| < \delta \implies |f(x) - L| < \epsilon$$

and we write

$$\lim_{x \to a} f(x) = L$$

Proposition 15. (Properties of Limits.) Let $D \in \mathbb{R}$ and let a be a limit point of D. Suppose $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$. Let $\alpha \in \mathbb{R}$.

(1) If $\lim_{x \to a} f(x) = L$ and $\lim_{x \to a} g(x) = M$ then

(a)
$$\lim_{x \to a} \alpha = \alpha$$

(b)
$$\lim_{x \to a} [f(x) + g(x)] = L + M$$

(c)
$$\lim_{x \to a} [f(x) - g(x)] = L - M$$

(d)
$$\lim_{x \to a} [f(x) \cdot g(x)] = L \cdot M$$

(e)
$$\lim_{x \to a} [\alpha \cdot f(x)] = \alpha \cdot L$$

(2) If $h : D \to \mathbb{R}$ and $h(x) \neq 0 \ \forall \ x \in D$ and $\lim_{x \to a} h(x) = H \neq 0$, then

$$\lim_{x \to a} \frac{1}{h(x)} = \frac{1}{H}$$

Note that properties (2) and (1)(d) combined imply

$$\lim_{x \to a} \frac{f(x)}{h(x)} = \frac{L}{H}$$

1.2.2 Homework 4

Definition 1.2.4. (Continuity.) Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ and $a \in D$. Then f is **continuous** at a if $\lim_{x \to a} f(x)$ exists and

$$\lim_{x \to a} f(x) = f(a)$$

Remark 2. if f is continuous at a, then we can say $\forall \ \epsilon > 0 \ \exists \ \delta > 0$ such that

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

that is, we don't need to say $0 < |x - a| < \delta$.

Definition 1.2.5. If $B \subseteq D$, then f is **continuous on B** if f is continuous at every $b \in B$.

Theorem 16. (Intermediate Value Theorem.) Let f be continuous on $[a, b]$ and suppose $f(a) < f(b)$. $\forall \ d$ such that

$$f(a) < d < f(b)$$

$\exists \ c \in \mathbb{R}$ where

$$a < c < b, \ f(c) = d.$$

1.3 Real Numbers (Chapter 1 of Pugh [2015]; also from Homework 6 of Math 4650 at Cal State LA)

Definition 1.3.1. (Uniform Continuity (from Math 4650).) Let $D \subseteq \mathbb{R}$ and let $f : D \to \mathbb{R}$. We say that f is **uniformly continuous** on D if $\forall \ \epsilon > 0 \ \exists \ \delta > 0$ where

$$x, y \in D \text{ and } 0 < |x - y| < \delta \implies |f(x) - f(y)| < \epsilon$$

Theorem 17. (Uniform continuity implies continuity; from Math 4650.) Suppose $f : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}$. If f is uniformly continuous on D, then f is continuous at every $a \in D$.

Theorem 18 (**Theorem 2.42 in Pugh [2015]**)**.** Every continuous function defined on a compact set is uniformly continuous.

Definition 1.3.2. Let S be any set, and let $s, s', s'' \in S$. An **equivalence relation** on S is a relation $s \sim s'$ that holds between some members $s, s' \in S$ and satisfies three properties:

(a) $s \sim s$

(b) $s \sim s'$ implies $s' \sim s$.

(c) $s \sim s'$ and $s' \sim s''$ implies that $s \sim s''$.

An **equivalence class** containing s consists of all elements $s' \in S$ equivalent to s. It is denoted by $[s]$. The element s is **representative** of its equivalence class. Equivalence classes are disjoint subsets.

Definition 1.3.3 (Well-defined; from *Logical Prerequisites* of Lang [2005]; p. 11 of pdf, p. X of book). Let A be a set with an equivalence relation and let E be an equivalence class of elements in A. We sometimes try to define a map of E into some set B. To define such a map f, we sometimes first give its value on a representative of E and then show that it is independent of the choice of representative $x \in E$. In that case we say f is **well-defined**.

1.4 A Taste of Topology (Chapter 2 of Pugh [2015]; also from Homework 5 of Math 4650 at Cal State LA)

Definition 1.4.1 (from Math 4650). Let $S \subseteq \mathbb{R}$. We say $x \in \mathbb{R}$ is an **interior point** of S if there exists an open interval (a, b) where

$$x \in (a, b) \text{ and } (a, b) \subseteq S.$$

Definition 1.4.2 (Open sets; from Math 4650). Let $S \subseteq \mathbb{R}$. We say S is **open** if every $x \in S$ is an interior point of S.

Definition 1.4.3 (Metric space; from Pugh [2015] Section 2.1, p. 67 of pdf, p. 57 of book.). A **metric space** is a set M, the elements of which are referred to as points of M, together with a **metric** d satisfying positive definiteness, symmetry, and the triangle inequality. The metric $d(x, y)$ is a real number defined for all points $x, y \in M$, and $d(x, y)$ is called the **distance** from the point x to the point y.

Strictly speaking, it is the pair (M, d) which is a metric space, but often M it self is referred to as a metric space.

The main examples of metric spaces are $\mathbb{R}^n$ and their subsets, using the Euclidean metric. A subset $M \subset \mathbb{R}^n$ becomes a metric space when we declare the distance between points of M to be their Euclidean distance apart as points in $\mathbb{R}^n$. We say that M **inherits** its metric from $\mathbb{R}^n$ and is a **metric subspace** of $\mathbb{R}^n$.

Definition 1.4.4 (Open sets; from Pugh [2015] Section 2.3.). Let M be a metric space and let S be a subspace of M. S is a **closed set** if it contains all its limit points.

Definition 1.4.5 (Closed sets; from Math 4650). Let $S \subseteq \mathbb{R}$. We say S is **closed** if $\mathbb{R} \setminus S$ is open.

Definition 1.4.6 (Closed sets; from Pugh [2015] Section 2.3.). Let M be a metric space and let S be a subspace of M. S is an **open set** if for every $p \in S$ there exists an $r > 0$ such that

$$d(p, q) < r \implies q \in S.$$

Theorem 19 (from Math 4650). A set is closed if and only if it contains all of its limit points.

(**Facts about open and closed sets; from Math 4650.**) Suppose $a, b \in \mathbb{R}$. Then

-

  Proposition 20 (from Math 4650). (a, ∞) is open.

  Proof. Let $x \in (a, \infty)$. Since $x > a$, $\exists\, \epsilon > 0 \mid a + \epsilon = x$. Then $a = x - \epsilon < x - \frac{\epsilon}{2} < x < x + \frac{\epsilon}{2} < \infty$. Therefore $x \in (x - \frac{\epsilon}{2}, x + \frac{\epsilon}{2}) \subseteq (a, \infty)$, so (a, ∞) is open. $\qquad\square$

-

  Proposition 21. $(-\infty, b)$ is open.

  Proof. Let $x \in (-\infty, b)$. Since $x < b$, $\exists\, \epsilon > 0 \mid b - \epsilon = x$. Then $-\infty < x - \frac{\epsilon}{2} < x < x + \frac{\epsilon}{2} < x + \epsilon = b$. Therefore $x \in (x - \frac{\epsilon}{2}, x + \frac{\epsilon}{2}) \subseteq (-\infty, b)$, so $(-\infty, b)$ is open. $\qquad\square$

-

  Proposition 22. (a, b) is open.

  Proof. In class notes. $\qquad\square$

-

  Proposition 23. If $a < b$, then $[a, b]$ is closed.

  Proof. Consider $\mathbb{R} \setminus [a, b] = (-\infty, a) \cup (b, \infty)$. By Proposition 21, $(-\infty, a)$ is open. By Proposition ra.hw5.5b, (b, ∞) is open. By Proposition 24, the union of two open sets is open. Therefore $\mathbb{R} \setminus [a, b]$ is open, so $[a, b]$ is closed. $\qquad\square$

-

  Proposition 24 (from Math 4650). If A and B are open, then $A \cup B$ is open.

  Proof. Since A is open, $\forall\, x_A \in A\, \exists\, (a_A, b_A) \subseteq A \mid x_A \in (a_A, b_A)$. Since B is open, $\forall\, x_B \in B\, \exists\, (a_B, b_B) \subseteq B \mid x_B \in (a_B, b_B)$.

  Let $x \in A \cup B$. If $x \in A$, then per above $\exists\, (a_A, b_A) \subseteq A \subseteq A \cup B \mid x_A \in (a_A, b_A)$. If $x \in B$, then per above $\exists\, (a_B, b_B) \subseteq B \subseteq A \cup B \mid x_B \in (a_B, b_B)$. Therefore $A \cup B$ is open. $\qquad\square$

-

  Proposition 25 (from Math 4650). If A and B are open, then $A \cap B$ is open.

  Proof. Since A is open, $\forall\, x_A \in A\, \exists\, (a_A, b_A) \subseteq A \mid x_A \in (a_A, b_A)$. Since B is open, $\forall\, x_B \in B\, \exists\, (a_B, b_B) \subseteq B \mid x_B \in (a_B, b_B)$.

  Let $x \in A \cap B$. Then $x \in A$ and $x \in B$, so $\exists\, (a_A, b_A) \subseteq A \mid x_A \in (a_A, b_A)$, and $\exists\, (a_B, b_B) \subseteq B \mid x_B \in (a_B, b_B)$. Let $a = \max\{a_A, a_B\}$, and $b = \min\{b_A, b_B\}$. Since $x > a$ and $x < b$, $x \in (a, b)$. Since $(a, b) \subseteq (a_A, b_A) \subseteq A$ and $(a, b) \subseteq (a_B, b_B) \subseteq B$, $(a, b) \subseteq A \cap B$. Therefore $A \cap B$ is open. $\qquad\square$

-

  Proposition 26 (from Math 4650). If A and B are closed, then $A \cup B$ is closed.

Proof. Since A is closed, $\mathbb{R} \setminus A$ is open. Since B is closed, $\mathbb{R} \setminus B$ is open. $\mathbb{R} \setminus (A \cup B) = (\mathbb{R} \setminus A) \cap (\mathbb{R} \setminus B)$. By Proposition 25 the intersection of two open sets is open. Therefore $\mathbb{R} \setminus (A \cup B)$ is open, so $A \cup B$ is closed.

$\square$

-

Proposition 27 (from Math 4650). If A and B are closed, then $A \cap B$ is closed.

Proof. Since A is closed, $\mathbb{R} \setminus A$ is open. Since B is closed, $\mathbb{R} \setminus B$ is open. $\mathbb{R} \setminus (A \cap B) = (\mathbb{R} \setminus A) \cup (\mathbb{R} \setminus B)$. By Proposition 24, the union of two open sets is open. Therefore $\mathbb{R} \setminus (A \cap B)$ is open, so $A \cap B$ is closed.

$\square$

-

Proposition 28 (from Math 4650). $\mathbb{R}$ is open and closed.

Proof. Let $\epsilon > 0$. Let $x \in \mathbb{R}$. Then $x - \epsilon, x + \epsilon \in \mathbb{R}$, and $x \in (x - \epsilon, x + \epsilon)$. Therefore $\mathbb{R}$ is open. $\mathbb{R}$ is closed because by Proposition 29, $\mathbb{R} \setminus \mathbb{R} = \emptyset$ is open. $\square$

-

Proposition 29 (from Math 4650). $\emptyset$ is open and closed.

Proof. To show that a set S is open, we most show that $\forall \, x \in S \, \exists \, S' \subseteq S | \, x \in S'$ where S' is open. Since there are no $x \in \emptyset$, this condition is satisfied for $\emptyset$. $\emptyset$ is closed because per Proposition 28, $\mathbb{R} \setminus \emptyset = \mathbb{R}$ is open.

$\square$

Proposition 30 (from Math 4650). Let $x_1, x_2, \ldots, x_n$ be real numbers. Let S be the finite set $S = \{x_1, x_2, \ldots, x_n\}$. Then S is closed.

Proof. Consider $\mathbb{R} \setminus S = (-\infty, x_1) \cup (x_1, x_2) \cup \ldots \cup (x_{n-1}, x_n) \cup (x_n, \infty)$. $(-\infty, x_1)$ is open by Proposition 21. (x_n, ∞) is open by Proposition 20.

Consider (x_i, x_{i+1}) where $i \in \{1, 2, 3, \ldots, n-1\}$. Let $x \in (x_i, x_{i+1})$. Then since $x > x_i$ and $x < x_{i+1}$, $\exists \, \epsilon > 0 \mid x_i + \epsilon = x$ and $\exists \, \delta > 0 \mid x_{i+1} - \delta = x$. Then $x_i = x - \epsilon < x - \frac{\epsilon}{2} < x < x + \frac{\delta}{2} < x + \delta = x_{i+1}$. Therefore $x \in (x - \epsilon, x + \delta) \subseteq (x_i, x_{i+1})$, so (x_i, x_{i+1}) is open.

Finally, since $\mathbb{R} \setminus S$, by Proposition 20 (and induction) $\mathbb{R} \setminus S$ is open. Therefore S is closed. $\square$

Proposition 31 (from Math 4650). Let $x_1, x_2, \ldots, x_n$ be real numbers. Let S be the finite set $S = \{x_1, x_2, \ldots, x_n\}$. Then S has no limit points.

Proof. Per Definition 1.2.2, we seek to show that (1) $\forall \, x_i \in S \, \exists \, \delta_i$ such that $\forall \, x_j \in D(x_j \neq x_i)$

$$|x_j - x_i| \geq \delta_i$$

and (2) $\forall \, x \in \mathbb{R} \setminus S \, \exists \, \delta_x$ such that $\forall \, x_i \in D$

$$|x_i - x| \geq \delta_x$$

(1) Let $x_i \in S$. Let $\delta_i = \frac{1}{2} \min\{|x_i - x_k| \mid x_k \neq x_i\}$. Then $\forall\ x_j \neq x_i \in S$,

$$|x_i - x_j| \geq |x_i - x_k| > \delta_i$$

(2) Let $x \in \mathbb{R} \setminus S$. Let $\delta_x = \frac{1}{2} \min\{|x - x_i| \mid x_i \in S\}$. Then

$$|x_i - x| \geq \min\{|x - x_i| \mid x_i \in S\} > \delta_x$$

$\square$

Definition 1.4.7 (from Math 4650). Let $S \subseteq \mathbb{R}$. An **open cover** of S is a collection $X = \{\mathcal{O}_\alpha \mid \alpha \in I\}$ where each set $\mathcal{O}_\alpha$ is an open subset of $\mathbb{R}$ such that

$$S \subseteq \bigcup_{\alpha \in I} \mathcal{O}_\alpha$$

(Here I is some set that indexes the $\mathcal{O}_\alpha$).

Definition 1.4.8 (from Math 4650). If $X' \subseteq X$ such that

$$S \subseteq \bigcup_{\mathcal{O}_\alpha \in X'} \mathcal{O}_\alpha$$

then X' is called a **subcover** of S contained in X. In addition, if X' is finite then we call X' a **finite subcover** of S contained in X.

Definition 1.4.9 (Compactness; from Math 4650.). Let $S \subseteq \mathbb{R}$. We say that S is **compact** if every open cover of S contains a finite subcover.

Definition 1.4.10 (Compactness; from Pugh [2015] Section 2.4.). A subset A of a metric space M is (sequentially) **compact** if every sequence (a_n) in A has a subsequence (a_{n_k}) that converges to a limit in A.

Definition 1.4.11 (from Math 4650). Let $S \subseteq \mathbb{R}$. We say that S is **bounded** if $\exists\ M > 0$ where $S \subseteq [-M, M]$.

Remark 3. S is bounded if and only if $|s| \leq M\ \forall\ s \in S$.

Theorem 32 (Heine-Borel Theorem; from Math 4650; Theorem 2.33 in Pugh [2015]). Let $S \subseteq \mathbb{R}^m$. S is compact if and only if S is closed and bounded.

Proposition 33 (from Math 4650). Let $x_1, x_2, \ldots, x_n$ be real numbers. Let S be the finite set $S = \{x_1, x_2, \ldots, x_n\}$. Then S is compact.

Proof. Let $\{O_\alpha\}$ be an open cover of S. By definition of open cover, $\forall\ i\ \exists\ O_{\alpha_i}$ such that $x_i \in O_{\alpha_i}$. Thus, $\{O_{\alpha_1}, O_{\alpha_2}, \ldots, O_{\alpha_n}\}$ is a finite subcover of S. $\square$

Proposition 34 (from Math 4650). Let A and B be compact subsets of $\mathbb{R}$. Then $A \cap B$ is compact.

Proof. Since $A \cap B \subseteq A$, $A \cap B \subseteq [-M_A, M_A]$. Therefore $A \cap B$ is bounded.

Since $A \cap B$ is closed and bounded, by the Heine-Borel Theorem (Theorem 32), $A \cap B$ is compact.

$\square$

Proposition 35 (from Math 4650). Let A and B be compact subsets of $\mathbb{R}$. Then $A \cup B$ is compact.

Proof. Let $M = \max\{M_A, M_B\}$. Note that $[-M_A, M_A] \subseteq [M, M]$ and $[-M_B, M_B] \subseteq [-M, M]$. This implies $A \subseteq [-M, M]$ and $B \subseteq [-M, M]$. Therefore $A \cup B \subseteq [-M, M]$.

Since $A \cup B$ is closed and bounded, by the Heine-Borel Theorem (Theorem 32), $A \cup B$ is compact.

$\square$

Theorem 36 (from Math 4650). Let $f : D \to \mathbb{R}$ be continuous on D. If $X \subseteq D$ and X is compact (closed and bounded), then

$$f(\bar{x}) = \{f(x) \mid x \in X\}$$

is compact (closed and bounded).

Corollary 36.1. Suppose $f : D \to \mathbb{R}$ where D is closed and bounded. Then there exists $a, b \in D$ where $f(a)$ is the min of f on D and $f(b)$ is the max of f on D.

Definition 1.4.12 (Section 2.3 of Pugh [2015]). A **topological space** is an ordered pair $(X, \mathcal{T})$, where X is a set and $\mathcal{T}$ is a collection of subsets of X satisfying the following axioms:

1. $\emptyset$ and X belong to $\mathcal{T}$.

2. Any union of elements of $\mathcal{T}$ is an element of $\mathcal{T}$

3. The intersection of any finite number of members of $\mathcal{T}$ belongs to $\mathcal{T}$.

The elements of $\mathcal{T}$ are called **open sets** and the collection $\mathcal{T}$ is called a **topology** on X.

Theorem 37 (Theorem 2.6 in Pugh [2015]). A metric space and all of its open subsets form a topological space.

Theorem 38 (Urysohn's subsequence principle; Exercise 1.6.8 in Tao [2010]). Let x_n be a sequence in a topological space X (e.g., $\mathbb{R}$), and let x be a point in X. Then the following are equivalent:

(a) x_n converges to x.

(b) Every subsequence of x_n converges to x.

(c) Every subsequence of x_n has a further subsequence that converges to x.

Definition 1.4.13. An **isomorphism** is a mapping between two structures of the same type that can be reversed by an inverse mapping.

Definition 1.4.14 (Section 2.2 of Pugh [2015]). Let M and N be topological spaces (see Definition 1.4.12). Let $f : M \to N$ be a bijection. If f is continuous and the inverse bijection $f^{-1} : N \to M$ is also continuous then f is a **homeomorphism**. (That is, a homeomorphism is an isomorphism of topological spaces, or a bicontinuous bijection.)

Definition 1.4.15. Let (M, d_M) and (N, d_N) be metric spaces. An **isometry** is a bijection $f : M \to N$ that preserves distance: for all $p, q \in M$ we have $d_N(f(p), f(q)) = d_M(p, q)$. An isometry of M is a homeomorphism $i : M \to M$ such that $d_M(i(x), i(y)) = d_M(x, y)$ for all $x, y \in M$. (That is, an isometry is an isomorphism of metric spaces.)

1.5   Functions of a Real Variable; Differentiation, Riemann Integration, and Series (Chapter 3 of Pugh [2015])

Theorem 39 (**Theorem 3.21 in Pugh [2015]**)**.** A bounded function is Riemann integrable if and only if for every $\epsilon > 0$ there exists a partition P_0 of $[a, b]$ such that $U(f, P_0) - L(f, P_0) < \epsilon$, where $U(f, P_0) := \sup_T\{R(f, P_0, T)\}$ is the upper sum of f with respect to P_0 and $L(f, P_0) := \inf_T\{R(f, P_0, T)\}$ is the lower sum.

1.6   Function Spaces (Chapter 4 of Pugh [2015])

1.6.1   Uniform Convergence and C^0 (Section 4.1 of Pugh [2015])

Definition 1.6.1 (**Pointwise convergence; from Section 4.1 of Pugh [2015]**)**.** Let $\mathcal{X}$ be a set, let $(\mathcal{Y}, d)$ be a metric space. Let $f_n : \mathcal{X} \to \mathcal{Y}$ be functions for $n \geq 1$. Let $f : \mathcal{X} \to \mathcal{Y}$ be another function. We say f_n **converges to** f **pointwise** if for all $x \in \mathcal{X}$, the sequence $\{f_n(x)\}_{n=1}^\infty$ converges to $f(x)$ as a sequence of points in $\mathcal{Y}$. I.e., if for all $\epsilon > 0$ and for all $x \in \mathcal{X}$, there exists N (maybe depending on ϵ and x) such that for $n \geq N$, we have $d(f_n(x), f(x)) < \epsilon$.

Definition 1.6.2 (**Uniform convergence; from Section 4.1 of Pugh [2015]**)**.** Let $\mathcal{X}$ be a set, let $(\mathcal{Y}, d)$ be a metric space. Let $f_n : \mathcal{X} \to \mathcal{Y}$ be functions for $n \geq 1$. Let $f : \mathcal{X} \to \mathcal{Y}$ be another function. We say f_n **converges uniformly to** f if for all $\epsilon > 0$, there exists N (maybe depending on ϵ bit not x) such that for $n \geq N$ and for all $x \in \mathcal{X}$, we have $d(f_n(x), f(x)) < \epsilon$.

Remark 4. Note that uniform convergence implies pointwise convergence. This fact implies that uniform limits are unique, since limits in metric space $(\mathcal{Y}, d)$ are unique, so the pointwise limits of sequences of functions are unique.

For an example of pointwise but not uniform convergence, see Example 1.6.1 below.

Theorem 40 (**Cauchy 1821; Theorem 4.1 from Pugh [2015]**)**.** Let (X, d) and (Y, d') be metric spaces. Let $f_n : X \to Y$ be continuous at $x_0 \in X$ for all $n \geq 1$. Suppose f_n converges uniformly to f for some $f : X \to Y$. Then f is continuous at x_0. (This implies that if all f_n are continuous everywhere and f_n converges uniformly to f then f is continuous everywhere.)

Proof. Let $\epsilon > 0$. Choose $N \geq 1$ such that we have

$$d'(f_n(x), f(x)) < \epsilon/3 \qquad \text{for } n \geq N \text{ and for all } x \in X. \tag{1.2}$$

The function f_N is continuous at x_0, so there exists $\delta > 0$ such that if $x \in X$ and $d(x, x_0) < \delta$, then $d'(f_N(x), f_N(x_0)) < \epsilon/3$. Then if $x \in X$, $n \geq N$, and $d(x, x_0) < \delta$, by (1.2)

$$d'(f(x), f(x_0)) \leq d'(f(x), f_n(x)) + d'(f_n(x), f_n(x_0)) + d'(f_n(x_0), f(x_0)) < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

$\square$

We can also show an alternative proof, but we need a few definitions and results.

Definition 1.6.3 (**Directed sets and nets; from Homework 1 in 425B**)**.** Let (X, d) be a metric space. Let A be a set with a binary relation $\preceq$ satisfying

(a) $a \preceq a$ for all $a \in A$ (reflexivity),

(b) if $a \preceq b$ and $b \preceq c$ then $a \preceq c$ (transitivity), and

(c) for any $a, b \in A$ there exists $c \in A$ with $a \preceq c$ and $b \preceq c$ (upper bounds exist).

Such a set A is called a **directed set**. A function $f : A \to X$ is called a **net** from A to X.

If f is a net from A to $\mathbb{R}$, we say that f **converges to a limit** L if for every $\epsilon > 0$ there exists an element $a_0 \in A$ such that for all $a \in A$ with $a_0 \preceq a$ we have $|f(a) - L| < \epsilon$. We write $\lim f = L$.

More generally, a net **converges to** $I \in X$ if for every $\epsilon > 0$ there exists $a_0 \in A$ such that for all $a \in A$ with $a_0 \preceq a$ we have $d(f(a), I) < \epsilon$.

Proposition 41. Limits of nets in $\mathbb{R}$ are unique: if $f : A \to \mathbb{R}$ is a net, $\lim f = L$, and $\lim f = L'$, then $L = L'$.

Proof. Let $L, L' \in \mathbb{R}$, with $\lim f = L$ and $\lim f = L'$. Then for all $\epsilon > 0$ there exists $N \in A$ such that for all $n \in A$ such that $N \preceq n$ we have $|f(n) - L| < \epsilon$, and there exists $N' \in A$ such that for all $n \in A$ such that $N' \preceq n$ we have $|f(n) - L'| < \epsilon$. Suppose $L \neq L'$, and let $\epsilon := \frac{1}{2}|L - L'| > 0$. Choose some $N^* \in A$ such that $N \preceq N^*$ and $N' \preceq N^*$ (such an N^* exists by Property 3 of the relation $\preceq$ from Definition 1.6.3). Then

$$2\epsilon = |L - L'| = |L - f(N^*) + f(N^*) - L'| \leq |f(N^*) - L| + |f(N^*) - L'| < \epsilon + \epsilon = 2\epsilon,$$

contradiction. Therefore $L = L'$.

$\square$

Proposition 42. For a closed interval $[a, b]$, the set A of partition pairs / tagged partitions (P, T) of $[a, b]$ is a directed set given the relation $(P, T) \preceq (P', T')$ when P' is a refinement of P.

Proof. We will check one at a time that the conditions of Definition 1.6.3 hold.

- **Reflexivity:** The definition on p.168 (Section 3.2) of Pugh [2015] says that a partition P' is a refinement of P if $P \subset P'$. If we interpret this to mean P is a subset of P' but not necessarily a proper subset (sometimes written $P \subseteq P'$), then it holds that $P \preceq P$ for all $(P, T) \in A$, satisfying reflexivity.

- **Transitivity:** Suppose we have $(P_1, T_1), (P_2, T_2), (P_3, T_3) \in A$ with $(P_1, T_1) \preceq (P_2, T_2)$ and $(P_2, T_2), \preceq (P_3, T_3)$. By definition, $P_1 \subset P_2$ and $P_2 \subset P_3$, so by transitivity of sets $P_1 \subset P_3$. Therefore $(P_1, T_1) \preceq (P_3, T_3)$ as desired.

- **Upper bounds exist:** Let $(P_1, T_1), (P_2, T_2) \in A$. Let $P^* := P_1 \cup P_2$ (this is a set of finite cardinality upper-bounded by $|P_1| + |P_2|$), and let T^* be any valid set of $|P^*| - 1$ sample points relative to P^*. Then since $P_1 \subset P^*$ and $P_2 \subset P^*$, we have $(P_1, T_1) \preceq (P^*, T^*)$ and $(P_2, T_2) \preceq (P^*, T^*)$.

□

Proposition 43. For a function $f : [a, b] \to \mathbb{R}$, the net from A to $\mathbb{R}$ given by $(P, T) \mapsto R(f, P, T)$ converges to a real number $I \in \mathbb{R}$ if and only if f is Riemann integrable with integral I.

Proof. • **Riemann integrability implies convergence of the net:** Suppose f is Riemann integrable with integral $I \in \mathbb{R}$. Let $\epsilon > 0$. By the definition in Section 3.2 of Pugh [2015], we know that there exists $\delta > 0$ such that for any partition pair (P^*, T^*) satisfying mesh $P^* < \delta$, $|R(f, P^*, T^*) - I| < \epsilon$. Fix one such partition pair (P_0, T_0). Consider any $(P, T) \in A$ with $(P_0, T_0) \preceq (P, T)$. Since P is a refinement of P_0, $P_0 \subset P$, so mesh $P \leq$ mesh $P_0 < \delta$. By the definition of Riemann integrability, we have that mesh $P < \delta \implies |R(f, P, T) - I| < \epsilon$. Therefore for all $(P, T) \in A$ with $(P_0, T_0) \preceq (P, T)$ we have $|R(f, P, T) - I| < \epsilon$ and the net converges to I.

• **Convergence of the net implies Riemann integrability:** Let $\epsilon > 0$. Suppose the net $(P, T) \mapsto R(f, P, T)$ converges to $I \in \mathbb{R}$. Then there exists $(P_0, T_0) \in A$ such that for all $(P, T) \in A$ with $(P_0, T_0) \preceq (P, T)$ (that is, $P_0 \subset P$), we have $|R(f, P, T) - I| < \epsilon/2$.

First we will show by contradiction that this implies f is bounded. Let $R := R(f, P_0, T_0)$. Since $(P_0, T_0) \preceq (P_0, T_0)$, we have $|R - I| < \epsilon/2$. Denote the elements of P_0 as $P_0 = \{x_0, \ldots, x_n\}$ and T_0 as $\{t_1, \ldots, t_n\}$. Suppose f is unbounded on $[a, b]$. Then there is also a subinterval $[x_{i_0-1}, x_{i_0}]$ on which it is unbounded. Choose a new set $T_0' = \{t_1', \ldots, t_n'\}$ with $t_i' = t_i$ for all $i \neq i_0$ and choose t_{i_0}' such that $|f(t_{i_0}') - f(t_{i_0})|\Delta x_{i_0} > \epsilon$, where $\Delta x_{i_0} := x_{i_0} - x_{i_0-1}$ (such a t_{i_0}' exists because the supremum of $\{|f(t)| : x_{i_0-1} \leq t \leq x_{i_0}\}$ is ∞). Let $R' := R(f, P_0, T_0')$. Then $|R - R'| > \epsilon$, and also $|R' - I| < \epsilon/2$ since $(P_0, T_0) \preceq (P_0, T_0')$. But

$$\epsilon < |R - R'| = |R - I + I - R'| \leq |R - I| + |R' - I| < \epsilon/2 + \epsilon/2 = \epsilon,$$

contradiction.

Next, let $T_U := \arg\sup_T\{R(f, P_0, T)\}$ and let $T_L := \arg\inf_T\{R(f, P_0, T)\}$, and denote $U(f, P_0) := R(f, P_0, T_U)$ and $L(f, P_0) := R(f, P_0, T_L)$. Then $(P_0, T_0) \preceq (P_0, T_L)$ and $(P_0, T_0) \preceq (P_0, T_U)$, so $|L(f, P_0) - I| = I - L(f, P_0) < \epsilon/2$ and $|U(f, P_0) - I| = U(f, P_0) - I < \epsilon/2$. Then

$$U(f, P_0) - L(f, P - 0) = U(f, P_0) - I + I - L(f, P_0) < \epsilon.$$

Applying Theorem 39 completes the proof.

□

Lemma 44 (Math 425b Homework 2). Let A be a directed set. Consider a sequence $f_n(a)$ of nets from A to a metric space Y. Assume that

(a) For each fixed n, the net $f_n(a)$ converges uniformly to a limit L_n in Y.

(b) The nets f_n converge uniformly (as a sequence of functions $A \to Y$) to some net $f : A \to Y$.

Then the limits $(L_n)_{n=1}^\infty$ form a Cauchy sequence in Y.

Proof. Let $\epsilon > 0$ be arbitrary. The limits $(L_n)_{n=1}^\infty$ form a Cauchy sequence in Y if there exists $N \in \mathbb{N}$ such that for all $n, k \geq N$ we have $d(L_n, L_m) < \epsilon$. Because the nets f_n converge uniformly, by the Cauchy

convergence criterion (Theorem 1.6 in Pugh [2015], or the completeness of $\mathbb{R}$, Theorem 1.5) we know that for all $a \in A$ there exists N such that for all $n, m \geq N$ we have $d(f_n(a), f_m(a)) < \epsilon/3$ for all $a \in A$.

Next, consider fixed integers $n, m \geq N$. By assumption, the net $f_n(a)$ converges to $L_n \in Y$, so there exists $a_1 \in A$ such that for all $a \in A$ with $a_1 \preceq a$, we have $d(f_n(a), L_n) < \epsilon/3$. Similarly, because $\lim f_m = L_m$ there exists $a_2 \in A$ such that for all $a \in A$ with $a_2 \preceq a$, we have $d(f_m(a), L_m) < \epsilon/3$. By the upper bound property of directed sets (from Definition 1.6.3), there exists $a_0 \in A$ with $a_1 \preceq a_0$ and $a_2 \preceq a_0$, and we have $d(L_n, f_n(a_0)) < \epsilon/3$ and $d(f_m(a_0), L_m) < \epsilon/3$. Putting this together, we have

$$d(L_n, L_m) \leq d(L_n, f_n(a_0)) + d(f_n(a_0), f_m(a_0)) + d(f_m(a_0), L_m) < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.$$

$\square$

Lemma 45 (**Math 425b Homework 2**). Let A be a directed set. Consider a sequence $f_n(a)$ of nets from A to a metric space Y. Assume that

(a) For each fixed n, the net $f_n(a)$ converges uniformly to a limit L_n in Y.

(b) The nets f_n converge uniformly (as a sequence of functions $A \to Y$) to some net $f : A \to Y$.

(c) The sequence $(L_n)_{n=1}^{\infty}$ converges to some limit $L \in Y$.

Then the limit of f (as a net from A to Y) exists and equals L.

Proof. Let $\epsilon > 0$. Suppose the sequence $(L_n)_{n=1}^{\infty}$ converges to some limit $L \in Y$. Then there exists $N \in \mathbb{N}$ such that there exists $n \geq N$ with $d(L_n, L) < \epsilon/3$.

Next, note that $d(f(a), L_n) \leq d(f(a), f_n(a)) + d(f_n(a), L_n)$. Because $f_n(a)$ converges to L_n, we know there exists $a_0 \in A$ such that for all $a_2 \in A$ satisfying $a_0 \preceq a$ we have $d(f_n(a_2), L_n) < \epsilon/3$. Choose one such a_2. Further, since f_n converges uniformly to f, there exists $a_0' \in A$ such that for all $a_1 \in A$ satisfying $a_0' \preceq a_1$, it holds that $d(f(a_1), f_n(a_1)) < \epsilon/3$.

Choose $a_0 \in A$ satisfying $a_1 \preceq a_0$ and $a_2 \preceq a_0$ (again, this a_0 exists by the upper bound property of directed sets from Definition 1.6.3). Then for all $a_0 \preceq a$,

$$d(f(a), L) \leq d(f(a), L_n) + d(L_n, L) \leq d(f(a), f_n(a)) + d(f_n(a), L_n) + d(L_n, L) < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.$$

$\square$

Now we are ready to present a different proof of Theorem 40.

Alternative proof of Theorem 40, using directed sets and nets. Let (X, d) and (Y, d') be metric spaces. Let (f_n) be a sequence of functions from X to Y. Fix $x \in X$, and let $\epsilon > 0$. Suppose each $f_n : X \to Y$ is continuous at x; that is, we know that for all $n \in \mathbb{N}$, there exists $\delta_n > 0$ such that for all $\tilde{x}_n \in X$ satisfying $d(\tilde{x}_n, x) < \delta_n$ it holds that $d'(f_n(\tilde{x}_n), f_n(x)) < \epsilon$. Further, suppose that the sequence of functions (f_n) converge uniformly to $f : X \to Y$. That is, there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have that

$$d'(f_n(x_0), f(x_0)) < \epsilon \qquad \forall x_0 \in X. \tag{1.3}$$

We then want to show that f is continuous at x; that is, there exists $\delta > 0$ such that for every $\tilde{x} \in X$ satisfying $d(\tilde{x}, x) < \delta$ we have $d'(f(x), f(\tilde{x})) < \epsilon$.

Consider the directed set $A = X \setminus \{x\}$ with ordering relation $a \preceq a'$ if $d(a', x) \leq d(a, x)$. Let each function f_n be restricted to a function from A to Y (a net from A to Y).

First, suppose x is a non-isolated point and consider a function $g : X \to Y$. Let $\epsilon_g > 0$. We will show that g is continuous at $x = X$ if and only if the restriction of g to A viewed as a net from A to Y converges as a net to $g(x)$. That is, we will show that

$$\exists\ \delta_g > 0 \text{ s.t. } d(\tilde{x}_g, x) < \delta_g \implies d'(g(\tilde{x}_g), g(x)) < \epsilon_g$$

if and only if

$$\exists\ \hat{x}_g \in A \text{ s.t. } \hat{x}_g \preceq x \implies d'(g(\hat{x}_g), g(x)) < \epsilon_g.$$

Suppose there exists $\delta_g > 0$ such that $d(\tilde{x}_g, x) < \delta_g \implies d'(g(\tilde{x}_g), g(x)) < \epsilon_g$. Then take any such $\tilde{x}_g$ (such a point exists since x is a non-isolated point) and let $\hat{x}_g = \tilde{x}_g$. Then for all $\hat{x} \in X$ satisfying $\hat{x}_g \preceq \hat{x}$, we have $d(\hat{x}, x) \leq d(\hat{x}_g, x) = d(\tilde{x}_g, x) < \delta$, so it holds that $d'(g(\hat{x}_g), g(x)) < \epsilon_g$. Clearly the converse of this argument holds as well.

Because of this fact, since $f_n(x)$ is continuous at $x \in X$ for all $n \in \mathbb{N}$, we know that the restriction of f_n to A as a net from A to Y converges as a net to $f_n(x)$.

The sequence $(f_n(x))_{n=1}^\infty$ converges to $L \in Y$ if there exists $N \in \mathbb{N}$ such that $n \geq N \implies d'(f_n(x), L) < \epsilon$. But this is exactly what we know from (1.3), with $L = f(x)$.

We now have that for each n, the net $f_n(a)$ converges to a limit L_n in Y, the nets f_n converge uniformly as a sequence of functions $A \to Y$ to $f : A \to Y$, and the sequence $(L_n)_{n=1}^\infty$ converges to $f(x) \in Y$. Therefore we know from Lemma 45 that the limit of f as a net from A to Y exists and equals L. We also know from Lemma 44 that the limits $(L_n)_{n=1}^\infty$ form a Cauchy sequence in Y. Therefore f is continuous at x.

$$\vdots$$

Then it follows that for every $\tilde{x} \in X$ satisfying $d(\tilde{x}, x) < \delta$ we have

$$d'(f(x), f(\tilde{x})) \leq \ldots < \epsilon.$$

$\square$

What about if you only have pointwise convergence? Then the proof of Theorem 40 won't work. Consider the sawtooth wave example: the limit is not continuous. There exist easier counterexamples as well.

Example 1.6.1 (From Section 4.1, p. 212 of Pugh [2015])**.** Let $X = [0,1]$, $Y = \mathbb{R}$, $f_n(x) = x^n$. Let $f : [0,1] \to \mathbb{R}$ be defined by

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

As before, f_n converges to f pointwise but not uniformly. f_n is continuous on $[0,1]$ for all n but f is not. See Figure 1.1.
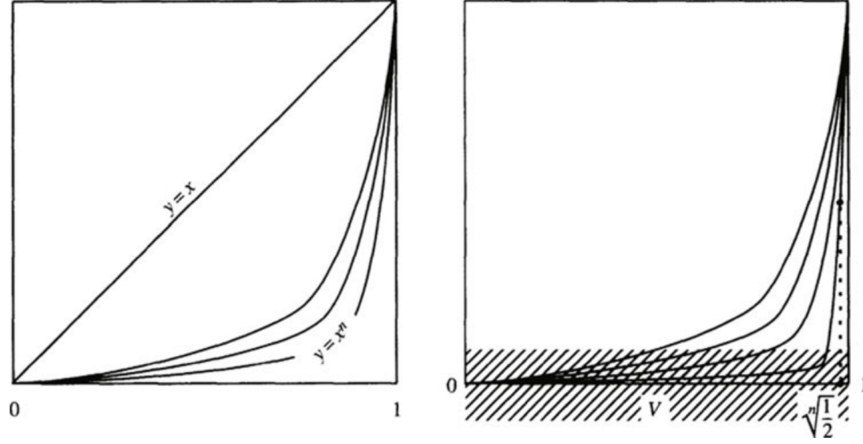
Figure 1.1: Figure 88 from Section 4.1, p. 213 of Pugh [2015]

Proposition 46. In Example 1.6.1, if we restrict f_n and f to the open interval from 0 to 1 (so that f is simply the 0 function on this interval) f_n does not converge uniformly to f on this interval.

Proof. Suppose f_n converges uniformly to 0 (the pointwise limit). Take $\epsilon = 1/2$. Can choose $N \ge 1$ such that $|x^n - 0| < 1/2$ for all $n \ge N$, and for all $x \in (0,1)$. Let $x = (3/4)^{1/N}$. (Nth roots exist by IVT.) Then $|x^N - 0| = 3/4$ which is not less than $1/2$.

$\square$

Proposition 47. Consider the intervals $E_1 := [0, 1/2) = [1 - (1/2)^0, 1 - (1/2)^1)$, $E_2 := [1/2, 1 - 1/2^2) = [1 - (1/2)^1, 1 - (1/2)^2)$, $[1 - 1/2^2, 1 - 1/2^3), \ldots$. On each E_j, f_n converges uniformly to f.

Proof. Observe that on interval E_j, we have uniform continuity by the following argument: for $\epsilon > 0$, let $N_j = \left\lceil \frac{\log(\epsilon)}{\log\left(1 - \frac{1}{2^j}\right)} \right\rceil$. Note that it follows that for $n > N_j$,

$$n > \frac{\log(\epsilon)}{\log\left(1 - \frac{1}{2^j}\right)} \qquad \Longleftrightarrow \qquad n \log\left(1 - \frac{1}{2^j}\right) < \log(\epsilon) \qquad \Longleftrightarrow \qquad \left(1 - \frac{1}{2^j}\right)^n < \epsilon.$$

Then if $n \ge N_j$, for all $x \in E_j$,

$$f_n(x) = x^n \le \left(1 - \frac{1}{2^j}\right)^n < \epsilon.$$

$\square$

Exercise 1. Let X be a set, (Y, d) be a metric space. Let $x_0 \in X$ be fixed. Let $f : X \to Y$ be a function. Then the following are equivalent:

1. There exists $m \in \mathbb{R}$ such that $d(f(x_0), f(x)) \leq m$ for all $x \in X$.

2. There exists $x_1 \in X$ and $m \in \mathbb{R}$ such that $d(f(x_1), f(x_2)) \leq m$ for all $x_2 \in X$.

3. There exists $m \in \mathbb{R}$ such that $d(f(x_1), f(x_2)) \leq m$ for all $x_1, x_2 \in X$.

Definition 1.6.4. If $f : X \to (Y, d)$ satisfies the statements in Exercise 1, then f is called **bounded.**

Definition 1.6.5 ($C_b(X, Y)$**; defined Section 4.1 of** Pugh **[2015], p. 224 of pdf**)**.** Let X be a set and (Y, d) be a metric space. Define $C_b(X, Y)$ as the set of bounded functions from X to Y.

We would like to define a metric on $C_b(X, Y)$ related to uniform convergence. Key point: metric $d : X \times X \to \mathbb{R}$, **not** $\mathbb{R} \cup \infty$. That is, any two points in the metric space must be finite distances away from each other (which is why we need boundedness). We will set this up with this lemma.

Lemma 48. If $f, g \in C_b(X, Y)$, then

$$\underbrace{\sup_{x \in X} d(f(x), g(x))}_{\text{will use for distance}} < \infty.$$

Proof. Let $x_0 \in X$. Choose M, N such that $d(f(x), f(x_0)) \leq M$ for all $x \in X$ and $d(g(x), g(x_0)) \leq N$ for all x (these exist since $C_b(X, Y)$ consists of bounded functions). Then for $x \in X$, we have

$$d(f(x), g(x)) \leq d(f(x), f(x_0)) + d(f(x_0), g(x_0)) + d(g(x_0), g(x)) \leq \underbrace{M + d(f(x_0), g(x_0)) + N}_{\text{finite quantity independent of } x}$$

Since this upper bound is finite, the result follows.

$\square$

Definition 1.6.6 (**Sup metric; defined Section 4.1 of** Pugh **[2015], p. 224 of pdf**)**.** For $f, g \in C_b(X, Y)$, let $d_\infty(f, g) := \sup_{x \in X} d(f(x), g(x))$.

Exercise 2. The metric space axioms hold for $d_\infty(f, g)$.

Theorem 49 (**Theorem 4.2 in** Pugh **[2015]**)**.** X set, (Y, d) metric space. If $f_n \in C_b(X, Y)$ for $n \geq 1$ and $f \in C_b(X, Y)$, then f_n converges uniformly to f if and only if $f_n \xrightarrow{d_\infty} f$).

Proof. f_n converges uniformly to f if and only if for every $\epsilon > 0$ there exists $N \geq 1$ such that for all $n \geq N$, $d(f_n(x), f(x)) < \epsilon/2$ for all $x \in X$. This is true if and only if for every $\epsilon > 0$ there exists $N \geq 1$ such that for all $n \geq N$,

$$\sup_{x \in X} d(f_n(x), f(x)) = d_\infty(f_n, f) \leq \epsilon/2 < \epsilon,$$

so $f_n \xrightarrow{d_\infty} f$.

$\square$

Does this adequately characterize uniform convergence of bounded functions? yes.

Proposition 50 (Uniform convergence preserves boundedness). If $f_n : X \to Y$ is bounded for all n and f_n converges uniformly to f, then f is bounded.

Proof. Let $x_0 \in X$. By uniform convergence, can choose $N \geq 1$ such that for all $n \geq N$ and all $x \in X$, we have $d(f_n(x), f(x)) < 1$. We know f_N is bounded; this implies there exists M such that $d(f_N(x), f_N(x_0)) \leq M$ for all $x \in X$. Note that

$$d(f(x), f(x_0)) \leq d(f(x), f_N(x)) + d(f_N(x), f_N(x_0)) + d(f_N(x_0), f(x_0)) \leq 1 + M + 1 = M + 2$$

which does not depend on x. This implies f is bounded.

$\square$

Corollary 50.1 (**similar to Theorem 4.2 in** Pugh [2015]). If $f_n \in C_b(X, Y)$ for $n \geq 1$, then (f_n) converges uniformly if and only if (f_n) converges in $C_b(X, Y)$. (functional analysis perspective of uniform convergence.)

Now study metric space properties of $C_b(X, Y)$. For now: if Y is complete then $C_b(X, Y)$ is complete: every Cauchy sequence in $C_b(X, Y)$ converges. But this result isn't unique to bounded functions.

Definition 1.6.7. X set, (Y, d) metric space. A sequence of functions $f_n : X \to Y$ is **uniformly Cauchy** if for every $\epsilon > 0$ there exists $N \geq 1$ such that for $n, m \geq N$ we have $d(f_n(x), f_m(x)) < \epsilon$ for all $x \in X$.

If all f_n are bounded, then a sequence is uniformly Cauchy if and only if the sequence is Cauchy in $C_b(X, Y)$.

Theorem 51. X set, (Y, d) metric space. If (Y, d) is complete, then any uniformly Cauchy sequence of functions $f_n : X \to Y$ is uniformly convergent.

Proof. Let $f_n : X \to Y$ for $n \geq 1$ be a uniformly Cauchy sequence. First we will show f_n converges pointwise. Then we will show the convergence is uniform.

Given $x \in X$, the sequence $(f_n(x))_{n=1}^\infty$ is a Cauchy sequence in Y since

$$d(f_n(x_0), f_m(x_0)) \leq \sup_{x \in X} \{d(f_n(x), f_m(x))\} = d_\infty(f_n, f_m).$$

Since Y is complete, $(f_n(x))_{n=1}^\infty$ converges to some point of Y. Call this point $f(x)$. (Limits are unique implies f is well-defined.) So by construction, $f_n \to f$ pointwise. Next we will show that it converges uniformly.

Let $\epsilon > 0$. By the uniform Cauchy condition, there exists $N \in \mathbb{N}$ such that if $n, m \geq N$, then $d_\infty(f_n, f_m) < \epsilon/2$.

We will now show that for all $n \geq N$ and all $x \in X$, we have $d(f_n(x), f(x)) < \epsilon$. Given $x \in X$, the sequence $(f_n(x))_{n=1}^{\infty}$ converges to $f(x)$ in Y. Thus, we can choose $M^*(x)$ such that for $m \geq M^*(x)$, we have $d(f_m(x), f(x)) < \epsilon/2$. Define $M(x) := \max\{M^*(x), N\}$. Now, for $n \geq N$ we have

$$d(f_n(x), f(x)) \leq d(f_n, f_{M(x)}) + d(f_{M(x)}(x), f(x)) < \underbrace{\epsilon/2}_{\text{uniform Cauchy condition}} + \epsilon/2 = \epsilon.$$

$\square$

Corollary 51.1 (Theorem 4.3 in Pugh [2015]). If (Y, d) is complete then $(C_b(X, Y), d_\infty)$ is complete.

Proof. Cauchy sequence (f_n) in $C_b(X, Y)$ is uniformly Cauchy by Theorem 51. So it is uniformly convergent. Since uniform convergence preserves boundedness, limit f must be in $C_b(X, Y)$, so f_n converges to a limit in the metric space C_b.

$\square$

Note: if (f_n) is uniformly convergent then it's uniformly Cauchy.

Definition 1.6.8. Assume both (X, d) and (Y, d') are metric spaces. $(C^0(X, Y) \subset C_b(X, Y)$ is the set of bounded continuous functions $X \to Y$. We can restrict d_∞ to $C_0(X, Y)$ and get a metric subspace.

Corollary 51.2 (Corollary to Theorem 40; Corollary 4.4 in Pugh [2015]). $C^0(X, Y)$ is a closed subset of $C_b(X, Y)$.

Proof. If we have a sequence in $C^0(X, Y)$ that converges in d_∞ to some $f \in C_b(X, Y)$, we want to show that $f \in C^0$. But f_n converges uniformly to f, so f is bounded and continuous by Theorem 40, so $f \in C^0(X, y)$.

$\square$

Corollary 51.3 (Corollary to Theorem 51). If (Y, d') is complete then $C^0(X, Y)$ is complete.

Proof. A closed subset of a complete metric space is complete as a metric subspace. (Closure comes from Corollary 51.2.)

$\square$

Theorem 52 (Dini's Theorem (Math 425B homework 1)). Let X be a compact metric space and let $f_n : X \to \mathbb{R}$ be a sequence of continuous functions. Assume that the functions f_n converge pointwise to $f : X \to \mathbb{R}$, that f is continuous, and that for all $x \in X$, the sequence $(f_n(x))_{n=1}^{\infty}$ is a decreasing sequence of real numbers. Then f_n converges uniformly to f.

Proof. Consider an arbitrary $\epsilon > 0$. For each $n \in \mathbb{N}$, consider the set $U_n := \{x : f_n(x) - f(x) < \epsilon\} \subseteq X$. First we will show that U_n is open. If U_n is empty, then it is (trivially) open. Suppose U_n is not empty and consider a point $u \in U_n$. $f_n(u) - f(u) = \epsilon_u \in [0, \epsilon)$; let $\epsilon' := \epsilon - \epsilon_u$, so $\epsilon' \in (0, \epsilon]$. Since f and f_n are continuous, $f_n - f$ is continuous, so there exists a $\delta > 0$ such that $t, u \in X$ and $d(t, u) < \delta \implies |f_n(t) - f(t) - (f_n(u) - f(u))| < \epsilon'$. But

$$|f_n(t) - f(t) - (f_n(u) - f(u))| < \epsilon' \iff f_n(u) - f(u) - \epsilon' < f_n(t) - f(t) < f_n(u) - f(u) + \epsilon'$$
$$\implies f_n(t) - f(t) < \epsilon_u + \epsilon' = \epsilon;$$

that is, all such t are in U_n. The open set $\{t : t \in X, d(t, u) < \delta\}$ is therefore a subset of U_n. U_n either consists of one set like this or the union of more than one such set; either way, U_n is open.

Next we will show that the U_n cover X; that is, $X \subseteq \bigcup_{n=1}^{\infty} U_n$. Since f_n converges pointwise to f, we know that for any $x \in X$ there exists a finite $N_x \in \mathbb{N}$ such that for $n \geq N_x$ we have $|f_n(x) - f(x)| = f_n(x) - f(x) < \epsilon$. Let $N^* := \max_x\{N_x\}$ (this quantity is well-defined since X is compact). Then for all $n \geq N^*$ we have that for all $x \in X$, $f_n(x) - f(x) < \epsilon$, so $X \subseteq \bigcup_{n=N^*}^{\infty} U_n \subseteq \bigcup_{n=1}^{\infty} U_n$.

Next, consider $n < n'$ for $n, n' \in \mathbb{N}$. We will show that since $U_n = \{x : f_n(x) - f(x) < \epsilon\}$ and $U_{n'} = \{x : f_{n'}(x) - f(x) < \epsilon\}$,

$$U_n \subseteq U_{n'} \qquad \forall n, n' \in \mathbb{N}, n < n' \tag{1.4}$$

by the following argument. Consider any $u \in U_n$. u satisfies $f_n(u) - f(u) < \epsilon$. Since $f_{n'}(u) \leq f_n(u)$, u also satisfies $f_{n'}(u) - f(u) < \epsilon$, so $u \in U_{n'}$[1].

X is compact, so there exists a reduction of $\bigcup_{n=1}^{\infty} U_n$ to a finite subcover of X; that is, there exists a finite set $\mathcal{S} \subset \mathbb{N}$ such that $\bigcup_{\mathcal{S}} U_n = X$. Let $s^* := \max\{\mathcal{S}\}$ (of course $s^* < \infty$ since $\mathcal{S}$ is finite). By (1.4), we have that for any $s \in \mathcal{S}$, $U_s \subseteq U_{s^*}$. Therefore $X = \bigcup_{\mathcal{S}} U_n = U_{s^*}$, so for all $n \geq s^*$ and all $x \in X$, $|(f_n(x) - f(x)| = f_n(x) - f(x) < \epsilon$.

$\square$

Proposition 53 (Uniform continuity is preserved under uniform convergence (Math 425b Homework 2)). Let (X, d) and (Y, d') be metric spaces. Let $f_n : X \to Y$ be uniformly continuous for all $n \geq 1$. Suppose f_n converges uniformly to f for some $f : X \to Y$. Then f is uniformly continuous. (This implies that if all f_n are uniformly continuous and f_n converges uniformly to f then f is uniformly continuous.)

Proof. Let $\epsilon > 0$. Choose $N \geq 1$ such that we have

$$d'(f_n(x), f(x)) < \epsilon/3 \qquad \text{for } n \geq N \text{ and for all } x \in X. \tag{1.5}$$

The function f_N is uniformly continuous, so there exists $\delta > 0$ such that

$$d'(f_N(x), f_N(x_0)) < \epsilon/3 \qquad \forall x, x_0 \in X \text{ with } d(x, x_0) < \delta. \tag{1.6}$$

---

[1]Note that U_n is not necessarily a proper subset of $U_{n'}$. For example, consider $f(x) := 0$ and the sequence of functions defined by $f_n(x) := 1/n$; note that these functions satisfy the assumptions of this theorem. Let $\tilde{n} := \min\{n \in \mathbb{N} : 1/n < \epsilon\}$. Then

$$U_n = \begin{cases} X, & n \geq \tilde{n}, \\ \emptyset, & n < \tilde{n}, \end{cases}$$

so for any $n, n' \in \{1, \ldots, \tilde{n} - 1\}$ or $n, n' \in \{\tilde{n}, \tilde{n} + 1, \ldots\}$, we have that $U_n = U'_n$.

This δ works for uniform continuity of f because for any $x, x_0 \in X$ with $d(x, x_0) < \delta$,

$$d'(f(x), f(x_0)) \leq d'(f(x), f_N(x)) + d'(f_N(x), f_N(x_0)) + d'(f_N(x_0), f(x_0)) < \underbrace{\frac{\epsilon}{3}}_{\text{by (1.5)}} + \underbrace{\frac{\epsilon}{3}}_{\text{by (1.6)}} + \underbrace{\frac{\epsilon}{3}}_{\text{by (1.5)}} = \epsilon.$$

$\square$

Definition 1.6.9 ($C^k(X, Y)$)**.** In general (say for $\mathbb{R} \to \mathbb{R}$): $C^k(\mathbb{R}, \mathbb{R})$ is the set of k times continuously differentiable functions. (Generalization of notation that C^0 means continuous.) Here, we assume bounded continuity to have a d_∞ norm. For $k = 0$, metric spaces are ok. For $k > 1$, have derivatives, so harder to generalize to metric spaces. Sometimes $C^\infty(\mathbb{R}, \mathbb{R})$ is called the set of **smooth** functions (infinitely differentiable).

If X is a set and V is a vector space (say $\mathbb{R}$ or $\mathbb{C}$), then the set of functions from X to V forms a vector space itself.

$$(f + g)(x) := f(x) + g(x)$$
$$(cf)(x) := c(f(x))$$

So if $f_n : X \to V$, makes sense to consider $\sum_{n=1}^{N} f_n$, the partial sum. But we need V to also be a metric space in order to talk about convergence, so we will use a normed vector space $(V, \|\cdot\|)$.

Definition 1.6.10 (**Normed vector space**)**.** $(V, \|\cdot\|)$ is a **normed vector space** if $\|\cdot\| : V \to \mathbb{R}$ satisfies

1. $\|v\| \geq 0$ for all v, $\|v\| = 0 \iff v = 0$.

2. $\|cv\| = |c|\|v\|$ for all $c \in \mathbb{R}$ or $\mathbb{C}$, $v \in V$.

3. $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$.

More generally: $f(V, \langle \cdot, \cdot \rangle)$ is an inner product space, can define a norm on V by $\|v\| = \sqrt{\langle v, v \rangle}$. But not every norm comes from an inner product. (can deduce norm axioms from inner product axioms.)

Exercise 3. If $(V, \|\cdot\|)$ is a normed vector space, then $d(v, w) := \|v - w\|$ is a metric on V.

So, for a vector (inner product) space it is easy to get a norm. So we can talk about $C_b(X, V)$, etc.

Have a vector space of all functions $X \to V$ with no restrictions.

Exercise 4. $C_b(X, V)$ is a vector subspace of all functions $X \to V$ since if f, g are bounded then $f + g$ is bounded and if f is bounded, then for $c \in \mathbb{R}$ or $c \in \mathbb{C}$, cf is bounded, and also a zero function $X \to V$ is bounded.

So $C_b(X, V)$ is itself a vector space. We have the metric d_∞ on the vector space $C_b(X, V)$. Question: is there a norm on $C_b(X, V)$ inducing the metric d_∞? Yes, the **sup-norm** or ∞-norm.

Definition 1.6.11. Let X be a set and $(V, \|\cdot\|)$ be a normed vector space. If $f \in C_b(X, V)$ then the **sup-norm** or ∞**-norm** is defined as

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

Note that $\|f\|_\infty = d_\infty(f, 0)$. Examples: see Figure 1.2.

Exercise 5. $\|f\|_\infty$ is a norm on $C_b(X, Y)$ and $d_\infty(f, g) = \|f - g\|_\infty$. (That is, the metric induced by $\|\cdot\|_\infty$ is d_∞.)
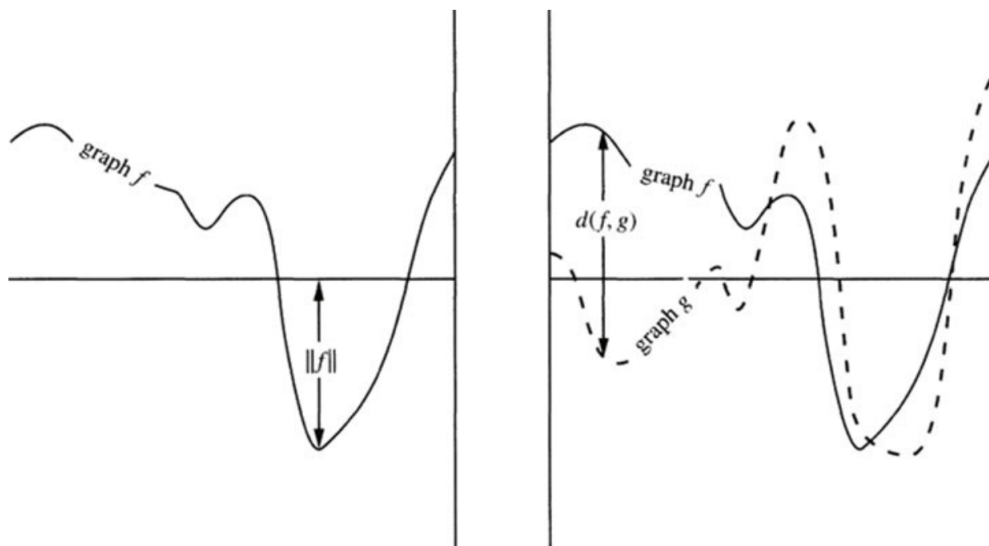
Figure 90 The sup norm of f and the sup distance between f and g

Figure 1.2: Illustration of sup norm, Figure 90 in section 4.1 of Pugh [2015].

We have special terminology for complete normed vector spaces (and inner product spaces).

Definition 1.6.12 (Banach space). A **Banach space** is a normed vector space $(V, \|\cdot\|)$ that is complete in the induced norm. (key concept in functional analysis.)

Definition 1.6.13 (Hilbert space). A **Hilbert space** is an inner product space that is complete in the induced metric (a Banach space for an inner product space).

Roughly: in physics, the "space of states" for quantum systems.

Now, if X is a set and $(V, \|\cdot\|)$ is a normed vector space, we can talk about series of functions $\sum_{n=1}^\infty f_n$ for $f_n : X \to V$.

Definition 1.6.14. Let X be a set and $(V, \|\cdot\|)$ be a normed vector space. Let $f_n : X \to V$ be functions for $n \in \mathbb{N}$.

1. We say that the series $\sum_{n=1}^\infty f_n$ **converges pointwise** to some $f : X \to V$ if the sequence of partial sums

$$\left(\sum_{n=1}^{N} f_n\right)_{N=1}^{\infty}$$

converges pointwise to f.

2. We say that the series $\sum_{n=1}^{\infty} f_n$ **converges uniformly** to some $f : X \to V$ if the sequence of partial sums

$$\left(\sum_{n=1}^{N} f_n\right)_{N=1}^{\infty}$$

converges uniformly to f.

Different types of convergence for a series $A = \sum_{n=1}^{\infty} v_n$ $(v_n \in V, (V, \|\cdot\|)$ a Banach space).

Consider also a series of functions $B = \sum_{n=1}^{\infty} f_n$ for $f_n : X \to V$ with $(V, \|\cdot\|)$ a Banach space. (Absolute convergence is easiest in a Banach space, since we can use completeness.)

To show $\sum_{n=1}^{\infty} v_n$ converges in V, it's enough to show the sequence of partial sums is Cauchy since V is complete. Similarly, to show that $\sum_{n=1}^{\infty} f_n$ converges uniformly to $f : X \to V$, it's enough to show the sequence of partial sums is uniformly Cauchy (again, because of completeness, which applies since we are in a Banach space).

Exercise 6. (A) Consider the series $\sum_{n=1}^{\infty} v_n$ with $(V, \|\cdot\|)$ a normed vector space and $v_n \in V$. Then

$$\left(\sum_{n=1}^{N} v_n\right)_{N=1}^{\infty}$$

is Cauchy if and only if for every $\epsilon > 0$ there exists N such that for $N \leq m, n$ we have

$$\left\|\sum_{k=m}^{n} v_k\right\| < \epsilon$$

Note:

$$\left\|\sum_{k=1}^{n} v_k - \sum_{k=1}^{m} v_k\right\| = \left\|\sum_{k=m+1}^{n} v_k\right\|$$

if $n > m$. Now fix bookkeeping.

(If seems trivial, don't worry about it, but can work out details if you want.)

(B) X set, $(V, \|\cdot\|)$ a normed vector space, $f_n : X \to V$. Then

$$\left(\sum_{n=1}^{N} f_n\right)_{N=1}^{\infty}$$

is uniformly Cauchy if and only if for every $\epsilon > 0$ there exists N such that for all $N \leq m < n$ and for all $x \in X$ we have

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

In the setting of $\sum_{n=1}^{\infty} f_n$, have uniform vs. pointwise distinction; in both settings, have absolute vs. conditional distinction.

Definition 1.6.15. Let $(V, \|\cdot\|)$ be a Banach space. A series

$$\sum_{n=1}^{\infty} v_n$$

of elements $v_n \in V$ **converges absolutely** if the series of real numbers

$$\sum_{n=1}^{\infty} \|v_n\|$$

converges in $\mathbb{R}$.

Proposition 54. Let $(V, \|\cdot\|)$ be a Banach space. If

$$\sum_{n=1}^{\infty} v_n$$

converges absolutely then it converges in V.

Proof. It suffices to show that

$$\left(\sum_{n=1}^{N} v_n \right)_{N=1}^{\infty}$$

is Cauchy. We know that

$$\left(\sum_{n=1}^{N} \|v_n\| \right)_{N=1}^{\infty}$$

converges in $\mathbb{R}$. Given $\epsilon > 0$, there exists N such that for $N \leq m < n$ we have

$$\left| \sum_{k=m}^{n} \|v_k\| \right| < \epsilon.$$

Then if $N \leq m < n$ we have

$$\left\| \sum_{k=m}^{n} v_k \right\| \leq \sum_{k=m}^{n} \|v_k\| = \left| \sum_{k=m}^{n} \|v_k\| \right| < \epsilon.$$

So

$$\left(\sum_{n=1}^{N} v_n \right)_{N=1}^{\infty}$$

is Cauchy and thus convergent.

$\square$

Definition 1.6.16. Let X be a set and let $(V, \|\cdot\|)$ be a Banach space. Let $f_n : X \to V$. Then

$$\sum_{n=1}^{\infty} f_n$$

converges absolutely if for all $x \in X$,

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

converges absolutely. (In other words, converging absolutely means converging absolutely pointwise.)

Definition 1.6.17 (less standard naming). Let X be a set and let $(V, \|\cdot\|)$ be a Banach space. Let $f_n : X \to V$. Then

$$\sum_{n=1}^{\infty} f_n$$

converges absolutely uniformly if

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

converges uniformly as a series of functions from X to $\mathbb{R}$. (note: $\|f_n(x)\|$ is a function $X \to \mathbb{R}$ whereas $\|f_n\|_{\infty} = \|f_n\|_{\sup}$ is a single real number, and it only exists if f is bounded.)

Proposition 55 (Absolute uniform convergence implies absolute convergence and uniform convergence). Let X be a set and let $(V, \|\cdot\|)$ be a Banach space. Let $f_n : X \to V$ be functions. If

$$\sum_{n=1}^{\infty} f_n$$

converges absolutely uniformly, then it converges absolutely (pointwise) and converges uniformly.

Proof. It is clear that absolute uniform convergence implies absolute pointwise convergence. So we want to show that

$$\left(\sum_{n=1}^{N} f_n\right)_{N=1}^{\infty}$$

converges uniformly. It is enough to show that

$$\left(\sum_{n=1}^{N} f_n\right)_{N=1}^{\infty}$$

is uniformly Cauchy. We know that

$$\left(\sum_{n=1}^{N} \|f_n\|\right)_{N=1}^{\infty}$$

is uniformly convergent, so it is uniformly Cauchy. Thus, given $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for $N \le m < n$ and $x \in X$, we have

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

Now if $N \le m < n$, then

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

Thus

$$\left(\sum_{n=1}^{N} f_n\right)_{N=1}^{\infty}$$

is uniformly Cauchy, and therefore uniformly convergent by the completeness of V.

$\square$

Now look at bounded functions $f_n : X \to V$, where $\|f_n\|_\infty$ makes sense in $\mathbb{R}$ (i.e. is finite). Now we can look at

$$\sum_{n=1}^{\infty} \|f_n\|_\infty$$

is a real number (series of real numbers).

Theorem 56 (**Weierstrass M-test; Theorem 4.5 in Pugh [2015], Theorem 7.10 in Rudin [1976], p. 148 of text/p. 157 of pdf**). Let X be a set and let $(V, \|\cdot\|)$ be a normed vector space. Let $f_n \in C_b(X, Y)$ be functions. If

$$\sum_{n=1}^{\infty} \|f_n\|_{\infty}$$

converges as a series in $\mathbb{R}$, then

$$\sum_{n=1}^{\infty} f_n$$

converges uniformly as a series of functions from $X \to \mathbb{R}$. (Further, if V is a Banach space, it converges absolutely uniformly, so absolutely and uniformly.)

Proof. Since $\mathbb{R}$ is complete, it suffices to show that

$$\left(\sum_{n=1}^{N} \|f_n\| \right)_{N=1}^{\infty}$$

is uniformly Cauchy. We know that

$$\left(\sum_{n=1}^{N} \|f_n\|_{\infty} \right)_{N=1}^{\infty}$$

is a Cauchy sequence in $\mathbb{R}$. That means that given $\epsilon > 0$, there exists N such that if $N \leq m < n$ then

$$\left| \sum_{k=m}^{n} \|f_k\|_{\infty} \right| < \epsilon.$$

If $N \leq m < n$ and $x \in X$, we have

$$\sum_{k=m}^{n} \|f_k(x)\| \leq \sum_{k=m}^{n} \sup_{x' \in X} \|f_k(x')\| = \sum_{k=m}^{n} \|f_k\|_{\infty} < \epsilon.$$

$\square$

Theorem 57 (**Weierstrass M-test, classical version, also appears in Pugh [2015]**). Let X be a set and let $(V, \|\cdot\|)$ be a Banach space. Let $f_n : X \to V$ be bounded and suppose that for all (sufficiently large) n, we have constants $M_n \in \mathbb{R}$ such that

$$\|f_n\|_{\infty} \leq M_n, \quad \text{and} \quad \sum_{n=1}^{\infty} M_n \text{ converges.}$$

Then $\sum_{n=1}^{\infty} f_n$ converges absolutely and uniformly.

Proof. Use previous version plus comparison test for series.

$\square$

The Weierstrass M-test can be used to deduce uniform convergence properties for series (particularly power series). Before applying this to power series, we will study uniform convergence and integrals/derivatives. (We are still building towards showing that power series can be differentiated term-by-term inside the disc of convergence; thus they're smooth functions. Thus, power series are C^{∞} in their disc of convergence.)

Integrals: let $[a, b] \in \mathbb{R}$. Let $\mathcal{R}[a, b]$ be the set of Riemann integrable functions from $[a, b]$ to $\mathbb{R}$ (or $\mathbb{C}$). Recall that (properly) Riemann integrable functions are bounded (Theorem 3.16 in Pugh [2015]). So, $\mathcal{R}[a, b] \subseteq C_b([a, b], \mathbb{R})$.

Proposition 58 (**Lebesgue's integrability condition**). $f \in C_b([a, b], \mathbb{R})$ is Riemann integrable if and only if $\operatorname{disc}(f)$ (the set of discontinuities of f) is a **null set** (i.e., a set of measure 0.)

Theorem 59 (**Theorem 4.6 in Pugh [2015]**). $\mathcal{R}[a, b]$ is a closed subset of $C_b([a, b], \mathbb{R})$, i.e.: if $f_n \in \mathcal{R}[a, b]$ and f_n converges uniformly to some $f \in C_b([a, b], \mathbb{R})$ (i.e., $f_n \xrightarrow{d_\infty} f$), then $f \in \mathcal{R}[a, b]$. Another way to say this: the uniform limit of Riemann integrable functions is Riemann integrable. Furthermore,

$$\int_a^b f(x) \ dx = \lim_{n \to \infty} \int_a^b f_n(x) \ dx; \tag{1.7}$$

i.e.,

$$\int_a^b \lim_{n \to \infty} f_n(x) \ dx = \lim_{n \to \infty} \int_a^b f_n(x) \ dx.$$

Proof of Theorem 59. Use Lebesgue's integrability condition: since we know that f_n is Riemann integrable, Lebesgue's integrability condition tells us that $\operatorname{disc}(f)$ is a null set. The countable union of null sets is a null set, so $\bigcup_{n=1}^{\infty} \operatorname{disc}(f_n)$ is a null set.

Therefore $\operatorname{disc}(f) \subseteq \bigcup_{n=1}^{\infty} \operatorname{disc}(f_n)$ is a sufficient condition to show that $f \in \mathcal{R}[a, b]$. We claim this is true. That is, we want to show that if $x \in \operatorname{disc}(f)$, then $x \in \operatorname{disc}(f)_n$ for some n; we will prove the contrapositive. Suppose $x \notin \operatorname{disc}(f)_n$ for any n for some $x \in [a, b]$; that is, f_n is continuous at x for all $n \in \mathbb{N}$. Since we assumed f_n converges uniformly to f, we have that f is continuous at x from Theorem 40. Therefore $x \notin \operatorname{disc}(f)$. Thus, this proves the claim, so $f \in \mathcal{R}[a, b]$.

Next we will show that (1.7) holds. By uniform convergence of $(f_n)_{n=1}^{\infty}$ to f, given $\epsilon > 0$ there exists N such that if $n \geq N$ then $d_\infty(f_n, f) < \epsilon/(b - a)$. Now for $n \geq N$,

$$\left| \int_a^b f(t) \ dt - \int_a^b f_n(x) \ dx \right| = \left| \int_a^b [f(x) - f_n(x)] \ dx \right| \leq \int_a^b |f(x) - f_n(x)| \ dx \leq \int_a^b d_\infty(f_n, f) \ dx$$

$$< \frac{\epsilon}{b - a} \cdot (b - a) = \epsilon.$$

$\square$

Alternative Proof of Theorem 59 (425b Homework 2). Let A be the set of tagged partitions (P, T) of $[a, b]$ with the refinement ordering $\preceq$ from Definition 1.6.3. Let $\epsilon > 0$. First we will show the nets $R(f_n, P, T)$ converge uniformly to $R(f, P, T)$. The nets converge uniformly if there exists $N \in \mathbb{N}$ such that for all $n \geq N$, for all $(P, T) \in A$ we have $|R(f_n, P, T) - R(f, P, T)| < \epsilon$.

Because the f_n converge uniformly to f, we know that there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have that

$$|f_n(x) - f(x)| < \frac{\epsilon}{b - a} \qquad \forall x \in [a, b]. \tag{1.8}$$

Therefore we have that for all (P, T) in A and for all $n \geq N$,

$$|R(f_n, P, T) - R(f, P, T)| = \left| \sum_{i=1}^{n} f_n(t_i) \Delta x_i - \sum_{i=1}^{n} f(t_i) \Delta x_i \right| = \left| \sum_{i=1}^{n} [f_n(t_i) - f(t_i)] \Delta x_i \right|$$

$$\leq \sum_{i=1}^{n} |f_n(t_i) - f(t_i)| \Delta x_i \leq \sum_{i=1}^{n} \frac{\epsilon}{b - a} \Delta x_i = \frac{\epsilon}{b - a} \sum_{i=1}^{n} \Delta x_i = \frac{\epsilon}{b - a} (b - a) = \epsilon.$$

By Proposition 43, since each function $f_n : [a, b] \to \mathbb{R}$ is Riemann integrable, each net $(P_n, T_n) \mapsto R(f_n, P_n, T_n)$ converges to a limit $I_n \in \mathbb{R}$. From Lemma 44, we know that because this is true and because we have shown that the nets converge uniformly to $R(f, P, T)$, the limits $(I_n)_{n=1}^{\infty}$ form a Cauchy sequence in $\mathbb{R}$. Therefore by the completeness of $\mathbb{R}$, the limits $(I_n)_{n=1}^{\infty}$ converge to some limit $I \in \mathbb{R}$, which means that by Lemma 45 the limit of $(P, T) \mapsto R(f, P, T)$ exists and equals I.

By Proposition 43, Riemann integrability of f is equivalent of convergence of the net $(P, T) \mapsto R(f, P, T)$, a net from A to $\mathbb{R}$. Therefore, because the net $(P, T) \mapsto R(f, P, T)$ converges to I, f is Riemann integrable with integral I.

$\square$

Corollary 59.1 (Theorem 4.8 in Pugh [2015]). If $f_n \in \mathcal{R}[a, b]$ and the series $\sum_{n=1}^{\infty}$ converges uniformly to some f, then $f \in \mathcal{R}[a, b]$ and $\int_a^b f(x) \, dx = \sum_{n=1}^{\infty} \int_a^b f_n(x) \, dx$. I.e.,

$$\int_a^b \sum_{n=1}^{\infty} f_n(x) dx = \sum_{n=1}^{\infty} \int_a^b f_n(x) \, dx.$$

Proof. Apply Theorem 59 to

$$\left(\sum_{n=1}^{N} f_n \right)_{N=1}^{\infty}.$$

$\square$

Corollary 59.2 (Corollary 4.7 in Pugh [2015]). If $f_n \in \mathcal{R}[a,b]$, f_n converges uniformly to f, then the function $\int_a^x f_n(t)\ dt$ converges uniformly to $\int_a^x f(t)\ dt$. (i.e., converges uniformly in $x \in [a,b]$.)

Proof. Similar to proof of Theorem 59. Given $\epsilon > 0$, choose N such that $d_\infty(f_n, f) < \epsilon/(b-a)$ for all $n \geq N$. If $n \geq N$, then

$$\left| \int_a^x f(t)\ dt - \int_a^x f_n(t)\ dt \right| = \left| \int_a^x [f(t) - f_n(t)]\ dt \right| \leq \int_a^x |f(t) - f_n(t)|\ dt \leq \int_a^x d_\infty(f_n, f)\ dt$$

$$< \frac{\epsilon}{b-a} \cdot (b-a) = \epsilon$$

(since $\int_a^x 1\ dx \leq \int_a^b 1\ dx = b - a$).

$\square$

Question: must the uniform limit of differentiable functions be differentiable? No.

Example 1.6.2. Let $f_n : [-1, 1] \to \mathbb{R}$ be defined by $f_n(x) := \sqrt{x^2 + 1/n}$. Let $f(x) := \sqrt{x^2} = |x|$. $f(\cdot)$ is not differentiable at 0, but each $f_n(x)$ is. (We will show below that f_n converges uniformly to f.) So, differentiability (at one point) is lost in the limit.

Proposition 60. f_n converges uniformly to f.

Proof. We will apply Dini's Theorem (Theorem 52). Note that $[-1, 1]$ is a compact metric space. $f_n(x) = \sqrt{x^2 + 1/n}$ is a sequence of continuous functions, and for all $x \in [-1, 1]$, the sequence $(\sqrt{x^2 + 1/n})_{n=1}^\infty$ is a decreasing sequence of real numbers since $1/n$ is decreasing in n and $\sqrt{\cdot}$ is a monotonic transformation in $\mathbb{R}_+$. We also have that $f(x) = |x|$ is continuous.

It only remains to be shown that the functions f_n converge pointwise to f. Let $\epsilon > 0$ and let $\tilde{x} \in [-1, 1]$. We want to find $N \in \mathbb{N}$ such that for all $n \geq N$ it holds that $\left| \sqrt{\tilde{x}^2 + 1/n} - \sqrt{\tilde{x}^2} \right| = \sqrt{\tilde{x}^2 + 1/n} - \sqrt{\tilde{x}^2} < \epsilon$. For all $n \geq 1$ we have

$$\sqrt{\tilde{x}^2 + 1/n} - \sqrt{\tilde{x}^2} < \epsilon \iff \tilde{x}^2 + \frac{1}{n} < \left(\epsilon + \sqrt{\tilde{x}^2} \right)^2 \iff \frac{1}{n} < \epsilon^2 + 2\epsilon\sqrt{\tilde{x}^2} + \tilde{x}^2 - \tilde{x}^2$$

$$\iff n > \frac{1}{\epsilon^2 + 2\epsilon\sqrt{\tilde{x}^2}}.$$

Therefore set $N := \left\lceil \left(\epsilon^2 + 2\epsilon\sqrt{\tilde{x}^2} \right)^{-1} + 1 \right\rceil$ to complete the proof.

$\square$

In fact, the uniform limit of smooth functions can be nowhere differentiable (even though it is continuous, since it is the uniform limit of continuous functions.)

Theorem 61 (Theorem 4.9 from Pugh [2015]). The uniform limit of a sequence of differentiable functions is differentiable provided that the sequence of derivatives also converges uniformly.

Theorem 62 (**Theorem 4.10 from** Pugh [2015]). A uniformly convergent series of differentiable functions can be differentiated term-by-term, provided that the derivative series converges uniformly. That is,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{k=0}^{\infty} f_k(x)\right) = \sum_{k=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}x}\left(f_k(x)\right)$$

Proof. Apply Theorem 61 to the sequence of partial sums.

$\square$

Theorem 63 (**Math 425b Homework 3**). Let $[a,b] \subset \mathbb{R}$ and let $f_n : [a,b] \to \mathbb{R}$ be continuously differentiable (C^1) functions. Suppose that the derivatives f_n' converge uniformly to a function $g : [a,b] \to \mathbb{R}$ and that for some $x_0 \in [a,b]$ the function values $f_n(x_0)$ converge to some limit $L \in \mathbb{R}$ (i.e., f_n converges "pointwise somewhere," a very weak condition). Then the functions f_n converge uniformly to a differentiable function $f : [a,b] \to \mathbb{R}$ with $f' = g$.

Proof. By the Fundamental Theorem of Calculus (Theorem 3.34 in Pugh [2015]), we have that for all $x \in [a,b]$, for a fixed $x_0 \in [a,b]$

$$\int_{x_0}^{x} f_n'(t)\ dt = f_n(x) - f_n(x_0) = f_n(x) - C_n \qquad \Longleftrightarrow \qquad f_n(x) = C_n + \int_{x_0}^{x} f_n'(t)\ dt$$

where $C_n := f_n(x_0)$.

Let $\epsilon > 0$. We have that $f_n : [a,b] \to \mathbb{R}$ are C^1 for all n and that the derivatives f_n' converge uniformly to g; that is, there exists $N' \in \mathbb{N}$ such that for all $n \geq N'$ we have $|f_n'(x) - g(x)| < \epsilon/(2|x - x_0|)$ for all $x \in [a,b]$. We also have that for some $x_0 \in [a,b]$, there exists $N'' \in \mathbb{N}$ such that for all $n \geq N''$, $|f_n(x_0) - L| = |C_n - L| < \epsilon/2$. Let $N := \max\{N', N''\}$. Then for all $n \geq N$ and for all $x \in [a,b]$ it holds that

$$\left| f_n(x) - \left(L + \int_{x_0}^{x} g(t)\ dt\right)\right| = \left|C_n + \int_{x_0}^{x} f_n'(t)\ dt - \left(L + \int_{x_0}^{x} g(t)\ dt\right)\right|$$

$$= \left|C_n - L + \int_{x_0}^{x} [f_n'(x) - g(x)]\ dx\right| \leq |C_n - L| + \left|\int_{x_0}^{x} [f_n'(x) - g(x)]\ dx\right|$$

$$\leq |C_n - L| + \int_{x_0}^{x} |[f_n'(x) - g(x)]|\ dx < \frac{\epsilon}{2} + |x_0 - x| \cdot \frac{\epsilon}{2|x_0 - x|} = \epsilon;$$

that is, $f_n(x)$ converges uniformly to $L + \int_{x_0}^{x} g(t)\ dt$. Let $f : [a,b] \to \mathbb{R}$ be defined by $f(x) := L + \int_{x_0}^{x} g(t)\ dt$. It remains to show that f is differentiable and $f' = g$. Again applying the Fundamental Theorem of Calculus, we have

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(L + \int_{x_0}^{x} g(t)\ dt\right) = \frac{\mathrm{d}}{\mathrm{d}x}(L) + \frac{\mathrm{d}}{\mathrm{d}x}\left(\int_{x_0}^{x} g(t)\ dt\right) = g(x).$$

$\square$

Definition 1.6.18. Let (X, d) and (Y, d') be metric spaces and let $f_n : X \to Y$ be a sequence of functions. Let $f : X \to Y$ be another function. We say that $(f_n)_{n=1}^\infty$ **converges compactly** to f if for any compact subset $K \subset X$, the restrictions of f_n to K converge uniformly to the restriction of f to K. We say $(f_n)_{n=1}^\infty$ **converges locally uniformly** to f if for any $x \in X$ there exists an open neighborhood U of x such that the restrictions of f_n to U converge uniformly to the restriction of f to U.

Definition 1.6.19. Let (X, d) be a metric space. We say that (X, d) is **locally compact** if for all $x \in X$ there exists an open set U and a compact set K such that $x \in U \subseteq K \subseteq X$.

Proposition 64 (Math 425b Homework 3). Let (X, d) be a locally compact metric space. Let $f_n, f : X \to Y$ be functions. The functions f_n converge compactly to f if and only if they converge locally uniformly to f.

Proof. Let $K \subset X$ be an arbitrary compact subset of X. We have that (X, d) is locally compact; that is, for all $x \in X$ there exist $x \in U \subset K$ with U open.

- **Compact convergence implies local uniform convergence:** Let $\epsilon' > 0$. Let K' be an arbitrary compact subset of X. We will assume that $(f_n)_{n=1}^\infty$ converges compactly to f; that is, the restrictions of f_n to K' converge uniformly to the restriction of f to K'. That is, there exists $N' \in \mathbb{N}$ such that for all $n \geq N'$, $d'(f_n(x), f(x)) < \epsilon'$ for all $x \in K'$. By local compactness, for all $x \in X$ there exists an open set U' with $x \subset U' \subset K'$ with U' open. Since $U' \subset K'$, it holds that for all $n \geq N'$, $d'(f_n(x), f(x)) < \epsilon'$ for all $x \in U'$. Therefore the restrictions of f_n to U' converge uniformly to the restriction of f to U'.

- **Local uniform convergence implies compact convergence:** Let $\epsilon'' > 0$. We will assume that $(f_n)_{n=1}^\infty$ converges locally uniformly to f; that is, for any $x'' \in X$ there exists an open neighborhood U'' of x'' such that the restrictions of f_n to U'' converge uniformly to the restriction of f to U''.

  We want to show that for any compact subset K'' of X there exists an $N'' \in \mathbb{N}$ such that for all $n \geq N''$, $d'(f_n(x), f(x)) < \epsilon''$ for all $x \in K''$.

  Consider an arbitrary compact subset K'' of X. For every $\tilde{x}'' \in K''$, by assumption there exists an open set $\tilde{U}''$ with $\tilde{x}'' \in \tilde{U}'' \subset \tilde{K}'' \subset X$ such that the restrictions of f_n to each $\tilde{U}''$ converge uniformly to the restriction of f to $\tilde{U}''$. That is, for each set there exists $\tilde{N}'' \in \mathbb{N}$ such that for all $n \geq N$ and for all $\tilde{x}'' \in \tilde{U}''$ it holds that $d'(f(\tilde{x}''), f_n(\tilde{x}'')) < \epsilon$. Note that the union of all such $\tilde{U}''$ (which might be an infinite number of sets) is an open cover of K''. Because K'' is compact, this open cover of K'' reduces to a finite subcover, so there exists a finite number of these sets that cover K''. Let the union of these sets (the finite subcover) be $\mathcal{U}$. Among each of these sets, let N'' be the maximum of all the $\tilde{N}''$ (this number is well-defined since there is a finite number of the $\tilde{N}''$). Then for all $x \in \mathcal{U}$ and for all $n \geq N''$ it holds that $d'(f_n(x), f(x)) < \epsilon$. Since $K'' \subset \mathcal{U}$, the result follows.

$\square$

Definition 1.6.20. For $f \in C^0([a, b], \mathbb{R})$, define $\|f\|_1 := \int_a^b |f(x)| dx$. (The norm axioms hold for $\|\cdot\|_1$.)

For $[a, b] \subset \mathbb{R}$, consider the vector space $C^0([a, b], \mathbb{R})$ (we could take functions into $\mathbb{C}$ instead and have a complex vector space). We have a norm $\|\cdot\|_\infty$ on $C^0([a, b], \mathbb{R})$ making it into a Banach space.

Proposition 65 (Math 425b Homework 3). If $f_n, f \in C^0([a, b], \mathbb{R})$ and f_n converge to f in $\|\cdot\|_\infty$ then f_n converges to f in $\|\cdot\|$. However, the opposite does not hold; converging in $\|\cdot\|_1$ does not imply convergence in $\|\cdot\|_\infty$.

Proof. Let $f_n, f \in C^0([a,b], \mathbb{R})$. Let $\epsilon > 0$. Suppose f_n converges to f in $\|\cdot\|_\infty$; that is, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $\|f_n - f\|_\infty = \sup_{x \in [a,b]} \{|f_n(x) - f(x)|\} < \epsilon/[2(b-a)]$. But then for all $n \geq N$,

$$\|f_n - f\|_1 = \int_a^b |f_n(x) - f(x)| \, dx \leq \int_a^b \sup_{x \in [a,b]} \{|f_n(x) - f(x)|\} \, dx = \frac{\epsilon}{2(b-a)} \cdot (b-a) = \frac{\epsilon}{2} < \epsilon.$$

For a counterexample showing that converging in $\|\cdot\|_1$ does not imply convergence in $\|\cdot\|_\infty$, define $f_n :$ $[0,1] \to \mathbb{R}$ by $f_n(x) = x^n$ and define $f : [0,1] \to \mathbb{R}$ by $f(x) = 0$. Clearly $f_n, f \in C^0([0,1], \mathbb{R})$. Let $\epsilon > 0$. Note that

$$\|f_n - f\|_1 < \epsilon \iff \int_0^1 x^n \, dx < \epsilon \iff \left[\frac{x^{n+1}}{n+1}\right]_0^1 < \epsilon \iff \frac{1}{n+1} < \epsilon \iff n > \frac{1}{\epsilon} - 1,$$

so for all $n \geq \lceil 1/\epsilon \rceil$ it holds that $\|f_n - f\|_1 < \epsilon$. Therefore f_n converges to f in $\|\cdot\|_1$. However,

$$\|f_n - f\|_\infty = \sup_{x \in [0,1]} \{|x^n|\} \geq |1^n| = 1 \qquad \forall n \in \mathbb{N},$$

so f_n does not converge to f in $\|\cdot\|_\infty$.

$\square$

Proposition 66 (Math 425b Homework 3). $(C^0([0,1], \mathbb{R}), \|\cdot\|_1)$ is not complete (i.e. not a Banach space).

Proof. Let $f_n : [0,1] \to \mathbb{R}$ be defined by

$$f_n(x) = \begin{cases} 0, & 0 \leq x \leq \frac{n-1}{2n}, \\ \frac{1-0}{1/2-(n-1)/2n}\left(x - \frac{n-1}{2n}\right), & \frac{n-1}{2n} < x \leq 1/2, \\ 1, & 1/2 < x \leq 1, \end{cases} = \begin{cases} 0, & 0 \leq x \leq \frac{n-1}{2n}, \\ 2nx - n + 1, & \frac{n-1}{2n} < x \leq 1/2, \\ 1, & 1/2 < x \leq 1, \end{cases}$$

for $n \in \mathbb{N}$. First we will show that these functions are Cauchy in d_1. Let $\epsilon > 0$. For $n, m \in \mathbb{N}$ with $n < m$, we have

$$\int_0^1 |f_n(x) - f_m(x)| \ dx$$

$$= \int_0^{(n-1)/2n} |0 - 0| \ dx + \int_{(n-1)/2n}^{(m-1)/2m} |2nx - n + 1 - 0| \ dx$$

$$+ \int_{(m-1)/2m}^{1/2} |2nx - n + 1 - (2mx - m + 1)| \ dx + \int_{1/2}^1 |1 - 1| \ dx$$

$$= \int_{(n-1)/2n}^{(m-1)/2m} (2nx - n + 1) \ dx + \int_{(m-1)/2m}^{1/2} |(n-m)(2x-1)| \ dx$$

$$= \left[nx^2 - (n-1)x \right]_{(n-1)/2n}^{(m-1)/2m} + (n-m) \left[x - x^2 \right]_{(m-1)/2m}^{1/2}$$

$$= \left[n \left(\frac{m-1}{2m} \right)^2 - (n-1) \left(\frac{m-1}{2m} \right) \right] - \left[n \left(\frac{n-1}{2n} \right)^2 - (n-1) \left(\frac{n-1}{2n} \right) \right]$$

$$+ (n-m) \left[\frac{1}{2} - \frac{1}{4} - \left(\frac{m-1}{2m} - \left[\frac{m-1}{2m} \right]^2 \right) \right]$$

$$= \frac{n}{4} \left[\left(\frac{m-1}{m} \right)^2 - \left(\frac{n-1}{n} \right)^2 \right] - \left(\frac{n-1}{2} \right) \left(\frac{m-1}{m} - \frac{n-1}{n} \right)$$

$$+ (n-m) \left[\frac{1}{4} - \frac{2m(m-1) - (m-1)^2}{4m^2} \right]$$

$$\vdots$$

$$\square$$

1.6.2   Power Series (Section 4.2 of Pugh [2015])

Note: if $z_0 \in \mathbb{C}$, can consider power series $\sum_{n=0}^\infty c_n (z - z_0)^n$ (where $c_n \in \mathbb{C}$): as a function of $z \in \mathbb{Z}$. We can still use

$$R = \frac{1}{\limsup\limits_{n \to \infty} |c_n|^{1/n}} \tag{1.9}$$

and if $|z - z_0| < R$ then the series converges absolutely, if $|z - z_0| > R$ then the series diverges (and this is the unique R with these properties). We can call $B_R(z_0)$ the open ball of radius R at z_0 and the disk of convergence.

Theorem 67 (**Theorem 4.11 from Pugh [2015]**). Let $\sum_{n=0}^\infty c_n (x - x_0)^n$ be a real power series. Let R be its radius of convergence, given by (1.9). If we have $r \in [0, R)$, then the series $\sum_{n=0}^\infty c_n (x - x_0)^n$ converges uniformly on $[-r + x_0, r + x_0]$.

Proof. Basic idea: use convergence of geometric series to study power series.

Chose β with $r < \beta < R$; such a β exists since $r < R$. Then $1/R < 1/\beta \iff \limsup\limits_{n \to \infty} |c_n|^{1/n} < 1/\beta$. Recall:

$$\limsup_{n\to\infty}|c_n|^{1/n} = \lim_{n\to\infty}\left(\sup_{n'\geq n}|c_{n'}|^{1/n'}\right)$$

(as n increases we have a smaller set that we are taking a supremum over, so the quantity we are taking the limit of is nonincreasing with n.) So $\limsup\limits_{n\to\infty}|c_n|^{1/n} < 1/\beta$ implies there exists an n_0 such that

$$\sup_{n'\geq n_0}|c_{n'}|^{1/n'} < \frac{1}{\beta};$$

i.e., for $n' \geq n_0$ we have

$$|c_{n'}|^{1/n'} < 1/\beta \iff |c_{n'}| < \beta^{-n'}.$$

Without loss of generality, let $x_0 = 0$. For $x \in [-r,r]$ and $n \geq n_0$ we have

$$|c_n x^n| < \beta^{-n}r^n = \underbrace{\left(\frac{r}{\beta}\right)^n}_{:=M_n}$$

(note that $r/\beta < 1$). By the Weierstrass M-test, if $\sum_{n=n_0}^{\infty} M_n$ is finite then $\sum_{n=n_0}^{\infty} c_n x^n$ converges uniformly on $[-r,r]$ (which is true if and only if $\sum_{n=0}^{\infty} c_n x^n$ converges uniformly on $[-r,r]$). But

$$\sum_{n=0}^{\infty}\left(\frac{r}{\beta}\right)^n = \frac{1}{1-r/\beta},$$

which is finite, proving the theorem.

$\square$

Corollary 67.1. $\sum_{n=0}^{\infty} c_n(x-x_0)^n$ converges compactly/locally uniformly on its interval of convergence $(-R+x_0, R+x_0)$.

Proof. Exercise. Show every compact subset $K \subset (-R+x_0, R+x_0)$ is contained in some $[-r+x_0, r+x_0]$.

$\square$

Theorem 68. Complex version: $\sum_{n=0}^{\infty} c_n(z-z_0)^n$ converges compactly/locally uniformly on its open disk of convergence $B_R(z_0)$.

Proof. Basically same; $K \subset \overline{B_r}(z_0) \subset B_R(z_0)$ for some $0 \leq r < R$ where $\overline{B_r}$ is the closure of B_r.

$\square$

Interlude: complex analysis approach to differentiating power series.

Definition 1.6.21. Let $D \subseteq \mathcal{C}$ be open. A function $f : D \to \mathcal{C}$ is **complex differentiable** or **holomorphic** if for each $z. \in D$ the limit of

$$\frac{\Delta f}{\Delta z} = \frac{f(z + \Delta z) - f(z)}{\Delta z}$$

exists as $\Delta z \to 0$ in $\mathbb{C}$. (The limit, if it exists, is a complex number.)

Theorem 69. if $\Omega \subset \mathbb{C}$ is open, $f_n : \Omega \to \mathbb{C}$ are **holomorphic** functions (complex differentiable) and f_n converge compactly/locally to some $f : \Omega \to \mathbb{C}$, then f is holomorphic and $f'(z) = \lim_{n \to \infty} f'_n(z)$.

Note that converging compactly/locally uniformly is weaker than uniform convergence. So somehow complex differentiability is easier to get than real differentiability. (Proven using different methods: Cauchy integral theorem, etc.)

Let's use this result. Apply this to power series. Note: $\sum_{n=0}^{N} c_n(z - z_0)^n$ is a (complex) polynomial. Some basic facts about complex differentiability: it's holomorphic and

$$\frac{\mathrm{d}}{\mathrm{d}z} \left(\sum_{n=0}^{N} c_n(z - z_0)^n \right)$$

can be differentiated using the usual rules for real number differentiation. Thus by the above theorem, complex power series can be (complex) differentiated term-by-term on their disk of convergence implies (not hard to prove) version. for real series, we'll give a different proof.

Theorem 70 (**Theorem 4.12 in** Pugh [**2015**]). Let $\sum_{k=0}^{\infty} c_k(x - x_0)^k$ be a real power series. Let R be its radius of convergence. Then $\sum_{k=0}^{\infty} c_k(x - x_0)^k$ is differentiable on $(-R + x_0, R + x_0)$ with derivative $\sum_{k=1}^{\infty} kc_k(x - x_0)^{k-1}$, which has the same radius of convergence.

In particular, the differentiated series

$$\sum_{k=1}^{\infty} kc_k(x - x_0)^{k-1}$$

and the integrated series

$$\sum_{k=0}^{\infty} \frac{c_k}{k+1}(x - x_0)^{k+1}$$

have the same radius of convergence R as the original series $\sum_{k=0}^{\infty} c_k(x - x_0)^k$.

(crucial fact about power series: you can differentiate them like you'd like to.)

To prove this, we will need some lemmas.

Lemma 71 (**Math 425b Homework 4**). Suppose a_n, b_n are sequences of real numbers with $a_n \geq 0$ and $b_n \geq 0$ for all n. Show that $\sup a_n b_n \leq \sup a_n \sup b_n$, as long as the right side of the inequality is not an indeterminate form $0 \times \infty$ or $\infty \times 0$.

Proof. Suppose $\sup_n a_n \in (-\infty, \infty), \sup_n b_n \in (-\infty, \infty)$. Let $n' \in \mathbb{N}$. Observe that $a_{n'}b_{n'} \leq \sup_n\{a_n\}b_{n'} \leq \sup_n\{a_n\}\sup_n\{b_n\}$. Since this statement is true for all n, it follows that $\sup_n\{a_n b_n\} \leq \sup_n\{a_n\}\sup_n\{b_n\}$.

Clearly $\sup_n a_n > -\infty$ and $\sup_n b_n > -\infty$ (in fact, $\sup_n a_n \geq 0$ and $\sup_n b_n \geq 0$), so the only other cases we need to consider are if one or both of the suprema equal infinity. But if this is true, then $\sup_n\{a_n\} \sup_n\{b_n\} = \infty$ and the inequality holds trivially.

$\square$

Lemma 72 (Math 425b Homework 4). Suppose a_n, b_n are sequences of real numbers with $a_n \geq 0$ and $b_n \geq 0$ for all n. Then

$$\limsup a_n b_n \leq \limsup a_n \limsup b_n,$$

as long as the right side of the inequality is not an indeterminate form $0 \times \infty$ or $\infty \times 0$ (and we can assume the limit of a product sequence is equal to the product of limits).

Proof. Recall that $\limsup a_n b_n = \lim_{N \to \infty} \sup_{n \geq N}\{a_n b_n\}$. For a fixed $N \in \mathbb{N}$, consider the sequences $\left(\tilde{a}_n^{(N)}\right)_{n=1}^{\infty}$ and $\left(\tilde{b}_n^{(N)}\right)_{n=1}^{\infty}$ defined by

$$\tilde{a}_n^{(N)} = \begin{cases} 0, & n < N, \\ a_n & n \geq N, \end{cases} \qquad \text{and} \qquad \tilde{b}_n^{(N)} = \begin{cases} 0, & n < N, \\ b_n, & n \geq N. \end{cases}$$

Note that $\sup \tilde{a}_n^{(N)} = \sup_{n \geq N} a_n$, etc. Since these are sequences of nonnegative real numbers for all $N \in \mathbb{N}$, we know from Lemma 71 that $\sup \tilde{a}_n^{(N)} \tilde{b}_n^{(N)} \leq \sup \tilde{a}_n^{(N)} \sup \tilde{b}_n^{(N)}$ for all $n \in \mathbb{N}$ and $N \in \mathbb{N}$ as long as the right side of the inequality is not an indeterminate form $0 \times \infty$ or $\infty \times 0$. Then

$$\sup \tilde{a}_n^{(N)} \tilde{b}_n^{(N)} \leq \sup \tilde{a}_n^{(N)} \sup \tilde{b}_n^{(N)}$$
$$\iff \lim_{N \to \infty} \left(\sup \tilde{a}_n^{(N)} \tilde{b}_n^{(N)}\right) \leq \lim_{N \to \infty} \left(\sup \tilde{a}_n^{(N)} \sup \tilde{b}_n^{(N)}\right)$$
$$\iff \lim_{N \to \infty} \left(\sup \tilde{a}_n^{(N)} \tilde{b}_n^{(N)}\right) \leq \lim_{N \to \infty} \left(\sup \tilde{a}_n^{(N)}\right) \lim_{N \to \infty} \left(\sup \tilde{b}_n^{(N)}\right)$$
$$\iff \lim_{N \to \infty} \left(\sup_{n \geq N} a_n \tilde{b}_n^{(N)}\right) \leq \lim_{N \to \infty} \left(\sup_{n \geq N} a_n\right) \lim_{N \to \infty} \left(\sup_{n \geq N} b_n\right)$$
$$\iff \limsup a_n b_n \leq \limsup a_n \limsup b_n,$$

where we used the assumption that the limit of a product sequence is equal to the product of limits.

$\square$

Lemma 73 (Math 425b Homework 4). Suppose a_n, b_n are sequences of real numbers with $a_n \geq 0$ and $b_n \geq 0$ for all n. Assume that $\lim_{n \to \infty} a_n = A$ with $0 < A < \infty$, and write $B = \limsup b_n \in [0, \infty]$. Show that

$$\limsup a_n b_n = AB.$$

Proof. If $B = \infty$, it holds trivially that $\limsup a_n b_n \leq AB$. Suppose $B \in [0, \infty)$. Since $\lim_{n \to \infty} a_n = A \in \mathbb{R}_+$ and $\limsup b_n = B \in \mathbb{R}_+$, it holds that $\limsup a_n = \lim a_n = A$ and likewise for b_n. Then by Lemma 72

$$\limsup a_n b_n \leq \limsup a_n \limsup b_n = \lim a_n \lim b_n = AB.$$

Now we will show the reverse inequality. First, suppose $B \in [0, \infty)$. Let $\epsilon > 0$. Because $\lim a_n = A$, there exists N_a such that for all $n \geq N_a$ we have $|a_n - A| < \epsilon \implies a_n > A - \epsilon$. Since $a_n, b_n \geq 0$ it follows that for all $N \geq N_a$

$$\sup_{n \geq N} a_n b_n \geq \sup_{n \geq N} (A - \epsilon) b_n = (A - \epsilon) \sup_{n \geq N} b_n = (A - \epsilon) B.$$

Therefore it follows from the ϵ-principle (Theorem 1.8 in Pugh [2015]) that $\sup_{n \geq N} a_n b_n \geq AB$ for all sufficiently large N, which in turn implies $\limsup a_n b_n \geq AB$.

Now suppose $B = \infty$. We have that there exists N_a' such that for all $n \geq N_a'$, $|A - a_n| < A/2 \implies a_n > A - A/2 = A/2$. Let $M \in \mathbb{R}_+$. Similarly, here exists N_b' such that for all $n \geq N_b'$, $b_n \geq 2M/A$. Let $N' := \max\{N_a', N_b'\}$, then for all $n \geq N'$ it holds that $a_n b_n > (A/2)(2M/A) = M$. Since M is arbitrary, it follows that in this case $\limsup a_n b_n \geq M$ for any $M \in \mathbb{R}_+$; that is, $\limsup a_n b_n = \infty$.

$\square$

Now we can prove the main result.

Proof of Theorem 70 (from Homework 4, Math 425b). We will rely on the identies

$$\lim_{k \to \infty} k^{1/k} = 1, \qquad \lim_{k \to \infty} \left(\frac{1}{k+1} \right)^{1/k} = 1 \tag{1.10}$$

and

$$\limsup \sqrt[k]{k|c_k|} = \limsup \sqrt[k]{|c_k|} = \limsup \sqrt[k]{\frac{|c_k|}{k+1}}, \tag{1.11}$$

which we will prove later. Assuming these hold, note that by Theorem 3.44 in Pugh [2015], the radius of convergence of the power series $\sum_{k=0}^{\infty} c_k (x - x_0)^k$ is

$$R = \frac{1}{\limsup\limits_{k \to \infty} \sqrt[k]{|c_k|}}.$$

We will show that $\sum_{k=1}^{\infty} k c_k (x - x_0)^{k-1} = \sum_{k'=0}^{\infty} (k'+1) c_{k'+1} (x - x_0)^{k'}$ and $\sum_{k=1}^{\infty} k c_k (x - x_0)^k$ have the same radii of convergence using the Root Test (Theorem 3.42 in Pugh [2015]). Let the radius of convergence of the first series be R_1' and the radius of convergence of the second series be R_1. Then in the first case, by Lemma 73

$$\limsup_{k' \to \infty} \sqrt[k']{|(k'+1) c_{k'+1}|} = \limsup_{k' \to \infty} (k'+1)^{1/k'} |c_{k'+1}|^{1/k'} = \limsup_{k' \to \infty} (k'+1)^{1/k'} \cdot \limsup_{k' \to \infty} |c_{k'+1}|^{1/k'}$$

$$= \lim_{k' \to \infty} \sup_{n' \geq k'} \left\{ (n'+1)^{1/n'} \right\} \cdot \lim_{k' \to \infty} \sup_{n' \geq k'} \left\{ |c_{n'+1}|^{1/n'} \right\} = \lim_{k' \to \infty} \sup_{n' \geq k'} \left\{ n'^{1/n'} \right\} \cdot \lim_{k' \to \infty} \sup_{n' \geq k'} \left\{ |c_{n'}|^{1/n'} \right\}$$

$$\implies R_1' = \frac{1}{\limsup\limits_{k'\to\infty} \sqrt[k']{|(k'+1)c_{k'+1}|}} = \frac{1}{\lim_{k'\to\infty}\sup_{n'\geq k'}\left\{\sqrt[n']{n'|c_{n'}|}\right\}}$$

and the radius convergence of the second is

$$R_1 = \frac{1}{\limsup\limits_{k\to\infty}\sqrt[k]{|kc_k|}} = \frac{1}{\lim_{k\to\infty}\sup_{n\geq k}\left\{\sqrt[n]{n|c_n|}\right\}} = R_1'.$$

Similarly, let the radius of convergence of $\sum_{k=0}^{\infty}\frac{c_k}{k+1}(x-x_0)^{k+1} = \sum_{k'=1}^{\infty}\frac{c_{k'-1}}{k'}(x-x_0)^{k'}$ be given by R_2' and the radius of convergence of $\sum_{k=0}^{\infty}\frac{c_k}{k+1}(x-x_0)^k$ be given by R_2. Then in the first case, by Lemma 73

$$\limsup\limits_{k'\to\infty}\sqrt[k']{\left|\frac{c_{k'-1}}{k'}\right|} = \limsup\limits_{k'\to\infty}\left\{k'^{-1/k'}|c_{k'-1}|^{1/k'}\right\} = \limsup\limits_{k'\to\infty}\left\{k'^{-1/k'}\right\}\limsup\limits_{k'\to\infty}\left\{|c_{k'-1}|^{1/k'}\right\}$$

$$= \lim\limits_{k'\to\infty}\sup_{n'\geq k'}\left\{n'^{-1/n'}\right\}\cdot\lim\limits_{k'\to\infty}\sup_{n'\geq k'}\left\{|c_{n'-1}|^{1/n'}\right\} = \lim\limits_{k'\to\infty}\sup_{n'\geq k'}\left\{(n'+1)^{-1/n'}\right\}\cdot\lim\limits_{k'\to\infty}\sup_{n'\geq k'}\left\{|c_{n'}|^{1/n'}\right\}$$

$$\implies R_2' = \frac{1}{\limsup\limits_{k'\to\infty}\sqrt[k']{\left|\frac{c_{k'-1}}{k'+1}\right|}} = \frac{1}{\lim_{k'\to\infty}\sup_{n'\geq k'}\left\{\sqrt[n']{(n'+1)^{-1}|c_{n'}|}\right\}}$$

and the radius convergence of the second is

$$R_2 = \frac{1}{\limsup\limits_{k\to\infty}\sqrt[k]{\left|\frac{c_k}{k+1}\right|}} = \frac{1}{\lim_{k\to\infty}\sup_{n\geq k}\left\{\sqrt[n]{(n+1)^{-1}|c_n|}\right\}} = R_2'.$$

Next, observe that the radius of convergence of $\sum_{k=1}^{\infty}kc_k(x-x_0)^k$ (and therefore the radius of convergence of the differentiated series $\sum_{k=1}^{\infty}kc_k(x-x_0)^{k-1} = \sum_{k'=0}^{\infty}(k'+1)c_{k'+1}(x-x_0)^{k'}$) is given by

$$R_1 = \frac{1}{\limsup\limits_{k\to\infty}\sqrt[k]{|kc_k|}} = \frac{1}{\limsup\limits_{k\to\infty}\sqrt[k]{k|c_k|}} = \frac{1}{\limsup\limits_{k\to\infty}\sqrt[k]{|c_k|}} = R,$$

where we used (1.11). Similarly, the radius of convergence of the series $\sum_{k=0}^{\infty}\frac{c_k}{k+1}(x-x_0)^{k+1}$ (and therefore the radius of convergence of the integrated series $\sum_{k=0}^{\infty}\frac{c_k}{k+1}(x-x_0)^{k+1}$) is

$$R_2 = \frac{1}{\limsup\limits_{k\to\infty}\sqrt[k]{|(k+1)^{-1}c_k|}} = \frac{1}{\limsup\limits_{k\to\infty}\sqrt[k]{(k+1)^{-1}|c_k|}} = \frac{1}{\limsup\limits_{k\to\infty}\sqrt[k]{|c_k|}} = R,$$

where we used (1.11).

It only remains to verify (1.10) and (1.11). To show that (1.10) holds, note that

$$\lim_{k \to \infty} k^{1/k} = \lim_{k \to \infty} \exp\left\{\frac{1}{k}\log(k)\right\} = \exp\left\{\lim_{k \to \infty}\left(\frac{\log(k)}{k}\right)\right\} = \exp\left\{\lim_{k \to \infty}\left(\frac{1}{k}\right)\right\} = 1,$$

where we applied L'Hopital's Rule. Next,

$$\lim_{k \to \infty}\left(\frac{1}{k+1}\right)^{1/k} = \lim_{k \to \infty}\exp\left\{\frac{1}{k}\log\left(\frac{1}{k+1}\right)\right\} = \exp\left\{\lim_{k \to \infty}\left[\frac{1}{k}\log\left(\frac{1}{k+1}\right)\right]\right\}$$
$$= \exp\left\{\lim_{k \to \infty}\left[\frac{k+1}{1} \cdot \frac{-1}{(k+1)^2}\right]\right\} = \exp\left\{\lim_{k \to \infty}\left[\frac{-1}{k+1}\right]\right\} = 1,$$

where we applied L'Hopital's Rule, verifying (1.10). We will conclude by verifying (1.11). Let $\limsup|c_k|^{1/k} :=$ B. Since $|c_k| \geq 0$ for all $k \in \mathbb{N}$, $B \in [0, \infty]$. We already have from (1.10) that $\lim k^{1/k} = \lim(k+1)^{-1/k} = 1$, so $\limsup k^{1/k} = \limsup(k+1)^{-1/k} = 1$. Since $k^{1/k} \geq 0$ for all $k \in \mathbb{N}$ and likewise with $(k+1)^{-1/k}$, by Lemma 73 we have

$$\limsup \sqrt[k]{k|c_k|} = \limsup k^{1/k}|c_k|^{1/k} = \limsup k^{1/k}\limsup|c_k|^{1/k} = \limsup|c_k|^{1/k}.$$

Similarly,

$$\limsup \sqrt[k]{\frac{|c_k|}{k+1}} = \limsup\left(\frac{1}{k+1}\right)^{1/k}|c_k|^{1/k} = \limsup\left(\frac{1}{k+1}\right)^{1/k}\limsup|c_k|^{1/k} = \limsup|c_k|^{1/k}.$$

$\square$

Proof of Theorem 70 (in-class sketch). Idea: first show the radius is unchanged when we pass from the original series to the derivative. Then show that the differentiated series converges uniformly on compact subsets (from M-test). Finally, use theorem on uniform convergence of derivatives.

Rigorously: first part will be on Homework 4. From there:

$$R = \frac{1}{\displaystyle\limsup_{n \to \infty}|nc_n|^{1/n}}.$$

Assume it holds that the radius is unchanged when we pass from the original series to the derivative. Without loss of generality, let $x_0 = 0$. For any $0 \leq r < R$, we know that $\sum_{n=1}^{\infty} nc_n x^{n-1}$ converges uniformly on $[-r, r]$, so it converges uniformly on $(-r, r)$. (For $x \in (-r, r)$, derivative of a function on $(-R, R)$ restricted to $(-r, r)$ is same as derivative of a function on $(-R, R)$).

Let $f_N := \sum_{n=0}^{N} c_n x^n$ on $(-r, r)$. Thus $f_N' = \sum_{n=0}^{N} nc_n x^{n-1}$ on $(-r, r)$. This converges uniformly to $g := \sum_{n=1}^{\infty} nc_n x^{n-1}$ on $(-r, r)$. Also, f_N converges to $f = \sum_{n=0}^{\infty} c_n x^n$ on $(-r, r)$. So f is differentiable on $(-r, r)$ with derivative g.

Since any $X \in (-R, R)$ is in some $(-r, r)$ for some $0 \leq r < R$, it follows that $\sum_{n=0}^{\infty} c_n x^n$ is differentiable on $(-R, R)$ with derivative $\sum_{n=1}^{\infty} nc_n x^{n-1}$.

$\square$

Definition 1.6.22. A function $f : (a, b) \to \mathbb{R}$ is **(real) analytic** if for all $x_0 \in (a, b)$, there exists a power series $\sum_{n=0}^{\infty} c_n (x - x_0)^n$ (centered at x_0) with positive radius of convergence such that for all $x \in \mathcal{U}$ with $\mathcal{U}$ open, where $\mathcal{U}$ is a subset of the intersection of the interval of convergence and (a, b), we have

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

Definition 1.6.23. $\Omega \subset \mathbb{C}$ is open: $f : \Omega \to \mathbb{C}$ is **(complex) analytic** if for all $z_0 \in \mathbb{C}$, there exists a series $\sum_{n=0}^{\infty} c_n (z - z_0)^n$ with a positive radius of convergence such that $f(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^n$ in an open neighborhood of z_0 contained in the intersection of Ω and the disk of convergence.

Remark 5. It is somewhat unintuitive that being analytic is necessary for being holomorphic. In real functional analysis, even being C^1 is a stronger condition than being differentiable. Note that being C^∞ is a weaker condition than being analytic; some functions are C^∞ but not analytic. So in complex functional analysis, this "hierarchy" collapses.

Theorem 74 (**Math 520**). $f : \Omega \to \mathbb{C}$ is holomorphic if and only if f is complex analytic.

What haven't we proved rigorously from calc 1 class?

- Transcendental functions (exp, sin, cos, etc.)

- Convergent power series are analytic in their interval/disc of convergence (not a tautology—not obvious that at any point in this interval, a new power series exists that expresses the function locally). (Proven in Section 4.6)

- "Given a smooth $f : (a, b) \to \mathbb{R}$, is f equal to its Taylor series in an open neighborhood of each $x_0 \in \mathbb{R}$?" e.g.

$$f(x) = \begin{cases} 0, & x \leq 0 \\ e^{-1/x}, & x > 0 \end{cases}$$

  Taylor series at 0: all derivatives equal 0, so all coefficients equal 0. So converges but not to function.

  The above question is equivalent to the question "is f analytic?" Since if $f(x) = \sum_{n=0}^{\infty} c_n (x - x_0)^n$ for x near x_0, then by term-by-term differentiation it is easy to see $f(x_0) = c_0, f'(x_0) = c_1, f''(x_0) = 2c_2, \ldots, f^{(r)}(x_0) = r! c_r$. So, having local derivative growth rate bounded comes if and only if f is analytic. (Proven in Section 4.6)

To finish 4.2; transcendental functions. Can define and prove basic properties using power series or differential equations (this is done in Section 4.5). (See Chapter 8 of Rudin [1976] for more content.)

Definition 1.6.24 (**Natural exponential function**).

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

Remark 6. Using techniques from 4.6 or 4.5 (Picard's Theorem), you can define e^x another way and show that the definitions agree. For example,

$$\log(x) := \int_1^x \frac{1}{t} \, dt$$

is a way to define the natural logarithm. Then can show that $\log : (0, \infty) \to \mathbb{R}$ is a bijection, then can define the exponential function to be the inverse of this function.

Another way to see the definitions are the same:

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

can prove term-by-term that exponential is differentiable (even smooth on $\mathbb{R}$) and that its derivative equals itself and $\exp(0) = 1$.

Or; if defining exp as the inverse of log, use the inverse deirvative formula with the same initial condition to get the same exp function.

Also: Picard's Theorem in 4.5 will imply that these two definitions of exp are the same. Very general way to understand exp, sin, cos, etc. determined by the ODEs they satisfy.

Definition 1.6.25. For $x \in \mathbb{R}$,

$$\sin(x) := \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}, \qquad \cos(x) := \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}, \qquad \tan(x) = \frac{\sin(x)}{\cos(x)}, \text{ etc.}$$

Exercise 7. $R = \infty$ for these power series. By term-by-term differentiation, get $\sin' = \cos, \cos' = -\sin$. So we get second-order differential equations $\sin'' = -\sin(0) = 0, \sin'(0) = 1$. Similarly, $\cos'' = -\cos, \cos(0) = 1, \cos'(0) = 0$.

Can also take these definitions for complex numbers and show $e^{i\theta} = \cos(\theta) + i\sin(\theta)$.

How can we deduce new power series from old ones? (Can integrate/differentiate old ones.) For example, power series for $\log(x)$. We know that (from the sum of an infinite geometric series)

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n, \qquad x \in (-1, 1)$$

with a radius of 1. We also know

$$\frac{1}{1+x} = \sum_{n=0}^{\infty} (-1)^n x^n, \qquad x \in (-1, 1)$$

Then

$$\log(1+x) = \int_0^x \frac{1}{1+t}\ dt = \int_0^x \sum_{n=0}^{\infty} (-1)^t x^t\ dt = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1}, \qquad x \in (-1,1).$$

Theorem 75 (**Theorem 4.13 in Pugh [2015]**)**.** Any analytic function is C^{∞}.

Remark 7. It turns out that the complex analogue of this statement is also true.

Proof. This follows immediately from term-by-term differentiation. In particular, an analytic function f is defined by a convergent power series. By Theorem 70, the derivative of f is given by a convergent power series with the same radius of convergence, so repeated differentiation is valid and f is smooth. $\qquad\square$

Theorem 76 (**From Math 520**)**.** $\Omega \subset \mathbb{C}$ open, $f : \Omega \to \mathbb{C}$ is holomorphic, $z_0 \in \Omega$, then Taylor series of f at z_0

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n$$

converges to f on any open ball $B_R(z_0)$ with $B_R(z_) \subset \Omega$. (so radius of convergence greater than or equal to R.) See Figure 1.3.
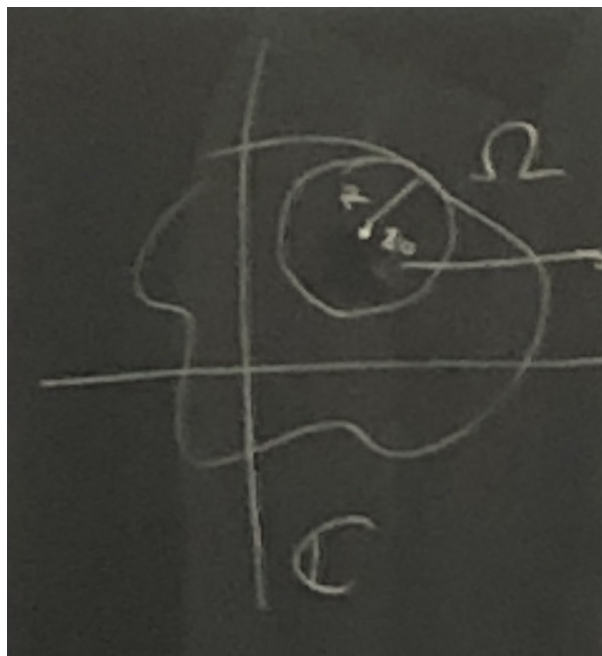
Figure 1.3: If $B_R(z_0)$ fits in Ω then the Taylor series converges to f at least on $B_R(z_0)$. See Theorem 76.

Pugh (end of section 4.2): compare $1/(1 + x^2)$ to e^x. Think about $1/(1 + z^2)$ in complex numbers; it blows up at $z = \pm i$. So the largest possible radius of convergence in the complex plane is 1. That's basically why the radius of convergence of the Taylor series of $1/(1 + x^2)$ centered at 0 is only 1.

1.6.3   Compactness and Equicontinuity in C^0 (Section 4.3 of Pugh [2015])

Recall: X set, Y complete metric space, then $C_b(X, Y)$ is complete (Corollary 51.1). If X is a metric space, then $C^0(X, Y)$ is complete (Corollary 51.3). How about if Y is compact? Is $C^0(X, Y)$ compact?

Example 1.6.3. $X = [n]$ with discrete metric. $C^0([n], \mathbb{R}) \approx \mathbb{R}^n$ with ℓ^1 metric, $d(x, y) = \max |x_i - y_i|$. Then $C^0([n], [-1, 1]) \approx [-1, 1]^n$ with ℓ^1 metric.

In this case, $[-1, 1]^n$ is compact in the usual metric, so compact here.

But: now let $X := \mathbb{N}$. Consider $(C^0(\mathbb{N}, [-1, 1]), d_\infty)$. Is this compact? No. Counterexample: $f_n(x) = \begin{cases} 1, & x = n \\ 0, & x \neq n \end{cases}$. Exercise; there does not exist a uniformly convergent subsequence for this sequence. So $C^0(\mathbb{N}, [-1, 1])$ is not compact.

One way to think about this: $C^0(\mathbb{N}, [-1, 1])$ is the closed unit ball in $C^0(\mathbb{N}, \mathbb{R})$. This is a Banach space under $\|\cdot\|_\infty$. This is an infinite dimensional vector space. In a finite-dimensional normed vector space, can show that the closed unit ball is compact. But in infinite dimensional vector spaces that usually goes wrong.

Definition 1.6.26 (**Equicontinuity; definition 7.22 in Rudin [1976], p.165 of pdf, p.156 of book**). Let $(X, d), (Y, d')$ be metric spaces. Let $\mathcal{E}$ be a set of functions from X to Y. $\mathcal{E}$ is called **equicontinuous** if for every $\epsilon > 0$ there exists $\delta > 0$ depending only on ϵ such that for all $x, x' \in X$ with $d(x, x') < \delta$ and for all $f \in \mathcal{E}$ we have $d'(f(x), f(x')) < \epsilon$.

Note: if $\mathcal{E}$ is equicontinuous and $f \in \mathcal{E}$ then f is uniformly continuous. So being equicontinuous is like being uniformly continuous together.

Can also formulate "pointwise equicontinuity:" δ can depend on a fixed point x_0 (but cannot depend on $f \in \mathcal{E}$).

Sequence of functions $(f_n)_{n=1}^\infty$: can take $\mathcal{E} = \{f_n : n \geq 1\}$ and ask if $\mathcal{E}$ is equicontinuous. Write explicitly what this means:

Definition 1.6.27. $f_n : X \to Y$, with X, Y metric spaces. $(f_n)_{n=1}^\infty$ is an equicontinuous sequence if for every $\epsilon > 0$ there exists $\delta > 0$ such that if $x, x' \in X$, $d(x, x') < \delta$ then for every $n \in \mathbb{N}$, $d'(f_n(x), f_n(x')) < \epsilon$.

Last time: finished Section 4.2 of Pugh [2015]. Learned a trick for determining the radius of convergence of an infinite series geometrically if you go to $\mathbb{C}$.

Next we discussed that in infinite dimensional normed vector spaces, the closed unit ball might not be compact. (In fact, the closed unit ball is compact if and only if it is finite-dimensional. Recall: definition of dimension is the number of elements in a basis for the space.) Next we discussed how to ensure compactness of subsets of $C^0(X, \mathbb{R})$ where X is a compact metric space. One way is if the space is totally bounded and complete, but this is harder than applying Heine-Borel (Theorem 2.33 in Pugh [2015], Theorem 32 in these notes) to the equivalent question for subsets of $\mathbb{R}^m$, where you just need to show the set is closed and bounded. Can we get an analogous result to Heine-Borel for subsets of $C^0(X, \mathbb{R})$ (where X is a compact metric space)? Yes: it turns out $\mathcal{E} \subset C^0(X, \mathbb{R})$ is compact if and only if it is closed, bounded, and equicontinuous. (We will show this in Theorem 89.)

Example 1.6.4 (**Non-example: sequence of functions that is not equicontinuous**)**.** Let $X = [0, 1]$ and $Y = \mathbb{R}$. Let $f_n(x) = x^n$. We will show that the set $\{f_n\}_{n=1}^{\infty}$ is not equicontinuous. Assume there exists $\delta \in (0, 1)$ such that if $s, t \in [0, 1]$ and $|s - t| < \delta$ and $n \geq 1$, then $|f_n(s) - f_n(t)| < 1/2$ (that is, let $\epsilon = 1/2$). We know $\lim_{n \to \infty} (1 - \delta/2)^n = 0$, so there exists n^* with $(1 - \delta/2)^{n^*} < 1/2$. Then

$$|f_{n^*}(1 - \delta/2) - f_{n^*}(1)| = \left| \underbrace{(1 - \delta/2)^{n^*}}_{<1/2} - \underbrace{1^{n^*}}_{=1} \right| > \frac{1}{2},$$

contradiction. Therefore $\{f_n\}_{n=1}^{\infty}$ is not equicontinuous.

Note: these functions are a sequence of functions in the closed unit ball of $(C^0([0, 1], \mathbb{R}), \|\cdot\|_{\infty})$ with no convergent subsequence. (This is true because if the limit existed it would be a uniform limit since we are using the $\|\cdot\|_{\infty}$ norm, and then the uniform limit would equal the pointwise limit since each of the individual functions are continuous. But the pointwise limit is not continuous.)

From this example we get an intuition that equicontinuity is related to the slopes of functions in the subset being bounded in some sense. How do we ensure equicontinuity? We can do it if we can bound the slopes of the functions, in the sense of the following definition.

Definition 1.6.28. Let (X, d) and (Y, d') be metric spaces. Let $f : X \to Y$. Then

1. $M \in \mathbb{R}$ is a **Lipschitz constant** for f if

$$\forall x, x' \in X, \text{ we have } d'(f(x), f(x')) \leq M \cdot d(x, x').$$

2. $M \in \mathbb{R}$ is a **uniform Lipschitz constant** for a set of functions $\mathcal{E} \subseteq C^0(X, Y)$ if

$$\forall f \in \mathcal{E} \text{ and } \forall x, x' \in X \text{ we have } d'(f(x), f(x')) \leq M \cdot d(x, x').$$

Think of M as a "global steepness bound" for f. Note that f having a Lipschitz constant M (that is, f is **Lipschitz continuous**) implies that f is uniformly continuous. (Proof: exercise. Hint: given ϵ, take $\delta = \epsilon/M$.)

Proposition 77. If a uniform Lipschitz constant M exists for a $\mathcal{E} \subset C^0(X, Y)$, then $\mathcal{E}$ is equicontinuous.

Proof. Let $\epsilon > 0$. Take $\delta := \epsilon/M$. If $x, x' \in X$, $d(x, x') < \delta$, and $f \in \mathcal{E}$, then

$$d'(f(x), f(x')) \leq M d(x, x') < M \cdot \frac{\epsilon}{M} = \epsilon.$$

$\square$

This makes sense when X, Y are metric spaces. Lipschitz constants are global bounds on steepness. What about local bounds?

Corollary 77.1. If $\mathcal{E}$ is a set of differentiable functions from $[a, b]$ to $\mathbb{R}$ (or $\mathbb{C}$) and there exists $M \in \mathbb{R}$ such that $|f'(\theta)| \leq M$ for all $f \in \mathcal{E}$ and for all $\theta \in [a, b]$, then $\mathcal{E}$ is equicontinuous.

Proof. By the Mean Value Theorem, M is a uniform Lipschitz constant for $\mathcal{E}$. This implies that $\mathcal{E}$ is equicontinuous. (Details: exercise.)

$\square$

Example 1.6.5 (Sanity check). Let $X = [0,1]$. Let $Y = \mathbb{R}$. Let $f_n(x) = x^n$. Based on what we've just done, there shouldn't be a uniform bound on $|f_n'(\theta)|$ for all n, θ, because if there were, this would imply equicontinuity.

Let's check: we have $f_n'(x) = nx^{n-1}$. Note that $\|f_n'\|_\infty = n$. Question: does there exist any M such that $\|f_n'\|_\infty \leq M$ for all n? Of course not: there is no M such that for all $n \in \mathbb{N}$, $n \leq M$. So everything is consistent.

A couple more remarks: Generalized Heine-Borel: we will show in Theorem 89 (Theorem 4.18 in Pugh [2015]) that $\mathcal{E} \subset C^0(X, \mathbb{R})$ with X is compact if and only if $\mathcal{E}$ is closed, bounded, and equicontinuous. Is this really a generalization—does it recover the usual Heine-Borel theorem (Theorem 2.33 in Pugh [2015], Theorem 32 in these notes)? Yes. Consider $X = \{1, \ldots, n\}$. Any subset of $C^0(\{1, \ldots, n\}, \mathbb{R})$ is equicontinuous. (We can think of the set $C^0(\{1, \ldots, n\}, \mathbb{R})$ as being an alternative way of expressing $\mathbb{R}^n$ itself.) Thus, $\mathcal{E} \subset \mathbb{R}^n$ is compact if and only if $\mathcal{E}$ is closed and bounded. That is, Heine-Borel can be thought of as an application of Theorem 89 in the special case of the set $C^0(\{1, \ldots, n\}, \mathbb{R})$.

Proposition 78 (Math 425b Homework 5). Let $\{f_1, \ldots, f_k\}$ be a finite set of uniformly continuous functions from X to Y, where (X, d) and (Y, d') are metric spaces. Then $\{f_1, \ldots, f_k\}$ is equicontinuous.

Proof. Let $\epsilon > 0$. Because all the f_i are continuous, for every $i \in \{1, \ldots, k\}$ there exists δ_i such that for all $x \in X$, $d(x, x') < \delta_i \implies d'(f_i(x), f_i(x')) < \epsilon$. Let $\delta := \min_{i \in \{1, \ldots, k\}} \delta_i$. Then for all $i \in \{1, \ldots, k\}$, $(d(x, x') < \delta_i \implies d'(f_i(x), f_i(x')) < \epsilon$.

$\square$

Proposition 79 (Math 425b Homework 5). Let $\{f_\alpha\}_{\alpha \in A}$ be an equicontinuous set of functions from X to Y, and let g be a uniformly continuous function from Y to Z, where (X, d_X), (Y, d_Y), and (Z, d_Z) are metric spaces. Then the set of functions $\{g \circ f_\alpha\}_{\alpha \in A}$ is equicontinuous.

Proof. Let $\epsilon > 0$. Because g is uniformly continuous, there exists $\delta > 0$ such that for any $y, y' \in Y$, $d_Y(y, y') < \delta \implies d_Z(g(y), g(y')) < \epsilon$. Because $\{f_\alpha\}_{\alpha \in A}$ is equicontinuous, there exists an $\eta > 0$ such that for all $\alpha \in A$, $d_X(x, x') < \eta \implies d_Y(f_\alpha(x), f_\alpha(x')) < \delta$. Therefore for all $\alpha \in A$, $d_X(x, x') < \eta \implies d_Z(g(f_\alpha(x)), g(f_\alpha(x'))) < \epsilon$, so $\{g \circ f_\alpha\}_{\alpha \in A}$ is equicontinuous.

$\square$

Proposition 80 (Math 425b Homework 5; Exercise 4.8 in Pugh [2015]). The sequence of functions $f_n : \mathbb{R} \to \mathbb{R}$ defined by

$$f_n(x) = \cos(n + x) + \log\left(1 + \frac{1}{\sqrt{n+2}} \sin^2(n^n x)\right)$$

is equicontinuous.

Proof. First we will show that the sequence $g_n(x) = \frac{1}{\sqrt{n+2}} \sin^2(n^n x)$ is equicontinuous. Let $\epsilon > 0$, and let $N := \lceil 4\epsilon^{-2} - 2 \rceil + 1$ so that $(n+2)^{-1/2} < \epsilon/2$ for all $n \geq N$. By Proposition 78, the functions $\{f_1, \ldots, f_{N-1}\}$ are equicontinuous; that is, there exists $\delta > 0$ such that for all $x \in \mathbb{R}$ $|x - x'| < \delta \implies |g_n(x) - g_n(x')| < \epsilon$ for all $n \in \{1, \ldots, N-1\}$. Observe that for all $n \geq N$

$$g_n(x) = \frac{\sin^2(n^n x)}{\sqrt{n+2}} < \frac{\epsilon}{2} \cdot \sin^2(n^n x) \leq \frac{\epsilon}{2}$$

for all $x \in \mathbb{R}$ since $\sin^2(x) \in [0, 1]$ for all $x \in \mathbb{R}$. Therefore for all $n \geq N$ and all $x, x' \in \mathbb{R}$ it holds that

$$|g_n(x) - g_n(x')| < \left| \frac{\epsilon}{2} + \frac{\epsilon}{2} \right| = \epsilon.$$

In particular, this holds if $|x - x'| < \delta$, so $(g_n)_{n=1}^\infty$ is equicontinuous. Next we will show that $(\cos(n + x))_{n=1}^\infty$ is an equicontinuous sequence on $[0, 2\pi]$. By Proposition 77, a sufficient condition to show this is that there exists $M \in \mathbb{R}$ such that

$$\left| \frac{\mathrm{d}}{\mathrm{d}\theta} \cos(n + x) \right| \leq M \qquad \forall n \in \{1, 2, \ldots\}, \forall x \in [0, 2\pi].$$

This follows since

$$\left| \frac{\mathrm{d}}{\mathrm{d}\theta} \cos(n + x) \right| = |\sin(n + x)| \leq 1 \qquad \forall n \in \mathbb{N}, \forall x \in \mathbb{R}.$$

Then by the periodicity of $\cos(\cdot)$, $(\cos(n + x))_{n=1}^\infty$ is an equicontinuous sequence everywhere.

By the fact that the sum of two equicontinuous sequences is equicontinuous, the result follows if we can show that $\left(\log \left[1 + \frac{\sin^2(n^n x)}{\sqrt{n+2}} \right] \right)_{n=1}^\infty$ is equicontinuous. We showed above that g_n is equicontinuous, so since the sum of two equicontinuous sequences is equicontinuous, $\left(1 + \frac{\sin^2(n^n x)}{\sqrt{n+2}} \right)_{n=1}^\infty$ is equicontinuous. By Proposition 79, we would be done if $\log(\cdot)$ were uniformly continuous, but that is not true. However, it holds that

$$1 \leq 1 + \frac{\sin^2(n^n x)}{\sqrt{n+2}} \leq 2 \qquad \forall n \in \mathbb{N},$$

so by Proposition 79 it is enough to show that $\log(\cdot)$ is uniformly continuous on $[1, 2]$.

We claim this is true. As in the proof of Proposition 77, if there exists $M \in \mathbb{R}$ such that $\left| \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\theta) \right| \leq M$ for all $\theta \in [1, 2]$, then by the Mean Value Theorem M is a uniform Lipschitz constant for $\log(\cdot)$ on $[1, 2]$ and then $\log(\cdot)$ is uniformly continuous on $[1, 2]$. Note that for $\theta \in [1, 2]$,

$$\left| \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\theta) \right| = \frac{1}{\theta} \in \left[\frac{1}{2}, 1 \right]$$

so $\left| \frac{\mathrm{d}}{\mathrm{d}\theta} \log(\theta) \right| \leq 1$ for all $\theta \in [1, 2]$ and the result follows.

$\square$

Proposition 81 (**Math 425b Homework 5;** Pugh [2015] **Exercise. 4.22**)**.** A sequence of smooth equicontinuous functions $f_n : [a, b] \to \mathbb{R}$ does not necessarily have uniformly bounded derivatives.

Proof. Consider $f_n : [0, 2\pi] \to \mathbb{R}$ defined by

$$f_n(x) := \frac{\sin^2(n^n x)}{\sqrt{n + 2}} \qquad \forall n \in \mathbb{N}.$$

Clearly $f_n \in C^\infty([0, 2\pi], \mathbb{R})$ so f_n is smooth for all $n \in \mathbb{N}$. We showed in the proof of Proposition 80 that $(f_n)_{n=1}^\infty$ is equicontinuous. However,

$$\frac{\mathrm{d}}{\mathrm{d}x} f_n(x) = \frac{2\sin(n^n x) \cdot n^n}{\sqrt{n + 2}} - \frac{1}{2(n + 2)^{3/2}} \cdot \sin^2(n^n x),$$

which is a sequence that is not uniformly bounded over n since if $x = \frac{\pi}{2n^n}$ (so $\sin(n^n x) = 1$)

$$\begin{aligned}
\lim_{n \to \infty} \frac{\mathrm{d}}{\mathrm{d}x} f_n(x) &= \lim_{n \to \infty} \left[\frac{2\sin(n^n x) \cdot n^n}{\sqrt{n + 2}} - \frac{1}{2(n + 2)^{3/2}} \cdot \sin^2(n^n x) \right] \\
&= \lim_{n \to \infty} \left[\frac{2 \cdot n^n}{\sqrt{n + 2}} - \frac{1}{2(n + 2)^{3/2}} \right] \\
&= \infty.
\end{aligned}$$

$\square$

Before proving the Arzela-Ascoli Theorem (Theorem 4.14 in Pugh [2015]) we will recall a concept from Math 425a and prove some lemmas.

Definition 1.6.29 (**Dense metric spaces (see also Definition 1.6.30 for a special case)**)**.** Let (X, d) be a metric space. $A \subset X$ is **dense** if any of the following equivalent statements hold:

1. For all $x \in X$ and all $\epsilon > 0$, there exists $a \in A$ with $d(x, a) < \epsilon$.

2. Each point in X is the limit of a sequence in A.

3. $\overline{A} = X$ (where $\overline{A}$ denotes the closure of X).

Example 1.6.6. Some examples: $\mathbb{Q} \subset \mathbb{R}$ is countable and dense. $\mathbb{Q} \cap [a, b] \subset [a, b]$ is countable and dense.

Lemma 82. If X is a compact metric space then X has a countable dense subset A. (That is, X is **seperable**.)

Proof. X is totally bounded, so there exists a finite subset A_n of X such that

$$\bigcup_{a \in A_n} B_{1/n}(a) = X.$$

Let $A := \bigcup_{n \geq 1} A_n$, a countable subset of X. Want to show: $\overline{A} = X$. Use condition 1 from Definition 1.6.29: if $x \in X$ and $\epsilon > 0$, then choose n sufficiently large such that $1/n < \epsilon$. Thus there exists $a \in A_n \subset A$ with $x \in B_{1/n}(a) \subset B_\epsilon(a)$, so $d(x, a) < \epsilon$. Thus $\overline{A} = X$.

$\square$

Lemma 83 (Math 425b Homework 5). Let X be a compact metric space and let $A = \{a_n\}_{n \geq 1}$ be a countable dense subset of X. Then for any $\delta > 0$ there exists M such that for any $x \in X$ we have $d(x, a_i) < \delta$ for some i with $1 \leq i \leq M$.

Proof. Let $\delta > 0$. Because $A = \{a_1, \ldots\}$ is dense in X, $\bigcup_{i=1}^{\infty} B_\delta(a_i)$ covers X. Because X is a compact metric space, by Theorem 2.63 in Pugh [2015], $\bigcup_{i=1}^{\infty} B_\delta(a_i)$ has a finite subcover. That is, there are finitely many points $x_1, \ldots, x_M \in A$ such that

$$X \subset \bigcup_{i=1}^{M} B_\delta(a_i).$$

Therefore for some $i \in \{1, \ldots, M\}$ it holds that $d(x, a_i) < \delta$.

$\square$

Theorem 84 (Arzela-Ascoli Propagation Theorem (Theorem 4.16 in Pugh [2015])). Let X be a compact metric space. Let $A \subseteq X$ be a countable dense subset of X (such a set A is guaranteed to exist by Lemma 82). Let Y be a complete metric space. Let $(f_n)_{n=1}^{\infty}$ be an equicontinuous sequence of functions $X \to Y$. If $(f_n(a))_{n=1}^{\infty}$ converges as a sequence in Y for all points $a \in A$, then f_n converge uniformly to some f on X.

Proof. Let $\epsilon > 0$. Then equicontinuity implies that for all $n \in \mathbb{N}$ there exists $\delta = \delta_\epsilon$ such that $d(x, x') < \delta \implies d(f_n(x), f_n(x')) < \epsilon/3$ for all $n \in \mathbb{N}$. Since A is countable, call its elements $\{a_1, \ldots, a_n, \ldots\}$. Because X is compact, Lemma 83 guarantees that for some $J \in \mathbb{N}$, for all $x \in X$ there exists $j \in \{n_1, \ldots, n_J\} \subset \{1, 2, \ldots\}$ such that $d(x, a_j) < \delta$. By the pointwise convergence assumption, for any $j \in \{n_1, \ldots, n_J\}$, $(f_n(a_j))_{n=1}^{\infty}$ is convergent in Y. So it is Cauchy, so there exists N_j such that for all $n, m \geq N_j$ and for all $j \in \{n_1, \ldots, n_J\}$ it holds that

$$d(f_n(a_j), f_m(a_j)) < \epsilon/3. \tag{1.12}$$

Let $N := \max_{j \in \{n_1, \ldots, n_J\}} \{N_j\}$; this quantity is well-defined since it is the maximum over a finite countable set. Then for any $n, m \geq N$, for all $x \in X$ there exists a_j such that $d(x, a_j) < \delta$; for this a_j it holds that

$$
\begin{aligned}
d(f_n(x), f_m(x)) \leq & d(f_n(x), f_n(a_j)) + d(f_n(a_j), f_m(a_j)) + d(f_m(a_j), f_m(x)) \\
< & \underbrace{\frac{\epsilon}{3}}_{\text{by equicontinuity}} + \underbrace{\frac{\epsilon}{3}}_{\text{by (1.12)}} + \underbrace{\frac{\epsilon}{3}}_{\text{by equicontinuity}} \\
= & \epsilon.
\end{aligned}
$$

Since Y is complete and (f_n) are uniformly Cauchy, f_n is uniformly convergent.

$\square$

Theorem 85 (Bolzano-Weierstrass Theorem; Theorem 2.31 in Pugh [2015]). Every bounded sequence in $\mathbb{R}^m$ has a convergent subsequence.

Theorem 86 (Theorem 7.23 in Rudin [1976], p.156, p.165 of pdf). If (f_n) is a pointwise bounded sequence of complex functions on a countable set E, then (f_n) has a subsequence f_{n_k} such that $(f_{n_k}(x))$ converges for every $x \in E$.

Proof. Let $(x_i), i \in \{1, 2, \ldots\}$ be the points of E, arranged in a sequence. Since $(f_n(x_1))$ is bounded, by the Bolzano-Weierstrass Theorem (Theorem 85) there exists a subsequence (which we will denote by $S_1 = (f_{1,k}) = \{f_{1,1}, f_{1,2}, f_{1,3}, \ldots\}$) of (f_n) such that $(f_{1,k}(x_1))$ converges as $k \to \infty$.

Now consider the sequence $(f_{1,k}(x_2))$. This sequence is also bounded (it is a subsequence of a bounded sequence), so it has its own subsequence $S_2 = (f_{2,k}) = \{f_{2,1}, f_{2,2}, f_{2,3}, \ldots\}$) such that $(f_{2,k}(x_2))$ converges as $k \to \infty$.

We can continue in this way, creating sequences S_n such that (a) S_n is a subsequence of S_{n-1} for $n \in \{2, 3, 4, \ldots\}$, (b) $(f_{n,k}(x_n))$ converges as $k \to \infty$, and (c) the order in which the functions appear is the same in each sequence (we only remove functions, so as n increases functions may move to the left in the sequence but never to the right).

Now go down the diagonal of the array, considering the sequence $S = \{f_{1,1}, f_{2,2}, f_{3,3}, \ldots\}$. This sequence (except possibly its first $n - 1$ terms) is a subsequence of S_n for all $n = 1, 2, 3, \ldots$. Therefore $(f_{n,n}(x_i))$ converges as $n \to \infty$ for all $x_i \in E$.

$\square$

Theorem 87 (Arzela-Ascoli Theorem (Theorem 7.25 in Rudin [1976], p. 158; simliar to Theorem 4.14 in Pugh [2015])). Let X be a compact metric space. Let $(f_n)_{n=1}^\infty$ be a sequence of functions $f_n : X \to \mathbb{R}$ (or $\mathbb{C}$). Assume

1. $\{f_n\}$ is equicontinuous, and

2. $\{f_n\}$ is pointwise bounded (for all $x \in X$, there exists M_x such that for all $n \in \mathbb{N}$, $|f_n(x)| \leq M_x$).

Then

1. $\{f_n\}$ are uniformly bounded (there exists M such that for all n and for all x, $|f_n(x)| \leq M$), and

2. $\{f_n\}$ have a uniform convergent subsequence.

Remark 8. This version is a little stronger than the version stated in Pugh [2015]. It's also a traditional version that is not maximally general.

Exercise: $\{f_n\}$ having a uniform convergent subsequence is equivalent to the sequence $(f_n)_{n=1}^\infty$ being **relatively compact** (its closure is compact) in $C^0(X, \mathbb{R})$.

Proof. 1. By equicontinuity there exists δ such that for any $x, x' \in X$ if $d(x, x') < \delta$ then $|f_n(x) - f_n(x')| < 1$ for all n. Because X is a compact metric space, by Theorem 2.63 in Pugh [2015] X has a finite subcover. That is, we can pick a finite $J \in \mathbb{N}$ such that $X \subset \bigcup_{j=1}^J B_\delta(a_j)$. Because f_n is

pointwise bounded by assumption, there exists M_j such that $|f_n(a_j)| < M_j$ for all $n \in \mathbb{N}$. Let $M := \max_{j \in \{n_1, \ldots, n_J\}} M_j$. Then for all $x \in X$,

$$|f_n(x)| \leq |f_n(x) - f_n(a_j)| + |f_n(a_j)| < M + 1$$

which is independent of x. Therefore the $\{f_n\}$ are uniformly bounded.

2. Let $A = \{a_1, \ldots, a_n, \ldots\} \subseteq X$ be a dense, countable subset of X (such a set A is guaranteed to exist by Lemma 82). By Theorem 86, (f_n) has a subsequence $(f_{n_i}) = (g_i)$ (for simplicity of notation) such that $(g_i(x))$ converges for every $x \in A$. We will show that (g_i) converges uniformly on X. Let $\epsilon > 0$. By equicontinuity there exists $\delta > 0$ such that if $d(x, x') < \delta$ then

$$|f_n(x) - f_n(x')| < \epsilon/3 \qquad \forall n. \tag{1.13}$$

Because $A = \{a_1, \ldots\}$ is dense in X, $\bigcup_{j=1}^{\infty} B_\delta(a_j)$ covers X. Because X is compact, by Theorem 2.63 in Pugh [2015] $\bigcup_{i=1}^{\infty} B_\delta(a_i)$ has a finite subcover. That is, there are finitely many points $a_1, \ldots, a_J \in A$ such that

$$X \subset \bigcup_{j=1}^{J} B_\delta(a_j).$$

Since $(g_i(x))$ converges for every $x \in A$, there exists $N \in \mathbb{N}$ such that

$$|g_m(x_j) - g_n(x_j)| < \frac{\epsilon}{3} \qquad \forall m, n \geq N, \text{ and any } j \in [J]. \tag{1.14}$$

If $x \in X$, $x \in B_\delta(x_j)$ for some j, so for every i there exists some $j \in [J]$ such that $|x - x_j| < \delta$ and therefore $|g_i(x) - g_i(x_j)| < \epsilon/3$. If $i \geq N$ and $k \geq N$, it follows that for some $j \in [J]$

$$|g_i(x) - g_k(x)| \leq \underbrace{|g_i(x) - g_i(x_j)|}_{\text{by (1.13)}} + \underbrace{|g_i(x_j) - g_k(x_j)|}_{\text{by (1.14)}} + \underbrace{|g_k(x_j) - g_k(x)|}_{\text{by (1.13)}} < \epsilon.$$

$\square$

Theorem 88 (**Arzela-Ascoli Theorem for functions into compact metric spaces**). Let X and Y be compact metric spaces. Then any equicontinuous sequence of functions $f_n : X \to Y$ has a uniformly convergent subsequence.

Proof. This follows immediately from Theorem 87. (Note that the functions must be pointwise bounded since X and Y are compact.)

$\square$

Now we will discuss some applications of Arzela-Ascoli (Theorem 87).

Theorem 89 (**Generalized Heine-Borel in a Function Space, Theorem 4.18 in Pugh [2015]**). Let X be a compact metric space. Let $\mathcal{E} \subset (C^0(X, \mathbb{R}), \|\cdot\|_\infty)$. Then $\mathcal{E}$ is compact if and only if $\mathcal{E}$ is closed, bounded, and equitcontinuous.

Proof. $\implies$: Let $\epsilon > 0$. Suppose $\mathcal{E}$ is compact. Then by Theorem 2.65 in Pugh [2015] it is closed and totally bounded, so there is a finite covering of $\mathcal{E}$ by neighborhoods in C^0 having radius $\epsilon/3$, say $B_{\epsilon/3}(f_k)$, with $k \in [n]$. So if $f \in \mathcal{E}$ then for some k we have $f \in B_{\epsilon/3}(f_k)$. Also, by equicontinuity each f_k is uniformly continuous, so there exists a $\delta > 0$ such that $|s - t| < \delta \implies \|f_k(s) - f_k(t)\|_\infty < \epsilon/3$. Then $|s - t| < \delta$ implies

$$|f(s) - f(t)| \leq |f(s) - f_k(s)| + |f_k(s) - f_k(t)| + |f_k(t) - f(t)| < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.$$

Therefore $\mathcal{E}$ is equicontinuous.

$\impliedby$: Assume that $\mathcal{E}$ is closed, bounded, and equicontinuous. If (f_n) is a sequence in $\mathcal{E}$, by the fact that $\mathcal{E}$ is bounded then (f_n) are uniformly bounded pointwise. Then by Arzela-Ascoli (Theorem 87) there exists a subsequence (f_{n_k}) that converges uniformly to a limit. Because $\mathcal{E}$ is closed, the limit lies in $\mathcal{E}$. Therefore $\mathcal{E}$ is compact (see Definition 1.4.10).

$\square$

Proposition 90 (Corollary 4.17 in Pugh [2015]). Suppose $f_n : [a, b] \to \mathbb{R}$ is a sequence of differentiable C^1 functions whose derivatives are uniformly bounded. If there exists one point $x_0 \in [a, b]$ such that the sequence $(f_n(x_0))$ is bounded as $n \to \infty$, then (f_n) has a convergent subsequence such that f_n converges uniformly to f on the whole interval $[a, b]$ and $f' = g$ **not sure about this convergence of derivative part**.

Proof. Let M be a bound for the derivatives $|f_n'(x)|$, valid for all $n \in \mathbb{N}$ and all $x \in [a, b]$. Equicontinuity of (f_n) follows from the Mean Value Theorem:

$$|s - t| < \delta \implies |f_n(s) - f_n(t)| = |f_n'(\theta)||s - t| \leq M\delta$$

for some θ between s and t. Thus given $\epsilon > 0$ the choice $\delta = \epsilon/(M + 1)$ shows that (f_n) is equicontinuous: in particular, for all $n \in \mathbb{N}$ it holds that

$$|s - t| < \frac{\epsilon}{M + 1} \implies |f_n(s) - f_n(t)| = |f_n'(\theta)||s - t| \leq M \cdot \frac{\epsilon}{M + 1} < \epsilon.$$

Let C be a bound for $|f_n(x_0)|$, valid for all $n \in \mathbb{N}$ (by assumption). Then

$$|f_n(x)| \leq |f_n(x) - f_n(x_0)| + |f_n(x_0)| \leq M|x - x_0| + C \leq M|b - a| + C$$

shows that the sequence (f_n) is bounded in C^0. Then by Arzela-Ascoli (Theorem 87) there exists a subsequence (f_{n_k}) that converges uniformly to a limit on the whole interval. **not sure about convergence of derivative**

$\square$

Proposition 91 (Math 425b Homework 5; Pugh [2015] Exercise. 4.23 (a) and (b)). Let (M, d) be a compact metric space, and let (i_n) be a sequence of isometries $i_n : M \to M$. Then there exists a subsequence i_{n_k} that converges to an isometry i as $k \to \infty$, and the space of self-isometries of M is compact.

Proof. Let $\epsilon > 0$. Then for all $n \in \mathbb{N}$ and for all $m, m' \in M$

$$d(m, m') < \epsilon \implies d(i_n(m), i_n(m')) = d(m, m') < \epsilon,$$

so $(i_n)_{n=1}^{\infty}$ is equicontinuous. Then by Theorem 88, $(i_n)_{n=1}^{\infty}$ has a subsequence $(i_{n_1}, i_{n_2}, \ldots)$ converging uniformly to a limit $i : M \to M$.

It remains to show that this limit is an isometry. By the definition of continuity in Section 2.2 of Pugh [2015], the limit of a continuous function of convergent sequences is the function of the limits of the sequences. Metrics are continuous by Theorem 2.20 in Pugh [2015]. So a subsequence $(d(i_{n_1}(m), i_{n_1}(m')), d(i_{n_2}(m), i_{n_2}(m')), \ldots)$ converges to $d(i(m), i(m'))$. The sequence itself can be written

$$(d(i_{n_1}(m), i_{n_1}(m')), d(i_{n_2}(m), i_{n_2}(m')), \ldots) = (d(m, m'), d(m, m'), \ldots)$$

since each i_n is an isometry. This constant sequence converges to $d(m, m')$, so $d(i(m), i(m')) = d(m, m')$ and i is an isometry.

Because we have shown that every sequence of isometries (i_n) from M to M has a subsequence that converges to an isometry from M to M, we have shown that the space of self-isometries of M is compact.

$\square$

1.6.4   Uniform Approximation in C^0 (Section 4.4 of Pugh [2015])

Definition 1.6.30 (Dense sets of functions; special case of Definition 1.6.29). Let X be a set of functions. X is **dense** in $C^0([a, b], \mathbb{R})$ if for each $f \in C^0$ and each $\epsilon > 0$ there exists a function $p \in X$ such that for all $x \in [a, b]$, $|f(x) - p(x)| < \epsilon$.

Theorem 92 (Weierstrass Approximation Theorem; Thereom 4.19 in Pugh [2015]). The set of polynomials is dense in $C^0([a, b], \mathbb{R})$.

(Alternatively: Let $f \in C^0([0, 1], \mathbb{R})$ (or $C^0([0, 1], \mathbb{C})$). Then the polynomials $p_n(x) = \sum_{k=0}^{n} f(k/n) f_k(x)$ converge uniformly to f on $[0, 1]$.)

Remark 9. See p. 229 of Pugh [2015] (p. 239 of pdf), Proof #1.

We will sketch the proof that the result generalizes from $[0, 1]$ to $[a, b]$. We have isomorphisms (mappings between two structures of the same type that can be reversed by inverse mappings; see Definition 1.4.13) of $\mathbb{R}$-vector spaces $C^0([0, 1], \mathbb{R})$ to and from $C^0([a, b], \mathbb{R})$. Suppose we have $f \in C^0([0, 1], \mathbb{R})$; we can send it to $x \mapsto f([x - a]/[b - a])$ in $C^0([a, b], \mathbb{R})$. Conversely, we can send a function $g \in C^0([a, b], \mathbb{R})$ to $C^0([0, 1], \mathbb{R})$ via the map $x \mapsto g(a + x[b - a])$. It can be checked that these isomorphisms are linear, compose to the identity in either order, and preserve $\|\cdot\|_{\infty}$. They also send polynomials to polynomials.

First we will show a lemma.

Lemma 93. $x(1 - x) \leq 1/4$ for all $x \in \mathbb{R}$.

Proof. $f(x) = x(1-x) = -x^2 + x$ is differentiable; $f'(x) = -2x + 1 = 0 \iff x = 1/2$. We see that f' is continuous on $\mathbb{R}$ and $f'(x) > 0$ for $x \in (-\infty, 1/2)$ and $f'(x) < 0$ for $x \in (1/2, \infty)$, so $f(x)$ has a global maximum at $x = 1/2$; that is, $x(1-x) \leq f(1/2) = 1/2(1 - 1/2) = 1/4$ for all $x \in \mathbb{R}$.

$\square$

Proof of Weierstrass Approximation Theorem (Theorem 92). For a given $n \geq 1$ and for $k \in \{0, \ldots, n\}$, let r_k be the Bernstein basis polynomial

$$r_k(x) = \binom{n}{k} x^k (1-x)^{n-k}.$$

Define

$$p_n(x) := \sum_{k=0}^{n} f(k/n) r_k(x).$$

We want to show that the polynomial functions p_n converge uniformly to f on $[0,1]$. Let $\epsilon > 0$. Note that f is uniformly continuous and bounded because it is a continuous function on a compact set. So, there exists $\delta > 0$ such that for any $x, y \in [0,1]$,

$$|x - y| < \delta \implies |f(x) - f(y)| < \epsilon/2, \tag{1.15}$$

and there exists $M \in (0, \infty)$ such that

$$|f(x)| \leq M \tag{1.16}$$

for all $x \in [0,1]$. Let $N \geq M/(\epsilon\delta^2)$. We want to show that for any $n \geq N$, $|p_n(x) - f(x)| < \epsilon$ for all $x \in [0,1]$.

Fix $x \in [0,1]$. Since $\sum_{k=0}^{n} r_k(x) = 1$, we have $f(x) = f(x)\sum_{k=0}^{n} r_k(x) = \sum_{k=0}^{n} f(x) r_k(x)$. Then

$$|p_n(x) - f(x)| = \left| \sum_{k=0}^{n} f(k/n) r_k(x) - \sum_{k=0}^{n} f(x) r_k(x) \right| = \left| \sum_{k=0}^{n} [f(k/n) - f(x)] \, r_k(x) \right|. \tag{1.17}$$

Let $K_1 := \{k \in \{0, \ldots, n\} : |k/n - x| < \delta\}$, and let $K_2 := \{0, \ldots, n\} \setminus K_1$. Observe that for all $k \in K_1$, $|f(k/n) - f(x)| < \epsilon/2$ by (1.15), so

$$\left| \sum_{k \in K_1} [f(k/n) - f(x)] \, r_k(x) \right| \leq \sum_{k=0}^{n} |[f(k/n) - f(x)] \, r_k(x)| < \sum_{k=0}^{n} \frac{\epsilon}{2} |r_k(x)| = \frac{\epsilon}{2} \sum_{k=0}^{n} r_k(x) = \frac{\epsilon}{2} \tag{1.18}$$

where we used $r_k(x) \geq 0$ for all k. Next, by (1.16) $|f(k/n) - f(x)| \leq |f(k/n)| + |f(x)| \leq 2M \leq 2\epsilon\delta^2 n$ (using the definition of N and $n \geq N$), so

$$\left| \sum_{k \in K_2} [f(k/n) - f(x)] \, r_k(x) \right| \le \sum_{k \in K_2} 2\epsilon\delta^2 n r_k(x).$$

Since $k \in K_2 \iff |k/n - x| \ge \delta \iff |k - nx| \ge \delta n \implies (k - nx)^2 \ge (n\delta)^2$, we have (using $\sum_{k=1}^{n} (k - nx)^2 r_k(x) = nx(1-x)$)

$$\sum_{k \in K_2} 2\epsilon\delta^2 n r_k(x) = \frac{2\epsilon}{n} \sum_{k \in K_2} (n\delta)^2 r_k(x) \le \frac{2\epsilon}{n} \sum_{k=1}^{n} (k - nx)^2 r_k(x) = \frac{2\epsilon}{n} nx(1-x) \le 2\epsilon \cdot \frac{1}{4} = \frac{\epsilon}{2}$$

where we applied Lemma 93 in the last step, yielding

$$\left| \sum_{k \in K_2} [f(k/n) - f(x)] \, r_k(x) \right| \le \frac{\epsilon}{2}. \tag{1.19}$$

Therefore using (1.18) and (1.19) we can write (1.17) as

$$\begin{aligned}
|p_n(x) - f(x)| &= \left| \sum_{k \in K_1} [f(k/n) - f(x)] \, r_k(x) + \sum_{k \in K_2} [f(k/n) - f(x)] \, r_k(x) \right| \\
&\le \left| \sum_{k \in K_1} [f(k/n) - f(x)] \, r_k(x) \right| + \left| \sum_{k \in K_2} [f(k/n) - f(x)] \, r_k(x) \right| \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon.
\end{aligned}$$

$\square$

Definition 1.6.31 (**Definition 7.28 in** Rudin [1976]**, p. 170 of pdf, p.161**)**.** A family A of complex functions defined on a set X is a **function algebra** if A is closed under addition, multiplication, and scalar multiplication (and does not necessarily include a multiplicative identity). That is, if $f, g \in A$ and $c \in \mathbb{C}$, then $f + g \in A$, $fg \in A$, and $cf \in A$. (We also consider algebras of real functions; in this case, the conditions only must hold for all $c \in \mathbb{R}$.)

If A has the property that $f \in A$ whenever $f_n \in A$ for $n \in \mathbb{N}$ and f_n converges uniformly to f on X, then A is said to be **uniformly closed.**

Let B be the set of all functions that are limits of uniformly convergent sequences of members of A. Then B is called the **uniform closure** of A.

In-class definition: a "non-unital subalgebra of $C^0(X, \mathbb{R})$". Also an $\mathbb{R}$-algebra: a ring (see Definition **??**) that's also an $\mathbb{R}$ vector space; structures are compatible. e.g. $C^0(X, \mathbb{R})$ is an $\mathbb{R}$-algebra. A sub-algebra is closed under addition, scalar multiplication, and ring multiplication, and contains a scalar multiplicative identity. A non-unital sub-algebra is the same but it might not contain a multiplicative identity. (Note that the multplicative identity for $C^0(X, \mathbb{R})$ is the constant function $f(x) = 1$.)

Example 1.6.7. The set of all polynomials is an algebra. The Weierstrass Approximation Theorem may be stated by saying that the set of continuous functions on $[a, b]$ is the uniform closure of the set of polynomials on $[a, b]$.

Definition 1.6.32 (Definition 7.30 in Rudin [1976], p. 171 of pdf, p.162). Let A be a family of functions on a set X. A **vanishes nowhere** if $\forall x \in X$ there exists $f \in A$ with $f(x) \neq 0$. (That is, there is no $x \in X$ with $f(x) = 0$ for all $f \in A$.)

Definition 1.6.33 (Definition 7.30 in Rudin [1976], p. 171 of pdf, p.162). Let A be a family of functions on a set X. A **separates points** on X if $\forall x \neq x' \in X$ there exists $f \in A$ with $f(x) \neq f(x')$.

Example 1.6.8. The algebra of all polynomials in one variable clearly vanishes nowhere and separates points on $\mathbb{R}$. An example of an algebra that does not separate points is the set of all even polynomials, say on $[-1, 1]$ since $f(-x) = f(x)$ for every even function f.

Theorem 94 (Stone-Weierstrass Theorem (Theorem 4.20 in Pugh [2015], Theorem 7.32 in Rudin [1976], p. 171 of pdf, p.162). Let X be a compact metric space. Let $A \subset C^0(X, \mathbb{R})$ or $C^0(X, \mathbb{C})$. Then if A is a function algebra that vanishes nowhere in X and A separates points on X, then A is dense in $(C^0(X, \mathbb{R}), \|\cdot\|_\infty)$.

Equivalent conclusion: $\overline{A} = C^0(X, \mathbb{R})$ in uniform metric. That is, the uniform closure B of A consists of all real continuous functions on X.

Remark 10. The Weierstrass Approximation Theorem (Theorem 92) is a special case of the Stone-Weierstrass Theorem.

Theorem 95 (Complex Stone-Weierstrass Theorem; similar to Theorem 7.33 in Rudin [1976], p. 174 of pdf, p.165). If $A \subset C^0(X, \mathbb{C})$ is a function algebra that is closed under complex conjugation, A vanishes nowhere, and A separates points, then A is dense in $(C^0(X, \mathbb{C}), \|\cdot\|_\infty)$ (that is, $\overline{A} = C^0(X, \mathbb{C})$.

Remark 11. Why is the complex conjugate assumption needed? Let $X := \{|z| \leq 1 | z \in \mathbb{C}\}$. Note that X is compact. Let A be the space of complex polynomials; that is, $A = \text{span}_\mathbb{C}(1, z, z^2, \ldots)$. It can be shown that A is a function algebra, separates points, and vanishes nowhere. Let $f(z) = \overline{z}$; this is not holomorphic. If $p_n \in A$ and p_n converges uniformly to f on X, then p_n converges uniformly to f on $\{|z| < 1. | z \in \mathbb{C}\}$. Thus p_n (which are holomorphic) converges uniformly to f on any compact $K \subset \{|z| = 1 \mid z \in \mathbb{C}\}$. This implies (by Theorem 69) that f is holomorphic, contradiction. So we need A to be closed under complex conjugation.

Now we will prove Theorem 94. We will need to use the Weierstrass Approximation Theorem (Theorem 92) for $f(x) = |x|$ on $[-1, 1]$ (could prove what we need direcdtly for this function, but we might as well use the result since we have it). We will show some lemmas first

Lemma 96 (Lemma 4.22 in Pugh [2015]; Similar to Theorem 7.29 in Rudin [1976], p. 170 of pdf, p.161). If X is compact and $A \subset C^0(X, \mathbb{R})$ is a function algebra, then so is its closure $\overline{A}$.

Proof. We will show $\overline{A}$ is closed under addition, scalar multiplication, and function multiplication. (We will prove that it is true for addition; the other two cases are similar.) Let $f, g \in \overline{A}$. We have $f_n \in A$ with f_n converging uniformly to f and $g_n \in A$ with g_n converging uniformly to g. A is closed under addition by assumption, so $f_n + g_n \in A$. The limits of nets are linear (exercise), so we have that $f_n + g_n$ converges uniformly to $f + g$. Thus $f + g$ is a uniform limit of elements of A, so $f + g \in \overline{A}$.

The proof of closure under scalar multiplication and function multiplication are similar.

$\square$

Lemma 97. Let $f \in C^0([a,b], \mathbb{R})$ and let $x_0 \in [a,b]$ with $f(x_0) = 0$. Given $\epsilon > 0$, there exists a polynomial p on $[a,b]$ with $p(x_0) = 0$ and $|f(x) - p(x)| < \epsilon$ for all $x \in [a,b]$.

Proof. Given $\epsilon > 0$, choose a polynomial q such that $|f(x) - q(x)| < \epsilon/2$ for all $x \in [a,b]$. q might not vanish at x_0, but $p := q - q(x_0)$ is a polynomial that vanishes at x_0 by construction. Since $|q(x) - f(x)| < \epsilon/2$ for all x, plug in x_0 and get $|q(x_0)| M \epsilon/2$. Now, for $x \in [a,b]$, we have

$$|p(x) - f(x)| = |q(x) - q(x_0) - f(x)| \leq |q(x) - f(x)| + |q(x_0)| < \epsilon/2 + \epsilon/2 = \epsilon.$$

$\square$

Lemma 98 (step 1 of proof of Theorem 7.32 in Rudin [1976], p. 171 of pdf, p.162). If $A \subset C^0(X, \mathbb{R})$ is a function algebra and $f \in \overline{A}$ (the closure of A) then $|f| \in \overline{A}$.

Proof. Consider the absolute value function $|\cdot| : [-\|f\|_\infty, \|f\|_\infty]$. (Note that $[-\|f\|_\infty, \|f\|_\infty$ is a compact subset of $\mathbb{R}$.) The function $|\cdot| : [-\|f\|_\infty, \|f\|_\infty] \to \mathbb{R}$ is continuous since $|\cdot| : \mathbb{R} \to \mathbb{R}$ is continuous. We can factor $|f| : X \to \mathbb{R}$ as $X \xrightarrow{f} [-\|f\|_\infty, \|f\|_\infty] \xrightarrow{|\cdot|} \mathbb{R}$. The function $|\cdot| : [-\|f\|_\infty, \|f\|_\infty] \to \mathbb{R}$ vanishes at $x = 0$. So, given $\epsilon > 0$, we can choose a polynomial p such that $p(0) = 0$ and $|p(x) - |x|| < \epsilon$ for $x \in [-\|f\|_\infty, \|f\|_\infty]$. We can write $p(x) = c_1 x + \ldots + c_k x^k$ (no constant term since $p(0) = 0$). The function $g : X \xrightarrow{f} [-\|f\|_\infty, \|f\|_\infty] \xrightarrow{p} \mathbb{R}$ is given by $g(x) = p(f(x)) = \sum_{i=1}^{k} c_i f^i(x)$. So $g = \sum_{i=1}^{k} c_i f^i$. Since $f \in \overline{A}$ and $\overline{A}$ is a function algebra, we have $g \in \overline{A}$. We want to show that $\big||f(x)| - g(x)\big| < \epsilon$ for all $x \in X$. Indeed, for $x \in X$ we have

$$\big||f(x)| - g(x)\big| = \big||f(x)| - p(f(x))\big| = \big||y| - p(y)\big|$$

where $y = f(x) \in [-\|f\|_\infty, \|f\|_\infty]$. But we chose p such that $\big||y| - p(y)\big| < \epsilon$ for all $y \in [-\|f\|_\infty, \|f\|_\infty]$. So $\big||f(x)| - g(x)\big| < \epsilon$ as desired.

We've shown that if $f \in \overline{A}$, then there exists $g \in \overline{A}$ with $d_\infty(|f|, g) < \epsilon$. This eimplies there exists a seqqence of functions in $\overline{A}$ converging uniformly to $|f|$, so $|f| = \overline{\overline{A}} = \overline{A}$. So we've proven that if $f \in \overline{A}$ then $|f| \in \overline{A}$.

$\square$

Corollary 98.1 (step 2 of proof of Theorem 7.32 in Rudin [1976], p. 171 of pdf, p.162). If $A \in C^0(X, \mathbb{R})$ is a function algebra and $f, g \in \overline{A}$ then $\max\{f, g\} \in \overline{A}$ (the pointwise max). (Similarly, $\min\{f, g\} \in \overline{A}$.) (closed under maxes and mins—it's a **lattice**.)

Proof. Observe that

$$\max\{f, g\} = \frac{f + g}{2} + \frac{|f - g|}{2}.$$

Each of these functions is in $\overline{A}$ (since the absolute value function is by Lemma 98) since $\overline{A}$ is a function algebra, and their sum is as well. Similarly,

$$\min\{f, g\} = \frac{f+g}{2} - \frac{|f-g|}{2} \in \overline{A}.$$

$\square$

Lemma 99 (**Theorem 7.31 in Rudin [1976], p. 171 of pdf, p.162; Lemma 4.21 in Pugh [2015]**).
Suppose $A \in C^0(X, \mathbb{R})$ satisfies all the hypotheses of the Stone-Weierstrass Theorem. Let $x_1 \neq x_2 \in X$ and let $c_1, c_2 \in \mathbb{R}$. Then there exists $f \in A$ such that $f(x_1) = c_1$ and $f(x_2) = c_2$. (Can specify values at any two points of X; stronger condition than separating points. Some map from A to $\mathbb{R}^2$ is surjective.)

Proof. Consider the linear transformations $\Phi : A \to \mathbb{R}^2$ defined by $\Phi(f) = (f(x_1), f(x_2))$, where A is a function algebra, so a vector subspace of $C^0(X, \mathbb{R})$. Check that Φ is linear:

$$\Phi(f+g) = ((f+g)(x_1), (f+g)(x_2)) = (f(x_1)+g(x_1), f(x_2)+g(x_2)) = (f(x_1), f(x_2))+(g(x_1), g(x_2)) = \Phi(f)+\Phi(g).$$

Similarly, $\Phi(cf) = c\phi(f)$ for $c \in \mathbb{R}$. The image of Φ must be a vector subapce of $\mathbb{R}^2$, so it must have dimension0, 1, or 2. We will show that the dimension is 2, which is true if and only if Φ is surjective, which is equivalent to the lemma holding. We want to show that the rank$(\Phi) = 2$ (the dimension of the image), not 0 or 1.

First we will show rank$(\Phi) \neq 0$. If rank$(\Phi) = 0$, then Φ is the 0 map; that is, for any $f \in A$, $\Phi(f) = (f(x_1), f(x_2)) = (0, 0)$. But this contradicts that f vanishes nowhere.

Next suppose rank$(\Phi) = 1$. Then the image of Φ is a line through the origin in $\mathbb{R}^2$. There are three possible cases:

- **Case 1:** This line is the $x - axis$. Then for all $f \in A$, $\Phi(f) = (f(x_1), f(x_2)) = (k, 0)$ for some $k \in \mathbb{R}$. But this contradicts that A vanishes nowhere.

- **Case 2:** This line is the $y - axis$. This leads to a similar contradiction.

- **Case 2:** The line is some other line. Then for $f \in A$, we have $f(x_1) = 0 \iff f(x_2) = 0$. A separates points means there exists $f \in A$ with $f(x_1) \neq f(x_2)$ (so neither $f(x_1)$ nor $f(x_2)$ is zero). Idea: consider $f - f(x_1)$. If this function is in A, then it vanishes at x_1 (by construction) but not at x_2 (because $f(x_1) \neq f(x_2)$). But why is $f - f(x_1) \in A$? ($f(x_1)$ is a constant function—it might not be in A.) Consider $f \cdot (f - f(x_1)) = f^2 - f \cdot f(x_1) \in A$ because A is closed under multiplication, subtraction, and scalar multiplication. Let $g := f \cdot (f - f(x_1)) \in A$. We have $g(x_1) = 0$ by construction; however,

$$g(x_2) = \underbrace{f(x_2)}_{\neq 0} \cdot \underbrace{(f(x_2) - f(x_1))}_{\neq 0}.$$

This contradicts what we showed earlier: for $f \in A$ it holds that $f(x_1) = 0 \iff f(x_2) = 0$. Thus we have ruled out all possibilities except rank$(\Phi) = 2$; that is, for all $(c_1, c_2) \in \mathbb{R}^2$ there exists $f \in A$ with $(f(x_1), f(x_2)) = (c_1, c_2)$.

$\square$

Now we are ready to prove the Stone-Weierstrass Theorem.

Proof of Theorem 94. Let $F \in C^0(X, \mathbb{R})$. Let $\epsilon > 0$. It is sufficient to find $G \in \overline{A}$ such that $\|F - G\|_\infty < \epsilon$; that is, $|F(x) - G(x)| < \epsilon$ for all $x \in X$. If this is true for all ϵ, then $F \in \overline{\overline{A}} = \overline{A}$ (because it's the uniform limit of elements of $\overline{A}$).

For any two points $p \neq q \in X$, by Lemma 99 there exist $H_{p,q} \in A$ such that $H_{p,q}(p) = F(p)$ and $H_{p,q}(q) = F(q)$.

Also pick $H_{p,p}$ for each $p \in X$ such that $H_{p,p}(p) = F(p)$.

Claim: Given a fixed p, consider $H_{p,q}$ for various q. In an open neighborhood of q, we have $H_{p,q}(x) > F(x) - \epsilon$. To see why, consider the continuous function $x \mapsto H_{p,q}(x) - F(x) + \epsilon$. This function takes value $\epsilon > 0$ at $x = q$. Thus, there exists an open neighborhood $U_{p,q}$ of q such that $H_{p,q} - F + \epsilon > 0$ on $U_{p,q}$; i.e., $H_{p,q}(x) > F(x) - \epsilon$ on $U_{p,q}$. For our fixed p, the set $\{U_{p,q} : q \in X\}$ is an open cover of X. (Important for this to work: we picked $H_{p,p}$.) Since X is compact, there exists a finite subcover; that is, for some $x \in \mathbb{N}$ there exist $q_1, \ldots, q_k \in X$ with $X = \bigcup_{i=1}^k U_{p,q_i}$. Define $G_p : X \to \mathbb{R} := \max_{i \in [k]} \{H_{p,q_i}\}$. Note that $G_p \in \overline{A}$ by Corollary 98.1.

So for any $x \in X$, $x \in U_{p,q_i}$ for some $i \in [k]$, which means $G_p(x) \geq H_{p,q_i}(x) > F(x) - \epsilon$. So $G_p > F - \epsilon$ everywhere. Now run the same argument for $\{G_p\}$. We claim that on an open neighborhood V_p of p, we have $G_p < F + \epsilon$; i.e., $G_p(x) < F(x) + \epsilon$ for all $x \in V_p$. To see why, observe that the function $x \mapsto F(x) + \epsilon - G_p(x)$ is continuous and takes the value $\epsilon > 0$ at $x = p$, so there exists V_p as desired.

The family of sets $\{V_p\}_{p \in X}$ is an open cover of X. X is compact, so there exists a finite subcover $X = \bigcup_{i=1}^\ell V_{p_i}$. Let $G := \min_{i \in [\ell]} \{G_{p_i}\}$. (Again, $G \in \overline{A}$ by Corollary 98.1.) Any point $x \in X$ is in some V_p, so $G(x) \leq G_{p_i}(x) < F(x) + \epsilon$. Also, $G(x) = G_{p_i}(x)$ for some i, and $G_{p_i}(x) > F(x) - \epsilon$ by construction of G_p. Thus for all $x \in X$, we have $F(x) - \epsilon < G(x) < F(x) + \epsilon$.

$\square$

Proof of Theorem 95. Let $A_R := \{$real and imaginary parts of functions $G \in A\} \subset C^0(X, \mathbb{R})$. (Note: if $G \in A$ then $\mathrm{Re}(G) = \frac{G + \overline{G}}{2} \in A$, $\mathrm{Im}(G) = \frac{G - \overline{G}}{2i} \in A$.)

So $A_R \subset A \cap C^0(X, \mathbb{R})$. It's clear that if $G \in A \cap C^0(X, \mathbb{R})$ then $G = \mathrm{Re}(G)$ so $G \in A_R$. Therefore $A \cap C^0(X, \mathbb{R})$, so we have $A_R = A \cap C^0(X, \mathbb{R})$. We want to show that Theorem 94 applies to A_R.

First, observe that A_R is a function algebra since $A \subset C^0(X, \mathbb{C})$ and $\subset C^0(X, \mathbb{R}) \subset C^0(X, \mathbb{C})$ are closed under addition, scalar multiplication, and function multiplication.

Next, A_R vanishes nowhere because if there exists $x \in X$ with $(\mathrm{Re}\, G)(x) = 0$, $(\mathrm{Im}\, G)(x) = 0$ for all $G \in A$. therefore $G(x) = 0$ for all $G \in A$, so A vanishes nowhere.

Next, A_R separates points: we know that if $x_1 \neq x_2 \in X$,. then there exists $G \in A$ with $G(x_1) \neq G(x_2)$. This implies that either $(\mathrm{Re}\, G)(x_1) \neq (\mathrm{Re}\, G)(x_2)$ (and note that $(\mathrm{Re}\, G) \in A_R$) or $(\mathrm{Im}\, G)(x_1) \neq (\mathrm{Im}\, G)(x_2)$ (and note that $(\mathrm{Im}\, G) \in A_R$). Thus, A_R separates points.

Now let $F \in C^0(X, \mathbb{C})$; write $F = u + iv$ for $u, v \in C^0(X, \mathbb{R})$. Given $\epsilon > 0$, there exists $G_1 \in A_R$ with $|U(x) - G_1(x)| < \epsilon/2$ for all $x \in X$. Also, there exists $G_2 \in A_R$ with $|V(x) - G_2(x)| < \epsilon/2$ for all $x \in X$.

Now: $G_1 \in A_R \subset A$, $G_2 \in A_R \subset A$, so $G_1 + iG_2 \in A$, and for all $x \in X$,

$$\left| \underbrace{F(x)}_{U(x)+iV(x)} - (G_1 + iG_2)(x) \right| \leq U(x) - G_1(x)| + |V(x) - G_2(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

$\square$

Now we will discuss some applications of Theorem 94.

Example 1.6.9 (**Example on p. 239 of Pugh [2015] (p. 249 of pdf)**). Let D^2 be the closed unit disk in $\mathbb{R}^2$. Let $F : D^2 \to \mathbb{R}^2$ be continuous. We want $G : D^2 \to \mathbb{R}^2$ approximating F uniformly, such that $G(x) = 0$ for only finitely many x.

Proof. Let $A \subset C^0(D^2, \mathbb{R})$ be the set of two-variable polynomials; known as $\mathbb{R}[x, y]$. We can also write this as $\text{span}_{\mathbb{R}}\{x^i y^j \mid i, j \in \mathbb{Z}_+\}$. Since A is a span, it is closed under addition and scalar multiplication. Also, $(x^i y^j)(x^k y^\ell) = x^{i+k} y^{j+\ell}$. Apply this to linear combinations, then A is closed under function multiplication. So A is a function algebra.

Now check the assumptions of Theorem 94. A vanishes nowhere since it contains constant functions. We will check if A vanishes nowhere. If $(x_1, y_1) \neq (x_2, y_2) \in D^2$, then $(x - x_1) \in A$ vanishes at (x_1, y_1) but not (x_2, y_2). Otherwise if $y_1 \neq y_2$ then $(y - y_1) \in A$ vanishes at (x_1, y_1) but not (x_2, y_2). Therefore we can apply Theorem 94 on A.

Write $F(x) = (F_1(x), F_2(x))$, with $F_i \in C^0(D^2, \mathbb{R})$. Then there exists $p_1 \in A$ with $\|p_1 - F_1\|_\infty < \epsilon/\sqrt{2}$, and there exists $p_2 \in A$ with $\|p_2 - F_2\|_\infty < \epsilon/\sqrt{2}$. This implies that $(p_1, p_2) : D^2 \to \mathbb{R}^2$ satisfies $\|(p_1, p_2) - F\|_\infty < \sqrt{\epsilon^2/2 + \epsilon^2/2} = \epsilon$.

We can assume without loss of generality that the coordinate functions $F_1(x), F_2(x)$ are polynomials. So if F_1, F_2 jointly vanish at a finite number of points, we're done. Given the same ϵ as before, $(f_1 + \delta, F_2)$ is ϵ-close to F_1, F_2 in $\|\cdot\|_\infty$ as long as $\delta < \epsilon$. This is true for infinitely many δ. Let $F_\delta = F_1 + \delta$.

$\mathbb{R}[x, y]$ is a UFD (**unique factorization domain**); thus, F_δ has a unique factorization into irreducible polynomials over $\mathbb{R}$. Similarly, F_2 has a unique factorization into irreducible polynomials over $\mathbb{R}$. And crucially, if $\delta \neq \delta'$ then F_δ and $F_{\delta'}$ have no common factors of positive degree in $\mathbb{R}[x, y]$ (because if they did then the constant function $\delta - \delta'$ would have polynomial factors of degree greater than 0 which is a contradiction of the basic algebraic properties of this ring (see Definition **??**). F_2 has finitely many factors, so they are factors of finitely many F_δs, so there exists $\delta < \epsilon$ with $F_1 + \delta$ and F_2 having no common factors.

A fact from algebra is that given $P, Q \in \mathbb{R}[x, y]$, can compute the resultant $\text{Res}(P, Q) \in \mathbb{R}[x]$, which is nonzero if P and Q have no common factors. If $\text{Res}(P, Q)(x) \neq 0$ then $P(x, \cdot)$ and $Q(x, \cdot)$ have no common roots y. It is true that $\text{Res}(P, Q)(x) \neq 0$ for all but finitely many x since $\text{Res}(P, Q)(x)$ is a polynomial. For given x with $\text{Res}(P, Q)(x) = 0$, consider $P(X, \cdot)$ and $Q(X, \cdot)$ as polynomials in $\mathbb{R}[y]$. They each have finitely many roots, so there exist finitely many (x, y) with $P(x.y) = Q(x, y) = 0$. This finishes the application.

$\square$

1.6.5   Contractions and ODEs (Section 4.5 of Pugh [2015])

Definition 1.6.34. Let X be a set, let $f : X \to X$ be a function. $x \in X$ is called a **fixed point** of f if $f(x) = x$.

Definition 1.6.35 (**Contractions; p. 240 of Pugh [2015]; Definition 9.22 in Rudin [1976], p. 229 of pdf, p. 220**). Let M be a metric space. A **contraction** of M is a mapping $f : M \to M$ such that for some constant $k < 1$ and for all $x, y \in M$ we have

$$d(f(x), f(y)) \leq k d(x, y);$$

i.e., f has a Lipschitz constant less than 1. (See also Definition **??**. For more on contraction mappings, see Section **??**.)

Note: f is a "weak contraction" if $d(f(x), f(y)) < d(x, y)$ for all $x, y \in X$. (Doesn't necessarily imply that f is a contraction mapping; we will se an example later.)

example: let $X = [1, \infty)$. Since $X \subset \mathbb{R}$, X is complete. Define $f : X \to X$ by $f(x) = x + 1/x$. f has no fixed point ($x + 1/x > x$ for $x \in X$). f i a weak contraction, i.e., for $x_1 \neq x_2 \in X$, we have $|f(x_1) - f(x_2)| < |x_1 - x_2|$. See why: if $x_1 > x_2 \geq 1$ then

$$|(x_1 + 1/x_1) - (x_2 + 1/x_2)| = |x_1 - x_2 + (1/x_1 - 1/x_2)|$$

Note that $1/x_2 - 1/x_1 \leq x_1 - x_2$ (equivalently, $x_1 - x_2 \leq x_1 x_2 (x_1 - x_2)$; equivalently, $1 \leq x_2 x_2$ since $x_1, x_2 \geq 1$. So f is a weak contraction but has no fixed point.

Theorem 100 (**Banach contraction principle, Banach 1922; Theorem 4.24 in Pugh [2015], p. 240; Theorem 9.23 in Rudin [1976], p. 229 of pdf, p. 220**). Let (X, d) be a complete metric space. Let $F : X \to X$ be a contraction. Then there exists a unique fixed point $x_0 \in X$ of f, and for any $x \in X$, the sequence of iterates $(f^n(x))_{n=1}^{\infty}$ (where $f^n(x) = f \circ f \circ \ldots \circ f(x)$) converges to x_0.

Before proving Theorem 100, we will discuss one other result and lay out a sketch of a proof (the full proof requires graduate-level algebraic topology).

Theorem 101 (**Brouwer Fixed-Point Theorem; Theorem 4.24 in Pugh [2015], p. 240**). Any continuous map $f : D^m \to D^n$ has at least one fixed point (where D^m is the closed unit ball in $\mathbb{R}^m$).

Proof. First begin with $n = 1$. Then $f : [-1, 1] \to [-1, 1]$ is continuous. Let $g(x) := f(x) - x$. Note that $g(-1) \geq 0$, $g(1) \leq 0$. So by the intermediate value theorem, there exists $x \in [-1, 1]$ with $g(x) = 0$; i.e., $f(x) = x$.

For a general n, if $f : D^n \to D^n$ has no fixed point, then define $g : D^n \mapsto \partial(D^n) = S^{n-1}$ by $g(x) = \frac{f(x) - x}{\|f(x) - x\|}$. Then $g|_{S^n} : S^{n-1} \to S^{n-1}$ induces a map on $(n-1)$th homology group. $\ldots$ (requires graduate level algebraic topology). $\square$

Now we will prove Theorem 100.

Proof of Theorem 100. Let (X, d) be a complete metric space and let $f : X \to X$ be a contraction with $k < 1$. First we will show that any fixed point must be unique.

If $x_1 \neq x_2 \in X$ and $f(x_1) = x_1$ and $f(x_2) = x_2$, then because f is a contraction mapping, $d(f(x_1), f(x_2)) \leq kd(x_1, x_2) < d(x_1, x_2)$. However, because $f(x_1) = f(x_2)$, it must hold that $d(f(x_1), f(x_2)) = d(x_1, x_2)$, contradiction.

To prove the rest of the theorem, start with some $x \in X$ and show that (1) the sequence $(f^n(x))_{n=1}^\infty$ converges to something and (2) the limit is a fixed point of f. Since X is complete, to show (1) it's enough to show that $(f^n(x))_{n=1}^\infty$ is Cauchy. Write $x_n := f^n(x)$ for notational convenience. We want to show that $(x_n)_{n=1}^\infty$ is Cauchy. We claim that $d(x_n, x_{n+1}) \leq k^n d(x_0, x_1)$ for all n. To see this, observe that $d(x_n, x_{n+1}) = d(f(x_{n-1}), f(x_n)) \leq kd(x_{n-1}, x_n) \leq k \cdot k^{n-2} d(x_1, x_2)$, where the last step follows by induction.

Now: $\sum_{k=1}^\infty k^n$ converges because $k < 1$, so this is a geometric series. Therefore the sequence of partial sums converges, so it is Cauchy. Given $\epsilon > 0$, choose N such that if $m, \ell \geq N$ then $\sum_{n=m}^{\ell-1} k^n < \epsilon / d(x_0, x_1)$. Then if $\ell > m \geq N$, we have

$$
\begin{aligned}
d(x_m, x_\ell) &\leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \ldots + d(x_{\ell-1}, x_\ell) \\
&\leq k^m d(x_0, x_1) + k^{m+1} d(x_0, x_1) + \ldots + k^{\ell-1} d(x_0, x_1) \\
&= \left(\sum_{n=m}^{\ell-1} k^n \right) d(x_0, x_1) \\
&< \frac{\epsilon}{d(x_0, x_1)} \cdot d(x_0, x_1) = \epsilon.
\end{aligned}
$$

So $(x_n)_{n=0}^\infty$ is a Cauchy sequence in X, so it converges to some point $x' \in X$. Now we will show that x' is a fixed point of f. We have $(x_n) \to x'$. f is continuous, so it sends convergent sequences to convergent sequences. That is, $(f(x_n))$ converges to $f(x')$, which can be written as (x_{n+1}) converges to $f(x')$. This is the same sequence, so x_n converges to x' and also $f(x')$. By uniqueness of limits, $f(x') = x'$, so x' is a (the) fixed point of f.

$\square$

Next we will discuss some basic results on ordinary differential equations (ODEs). (For more on differential equations, see Section **??** of these notes.) Picard's Theorem gives us the existence and uniqueness of solutions to initial value problems for systems of first order ODEs given a Lipschitz continuity assumption.

Also consider systems of first order ODEs. In particular, we can phrase a second order ODE as a system of first order ODEs. For example, suppose we have $x'' - 4x' + 3x = 0$. Let $u(t) = x(t), v(t) = x'(t)$. Then we have

$$
u'(t) = v(t), \qquad v'(t) - 4v(t) + 3u(t) = 0.
$$

Thus given a solution $x(t)$ to $x'' - 4x' + 3x = 0$, we get $u(t), v(t)$ satisfying

$$
\begin{bmatrix} u \\ v \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ 4 & -3 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \tag{1.20}
$$

Conversely, given (u, v) solving this system, then $x(t) := u(t)$ solves the original system. So, existence/uniqueness of solutions for first order systems implies existence/uniqueness for higher order systems too.

What is a first order system in general? (maybe nonlinear, etc.)

Example 1.6.10.

$$\begin{bmatrix} u \\ v \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ 4 & -3 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}.$$

Call

$$\gamma(t) := \begin{bmatrix} u \\ v \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ 4 & -3 \end{bmatrix},$$

a path in $\mathbb{R}^2$. We can write

$$\gamma'(t) = \begin{bmatrix} 0 & 1 \\ 4 & -3 \end{bmatrix} \gamma(t).$$

For $w \in \mathbb{R}^2$, let

$$F(w) = \begin{bmatrix} 0 & 1 \\ 4 & -3 \end{bmatrix} w.$$

Our first order system is $\gamma' = F(\gamma)$. Geometrically, F is a vector field (see Definition 1.7.16) on $\mathbb{R}^2$. Then $\gamma(t)$ solves the system if and only if the velocity vector $\gamma'(t)$ agrees with the vector $F(\gamma(t))$ for all t.

(This is an example of a linear F; in general F need not be linear. In the nonlinear case, of course we won't be able to use matrices to express F.)

Definition 1.6.36. Let $U \subset \mathbb{R}^m$ and let $F : U \to \mathbb{R}^m$ be a function. A path $\gamma : (a, b) \to U$ **solves the system corresponding to** F if

1. γ is differentiable, and

2. $\gamma'(t) = F(\gamma(t))$ for all $t \in (a, b)$.

Furthermore, if $t_0 \in (a, b)$ and $\gamma(t_0) = p \in U$, then γ is said to solve the **initial value problem** (IVP)

$$\gamma' = F(\gamma), \qquad \gamma(t_0) = p.$$

(This type of system is "autonomous.")

More generally: suppose $\Omega \subset \mathbb{R} \times \mathbb{R}^m$ is open. Let $F : \Omega \to \mathbb{R}^m$ be a function (think $F(t, \gamma(t))$ instead of $F(\gamma(t))$). A solution to the first order system of ODEs corresponds to F with initial value p at time t_0 is defined to be a differentiable function $\gamma : (a, b) \to \mathbb{R}^m$ (with $t_0 \in (a, b)$) such that

1. $(t, \gamma(t)) \in \Omega$ for all $t \in (a, b)$,

2. $\gamma'(t) = F(t, \gamma(t))$ for all $t \in (a, b)$, and

3. $\gamma(t_0) = p$.

For example, a non-autonomous ODE is $x'' - t^2 x' + e^t x = e^{3t}$. (Picard's Theorem in this general setting is just messier, not fundamentally harder. However, we will not address it in class since it is messy.)

Theorem 102 (Picard's Theorem). Let Ω be an open subset of $\mathbb{R} \times \mathbb{R}^m$ and let $F : \Omega \to \mathbb{R}^m$ be a function ($F(t, y)$ for $t \in \mathbb{R}$ and $y \in \mathbb{R}^m$) such that F is continuous and F is "locally uniformly Lipschitz in y" (for all $(t_0, y_0) \in \Omega$, there exists an open neighborhood $V \subset \Omega$ of (t_0, y_0) and there exists $L \geq 0$ such that for (t, y) (t, y') in V (same t), we have $\|F(t, y) - F(t, y')\| \leq L\|y - y'\|$) (the "uniformly" means "uniformly in t"). Then for all $(t_0, y_0) \in \Omega$, there exists an open interval (a, b) of $\mathbb{R}$ containing t_0 and a solution γ to the initial value problem $\gamma'(t) = F(t, \gamma(t)), \gamma(t_0) = y_0$ defined on (a, b).

Further, if γ and $\tilde{\gamma}$ are two local solutions to the initial value problem, then $\gamma = \tilde{\gamma}$ on some (possibly smaller) open neighborhood of t_0 in $\mathbb{R}$.

(In particular: say $m = 1$. So we have $F(t, y)$ where $y \in \mathbb{R}$. $\Omega \subset \mathbb{R} \times \mathbb{R}$ is open, $F : \Omega \to \mathbb{R}$ such that F is continuous and $\frac{\partial F}{\partial y}$ is continuous on Ω. This implies that on a compact subset $[t_0 - \epsilon_1, t_0 + \epsilon_1] \times [y_0 - \epsilon_2, y_0 + \epsilon_2] \subset \Omega$, $\frac{\partial F}{\partial y}$ is bounded (say by L); L works in general then as a local uniform Lipschitz constant near (t_0, y_0). Then the assumptions of Picard's Theorem hold, so initial value problems of the form $\gamma'(t) = F(t, \gamma(t)), \gamma(t_0) = y_0$ have unique solutions locally.)

Theorem 103 (Picard-Lindelof Theorem for autonomous systems (Special version we'll prove: Theorem 4.25 in Pugh [2015])). Let $U \subset \mathbb{R}^m$ be open and let $F : U \to \mathbb{R}^m$ be a function such that F is **locally Lipschitz continuous** (which implies continuity; see Definition 1.7.21): for all $y_0 \in U$ there exists an open neighborhood V of y_0 in U and there exists $L \geq 0$ such that $\|F(y) - F(y')\| \leq L\|y - y'\|$ for all $y, y' \in V$ (don't need to say "uniformly Lipschitz" because there's no separate t variable that F has to be uniformly Lipschitz over). Let $t_0 \in \mathbb{R}$ and let $y_0 \in U$. Then initial value problems of the form $\gamma'(t) = F(\gamma(t)), \gamma(t_0) = y_0$ have unique solutions locally (existence + uniqueness).

We will need some lemmas and definitions. For starters, we'll need integrals of functions $\gamma : \mathbb{R} \to \mathbb{R}^m$ (vector-valued functions).

Definition 1.6.37. Let $F : [a, b] \to \mathbb{R}^m$ ($F = \begin{bmatrix} F_1 & F_2 & \cdots & F_m \end{bmatrix}^T$, so $F_i : [a, b] \to \mathbb{R}$). F is called **Riemann integrable** if each F_i is Riemann integrable. If so, then

$$\int_a^b F(t)dt := \begin{bmatrix} \int_a^b F_1(t)dt \\ \int_a^b F_2(t)dt \\ \vdots \\ \int_a^b F_m(t)dt \end{bmatrix}.$$

Proposition 104. Let A be a set of tagged partitions (P, T) of $[a, b]$. For $(P, T) \in A$, let

$$R(F, P, T) := \begin{bmatrix} R(F_1, P, T) \\ R(F_2, P, T) \\ \vdots \\ R(F_m, P, T) \end{bmatrix} = \sum_{i=1}^n F(t_i)(x_i - x_{i-1}) \qquad \text{if } (P, T) = (x_0 < \ldots < x_n, t_1 < \ldots < t_n).$$

Assignment $(P, T) \mapsto R(F, P, T)$ is a net in $\mathbb{R}^m$, mapping A to $\mathbb{R}^m$ (recall Definition 1.6.3). Then

1. F is Riemann integrable if and only if this net converges; if so, $\int_a^b F(t)dt$ is the limit of the net in $\mathbb{R}^m$. (also equivalent to using a mesh-based condition, as Riemann did, or a Darboux-style condition.)

2. Continuous (vector-valued) functions $F : [a, b] \to \mathbb{R}^m$ are Riemann integrable.

3. If F' exists and is continuous (F is C^1), then $\int_a^b F'(t)dt = F(b) - F(a)$.

4.
$$\left\| \int_a^b F(t)dt \right\|_2 \leq M(b - a)$$

where $M := \sup_{t \in [a,b]} \| F(t) \|_2$.

Proof.   1. F is Riemann integrable with integral $\boldsymbol{I} = (I_1, \ldots, I_m)^T$ if and only if each F_i is Riemann integrable with integral I_i. This is true if and only if $\lim_{(P,T)} R(F_i, P, T)$ exists and equals I_i for $i \in [m]$, which is true if and only if $\lim_{(P,T)} R(F, P, T)$ exists and equals $\boldsymbol{I}$. (We know from Corollary 2.19 in Pugh [2015] that a sequence in $\mathbb{R}^m$ converges to $\boldsymbol{I}$ if and only if each component converges to I_i in $\mathbb{R}$. True for nets too with the same proof.)

2. If F is continuous then each F_i is continuous, which implies each F_i is Riemann integrable, which by definition means that F is Riemann integrable.

3.
$$\int_a^b F'(t)dt = \int_a^b \begin{bmatrix} F_1'(t) \\ \vdots \\ F_m'(t) \end{bmatrix} dt = \begin{bmatrix} \int_a^b F_1(t)dt \\ \int_a^b F_2(t)dt \\ \vdots \\ \int_a^b F_m(t)dt \end{bmatrix} = \begin{bmatrix} F_1(b) - F_1(a) \\ \vdots \\ F_m(b) - F_m(a) \end{bmatrix} = F(b) - F(a).$$

4.
$$\left\| \int_a^b F(t)dt \right\|_2 = \left\| \lim_{(P,T)} R(F, P, T) \right\|_2 = \lim_{(P,T)} \| R(F, P, T) \|_2$$

where the limit may come out from the norm because norms are continuous: $\| \cdot \|$ sends convergent sequences to convergent sequences, so preserves limits of sequences; same applies here for nets, with the same proof. We have

$$\| R(F, P, T) \|_2 \leq \sum_{i=1}^n \| R(t_i) \|_2 (x_i - x_{i-1}) \leq M \sum_{i=1}^n (x_i - x_{i-1}) = M(b - a).$$

$\square$

Lemma 105. Let $\Omega \subset \mathbb{R} \times \mathbb{R}^m$ be open. Let $F : \Omega \to \mathbb{R}^m$ be continuous. Let $\gamma : (a, b) \to \mathbb{R}^m$ be a continuous function such that $(t, \gamma(t)) \in \Omega$ for all $t \in (a, b)$. Let $t_0 \in (a, b)$. Then γ is differentiable on (a, b) with $\gamma'(t) = F(t, \gamma(t))$ for all t if and only if $\gamma(t) = \gamma(t_0) + \int_{t_0}^t F(s, \gamma(s))ds$.

Proof. Assuming the first statement holds, by the Fundamental Theorem of Calculus,

$$\gamma(t) - \gamma(t_0) = \int_{t_0}^{t} \gamma'(s)ds = \int_{t_0}^{t} F(s, \gamma(s))ds,$$

proving (2). (We can apply the FTC because γ' is continuous since $\gamma'(s) = F(s, \gamma(s))$, and both γ and F are continuous.)

Assuming (2) holds, γ is differentiable since $\gamma(t_0) + \int_{t_0}^{t} F(s, \gamma(s))ds$ is differentiable (recall $F(s, \gamma(s))$ is continuous). Then by the Fundamental Theorem of Calculus, $\gamma'(t) = F(t, \gamma(t))$.

$\square$

Now we are ready to prove Picard's Theorem.

Proof of Theorem 103. Without loss of generality, F is Lipschitz on U (otherwise, replace U with an open neighborhood V of y_0 in U on which F is Lipschitz). Let L be a Lipschitz constant for F on U; i.e., $\|F(y_1) - F(y_2)\| \leq L\|y_1 - y_2\|$ for all $y_1, y_2 \in U$. For convenience, assume without loss of generality $t_0 = 0$.

Choose $r > 0$ such that $\overline{B_r(y_0)} \in U^2$. Let $N := \overline{B_r(y_0)}$. N is compact and F is continuous, so F attains a finite maximum on N; that is, there exists M such that $\|F(x)\|_2 \leq M$ for all $x \in N$. Choose $\tau > 0$ such that $\tau < \min\left\{\frac{r}{M}, \frac{1}{L}\right\}$. (It is possible to remove the condition that $\tau < 1/L$, but it makes the proof harder.)

We'll show (1) that there exists a solution γ on $[-\tau, \tau]$ with $\gamma(0) = y_0$ and (2) any two solutions to this IVP on $[-\tau, \tau]$ are equal (which implies local uniqueness). Let $\mathcal{C} = C^0([-\tau, \tau], N)$. Since N is compact, $\mathcal{C}$ is a complete metric space with the uniform metric d_∞. Therefore if we can find a contraction mapping then a unique fixed point will exist. We'll define $\Phi : \mathcal{C} \to \mathcal{C}$ to be a contraction such that $\gamma \in \mathcal{C}$ is a fixed point of Φ if and only if γ solves the integral equation $\gamma(t) = y_0 + \int_0^t F(\gamma(s))ds$.

How to find Φ: Picard iteration. For $\gamma \in \mathcal{C}$, let $\Phi(y) := y_0 + \int_0^t F(\gamma(s))ds$. (Note: $\Phi(\gamma)(0) = y_0$, so it has the right initial value.)

We claim that $\Phi(\gamma) \in \mathcal{C}$; that is, $\Phi : \mathcal{C} \to \mathcal{C}$. Observe that $\Phi(\gamma)$ is some function $[-\tau, \tau] \to \mathbb{R}^m$; we want to show that $\Phi(\gamma)(t) \in N$ for all $t \in [-\tau, \tau]$. Indeed,

$$\|\Phi(\gamma)(t) - y_0\|_2 = \text{ (by definition) } \left\|\int_0^t F(\gamma(s))ds\right\|_2 \leq M(t - 0) = Mt \leq M\tau < r$$

since $\|F(x)\|_2 \leq M$ for all $x \in N$ and $\gamma(s) \in N$ for all s because $\gamma \in \mathcal{C}$ by assumption. This implies $\Phi(\gamma)(t) \in N$ for all $t \in [-\tau, \tau]$; i.e., $\Phi(\gamma) \in \mathcal{C}$. (Also note: $\Phi(\gamma)$ is continuous by the Fundamental Theorem of Calculus.)

It remains to show that Φ is a contraction; then it has a unique fixed point. We claim Φ is a contraction with constant $k := \tau L$ (which is less than 1 by assumption that $\tau < 1/L$). Let $\gamma, \sigma \in \mathcal{C}$. We want to show that $d_{\text{sup}}(\Phi(\gamma), \Phi(\sigma)) \leq k d_{\text{sup}}(\gamma, \sigma)$ (note that norms don't exist since this is only a metric space, not a vector space). We have

---

[2] We can fit a closed ball inside U because clearly we can fit an open ball inside U; if this open ball is of radius $2r$, then we can fit a closed ball of radius r.

$$d_{\sup}(\Phi(\gamma), \Phi(\sigma)) = \sup_{t \in [-\tau, \tau]} \left\| y_0 + \int_0^t F(\gamma(s))ds - y_0 - \int_0^t F(\sigma(s))ds \right\|_2$$

$$= \sup_{t \in [-\tau, \tau]} \left\| \int_0^t [F(\gamma(s)) - F(\sigma(s))] \, ds \right\|_2$$

$$\leq \underbrace{(t - 0)}_{\leq \tau} \sup_{t \in [-\tau, \tau]} \|F(\gamma(s)) - F(\sigma(s))\|_2$$

$$\leq \tau \sup_{t \in [-\tau, \tau]} L \|\gamma(t) - \sigma(t)\|_2$$

$$= \tau L d_{\sup}(\gamma, \sigma),$$

so $d_{\sup}(\Phi(\gamma), \Phi(\sigma)) \leq k d_{\sup}(\gamma, \sigma)$, so Φ is a contraction. Therefore Φ has a unique fixed point γ, and $\Phi(\gamma) = \gamma$ implies $\phi(t)$ is differentiable and the unique solution to the IVP locally.

Any other solution $\sigma(t)$ defined on $[-\tau, \tau]$ is also a fixed point of Φ. Since a contraction mapping has a unique fixed point, $\gamma = \sigma$.

$\hfill \square$

Definition 1.6.38 (p. 246 of Pugh **[**2015**], p. 256 of pdf).** The **flow** associated with a first order system is defined as follows. Assume $U = \mathbb{R}^m$ and $F : \mathbb{R}^m \to \mathbb{R}^m$ is globally Lipschitz. We can show that solutions to $\gamma'(t) = F(\gamma(t))$ exist for all time and are unique (problem from Homework 9). Given t, can consider $y_0 \mapsto \gamma(t)$ if γ is a solution with $\gamma(0) = y_0$. In particular, define $\phi : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ by $\phi(t, y) := \gamma(t)$ where $\gamma : \mathbb{R} \to \mathbb{R}^m$ is the unique solution to $\gamma' = F(\gamma), \gamma(0) = y$. $\phi(t, y)$ is called the **flow** associated with the vector field F (see Definition 1.7.16).

Think of $\phi(t, y)$ as the flow of y along the vector field F for t seconds. It turns out we can show ϕ is continuous jointly in t and y; that is, it has continuous dependence on its initial conditions (see Proposition 108 below). It also turns out that if F is C^k (all partial derivatives up to the kth order exist and are continuous), then ϕ is C^k, for $1 \leq k \leq \infty$. (Note that this implies F is locally Lipschitz; see Definition 1.7.21.)

Local version works too: say $F : U \to \mathbb{R}^m$ is locally Lipschitz, and let $y_0 \in U$. It turns out this implies that there exists an open neighborhood V of y_0 and there exists $t_0 \in \mathbb{R}_{++}$ such that (1) F is Lipschitz on V and (2) for every $y \in V$, the solution to the IVP $\gamma' = F(\gamma), \gamma(0) = y$ is defined at least on $(-t_0, t_0)$. Thus we can define $\phi : (-t_0, t_0) \times V \to \mathbb{R}^m$ where $\phi(t, y) := \gamma(t)$ for the unique solution γ to the IVP $\gamma' = F(\gamma), \gamma(0) = y$. In this case, ϕ is continuous and if F is C^k for some $1 \leq k \leq \infty$, then ϕ is C^k.

Here's a basic fact about flows: intuitively, after flowing for time t_1 and then flowing from that position for t_2, you should be in the same place as if you just initially flowed for time $t_1 + t_2$. More precisely, assume we're in the global case ($F : \mathbb{R}^m \to \mathbb{R}^m$ is Lipschitz), although the local version is true too. Let $\phi : \mathbb{R} \times \mathbb{R}^m$ be the flow associated with F.

Proposition 106. For all $t_1, t_2 \in \mathbb{R}$ and any $y \in \mathbb{R}^m$, we have $\phi(t_1 + t_2, y) = \phi(t_2, \phi(t_1, y))$.

Proof. Fix t_1, y and let $\gamma(t) := \phi(t + t_1, y)$. (Note that this is a solution: $\gamma' = F(\gamma)$, since $\frac{\partial}{\partial t}(t + t_1) = 1$.) Let $\zeta(t) := \phi(t, \phi(t_1, y))$. Observe that $\zeta' = F(\zeta)$ and $\zeta(0) = \phi(t_1, y)$ by construction. Therefore $\zeta(0) = \phi(t_1, y)$. By global uniqueness (proven on Homework 9), we have $\phi(t_2) = \zeta(t_2)$ for all t_2. Therefore $\phi(t_1 + t_2, y) = \phi(t_2, \phi(t_1, y))$ for all t_1, t_2, y.

$\square$

Note that applying Proposition 106 yields $\phi(-t, \phi(t, y)) = \phi(0, y) = y$. So $\phi(t, \cdot)$ is invertible with inverse $\phi(-t, \cdot)$. By continuous dependence on initial conditions, $\phi(t, \cdot)$ and its inverse are continuous. That is, $\phi(t, \cdot)$ is a homeomorphism (see Definition 1.4.14) from $\mathbb{R}^m$ to $\mathbb{R}^m$.

We can view the flow construction as giving a map $\Phi : \mathbb{R} \mapsto \mathrm{Homeo}(\mathbb{R}^m)$, $t \mapsto \phi(t, \cdot)$. By Proposition 106, we have $\Phi(t_1 + t_2) = \Phi(t_2) \circ \Phi(t_1)$ because $\Phi(t_1 + t_2)(y) = \phi(t_1 + t_2, y) = \phi(t_2, \phi(t_1, y)) = \phi(t_2, \Phi(t_1)(y)) = \Phi(t_2)(\Phi(t_1)(y))$. So we have group stuctures on $\mathbb{R}$ and $\mathrm{Homeo}(\mathbb{R}^m)$. (on $\mathbb{R}$, the group structure is addition of numbers; on $\mathrm{Homeo}(\mathbb{R}^m)$, the group structure is composition of homeomorphisms.) Then the Proposition says $\Phi : \mathbb{R} \to \mathrm{Homeo}(\mathbb{R}^m)$ is a group homomorphism.

If F were smooth, then $\Phi : \mathbb{R} \mapsto \mathrm{Diffeo}(\mathbb{R}^m)$, where $\mathrm{Diffeo}(\mathbb{R}^m)$ is a set of smooth functions with smooth inverses. In some situations, we have even more structure in the self-maps of $\mathbb{R}^m$. For instance, below we have $\Phi : \mathbb{R} \to \mathrm{Symplectomorphisms}(\mathbb{R}^{2n})$.

Next we will apply flows using an example from physics.

Example 1.6.11. Let $H : \mathbb{R}^{2n} \to \mathbb{R}$ be a C^∞ function; call it the **Hamiltonian** function. We will think of $\mathbb{R}^m$ as consisting of coordinates $\underbrace{q_1, \ldots, q_n}_{\text{position coordinates}}, \underbrace{p_1, \ldots, p_n}_{\text{momentum}}$. We can refer to as $\mathbb{R}^{2n}$ as the **phase space** of a classical physical system, so a point in $\mathbb{R}^{2n}$ is the state of the system. Our observable quantities are $\mathbb{R}$-valued functions on $\mathbb{R}^{2n}$. We want to learn how the states and observable quantities evolve over time. (More generally, a phase space is a kind of symplectic manifold—see Definition 1.7.10.)

Let the **Hamiltonian vector field** X_H be the vector field on $\mathbb{R}^{2n}$ defined by

$$X_H(q_1, \ldots, q_n, p_1, \ldots, p_n) := \begin{bmatrix} \frac{\partial H}{\partial p_1}(q_1, \ldots, q_n, p_1, \ldots, p_n) \\ \vdots \\ \frac{\partial H}{\partial p_n}(q_1, \ldots, q_n, p_1, \ldots, p_n) \\ -\frac{\partial H}{\partial q_1}(q_1, \ldots, q_n, p_1, \ldots, p_n) \\ \vdots \\ -\frac{\partial H}{\partial q_n}(q_1, \ldots, q_n, p_1, \ldots, p_n) \end{bmatrix},$$

the symplectic analogue of the gradient ∇H. (X_H corresponds to F in our general setup of ODEs, where F is a vector field on $U \subset \mathbb{R}^m$ with $m = 2n$.) Locally, X_H defines a flow $\phi_H : (-t_0, t_0) \times V \mapsto \mathbb{R}^{2n}$. (We'll assume it's global for now.) So $\phi_H : \mathbb{R} \times \mathbb{R}^{2n} \to \mathbb{R}^{2n}$.

Theorem 107 (Result from classical physics). ϕ_H determines how the state of the system evolves over time. $\phi_H : \mathbb{R} \times \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is equivalent to (via "currying") $\Phi_H : \mathbb{R} \to \mathrm{Homeo}(\mathbb{R}^{2n})$ (actually, maps to **symplectomorphisms** of $\mathbb{R}^{2n}$, which are even more structured: they are diffeomorphisms that preserve symplectic structure.)

Definition 1.6.39 (Diffeomorphism). A **diffeomorphism** is an isomorphism (mapping between two structures of the same type that can be reversed by an inverse mapping; see Definition 1.4.13) of spaces equipped with a differential structure, typically differentiable manifolds (see Definition 1.7.10).

A C^r diffeomorphism is a C^r bijection whose inverse is C^r for $1 \leq r \leq \infty$ (differentiable function with a differentiable inverse).

(Note that a diffeomorphism is a homeomorphism—see Definition 1.4.14.)

Note that since differentiability implies continuity, every diffeomorphism is a homeomorphism.

Example 1.6.12. Just because a function is C^r does not mean its inverse is C^r. For example, consider $f(x) = x^3$. It is a smooth homeomorphism from $\mathbb{R}$ to $\mathbb{R}$ but it is not a diffeomorphism because its inverse fails to be differentiable at the origin.

Proposition 108. The function $y_0 \mapsto \gamma(t)$ depends continuously on y_0. (continuous dependence on initial conditions).

1.7 Multivariable Calculus (Chapter 5 of Pugh [2015])

1.7.1 Linear Algebra (Section 5.1 of Pugh [2015])

(For more notes on linear algebra, see Section **??**.)

Theorem 109. Linear transformations are unbounded in sup norm (except $T = 0$).

Proof. Suppose T is a linear transformation with $T(v_0) \neq 0$ and $\|T(v)\| \leq M \in \mathbb{R}_+$ for all v. Then

$$\left\| T\left(\frac{(M+1)v_0}{\|T(v_0)\|} \right) \right\| = \frac{(M+1)}{\|T(v_0)\|} \|T(v_0)\| = M + 1,$$

contradiction.

$\square$

For this reason, it's not so useful to look at $\|\cdot\|_\infty$ for linear transformations. Is there a more useful norm on linear transformations? Yes: the operator norm (basic ingredient of functional analysis).

Definition 1.7.1 (p. 279 of Pugh [2015], p. 289 of pdf). Let V, W be normed vector spaces (possibly infinite-dimensional). Let $V : V \to W$ be a linear transformation. Define

$$\|T\|_{\text{op}} := \sup_{v \neq 0 \in V} \frac{\|T(v)\|}{\|v\|} \in \mathbb{R}_+ \cup \{\infty\}.$$

(It turns out the following expressions are equivalent:)

$$\|T\|_{\text{op}} = \sup_{v \in V, \|v\|=1} \|T(v)\|$$
$$= \inf \{c \in \mathbb{R} | \|T(v)\| \leq c\|v\| \forall v \in V\}.$$

Also, an immediate consequence of the definition is that $\|T(v)\| \leq \|T\|_{\text{op}}\|v\|$ for all $v \in V$.

A question we would like to know the answer to: which T have finite operator norm?

Theorem 110 (**Theorem 5.2 in Pugh [2015]**)**.** Let V, W be normed vector spaces (especially infinite dimensional spaces; we'll talk about finite dimensional spaces soon; easier) and let $T : V \to W$ be a linear transformation. The following are equivalent:

1. $\|T\|_{\mathrm{op}} < \infty$.

2. T is Lipschitz continuous.

3. T is uniformly continuous.

4. T is continuous.

5. T is continuous at the origin.

(Note that this shows that continuity at one point comes if and only if T is continuous everywhere. So T being linear is a strong restriction.

Proof. **1 implies 2:** Given (1), we claim that $\|T\|_{\mathrm{op}}$ is a Lipschitz constant for T (given norms on V, W). Indeed: if $v_1 \neq v_2 \in V$, then

$$\|T\|_{\mathrm{op}} \geq \frac{\|T(v_2 - v_1)\|}{\|v_2 - v_1\|} = \frac{\|T(v_2) - T(v_1)\|}{\|v_2 - v_1\|},$$

so $\|T(v_2) - T(v_1)\| \leq \|T\|_{\mathrm{op}} \|v_2 - v_1\|$ if $v_1 \neq v_2$ (also true if $v_1 = v_2$). Therefore $\|T\|_{\mathrm{op}}$ is a Lipschitz constant for T (recall Definition 1.6.28).

2 implies 3 implies 4 implies 5: already shown by definitions of these terms.

5 implies 1: Assume T is continuous at the origin. Then for $\epsilon = 1$ there exists $\delta > 0$ such that if $v \in V$ with $\|v - 0\| < \delta$ then $\|T(v) - 0\| < 1$. We claim that $\|T\|_{\mathrm{op}} \leq 2/\delta$. To see this, note that for any $v \in V$ and $v \neq 0$, we have

$$\left\| \frac{\delta}{2} \frac{v}{\|v\|} \right\| = \frac{\delta}{2},"$$

Thus, by continuity for any $v \in V$

$$\left\| T\left(\frac{\delta}{2} \frac{v}{\|v\|} \right) \right\| < 1 \qquad \Longleftrightarrow \qquad \frac{\delta}{2\|v\|} \|T(v)\| < 1 \qquad \Longleftrightarrow \qquad \frac{\|T(v)\|}{\|v\|} < \frac{2}{\delta}.$$

so $\sup_{v \in V} \frac{\|T(v)\|}{\|v\|} = \|T\|_{\mathrm{op}} \leq \frac{2}{\delta}$.

$\square$

Definition 1.7.2. If $T : V \to W$ satisfies any of the equivalent properties in Theorem 110, T is called a **bounded linear transformation/map/operator** (usually for the $V = W$ case) or a **continuous linear transformation/map/operator**.

Definition 1.7.3. Let $\mathcal{L}(V, W)$ be the set of bounded linear transformations from V to W. (It's a vector space itself, and it has a norm $\|\cdot\|_{\text{op}}$. Exercise: the norm axioms are satisfied. Check that $\|T + S\|_{\text{op}} \leq \|T\|_{\text{op}} + \|S\|_{\text{op}}$, NOT $\frac{\|T(v_1 + v_2)\|}{\|v_1 + v_2\|\}}$.

Fact we won't prove: if V, W are Banach spaces than $\mathcal{L}(V, W)$ is a Banach space. $\mathcal{L}(V, W)$: can add, do scalar multiplication (vector space structure), and also can compose linear transformations. A natural question: how does that interact with operator norm?

Proposition 111. If $S : V \to W$ and $T : W \to Z$ are linear transformations with V, W, Z normed vector spaces, then

$$\|T \circ S\|_{\text{op}} \leq \|T\|_{\text{op}} \|S\|_{\text{op}}.$$

Proof.

$$\|T \circ S\|_{\text{op}} = \sup_{v \neq 0 \in V} \left\{ \frac{\|T(S(v))\|}{\|v\|} \right\} = \max \left\{ \sup_{\{v \in V : \neq 0, S(v) \neq 0\}} \left\{ \frac{\|T(S(v))\|}{\|v\|} \right\}, \sup_{\{v \in V : \neq 0, S(v) = 0\}} \left\{ \frac{\|T(S(v))\|}{\|v\|} \right\} \right\}$$

Observe that if $S(v) = 0$ then $T(S(v)) = 0$ so $\sup_{\{v \in V : \neq 0, S(v) = 0\}} \left\{ \frac{\|T(S(v))\|}{\|v\|} \right\} = 0$. Therefore since the operator norm is nonnegative, $\sup_{\{v \in V : \neq 0, S(v) \neq 0\}} \left\{ \frac{\|T(S(v))\|}{\|v\|} \right\} \geq \sup_{\{v \in V : \neq 0, S(v) = 0\}} \left\{ \frac{\|T(S(v))\|}{\|v\|} \right\}$ and we have

$$
\begin{aligned}
\|T \circ S\|_{\text{op}} &= \sup_{\{v \in V : \neq 0, S(v) \neq 0\}} \left\{ \frac{\|T(S(v))\|}{\|v\|} \right\} \\
&= \sup_{\{v \in V : \neq 0, S(v) \neq 0\}} \left\{ \frac{\|T(S(v))\|}{\|S(v)\|} \cdot \frac{\|S(v)\|}{\|v\|} \right\} \\
&\leq \sup_{w \neq 0 \in W} \left\{ \frac{\|T(w)\|}{\|w\|} \right\} \cdot \sup_{v \neq 0 \in V} \left\{ \frac{\|S(v)\|}{\|v\|} \right\} \\
&= \|T\|_{\text{op}} \|S\|_{\text{op}}.
\end{aligned}
$$

$\square$

Restrict to case $V = W = Z$, so S, T are operators on V. $\mathcal{L}(V, V)$ is not just a vector space; it's an algebra (ring (see Definition **??** and vector space at the same time) over $\mathbb{R}$ or $\mathbb{C}$.

Definition 1.7.4. An algebra A over $\mathbb{R}$ or $\mathbb{C}$ is called a **Banach algebra** if as a vector space it has a complete norm $\|\cdot\|$ such that $\|a_1 a_2\| \leq \|a_1\| \|a_2\|$ for all $a_1, a_2 \in A$.

Corollary 111.1 (Corollary to Proposition 111)**.** If $(V, \|\cdot\|)$ is a Banach space, then $(\mathcal{L}(V, V), \|\cdot\|_{\text{op}})$ is a Banach algebra.

Definition 1.7.5. If V is a vector space, a **linear functional** on V is a linear transformation $\alpha : V \to \mathbb{R}$ (similar definition for $\mathbb{C}$).

Definition 1.7.6 (**Dual Space**)**.** If V is a vector space, the **dual space** V^* to V is the vector space of linear functionals on V.

Definition 1.7.7 (**Covectors and dual spaces; definition from Chapter 11 of** Lee [2012]**, p. 287 of pdf, p.272 of book**)**.** Let V be a finite-dimensional real vector space. We define a **covector on V** to be a real-valued linear functional on V, that is, a linear map $\omega : V \to \mathbb{R}$. The space of all covectors on V is itself a real vector space under the obvious operations of pointwise addition and scalar multiplication. It is denoted by V^* and called the **dual space of V**.

Example 1.7.1. If $V = \mathbb{R}^n$ is the set of column vectors of size n, then the set of linear maps V^* from V to $\mathbb{R}$ is a set of row vectors of size n (then elements of V^* can be matrix multiplied by elements of V yielding real numbers).

What about if V is a normed vector space with $\dim(V) = \infty$? If V^* is the set of all linear functionals on V (disregarding norm), then V^* is "too big." Instead, define the continuous dual of V.

Definition 1.7.8. $V^* := \mathcal{L}(V, \mathbb{R})$ is a vector space of continuous linear functionals on V. V^* has the operator norm as its norm: for all $\alpha \in V^*$, $\|\alpha\|_{\mathrm{op}} = \sup_{v \in V, v \neq 0} \frac{|a(v)|}{\|v\|}$.

We will focus for now on the finite-dimensional case (much easier).

Definition 1.7.9. Let V be a vector space, and let $\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on V. Then $\|\cdot\|_1$ and $\|\cdot\|_2$ are **comparable** if there exist $c_1, c_2 > 0$ such that for all $v \in V$ we have $\|v\|_1 \leq c_1 \|v\|_2$ and $\|v\|_2 \leq c_2 \|v\|_1$.

Theorem 112 (**Comparability of norms; from Homework 10**)**.** Let V be a finite-dimensional vector space over $\mathbb{R}$. Any two norms on V are comparable.

Proof. First we will prove two lemmas.

Lemma 113. Let V be a finite-dimensional vector space over $\mathbb{R}$. Let $\dim(V) = n$ be a basis for V, let $\beta = \{e_1, \ldots, e_n\}$, and define a norm $\|\cdot\|_\beta$ on V by

$$\left\| \sum_{i=1}^n v^i e_i \right\|_\beta := \sqrt{\sum_{i=1}^n (v^i)^2}. \tag{1.21}$$

Any norm $\|\cdot\|_1 : V \to \mathbb{R}$ is continuous when V is given the metric induced by $\|\cdot\|_\beta$.

Proof. By the Triangle Inequality (satisfied by any norm by definition)

$$\|v\|_1 = \|v - w + w\|_1 \leq \|v - w\|_1 + \|w\|_1 \implies \|v\|_1 - \|w\|_1 \leq \|v - w\|_1$$

Also,

$$\|w\|_1 = \|w - v + v\|_1 \leq \|w - v\|_1 + \|v\|_1 \implies -\|w - v\|_1 \leq \|v\|_1 - \|w\|_1 \iff -\|v - w\|_1 \leq \|v\|_1 - \|w\|_1$$

Therefore $-\|v - w\|_1 \leq \|v\|_1 - \|w\|_1 \leq \|v - w\|_1$, so

$$\big| \|v\|_1 - \|w\|_1 \big| \leq \|v - w\|_1. \tag{1.22}$$

Let $\epsilon > 0$. We will find $\delta > 0$ such that $\|v - w\|_\beta < \delta \implies | \|v\|_1 - \|w\|_1 | < \epsilon$, showing that $\|\cdot\|_1 : V \to \mathbb{R}$ is continuous when V is given the metric induced by $\|\cdot\|_\beta$. Write v as $v = \sum_{i=1}^n v^i e_i$ and likewise $w = \sum_{i=1}^n w^i e_i$, so

$$\|v - w\|_1 = \left\|\sum_{i=1}^n (v^i - w^i) e_i\right\|_1 \leq \sum_{i=1}^n \left\|(v^i - w^i) e_i\right\|_1 = \sum_{i=1}^n |v^i - w^i| \|e_i\|_1 = \sum_{i=1}^n |v^i - w^i|,$$

so

$$\sum_{i=1}^n |v^i - w^i| < \epsilon \implies \|v - w\|_1 < \epsilon. \tag{1.23}$$

Suppose $|v_i - w_i| < 1$ for all i; then $(v_i - w_i)^2 < \delta \implies |v_i - w_i| < \delta$. Therefore if $|v_i - w_i| < 1$ for all i and letting $\delta = \sqrt{\epsilon}$,

$$
\begin{aligned}
\|v - w\|_\beta < \delta \qquad &\Longleftrightarrow \qquad \left\|\sum_{i=1}^n (v^i - w^i) e_i\right\|_\beta < \sqrt{\epsilon} \\
&\Longleftrightarrow \qquad \sqrt{\sum_{i=1}^n (v^i - w^i)^2} < \sqrt{\epsilon} \\
&\Longleftrightarrow \qquad \sum_{i=1}^n (v^i - w^i)^2 < \epsilon \\
\text{(by assumption } |v_i - w_i| < 1 \text{ for all } i) \qquad &\Longrightarrow \qquad \sum_{i=1}^n |v^i - w^i| < \epsilon \\
\text{(by (1.23))} \qquad &\Longrightarrow \qquad \|v - w\|_1 < \epsilon \\
\text{(by (1.22))} \qquad &\Longrightarrow \qquad | \|v\|_1 - \|w\|_1 | < \epsilon.
\end{aligned}
$$

$\square$

Lemma 114 (**Transitivity of norm comparability**)**.** Let V be a finite-dimensional vector space over $\mathbb{R}$ and let $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_3$ be norms on V. If $\|\cdot\|_1$ is comparable to $\|\cdot\|_2$ and $\|\cdot\|_2$ is comparable to $\|\cdot\|_3$, then $\|\cdot\|_1$ is comparable to $\|\cdot\|_3$.

Proof. If $\|\cdot\|_1$ is comparable to $\|\cdot\|_2$ then there exists constants c_{12} and c_{21} such that for all $v \in V$

$$\|v\|_1 \leq c_{12}\|v\|_2, \qquad \|v\|_2 \leq c_{21}\|v\|_1.$$

Similarly, if $\|\cdot\|_2$ and $\|\cdot\|_3$ are comparable then

$$\|v\|_2 \leq c_{23}\|v\|_3, \qquad \|v\|_3 \leq c_{32}\|v\|_2$$

for every $v \in V$ for some constants c_{23} and c_{32}. It follows that for every $v \in V$

$$\|v\|_1 \le c_{12}\|v\|_2 \le c_{12}c_{23}\|v\|_3$$

and

$$\|v\|_3 \le c_{32}\|v\|_2 \le c_{32}c_{21}\|v\|_1.$$

$\square$

Now we can prove the main result. Again, let $\dim(V) = n$ be a basis for V, let $\beta = \{e_1, \ldots, e_n\}$, and consider the norm $\|\cdot\|_\beta$ on V defined by (1.21). Consider the restriction of $\|\cdot\|_1$ to the $\|\cdot\|_\beta$-unit sphere $S := \{v \in V \mid \|v\|_\beta = 1\}$. S is closed and totally bounded by construction, and every finite-dimensional normed vector space is complete, so by the Generalized Heine-Borel Theorem (Theorem 2.65 in Pugh [2015]) S is compact. Since S is bounded, it follows that for some $M \in (0, \infty)$

$$\|\tilde{v}\|_1 \le M \qquad \forall \tilde{v} \in S. \tag{1.24}$$

Further, since S does not contain the zero vector, it holds that there exists $m \in (0, M]$ such that

$$m \le \|\tilde{v}\|_1 \qquad \forall \tilde{v} \in S. \tag{1.25}$$

Now consider a general $v \in V$. We can express any $v \in V$ as $v = c\tilde{v}$ for some $c \in \mathbb{R}$ and some $\tilde{v} \in S$. By norm axioms, it holds that for any $c \in \mathbb{R}$ $\|cv\|_\beta = |c|\|v\|_\beta$, so in particular $\|v\|_\beta = \|c\tilde{v}\|_\beta = |c|\|\tilde{v}\|_\beta = |c|$. Then it follows from (1.24) that for any $v \in V$ there exists some $c \in \mathbb{R}$ such that

$$\|v\|_1 = \|c\tilde{v}\|_1 = |c|\|\tilde{v}\|_1 = \|v\|_\beta\|\tilde{v}\|_1 \le M\|v\|_\beta$$

and from (1.25) we have

$$\|v\|_1 = \|c\tilde{v}\|_1 = |c|\|\tilde{v}\|_1 = \|v\|_\beta\|\tilde{v}\|_1 \ge m\|v\|_\beta \qquad \Longleftrightarrow \qquad \|v\|_\beta \le \frac{1}{m}\|\tilde{v}\|_1.$$

Therefore $\|\cdot\|_\beta$ and $\|\cdot\|_1$ are comparable. By the same argument, $\|\cdot\|_\beta$ and $\|\cdot\|_2$ are comparable. Therefore by transitivity of norm comparability (Lemma 114), $\|\cdot\|_1$ and $\|\cdot\|_2$ are comparable.

$\square$

Corollary 114.1 (Corollary to Theorem 112)**.** $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous in the Euclidean norms if and only if f is continuous given any norms on $\mathbb{R}^n, \mathbb{R}^m$.

Corollary 114.2 (Corollary to Theorem 112)**.** Let V, W be finite-dimensional vector spaces over $\mathbb{R}$ or $\mathbb{C}$. Then any linear transformation $T : V \to W$ is continuous given any choice of norms for V, W.

Proof. Pick isomorphisms $V \cong \mathbb{R}^n$ and $W \cong \mathbb{R}^m$ (mappings between two structures of the same type that can be reversed by inverse mappings; see Definition 1.4.13); get norms on V, W by transforming Euclidean norms on $\mathbb{R}^n, \mathbb{R}^m$ to V, W. Under these isomorphisms, T acts by a matrix ($T(v) = Av$ for $v \in \mathbb{R}^n$).

Each entry of Av ($a_{i1}v_1 + \ldots + a_{in}v_n$) is a continuous function of $v \in \mathbb{R}^n$, so T is continuous in these norms, and therefore in any norms on V, W.

$$\square$$

So if V, W are finite-dimensional (say $\dim(V) = n$ and $\dim(W) = m$), then $\mathcal{L}(v, w)$ is the set of all linear transformations from V to W. We can choose bases for V and W and make a basis $\mathrm{mat}_{m \times n}(\mathbb{R}) = \mathbb{R}^{m \times n}$ (or $\mathbb{C}$) which is the set of $m \times n$ matrices with entries in $\mathbb{R}$ (or $\mathbb{C}$). We can also order the entries in some way and express these matrices as vectors in $\mathbb{R}^{m \cdot n}$.

Thus, we have many norms on $\mathcal{L}(V, W)$.

- We can pick the operator norm on $\mathcal{L}(V, W)$ for any choice of norms on V, W.

- We have many norms on $\mathbb{R}^{mn}$ (e.g. Euclidean norm, ℓ_1 norm, ℓ_∞ norm)

All of these norms are comparable.

Corollary 114.3 (Corollary to Theorem 112)**.** If $U \subset \mathbb{R}^k$ is open and $A : U \to \mathrm{mat}_{m \times n}(\mathbb{R})$ is a function ($A(x)$ is a matrix-valued function of $x \in U$), then

1. each entry of A is a continuous function from U to $\mathbb{R}$ ("A is continuous given ℓ_∞ norm on $\mathrm{mat}_{m \times n}(\mathbb{R})$") if and only if

2. for $x_0 \in U$, given $\epsilon > 0$, there exists $\delta > 0$ such that if $\|x - x_0\| < \delta$ then $\|A(x) - A(x_0)\|_{\mathrm{op}} < \epsilon$ ("A is continuous given its operator norm on $\mathrm{mat}_{m \times n}(\mathbb{R})$").

Now we will discuss some intuition for the $\|\cdot\|_{\mathrm{op}}$ on $\mathrm{mat}_{m \times n}(\mathbb{R})$.

Theorem 115 (**Spectral Theorem**)**.** Let $A \in \mathrm{mat}_{m \times n}(\mathbb{R})$ where $m = n$ and $A = A^T$. Then A has an orthonormal eigenbasis $v_1, \ldots, v_n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$.

Proposition 116. Let $A \in \mathrm{mat}_{m \times n}(\mathbb{R})$ where $m = n$ and $A = A^T$. Then $\|A\|_{\mathrm{op}} = |\lambda_1|$ (the absolute value of the largest eigenvalue).

Proof. For $v \in V$, write $v = c_1 v_1 + \ldots + c_n v_n$ (uniquely). Then

$$Av = \sum_{i=1}^n c_i A v_i = \sum_{i=1}^n c_i \lambda_i v_i.$$

By orthonormality, we then have

$$\|Av\|_2^2 = \sum_{i=1}^n |c_i|^2 |\lambda_i|^2 \leq |\lambda_1|^2 \sum_{i=1}^n |c_i|^2 = |\lambda_1|^2 \|v\|_2^2.$$

If $v \neq 0$, this means $\frac{\|Av\|_2}{\|v\|_2} \leq |\lambda_1|$, so $\|A\|_{\mathrm{op}} \leq |\lambda_1|$. On the other hand, $\frac{\|Av_1\|_2}{\|v_1\|_2} = |\lambda_1| \frac{\|v_1\|_2}{\|v_1\|_2} = |\lambda_1|$. So $\|A\|_{\mathrm{op}} \geq |\lambda_1|$.

$$\square$$

Now consider the general case: $A \in \mathrm{mat}_{m \times n}(\mathbb{R})$. Write A in its singular value decomposition $A = U\Sigma V^T$ where U is an $m \times m$ orthogonal matrix, V is an $n \times n$ orthogonal matrix, and Σ is the $m \times n$ matrix containing the singular values on the diagonal (see Section **??** for details on the singular value decomposition).

Proposition 117. Let $A \in \mathrm{mat}_{m \times n}(\mathbb{R})$. Then $\|A\|_{\mathrm{op}} = \sigma_1$ (the largest singular value).

Proof. For $v \in \mathbb{R}^n$, write $v = c_1 v_1 + \ldots + c_n v_n$ (uniquely), where v_i are the orthonormal basis vectors of $\mathbb{R}^n$ forming the columns of V. Let $\mathrm{rank}(A) = k \leq \min\{m, n\}$. Then

$$Av = U\Sigma V^T v = \sum_{i=1}^{k} c_i \underbrace{U}_{m \times k} \underbrace{\Sigma}_{k \times k} \underbrace{V^T}_{k \times n} v_i = \sum_{i=1}^{k} c_i U\Sigma e_i = \sum_{i=1}^{k} c_i \sigma_i u_i.$$

(where e_i is the standard basis vector for $\mathbb{R}^n$). By orthonormality, we then have

$$\|Av\|_2^2 = \sum_{i=1}^{k} |c_i|^2 \sigma_i^2 \leq \sigma_1^2 \sum_{i=1}^{n} |c_i|^2 = \sigma_1^2 \|v\|_2^2.$$

If $v \neq 0$, this means $\frac{\|Av\|_2}{\|v\|_2} \leq \sigma_1$, so $\|A\|_{\mathrm{op}} \leq \sigma_1$. On the other hand, $\frac{\|Av_1\|_2}{\|v_1\|_2} = \sigma_1 \frac{\|v_1\|_2}{\|v_1\|_2} = \sigma_1$. So $\|A\|_{\mathrm{op}} \geq \sigma_1$.

$\square$

1.7.2   Manifolds (Chapter 5 of Spivak [1971])

Definition 1.7.10 (**Manifold; definition from section 5.1 of Spivak [1971], p. 122 of pdf, p. 109 of book**). A subset M of $\mathbb{R}^n$ is called a k-**dimensional manifold** (in $\mathbb{R}^n$) if for every point $x \in M$ the following condition is satisfied: there is an open set U containing x, an open set $V \subset \mathbb{R}^n$, and a diffeomorphism $h : U \to V$ (recall Definition 1.6.39) such that

$$h(U \cap M) = V \cap (\mathbb{R}^k \times \{0\}) = \{y \in V : y^{k+1} = \ldots = y^n = 0\}.$$

(In other words, $U \cap M$ is, "up to diffeomorphism," simply $\mathbb{R}^k \times \{0\}$.)

Definition 1.7.11 (**Topological manifold; definition from Lee [2012], p. 17 of pdf, p. 2 of book**). Suppose M is a topological space (see Definition 1.4.12). We say that M is a **topological manifold of dimension** n or a **topological n-manifold** if it has the following properties:

1. M is a **Hausdorff space**: for every pair of distinct points $p, q \in M$, there are disjoint open subsets $U, V \subseteq M$ such that $p \in U$ and $q \in V$. (This is automatically satisfied for any metric space.)

2. M is **second-countable**: there exists a countable basis for the topology of M.

3. M is **locally Euclidean of dimension** n: each point of M has a neighborhood that is homeomorphic to an open subset of $\mathbb{R}^n$.
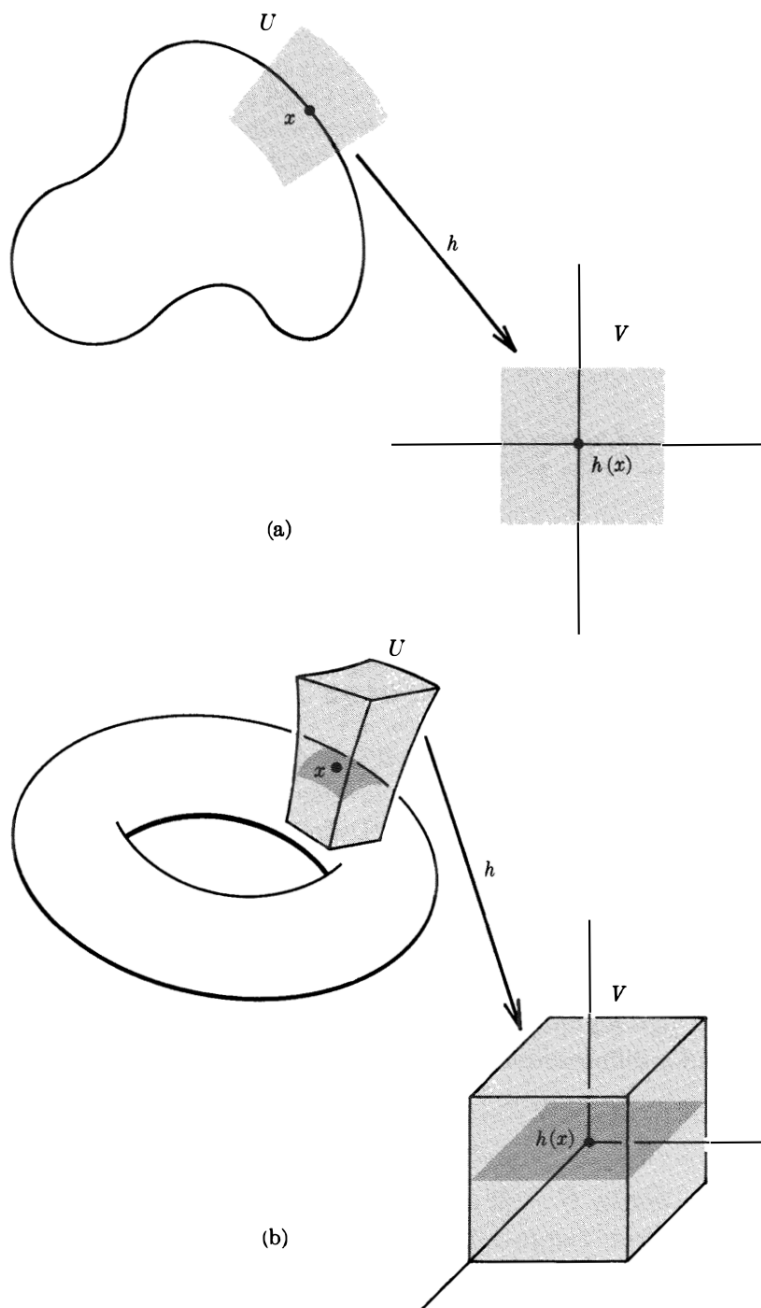
Figure 1.4: Figure 5.1 in Spivak [1971]. (a) A one-dimensional manifold in $\mathbb{R}^2$. (b) A two-dimensional manifold in $\mathbb{R}^3$.

The third property means, more specifically, that for each $p \in M$ we can find

- an open subset $U \subseteq M$ containing p,

- an open subset $\hat{U} \subseteq \mathbb{R}^n$, and

- a homeomorphism $\phi : U \to \hat{U}$.

Example 1.7.2. See Figure 1.4 for some examples of manifolds. There are two extreme cases of this definition. $\mathbb{R}^n$ itself is a topological n-manifold: it is Hausdorff because it is a metric space, and it is second-countable because the set of open balls with rational centers and rational radii is a countable basis for its topology.

and an open subset of $\mathbb{R}^n$ is an n-dimensional manifold.

Also, a point in $\mathbb{R}^n$ is a 0-dimensional manifold,

One common example of an n-dimensional manifold is the n-sphere S^n, defined as $\{x \in \mathbb{R}^{n+1} : |x| = 1\}$. The fact that S^n is an n-dimensional manifold can be verified by definition, or more easily by the following result.

Theorem 118 (**Theorem 5-1 in Spivak [1971], p. 124 of pdf, p. 111 of book**)**.** Let $A \subset \mathbb{R}^n$ be open and let $g : A \to \mathbb{R}^p$ be a differentiable function such that $g'(x) = (Dg)_x \in \mathbb{R}^{p \times n}$ has rank p whenever $g(x) = 0$. Then $g^{-1}(0)$ is an $(n-p)$-dimensional manifold in $\mathbb{R}^n$.

Proof. Immediate from Theorem 135.

$\square$

Note that $S^n = g^{-1}(0)$, where $g : \mathbb{R}^{n+1} \to \mathbb{R}$ is defined by $g(x) = |x|^2 - 1$.

Definition 1.7.12 (**Coordinate system or coordinate chart; definition from Spivak [1971], p. 124 of pdf, p. 111 of book**)**.** Let $M \subset \mathbb{R}^n$, let $x \in M$, and let U be an open set containing x. Let $W \subset \mathbb{R}^k$. A **coordinate system** around x is a one-to-one differentiable function $f : W \to \mathbb{R}^n$ such that

1. $f(W) = M \cap U$,

2. $f'(y)$ has rank k for each $y \in W$, and

3. $f^{-1} : f(W) \to W$ is continuous.

See Figure 1.6.

Definition 1.7.13 (**Coordinate system or coordinate chart; definition from Lee [2012], p. 19 of pdf, p. 4 of book**)**.** Let M be a topological n-manifold (see Definition 1.7.11). A **coordinate chart** (or just a **chart**) on M is a pair (U, ϕ), where U is an open subset of M and $\phi : U \to \hat{U}$ is a homeomorphism from U to an open subset $\hat{U} = \phi(U) \subseteq \mathbb{R}^n$ (see Figure 1.5). By the definition of a topological manifold, each point $p \in M$ is contained in the domain of some chart (U, ϕ).

If $\phi(p) = 0$, we say the chart is **centered at** p. If (U, ϕ) is any chart whose domain contains p, it is easy to obtain a new chart centered at p by subtracting the constant vector $\phi(p)$.
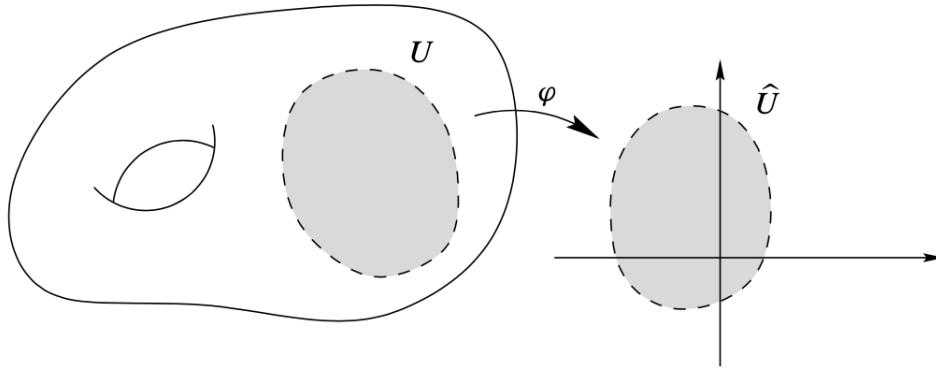
Fig. 1.2 A coordinate chart

Figure 1.5: Figure 1.2 in Lee [2012]; a coordinate chart.

Given a chart (U, ϕ), we call the set U a **coordinate domain**, or a **coordinate neighborhood** of each of its points.

The following alternative characterization of manifolds is important.

Theorem 119 (**Theorem 5-2 in Spivak [1971], p. 124 of pdf, p. 111 of book**). Let $M \subset \mathbb{R}^n$, let $x \in M$, and let U be an open set containing x. Let $W \subset \mathbb{R}^k$. M is a k-dimensional manifold if and only if for each $x \in M$ there is an open set U containing x and an open set $W \subset \mathbb{R}^k$ with a coordinate chart $f : W \to \mathbb{R}^n$.

1.7.3   Differentiability of functions from $\mathbb{R}^n \to \mathbb{R}^m$ (Section 5.2 of Pugh [2015])

(See also Section **??** for more on this material.)

Definition 1.7.14 (**Tangent space; from Section 4.2 of Spivak [1971], p. 86 of book, p. 99 of pdf**). Let $p \in \mathbb{R}^n$. The set of all pairs (p, v) for $v \in \mathbb{R}^n$ is denoted $\mathbb{R}^n_p$ and called the **tangent space** of $\mathbb{R}^n$ at p. This set is made into a vector space by defining

$$(p, v) + (p, w) = (p, v + w), \qquad a \cdot (p, v) = (p, av)$$

for $v, w \in \mathbb{R}^n$ and $a \in \mathbb{R}$. In 425b, we use the notation $(T\mathbb{R}^n)_p$.

Definition 1.7.15 (**Tangent space; from Section 3.1 of Lee [2012], p. 66 of pdf, p. 51 of book**). Given a point $a \in \mathbb{R}^n$, define the **geometric tangent space to** $\mathbb{R}^n$ **at** a, denoted by $\mathbb{R}^n_a$, to be the set $\{a\} \times \mathbb{R}^n = \{(a, v) : v \in \mathbb{R}^n\}$. A **geometric tangent vector** in $\mathbb{R}^n$ is an element of $\mathbb{R}^n_a$ for some $a \in \mathbb{R}^n$.

As a matter of notation, we abbreviate (a, v) as v_a or sometimes $v|_a$. n 425b, we use the notation $(T\mathbb{R}^n)_a$.

We think of the vector v_a as the vector v with its initial point at a. The set $\mathbb{R}^n_a$ is a real vector space under the natural operations $v_a + w_a = (v + w)_a$ and $c(v_a) = (cv)_a$. The vectors $e_i|_a, i \in [n]$, are a basis for $\mathbb{R}^n_a$.
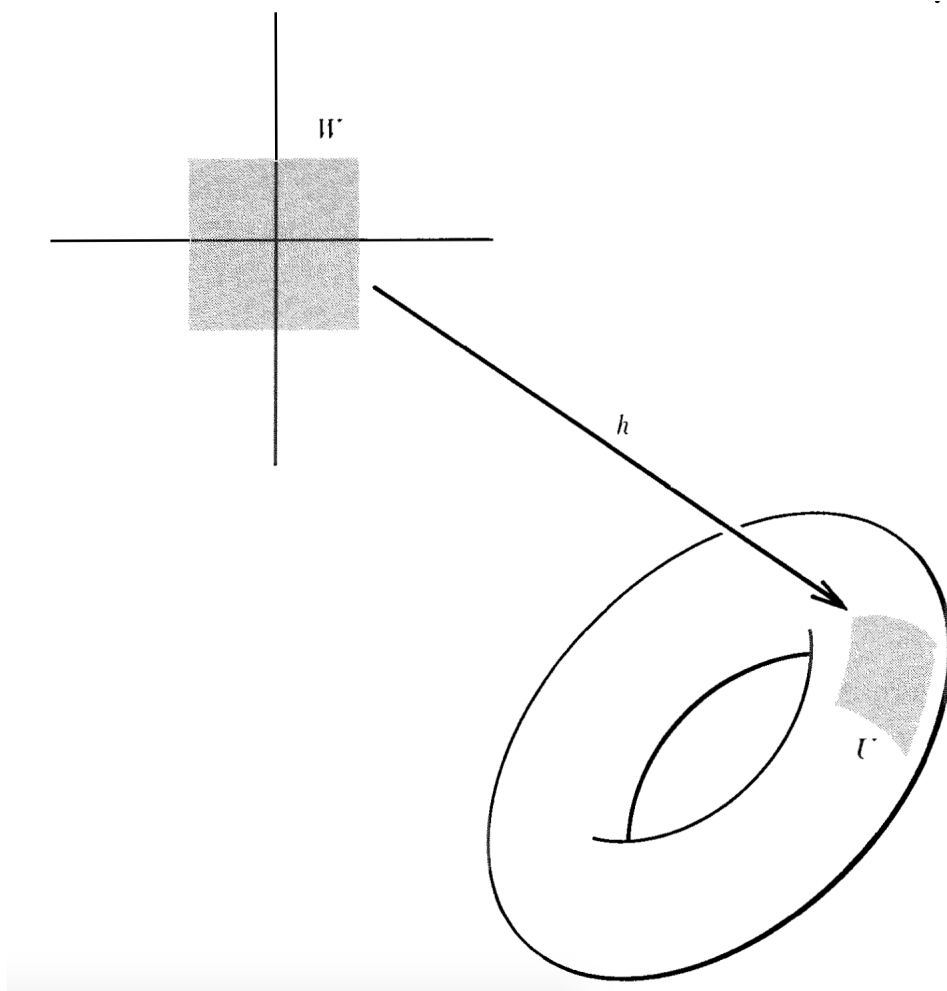
Figure 1.6: Figure 5-2 in Spivak [1971], for Theorem 119.

Any operations which is possible in a vector space may be performed in each $\mathbb{R}_p^n$. If a selection of a vector is made in each $\mathbb{R}_p^n$, we obtain a **vector field**.

Definition 1.7.16 (Vector field; Section 4.2 of Spivak [1971], p. 100 of pdf, p. 87 of book).
A **vector field** is a function F such that $F(p) \in \mathbb{R}_p^n$ for each $p \in \mathbb{R}^n$. For each p there are numbers $F^1(p), \ldots, F^n(p)$ such that

$$F(p) = F^1(p) \cdot (e_1)_p + \ldots + F^n(p) \cdot (e_n)_p.$$

We thus obtain n **component functions** $F^i : \mathbb{R}^n \to \mathbb{R}$. The vector field is called continuous, differentiable, etc. if the functions F^i are. Similar definitions can be made for a vector field defined only on an open subset of $\mathbb{R}^n$. Operations on vectors yield operations on vector fields when applied at each point separately. For example, if F and G are vector fields and f is a function, we define

$$(F + G)(p) = F(p) + G(p),$$
$$\langle F, G \rangle(p) = \langle F(p), G(p) \rangle,$$
$$(f \cdot F)(p) = f(p)F(p).$$

If $F_1, \ldots, F_{n-1}$ are vector fields on $\mathbb{R}^n$, then we can similarly define

$$(F_1 \times \ldots \times F_{n-1})(p) = F_1(p) \times \ldots \times F_{n-1}(p).$$

Matrix for $(DF)_p$ in standard bases for $\mathbb{R}^n$, $\mathbb{R}^m$ (**Jacobian**):

$$\begin{bmatrix} \frac{\partial F_1}{\partial x_1}(p) & \cdots & \frac{\partial F_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1}(p) & \cdots & \frac{\partial F_m}{\partial x_n}(p) \end{bmatrix}$$

We can use this as the definition and we'd be fine for C^1 functions (note: haven't yet defined C classes for multivariate functions). Important to know about another approach based on an idea of linear approximations.

Definition 1.7.17 (Total derivative; Section 5.2 of Pugh [2015], p. 282 (p. 292 of pdf)). Let $f : U \to \mathbb{R}^m$ be given where U is an open subset of $\mathbb{R}^n$. The function F is **differentiable** at $p \in U$ with **(total) derivative** (or **Frechet derivative**] $(DF)_p = T$ if $T\mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and

$$\lim_{\|v\| \to 0} \frac{F(p + v) - F(p) - T(v)}{\|v\|} = 0;$$

that is, (book definition)

$$R(v) := f(p + v) - f(p) - T(v) \qquad \implies \qquad \lim_{\|v\| \to 0} \frac{R(v)}{\|v\|} = 0.$$

Proposition 120 (Similar to Theorem 5.5 in Pugh [2015]). Let $U \subset \mathbb{R}^n$ be open. Let $F : U \to \mathbb{R}^m$ be a function and let $p \in U$. Let $F = \begin{bmatrix} F_1 \\ \vdots \\ F_m \end{bmatrix}$, $F_i : U \to \mathbb{R}$. Then

1. F is differentiable at p with derivative $T(v) = Av$ ($A \in \mathrm{Mat}_{m \times n}(\mathbb{R})$), if and only if

2. Each f_i is differentiable at p with derivative $T_i(v) = \begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$.

Proof. (1) is true if and only if $\lim_{\|v\| \to 0} \frac{F(p+v)-F(p)-T(v)}{\|v\|} = 0$, which is true if and only if each coordinate of $\frac{F(p+v)-F(p)-T(v)}{\|v\|} \to 0$ as $v \to 0$, which is true if and only if (2) is true.

$\square$

Next we will talk about uniqueness. We want to show that $T = (DF)_p$ is unique if it exists. This is easy if we use the chain rule (the book is quicker and more direct but a bit redundant).

Proposition 121 (Differentiability implies continuity; Theorem 5.6 in Pugh [2015]). Let $U \subset \mathbb{R}^n$ be open. Let $p \in U$, $F : U \to \mathbb{R}^m$. If F is differentiable at p with derivative T, then F is continuous at p.

Proof. We have $F(p+v) = F(p)+T(v)+R_F(v)$, where $\lim_{v \to 0} \frac{\|R_F(v)\|}{\|v\|} = 0$. This implies $\lim_{v \to 0} \|R_F(v)\| = 0$ (note: $\|R_F(v)\| \le \frac{\|R_f(v)\|}{\|v\|}$ for $\|v\| \le 1$). Therefore $\lim_{v \to 0} F(p+v) = F(p)+0+0$; that is, for every $\epsilon > 0$ there exists $\delta > 0$ such that $\|v\|_2 < \delta \implies \|F(p+v)-F(p)\|_2 < \epsilon$, so F is continuous at p.

$\square$

Theorem 122 (Chain rule; Theorem 5.9(c) in Pugh [2015]). Let $U \subset \mathbb{R}^n$ be open. Let $V \subset \mathbb{R}^m$ be open. Let $F : U \to V$ and let $G : V \to \mathbb{R}^k$ be functions such that (1) F is differentiable at p (with derivative A) and (2) G is differentiable at $F(p) = q \in V$ (with derivative B). Then $G \circ F$ is differentiable at p with derivative $B \circ A$.

Note that if the derivative is unique, then $(D(G \circ F))_p = (DG)_{f(p)} \circ (DF)_p$ (composing then linearly approximating is the same as linearly approximating then composing).

Proof. Let $R_F = R_{F,p,A}$, a function of $v \in \mathbb{R}^n$. $R_F(v) = F(p+v) - F(p) - Av$, so

$$F(p+v) = F(p) + Av + R_F(v). \tag{1.26}$$

Let $R_G = R_{G,q,B}$, a function of $w \in \mathbb{R}^m$. $R_G(w) = G(q+w) - G(q) - Bw$, so

$$G(q+w) = G(q) + Bw + R_G(w). \tag{1.27}$$

Let $R_{G \circ F} = R_{G \circ F, p, BA}$. We want to show that $\lim_{v \to 0} \frac{R_{G \circ F}}{\|v\|} = 0$.

We have

$$R_{G \circ F}(v) = (G \circ F)(p + v) - (G \circ F)(p) - BAv$$

$$= G(F(p + v)) - G(\underbrace{F(p)}_{=q}) - BAv$$

(by (1.26)) $$= G(F(p) + Av + R_F(v)) - G(q) - BAv$$

(by (1.27) and $F(p) = q$) $$= G(q) + B(Av + R_F(v)) + R_G(Av + R_F(v)) - G(q) - BAv$$

$$= BR_F(v) + R_G(Av + R_F(v)).$$

Thus,

$$\frac{\|R_{G \circ F}(v)\|}{\|v\|} \leq \frac{\|BR_F(v)\|}{\|v\|} + \frac{\|R_G(Av + R_F(v))\|}{\|v\|}. \tag{1.28}$$

We want to show both terms on the right in (1.28) go to 0 as $v \to 0$. First,

$$\frac{\|BR_F(v)\|}{\|v\|} \leq \|B\|_{\mathrm{op}} \frac{\|R_F(v)\|}{\|v\|} \to 0$$

as $v \to 0$ by differentiability of F (since $\|B\|_{\mathrm{op}} < \infty$). Next, for v with $Av + R_F(v) = 0$, we have

$$R_G(Av + R_F(v)) = R_G(0) = 0$$

since $R_G(w) = G(q + w) - G(q) - Bw \implies R_G(0) = G(q) - G(q) - 0 = 0$. Thus the second term in (1.28) is 0 if $v \neq 0$ and $Av + R_V(v) = 0$. For x with $Av + R_F(v) \neq 0$, we have

$$\frac{\|R_G(Av + R_F(v))\|}{\|v\|} = \frac{\|R_G(Av + R_F(v))\|}{\|Av + R_F(v)\|} \frac{\|Av + R_F(v)\|}{\|v\|}$$

As $v \to 0$ we have $Av + R_F(v) \to 0$, since matrix multiplication by A is continuous and $R_F(v) = F(p + v) - F(p) - Av$ is continuous at $v = 0$ (since differentiability of F at p implies continuity of F at p by Proposition 121). Since $\lim_{v \to 0}(Av + R_F(v)) = 0$ and $\lim_{w \to 0} \frac{\|R_G(w)\|}{\|w\|} = 0$, we have

$$\lim_{v \to 0} \frac{\|R_G(Av + R_F(v))\|}{\|Av + R_F(v)\|} = 0.$$

From here it is enough to show that $\lim_{v \to 0} \frac{\|Av + R_F(v)\|}{\|v\|}$ is bounded. Observe that

$$\frac{\|Av + R_F(v)\|}{\|v\|} \leq \frac{\|Av\|}{\|v\|} + \frac{\|R_F(v)\|}{\|v\|} \leq \|A\|_{\mathrm{op}} + 0.$$

$\square$

An important geometric interpretation that gives us the uniqueness of derivatives: $(DF)p$ is determined by how it acts on velocity vectors of curves (see notes).

Corollary 122.1. Let $U \subset \mathbb{R}^n$ be open. Let $p \in U$, $F : U \to \mathbb{R}^m$. Let $0 \in (a, b)$ and let $\gamma : (a, b) \to U$ be a differentiable function (curve) with $\gamma(t_0) = p$. Then $(DF)_p(\gamma'(t_0)) = (F \circ \gamma)'(t_0)$.

Proof. By the chain rule (Theorem 122), $F \circ \gamma$ is differentiable at t_0 with $(D(F \circ \gamma))_{t_0} = (DF)_{\gamma(t_0)}(D\gamma)_{t_0} = (DF)_p(D\gamma)_{t_0}$. Viewed as a $m \times 1$ matrix, we showed last time that $(D(F \circ \gamma))_{t_0} = (F \circ \gamma)'(t_0)$. Viewed as an $n \times 1$ matrix, we showed that $(D\gamma)_{t_0} = \gamma'(t_0)$.

$\square$

Corollary 122.2. If F is differentiable at $\gamma(t_0)) = p$ with derivative $T = (DF)_p$, then

$$T(v) = \lim_{t \to 0} \frac{F(p + tv) - F(p)}{t}.$$

(since the right side is independent of T, this shows T is unique.)

Proof. Let $\gamma(t) = p + tv$; then the right hand side is $(F \circ \gamma)'(t_0)$ (using $\gamma(t_0) = p$).

$\square$

If $\lim_{t \to 0} \frac{f(p+tv) - F(p)}{t}$ exists for all v (this is $(DF)_p(v)$) and this quantity is linear in v, is F (Frechet) differentiable at p? No. For example, $F : \mathbb{R}^2 \to \mathbb{R}$,

$$F(x, y) = \begin{cases} \frac{x^3 y}{x^4 + y^2}, & (c, y) \neq (0, 0), \\ (0, 0), & (x, y) = (0, 0). \end{cases}$$

Exercise (Pugh [2015] ex. 18 and 19, ch 5; more discussion on why to prefer Frechet): directional derivative exists and equals 0 for all v but F is not differentiable at $(0, 0)$.

Proposition 123 (Theorem 5.9(a) and 5.9(b) in Pugh [2015]). Let $U \subset \mathbb{R}^n$ be open. Let $p \in U$, $F, G : U \to \mathbb{R}^m$.

(a) If F and G are differentiable at p then $F + G$ is differentiable at p with $D(F + G)p = (DF)p + (DG)p$.

(b) If $c \in \mathbb{R}$ then cF is differentiable at p with $D(cF)_p = c(DF)_p$.

(c) Any constant function $F : U \to \mathbb{R}^m$ is differentiable with derivative 0 everywhere.

(d) If $F : U \to \mathbb{R}^m$ is affine linear ($F(x) = Ax + y_0$ for some $A \in \mathbb{R}^{m \times n}$, $y_0 \in \mathbb{R}^m$) then F is differentiable at all points of U with derivative A.

Proof. (a)

$$\frac{\|(F + G)(p + v) - (F + G)(p) - ((DF)_p + (DG)_p)(v)\|}{\|v\|}$$

$$\leq \frac{\|F(p + v) - F(p) - (DF)_p(v)\|}{\|v\|} + \frac{\|G(p + v) - G(p) - (DG)_p(v)\|}{\|v\|}.$$

Each of these terms goes to 0 as $v \to 0$ by differentiability of F and G.

(b)

$$\frac{\|(cF)(p+v) - (cF)(p) - D[cF])_p(v)\|}{\|v\|} \leq c \frac{\|F(p+v) - F(p) - (DF)_p(v)\|}{\|v\|}$$

which goes to 0 as $v \to 0$ by differentiability of F.

(c)

(d) If $F(x) = Ax + y_0$ then

$$\frac{F(p+v) - F(p) - Av}{\|v\|} = \frac{A(p+v) - A(p) - A(v)}{\|v\|} = \frac{0}{\|v\|} \to 0$$

as $v \to 0$.

$\square$

Definition 1.7.18. Let $U \subset \mathbb{R}^n$ be open. Let $p \in U$ and let $F : U \to \mathbb{R}^m$. Let $1 \leq i \leq n$. The ith **partial derivative** of F at p, $\frac{\partial F}{\partial x_i}(p)$, is defined to be

$$\frac{\partial F}{\partial x_i}(p) := \lim_{t \to 0} \frac{F(p + te_i) - F(p)}{t}$$

(if it exists) where e_i is the ith standard basis vector for $\mathbb{R}^n$ (the directional derivative of F in the direction e_i). If F is differentiable, it also equals $(DF)_p(e_i) = (F \circ \gamma)'(0)$ where $\gamma(t) = p + te_i$.

Remark 12. If $F : U \to \mathbb{R}^m$ and $\frac{\partial F}{\partial x_i}(p)$ exists for all $p \in U$, then $\frac{\partial F}{\partial x_i} : U \to \mathbb{R}^m$ (just the F). This implies we can (try to) take successive derivatives, e.g., we can do

$$\frac{\partial^2 F}{\partial x_j \partial x_i} := \frac{\partial}{\partial x_j}\left(\frac{\partial F}{\partial x_i}\right),$$

etc., and keep getting functions from U to $\mathbb{R}^m$ as long as the partial derivatives keep existing.

Definition 1.7.19. Let $F : U \to \mathbb{R}^m$ is of **class** C^k (we can also say F is a C^k **function**) if all partial derivatives of F of orders $\leq k$ exist and are continuous in U. (We use this notation with $k = \infty$ if this is true for all orders, not "infinite orders.") We call C^∞ functions **smooth**.

Theorem 124. If $F : U \to \mathbb{R}^m$ is of class C^1, then F is differentiable at all $p \in U$ and $(DF)_p : \mathbb{R}^n \to \mathbb{R}^m$ has standard basis matrix

$$\begin{bmatrix} \frac{\partial F_1}{\partial x_1}(p) & \cdots & \frac{\partial F_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1}(p) & \cdots & \frac{\partial F_m}{\partial x_n}(p) \end{bmatrix} \tag{1.29}$$

Remark 13. Note that the columns are $\frac{\partial F}{\partial x_i}(p)$ for $i \in [n]$. This makes sense: if $\lim_{t \to 0} \frac{F(p+v) - F(p)}{t}$ is linear in v, then the ith column of its standard basis matrix should be obtained by plugging $v = e_i$ into $\lim_{t \to 0} \frac{F(p+v) - F(p)}{t}$.

Proof. Let J be the matrix (1.29). Let $T(v) = Jv$ ($T : \mathbb{R}^n \to \mathbb{R}^m$ linear). Let $R(v) = F(p+v) - F(p) - Jv$. For $i \in [m]$, let $R_i(v)$ be the ith coordinate of $R(v)$. Then

$$R_i(v) = F_i(p+v) - F_i(p) - \begin{bmatrix} \frac{\partial F_i}{\partial x_1}(p) & \cdots & \frac{\partial F_i}{\partial x_n}(p) \end{bmatrix} v.$$

It suffices to show $\frac{R_i(v)}{\|v\|} \to 0$ as $v \to 0$ for $i \in [m]$. (Then each point of $\frac{R(v)}{\|v\|} \to 0$, so $\frac{R(v)}{\|v\|} \to 0$ as $v \to 0$.) Given $\epsilon > 0$, choose $\delta > 0$ such that if $\|v\| < \delta$ then (a) $p + v \in U$ and b $\left| \frac{\partial F_i}{\partial x_j}(p+d) - \frac{\partial F_i}{\partial x_j}(p) \right| < \epsilon/n$ for $j \in [n]$ (i is fixed). (Recall that using the C^1 assumption, $\frac{\partial F_i}{\partial x_j}$ is continuous at p for $j \in [n]$.)

We claim that if $\|v\| < \delta$ then $\frac{|R_i(v)|}{\|v\|} < \epsilon$, proving the theorem. To see this, let v be fixed with $\|v\| < \delta$ ($v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$). We have

$$R_i(v) = F_i(p+v) - F_i(p) - V_1 \frac{\partial F_i}{\partial x_1}(p) - \ldots - v_n \frac{\partial F_i}{\partial x_n}(p).$$

We can write

$$
\begin{aligned}
F_i(p+v) - F_i(p) =\ & F_i(p + V_1 e_1 + \ldots + v_n e_n) - F_i(p) \\
=\ & F_i(p + v_1 e_1 + \ldots + v_n e_n) - F_i(p + V_1 e_1 + \ldots + v_{n-1} e_{n-1}) \\
& + F_i(p + v_1 e_1 + \ldots + v_{n-1} e_{n-1}) - F_i(p + v_1 e_1 + \ldots + v_{n-2} e_{n-2}) \\
& + F_i(p + v_1 e_1 + \ldots + v_{n-2} e_{n-2}) - \ldots + F_i(p + v_1 e_1) - F_i(p)
\end{aligned}
$$

It suffices to show

$$\frac{|F_i(p + v_1 e_1 + \ldots + v_j e_j) - F_i(p + V_1 e_1 + \ldots + v_{j-1} e_{j-1})|}{\|v\|} < \epsilon/n$$

for $2 \leq j \leq n$. We will use the one-variable MVT. First: if $v_j = 0$, then $p_j = p_{j-1}$ and the whole thing is 0. Assume note. Use a path σ_j starting at p_{j-1} and ending at p_j (these paths σ form the "staircase path" from p to $p+v$ shown in Figure 108, p. 285 (p. 295 of pdf) in Pugh [2015]). Indeed: let $\sigma_j(t) := p_{j-1} + tv_j e_j$ for $t \in [0,1]$. $F_i \circ \sigma_j(t) = F_i(p_{j-1} + tv_j e_j)$; derivative quotient is

$$
\begin{aligned}
\lim_{t' \to t} \frac{F_i(p_{j-1} + t' v_j e_j) - F_i(p_{j-1} + t v_j e_j)}{t' - t} &= v_j \lim_{t' \to t} \frac{F_i(p_{j-1} + t' v_j e_j) - F_i(p_{j-1} + t v_j e_j)}{v_j t' - v_j t} \\
&= \lim_{t' \to t} \frac{F_i(p_{j-1} + t' v_j e_j) - F_i(p_{j-1} + t v_j e_j)}{t' - t} \\
&= \frac{\partial F_i}{\partial x_j}(p_{j-1} + t e_j)
\end{aligned}
$$

since $t' \to t \iff v_j t' \to v_j t$. $\frac{\partial F_i}{\partial x_j}(p_{j-1} + t e_j)$ exists for all $t \in [0,1]$, so we get $F_i \circ \sigma_j$ is a differentiable function from $[0,1] \to \mathbb{R}$ with derivative $v_j \frac{\partial F_i}{\partial x_j}(p_{j-1} + t e_j)$ at the point $t \in [0,1]$. Let $p_{ij} := p_{j-1} + t e_j$.

By the Mean Value Theorem, there exists $t \in [0,1]$ with $F_i(\sigma_j(1)) - F_i(\sigma_j(0)) = v_j \frac{\partial F_i}{\partial x_j}(p_{ij})$ $(\sigma_j(1) = p_j,$ $\sigma_j(0) = p_{j-1})$. Thus,

$$
\begin{aligned}
\frac{\left| F_i(p_j) - F_i(p_{j-1}) - \frac{\partial F_i}{\partial x_j}(p)v_j \right|}{\|v\|} &= \frac{1}{\|v\|} \left| v_j \left(\frac{\partial F_i}{\partial x_j}(p_{ij}) - \frac{\partial F_i}{\partial x_j}(p) \right) \right| \\
&\leq \left| \frac{\partial F_i}{\partial x_j}(p_{ij}) - \frac{\partial F_i}{\partial x_j}(p) \right| \\
&< \frac{\epsilon}{n}
\end{aligned}
$$

where we used $|v_i|/\|v\| \leq 1$, and the last step follows since $\|p_{ij} - p\| < \delta$ (indeed: $p_{ij} - p = v_1 e_1 + \ldots + v_{j-1}e_{j-1} + tv_j e_j$ for $t \in [0,1]$, and $v = v_1 e_1 + \ldots + v_n e_n$, so $\|p_{ij} - p_j\| \leq \|v\| < \delta$, finishing the proof.

$\square$

Example 1.7.3. Define $F : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$
F(x,y,z) := \begin{bmatrix} e^z \cos(y) \\ e^{-z} \sin(y) \\ x^2 + y^2 + z^3 \end{bmatrix}.
$$

Is F differentiable? Yes, F is C^1 (even C^∞) by inspection of each component. In particular,

$$
(DF)_{(x,y,z)} = \begin{bmatrix} 0 & -e^z \sin(y) & e^z \cos(y) \\ 0 & e^{-z} \cos(y) & -e^{-z} \sin(y) \\ 2x & 2y & 2z \end{bmatrix}.
$$

Theorem 125 (Mean Value Theorem for functions from $\mathbb{R}^n$ to $\mathbb{R}$). Let $U \subset \mathbb{R}^n$ be open. Let $F : U \to \mathbb{R}$ be differentiable. Let $p, q \in U$ such that $p + t(q - p) \in U$ for $t \in [0,1]$. Then there exists $\theta \in [0,1]$ with $F(q) - F(p) = (DF)_{p+\theta(q-p)}(q - p)$.

Proof. Apply the one-variable Mean Value Theorem to $F(p + t(q - p))$, a differentiable function from $[0,1] \to \mathbb{R}$. The derivative at θ is $(DF)_{p+\theta(q-p)}(q - p)$. Then

$$
F(\underbrace{p + 1(q - p)}_{q}) - F(\underbrace{p + 0(q - p)}_{p}) = \big((DF)_{p+\theta(q-p)}(q - p)\big)(1 - 0).
$$

$\square$

On the other hand, the case of vector-valued functions isn't so simple.

Example 1.7.4. Consider $F : \mathbb{R} \to \mathbb{R}^2$ defined by $F(t) = (\cos t, \sin t)$. Then observe that

$$
\underbrace{F(2\pi) - F(0)}_{0} =? (DF)_t(2\pi - 0) \qquad \text{for any } t?
$$

But

$$\begin{bmatrix} -\sin t \\ \cos(t) \end{bmatrix} \neq 0$$

for any t.

What do we do in this case?

Theorem 126 (**Mean value inequality; Theorem 5.11 in Pugh [2015], p. 298 of pdf, p. 288 of book**)**.** Let $U \subset \mathbb{R}^n$ be open. Let $F : U \to \mathbb{R}^m$ be differentiable. Let $p, q \in U$ such that the line segment $\{p + t(q - p) \mid t \in [0, 1]\} \subset U$. Then

$$\|F(q) - F(p)\| \leq \left(\sup_{t \in [0,1]} \|(DF)_{p+t(q-p)}\|_{\mathrm{op}} \right) \|q - p\|$$

(Note that this is upper-bounded by

$$\left(\sup_{x \in U} \|(DF)_x\|_{\mathrm{op}} \right) \|q - p\|,$$

which is the version shown in Pugh [2015].)

Proof. If we can show that for some $M \in \mathbb{R}$ $\langle F(q) - F(p), u \rangle \leq M\|q - p\|$ for all unit vectors $u \in \mathbb{R}^m$, then we're done because we can take $u = (F(q) - F(p))/\|F(q) - F(p)\|$ and get the statement of the theorem. Let $v := q - p$ and let

$$g(t) := \langle F(p + tv), u \rangle = \sum_{i=1}^{m} u_i F_i(p + tv)$$

(if $u = (u_1, \ldots, u_m)^\top$ and $F = (F_1, \ldots, F_m)^\top$). Each F_i is differentiable, and $t \mapsto p + tv$ is differentiable, so $F_i(p + tv)$ is a differentiable function of t. Sums and scalar multiples of differentiable functions are differentiable by Proposition 123, so $g(t)$ is differentiable. In particular,

$$g'(t) = \sum_{i=1}^{m} u_m (DF_m)_{p+tv}(v) = \langle (DF)_{p+tv}(v), u \rangle.$$

(Note that this equation also results from the Leibniz rule, although we didn't prove the Leibniz rule.) By the one-dimensional Mean Value Theorem for g, there exists $\theta \in [0, 1]$ with

$$g'(\theta)(1 - 0) = g(1) - g(0) = \langle F(q), u \rangle - \langle F(p), u \rangle = \langle F(q) - F(p), u \rangle$$

Observe that

95

$$g'(\theta)(1-0) = \langle (DF)_{p+\theta v}(v), u \rangle \leq \text{(by Cauchy-Schwarz) } \|(DF)_{p+\theta v}(v)\| \underbrace{\|u\|}_{q} \leq \underbrace{\|(DF)_{p+\theta v}\|_{\text{op}}}_{\leq M} \underbrace{\|v\|}_{=\|q-p\|} .$$

So we've shown $\langle F(q) - F(p), u \rangle \leq M\|q - p\|$ for all unit vectors $u \in \mathbb{R}^m$, proving the theorem.

$\square$

Theorem 127 (C^1 **Mean Value Theorem: may do later**).

Corollary 127.1. If $U \subset \mathbb{R}^n$ is open and connected and $F : U \to \mathbb{R}^m$ is differentiable with $(DF)_x = 0$ for all $x \in U$, then F is constant.

Proof. Exercise, use topological fact: any two points in U can be connected by a differentiable path γ since U is connected.

$\square$

Theorem 128 (**Differentiation under the integral sign**). Let $f : [a, b] \times (c, d) \to \mathbb{R}$ (or $\mathbb{R}^m$) be a continuous function. Assume $\frac{\partial f}{\partial y}(x, y)$ is also continuous. Let $F(y) = \int_a^b f(x, y)dx$. Then F is differentiable on (c, d) with $F'(y) = \int_a^b \frac{\partial f}{\partial y}(x, y)dx$; i.e., $\frac{\partial}{\partial y} \int_a^b f(x, y)dx = \int_a^b \frac{\partial}{\partial y} f(x, y)dx$.

Remark 14. Better theorems exist under Lebesgue theory.

Proof. In the book, uses C^1 Mean Value Theorem. We'll skip for now.

$\square$

1.7.4   Implicit and Inverse Functions Theorems (Section 5.4 of Pugh [2015])

We will start with the Implicit Function Theorem and use it to deduce the Inverse Function Theorem. Here is the basic idea: say we have $n + m$ variables $x_1, \ldots, x_n, y_1, \ldots, y_m$, and we impose m nonlinear constraints: we look at points $(x, y) := (x_1, \ldots, x_n, y_1, \ldots, y_m)$ such that $F(x, y) = z$ where $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$ is some function and $z \in \mathbb{R}^m$. (F has coordinates $F_1, \ldots, F_m$; $F(x, y) = z$ amounts to m scalar equations. So we have m constraints $F_i(x, y) = z_i$, $i \in [m]$.)

Example 1.7.5. Let $n = 1, m = 1, F(x, y) = x^2 + y^2, z = 1$. Then $(x, y) \in \mathbb{R}^2$ must lie on the unit circle.

Let $(x_0, y_0) \in F^{-1}(z)$ (the level set of F at level z). The main question for the Implicit Function Theorem is: locally near (x_0, y_0), can we write $F^{-1}(z)$ as a graph of y as a function of x (or x as a function of y)?

In this case, if $(x_0, y_0) \neq (\pm 1, 0)$, then we can write the local part of the circle as a graph of y as a function of x, namely $y = \pm\sqrt{1 - x^2}$ (pick which one depending on sign of y_0). However if $(x_0, y_0) \in \{(1, 0), (-1, 0)\}$, we can't express y locally as a function of x, although we can express x as a function of y (although not near $(0, \pm 1)$). So we have four open subsets of the circle where $F^{-1}(z)$ can be written locally as a graph.

Definition 1.7.20. Let $f : A \to B$ be a function. The **graph** of f is the set S of all pairs $(a, b) \in A \times B$ such that $b = f(a)$.

Theorem 129 (**Implicit Function Theorem (Dini 1876; Theorem 5.22 in** Pugh [2015])**).** Let $U \subset \mathbb{R}^{m+n}$ be open. Let $F : U \to \mathbb{R}^m$ be a class of C^r functions for $r \geq 1$. Let $(x_0, y_0) \in U$ ($x_0 \in \mathbb{R}^n, y_0 \in \mathbb{R}^m$), let $z_0 \in \mathbb{R}^m$ with $F(x_0, y_0) = z_0$, and $(DF)_{(x_0,y_0)} \in \mathbb{R}^{m \times (n+m)}$ have the form $\begin{bmatrix} A & B \end{bmatrix}$ with $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{m \times m}$ with B invertible[3]. Then there exists $r, \tau > 0$ such that $B_\tau(x_0) \times B_r(y_0) \subset U$ such that there exists a unique function $g : B_\tau(x_0) \to B_r(y_0)$ with $F^{-1}(z_0) \cap (B_\tau(x_0) \times B_r(y_0)) = \text{graph}(g) = \{(x, g(x)) : x \in B_\tau(x_0)\}$ (see Definition 1.7.20). This function g is of class C^r.

(For notational ease, assume without loss of generality that (x_0, y_0) is the origin in $\mathbb{R}^{n+m}$ and $z_0 = 0 \in \mathbb{R}^m$.)

We will prove some lemmas before proving this result.

Definition 1.7.21. A function g is **locally Lipschitz** at 0 if there exists L such that for $x \in B_\tau(0)$ small enough, we have $\|g(x) - g(0)\| \leq L\|x - 0\|$. (See also Theorem 103.)

Lemma 130. Let $U \subset \mathbb{R}^{m+n}$ be open. Let $F : U \to \mathbb{R}^m$ be a class of C^r functions for $r \geq 1$. Let $(\underbrace{0}_{\in \mathbb{R}^n}, \underbrace{0}_{\in \mathbb{R}^m}) \in U$, let $F^{-1}(\underbrace{0}_{\in \mathbb{R}^m}) = (\underbrace{0}_{\in \mathbb{R}^n}, \underbrace{0}_{\in \mathbb{R}^m})$, and $(DF)_{(0,0)} \in \mathbb{R}^{m \times (n+m)}$ have the form $\begin{bmatrix} A & B \end{bmatrix}$ with $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{m \times m}$ with $B = \begin{bmatrix} \frac{\partial F_i(0,0)}{\partial y_j} \end{bmatrix}$ invertible (equivalently, the linear transformation represented by B is an isomorphism from $\mathbb{R}^m$ to $\mathbb{R}^m$). Then there exists $r, \tau > 0$ such that $B_\tau(0) \times B_r(0) \subset U$ such that there exists a unique function $g : \underbrace{B_\tau(0)}_{\subset \mathbb{R}^n} \to \underbrace{B_r(0)}_{\subset \mathbb{R}^m}$ with $F^{-1}(0) \cap (B_\tau(0) \times B_r(0)) = \text{graph}(g) = \{(x, g(x)) : x \in B_\tau(x_0)\}$. This function g is locally Lipschitz at 0.

Proof. For any $(x, y) \in U$, let

$$R(x, y) := F(x, y) - F(0, 0) - \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$= F(x, y) - F(0, 0) - Ax - By$$

$$\implies \quad F(x, y) = Ax + By + R(x, y),$$

where $A := \begin{bmatrix} \frac{\partial F_i(x_0, y_0)}{\partial x_j} \end{bmatrix}$ and the last expression is the Taylor expression for F so R is sublinear. Then for $(x, y) \in U$, we have

$$F(x, y) = 0 \quad \Longleftrightarrow \quad Ax + By + R(x, y) = 0$$

$$\Longleftrightarrow \quad By = -Ax - R(x, y)$$

$$\Longleftrightarrow \quad y = -B^{-1}(Ax + R(x, y)). \tag{1.30}$$

If R does not depend on y, then (1.30) is an explicit formula for $g(x)$. In general, we hope to show that R depends very weakly on y, so that we can switch it to the left side of (1.30), absorbing it in the y-term.

Choose $r > 0$ such that $\underbrace{\overline{B_r(0)}}_{\subset \mathbb{R}^n} \times \underbrace{\overline{B_r(0)}}_{\subset \mathbb{R}^m} \subset U$. Let $K_x : \underbrace{\overline{B_r(0)}}_{\subset \mathbb{R}^m} \to \mathbb{R}^m$ be defined by $K_x(y) := -B^{-1}(Ax + R(x, y))$ (the right side of (1.30)) for a fixed $x \in B_r(0) \subset \mathbb{R}^n$.

---

[3]Note that any full rank matrix of size $m \times (m + n)$ can have this form, possibly after permuting columns.

In particular, we would like to find a fixed point of $K_x(y) := -B^{-1}(Ax + R(x,y))$, so we hope to show K_x contracts and that K_x maps a complete metric space onto itself (namely $\overline{B_r(0)}$) (so that we can apply the Banach Contraction Principle, Theorem 100). Note that $R(x,y)$ is a C^r function of (x,y) since F is. Also, $(DR)_{(0,0)} = 0$.

Let $\frac{\partial R}{\partial y}(x,y)$ be the "partial Jacobian" of R, using only the partial derivatives from y (so it has M columns—a function from U to $\mathbb{R}^{m \times m}$). Because R is at least C^1, every entry of $\frac{\partial R}{\partial y}(x,y)$ is continuous. Then by Corollary 114.3, we have that $\frac{\partial R}{\partial y}(x,y)$ is continuous if we use the operator norm to as the metric in $\mathbb{R}^{m \times m}$ (see the second part of Corollary 114.3 for the formal statement; written in shorthand notation as $\frac{\partial R}{\partial y}(x,y) : U \mapsto (\mathbb{R}^{m \times m}, \|\cdot\|_{\text{op}})$ is continuous). Therefore there exists $r > 0$ such that

$$(x,y) \in U, \|x\| \le r, \|y\| \le r \quad \implies \quad \left\| \frac{\partial R}{\partial y}(x,y) \right\| < \frac{1}{2\|B^{-1}\|_{\text{op}}}. \tag{1.31}$$

For notational ease, let $B_{(x,y)} := \left[\frac{\partial F_i(x,y)}{\partial y_j} \right] = \frac{\partial F}{\partial y}(x,y)$ (i.e., $(DF)_{(x,y)} = \begin{bmatrix} A_{(x,y)} & B_{(x,y)} \end{bmatrix}$). Note that $\det(B_{(0,0)}) \ne 0$ since B is invertible at $(0,0)$, and $\det : \mathbb{R}^{m \times m} \mapsto \mathbb{R}$ is a continuous function, so for sufficiently small r we have that $\det(B_{(x,y)}) \ne 0$ if $\|x\|, \|y\| \le r$.

Note that for $\|x\|, \|y_1\|, \|y_2\| \le r$ we have

$$\begin{aligned}
\|K_x(y_1) - K_x(y_2)\| &= \|-B^{-1}(Ax + R(x,y_1)) - [-B^{-1}(Ax + R(x,y_2))]\| \\
&= \|-B^{-1}[R(x,y_1) - R(x,y_2)]\| \\
&\le \|B^{-1}\|_{\text{op}} \|R(x,y_1) - R(x,y_2)\| \\
\text{(for some } t \in [0,1] \text{ by the MVT)} \quad &\le \|B^{-1}\|_{\text{op}} \left\| \frac{\partial R}{\partial x}(x, y_1 + t(y_2 - y_1)) \cdot (x - x) \right. \\
&\quad \left. + \frac{\partial R}{\partial y}(x, y_1 + t(y_2 - y_1)) \cdot (y_1 - y_2) \right\| \\
&\le \|B^{-1}\|_{\text{op}} \left\| \frac{\partial R}{\partial y}(x, y_1 + t(y_2 - y_1)) \right\|_{\text{op}} \|y_1 - y_2\| \\
\text{(by (1.31))} \quad &< \frac{1}{2} \|y_1 - y_2\|. \tag{1.32}
\end{aligned}$$

Therefore $K_x(y)$ brings points closer together. We will also show that it maps $\overline{B_r(0)}$ into itself. We would like to show that for sufficiently small x, the function $x \mapsto K_x(0) = -B^{-1}(Ax + R(x,0))$ makes y smaller. Since $x \mapsto K_x(0) = -B^{-1}(Ax + R(x,0))$ is continuous in x, there exists $\tau > 0$ such that if $\|x\| \le \tau$ then $\|K_x(0)\| \le r/2$. Then if $\|x\| \le \tau$ and $\|y\| \le r$, we have

$$\begin{aligned}
\|K_x(y)\| &= \|K_x(y) - K_x(0) + K_x(0)\| \\
&\le \|K_x(y) - K_x(0)\| + \|K_x(0)\| \\
\text{(by (1.32))} \quad &\le \frac{1}{2} \underbrace{\|y - 0\|}_{\le r} + \underbrace{\|K_x(0)\|}_{\le r/2} \\
&\le r. \tag{1.33}
\end{aligned}$$

We have established that if $\|x\| \leq \tau$ then K_x maps $\overline{B_r(0)} \subset \mathbb{R}^m$ into $\overline{B_r(0)} \subset \mathbb{R}^m$ and is a contraction. Since $\overline{B_r(0)} \subset \mathbb{R}^m$ is compact and therefore complete, we can now apply the Banach Contraction Principle, Theorem 100, to conclude that K_x has a unique fixed point in $\overline{B_r(0)}$. Call this point $g(x)$. By construction, $F^{-1}(0) \cap (\overline{B_\tau(0)} \times \overline{B_r(0)})$ is graph(g) (see Definiton 1.7.20).

Next, we will find a local Lipschitz constant for g. Note that $g(0)$, the unique fixed point of $K_0(y)$, is 0, since $(0,0)$ is the unique point in $(\overline{B_\tau(0)} \times \overline{B_r(0)}) \cap F^{-1}(0)$ with x coordinate 0. Also note that since R is sublinear, for small enough x we have $\|R(x,0)\| \leq \|A\|_{\text{op}}\|x\|$. Since $g(x) = K_x(g(x))$, for small enough x we have

$$
\begin{aligned}
\|g(x)\| &= \|K_x(g(x)) - K_x(g(0)) + K_x(g(0))\| \\
&\leq \|K_x(g(x)) - K_x(g(0))\| + \|K_x(0)\| \\
\text{(by (1.32) and def of } K_x(0)) \quad &\leq \frac{1}{2}\|g(x) - g(0)\| + \|B^{-1}(Ax + R(x,0))\| \\
&\leq \frac{1}{2}\|g(x)\| + \|B^{-1}\|_{\text{op}}\|Ax + R(x,0)\| \\
\text{(using the fact about small } x) \quad &\leq \frac{1}{2}\|g(x)\| + \|B^{-1}\|_{\text{op}} \cdot 2\|A\|_{\text{op}}\|x\| \\
\Longleftrightarrow \quad \|g(x)\| - \frac{1}{2}\|g(x)\| &\leq 2\|B^{-1}\|_{\text{op}}\|A\|_{\text{op}}\|x\| \\
\Longleftrightarrow \quad \|g(x)\| &\leq 4\|B^{-1}\|_{\text{op}}\|A\|_{\text{op}}\|x\|,
\end{aligned}
$$

so for small enough x, $4\|B^{-1}\|_{\text{op}}\|A\|_{\text{op}}$ is a local Lipschitz constant for g. Let $L := \|B^{-1}\|_{\text{op}}\|A\|_{\text{op}}$ so we can write the constant as $4L$. This implies continuity of g at 0, so there exists $\tau > 0$ (smaller than before) such that $g(B_\tau(0)) \subset B_r(0)$ (open balls). By construction, we have $F^{-1}(0) \cap (B_\tau(0) \times B_r(0)) = \text{graph}(g)$, because for $(x,y) \in B_\tau(0) \times B_r(0)$ we have $F(x,y) = 0 \iff y = g(x)$. Since two functions with the same graph are equal, g is the unique function with this property.

$\square$

Lemma 131. In Lemma 130, the unique function g is differentiable at zero, with derivative $-B^{-1}A$ where $(DF)_{(0,0)} = \begin{bmatrix} A & B \end{bmatrix} \in \mathbb{R}^{m \times (n+m)}$ $(A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{m \times m})$.

Proof. Recall: for $x \in B_\tau(0)$, $g(x)$ is the unique fixed point of $y \mapsto -B^{-1}(Ax + R(x,y)) = k_x(y)$ in $\overline{B_r(0)}$ (the closure of $B_r(0)$). We know $g(x) \in B_r(0)$. We then have $g(x) = K_x(g(x)) = -B^{-1}(Axx + R(x,g(x)))$. Thus,

$$
\begin{aligned}
\|g(x) - \underbrace{g(0)}_{=0} - (-B^{-1}A)x\| &= \|-B^{-1}[Ax + R(x,g(x))] + B^{-1}Ax\| \\
&= \|B^{-1}[R(x,g(x))]\| \\
&\leq \|B^{-1}\|_{\text{op}}\|R(x,g(x))\|.
\end{aligned}
$$

Thus, for $x \neq 0$

$$
\frac{\|g(x) - g(0) - (-B^{-1}A)x\|}{\|x\|} \leq \|B^{-1}\|_{\text{op}} \frac{\|R(x,g(x))\|}{\|(x,g(x))\|} \cdot \frac{\|(x,g(x))\|}{\|x\|}.
$$

Observe that $\frac{\|R(x,g(x))\|}{\|(x,g(x))\|} \to 0$ as $x \to 0$ because $(x, g(x)) \to 0$ as $x \to 0$ since g is continuous at 0; also using the fact that F is C^r, and therefore C^1, and therefore differentiable. Further, $\frac{\|(x,g(x))\|}{\|x\|}$ is bounded, because we have

$$\|(x, g(x))\| \leq C(\|x\| + \|g(x)\|) \leq C(\|x\| + 4L\|x\|)$$

for small enough x, where C is some constant that is independent of x and L is the constant from the previous lemma (the first inequality follows from comparability of norms). Therefore

$$\frac{\|(x, g(x))\|}{\|x\|} \leq C(1 + 4L)$$

for small enough x. So,

$$\frac{\|g(x) - g(0) - (-B^{-1}A)x\|}{\|x\|} \to \|B^{-1}\|_{\mathrm{op}} \cdot 0 \cdot C(1 + 4L) = 0$$

as $x \to 0$.

$\square$

Lemma 132 (appears in book, but weakly justified ("origin is as good as any other point")). In Lemma 130, the unique function g is differentiable on $B_\tau(0)$ with derivative $-B_{(x,g(x))}^{-1}A_{(x,g(x))}$, where $(DF)_{(x,y)} = \begin{bmatrix} A_{x,y} & B_{x,y} \end{bmatrix} \in \mathbb{R}^{m \times (n+m)}$, where $A_{x,y} \in \mathbb{R}^{m \times n}$ and $B_{x,y} \in \mathbb{R}^{m \times m}$.

Proof. Let $x_0 \in B_\tau(0)$ and let $y_0 = g(x_0)$. Let $\tilde{F}(x, y) := F(x + x_0, y + y_0)$. Note that $\tilde{F}(x, y)$ is C^r since F is C^r, and it is defined in $B_{\tau'}(0) \times B_{r'}(0)$ as long as $\tau' < \tau - \|x_0\|$, $r' < r - \|y_0\|$. We can apply the previous results to $\tilde{F} \colon B_{\tau'(0)} \times B_{r'}(0) \to \mathbb{R}^m$: there exists $\tau'' \leq \tau'$ and a function $\tilde{g} = B_{\tau''}(0) \mapsto B_{r'}(0)$ (really $B_{r''}(0)$ for some $r'' \leq r'$, but still maps onto the larger set $B_{r'}(0)$) such that $\tilde{F}(x, \tilde{g}(x)) = 0$ for all $x \in B_{\tau''}(0)$, and $\tilde{g}$ is differentiable at 0, with derivative

$$-\left(\frac{\mathrm{d}\tilde{F}}{\mathrm{d}y}(0,0)(\right)^{-1}\left(\frac{\mathrm{d}\tilde{F}}{\mathrm{d}x}(0,0)\right).$$

We claim that $g(x) = \tilde{g}(x - x_0) + y_0$. To see why, note that $g(x)$ is the unique fixed point on $B_r(0)$ with $F(x, g(x)) = 0$. But we also have $F(x, \tilde{g}(x - x_0) + y_0) = \tilde{F}(x - x_0,)$. ... Thus, $g(x) = \tilde{g}(x - x_0) + y_0$.

So, by the Chain Rule (Theorem 122), g is differentiable at x_0 with

$$(Dg)_{x_0} + (Dy)_0 = -\left(\frac{\partial \tilde{F}}{\partial}\right)$$

$\square$

Lemma 133. The unique function g from Lemma 130 is of class C^1.

Proof. We need to show each entry of $(Dg)_x$ is a continuous function of $x \in B_\tau(0)$. We have $(Dg)_x = -B_{x,g(x)}^{-1} A_{x,g(x)}$. The entries of $A_{x,y}$ and $B_{x,y}$ are partial derivatives of components of F, which is C^r and therefore C^1, so they're continuous functions of $(x,y) \in U$. g is differentiable on $B_\tau(0)$, so it is continuous on $B_\tau(0)$. Thus, the entries of $A_{x,g(x)}$ and $B_{x,g(x)}$ are continuous functions of $x \in B_\tau(0)$.

We can deal with $B_{x,g(x)}^{-1}$ using Cramer's rule:

$$B_{x,g(x)}^{-1} = \frac{1}{\det(B_{x,g(x)})} \cdot \left(\text{cofactor}(B_{x,g(x)})\right)^\top$$

where the cofactor matrix $\text{cofactor}(B_{x,g(x)})$ has each entry in row i and column j corresponding to the cofactor as when computing row/column expansions. Since $\det(B_{x,g(x)})$ is a polynomial function in the entries of $B_{x,g(x)}$, it is continuous in x. $\text{cofactor}(B_{x,g(x)})$ is also built out of determinants, which are polynomial functions of the entries, so it is continuous in x. So the entries of $B_{x,g(x)}^{-1}$ are continuous in x, which means the entries of $-B_{x,g(x)}^{-1} A_{x,g(x)}$ are continuous in x, so g is C^1.

$\square$

Lemma 134. The function g from Lemma 130 is C^r (assuming F is C^r).

Proof. We will induct on $r \geq 1$. We have the base case from the previous lemma. For a general r, suppose that g is class C^{r-1}. Note that it's enough to show that the entries of $(Dg)_x$ are C^{r-1} function of x, because the rth order partial derivatives are the $(r-1)$th order partials of the first order partials. We have $(Dg)_x = -B_{x,g(x)}^{-1} A_{x,g(x)}$. The entries of $A_{x,y}$ and $B_{x,y}$ are C^{r-1} since F is C^r. Also, g is C^{r-1} by above. Therefore the entries of $A_{x,g(x)}$ and $B_{x,g(x)}$ are C^{r-1} functions of x. By the same Cramer's rule reasoning as above, the result follows.

$\square$

Proof of Implicit Function Theorem (Theorem 129). Follows from Lemmas 130 and 134.

$\square$

for all $(x_0, y_0) \in F^{-1}(z_0)$,

i.e. $F^{-1}(z_0)$ is a "nonsingular level set." For any point $(x_0, y_0) \in F^{-1}(z_0)$, we have a homeomorphism $(x, g(x)) : B_\tau(x_0) \mapsto F^{-1}(z_0) \cap V$

an open subset of $F^{-1}(z_0)$ containing (x_0, y_0).

$x \mapsto (x, g(x)) \in F^{-1}(z_0) \cap V$. The map (id, g) is a homeomorphism since the projection $(x, y) \mapsto x$ restricted to $F^{-1}(z_0) \cap V$ is the inverse of (id, g) and is continuous.

The implicit function theorem tells us that F being C^∞ implies g is C^∞ implies (id, g)

Thus, these homemorphisms form a smooth atlas on $F^{-1}(z_0)$, giving $F^{-1}(z_0)$ the structure of a smooth manifold. (Note: the definition of a smooth manifold is only reasonable because the above is true—see Definition 1.7.10.)

The following is a useful result for manifolds (see Theorem 118) that is closely related to the Implicit Function Theorem.

Theorem 135 (**Theorem 2-13 in** Spivak **[1971], p. 56 of pdf, p.443 of book**). Let $f : \mathbb{R}^n \to \mathbb{R}^p$ be C^r for $r \geq 1$ in an open set containing x_0, where $p \leq n$. If $f(x_0) = 0$ and $(Df)_{x_0} \in \mathbb{R}^{p \times n}$ has rank p, then there is an open set $A \subset \mathbb{R}^n$ containing x_0 and a differentiable function $h : A \to \mathbb{R}^n$ with differentiable inverse such that

$$f \circ h(x^1, \ldots, x^n) = (x^{n-p+1}, \ldots, x^n).$$

Now we will discuss the Inverse Function Theorem.

Definition 1.7.22. If $U \subset \mathbb{R}^n$ is open and $V \subset \mathbb{R}^m$ is open, then $F : U \to V$ is a C^r **diffeomorphism** (recall Definition 1.6.39) (for $r \geq 1$) if (1) F is C^r, (2) F is invertible, and (3) F^{-1} is C^r.

Note that this is only possible when $n = m$, because $\text{Id}_a = F^{-1} \cdot F$. Take the total derivative of both sides and apply the chain rule (Theorem 122):

$$= \text{Id}_{\mathbb{R}^n} = (DF)_{f(p)} \circ (DF)_p \qquad \forall p \in U.$$

Similarly, $\text{Id}_{\mathbb{R}^m} = (DF)_p \circ (DF^{-1})_{F(p)}$ for all $p \in U$. This implies $(DF)_p$ is invertible for all p (inverse $(DF^{-1})_{F(p)}$), so $(DF)_p$ is square $n = m$.

So we see that F is "nonlinearly invertible" then $(DF)_p$ is "linearly invertible" for all p. Like in the Implicit Function Theorem, we want to go backwards locally.

Theorem 136 (**Inverse Function Theorem**). Let $U_0 \subset \mathbb{R}^n$ be open, let $f : U_0 \to \mathbb{R}^m$ be a C^r function for $r \geq 1$, $p \in U_0$. If $(DF)_p$ is invertible, then $n = m$ and there exists an open subset $U \subset U_0$, $V \subset \mathbb{R}^m$ containing p and $f(p)$ respectively such that f is a C^r diffeomorphism from U to V (i.e. f is a locally C^r diffeomorphism at p).

Proof. We will apply the Implicit Function Theorem. Define $F : U_0 \times \mathbb{R}^n \to \mathbb{R}^n$ by $F(x, y) = f(x) - y$ for $x \in U_0, y \in \mathbb{R}^n$. Let $q = f(p) \in \mathbb{R}^m$. We have $F(p, q) = q - q = 0$, so $(p, q) \in F^{-1}(0)$. The Jacobian matrix of F at (p, q) is

$$(DF)_{(p,q)} = \left[\frac{\partial F}{\partial x}(p, q) \quad \frac{\partial F}{\partial y}(p, q) \right]$$
$$= \left[(Df)_p \quad -\boldsymbol{I}_n \right],$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. Recall that $(Df)_p \in \mathbb{R}^{n \times n}$ is invertible by assumption, and of course $-\boldsymbol{I}_n$ is invertible. Then by the Implicit Function Theorem, there exist open neighborhoods $U_p \subset U_0$ and $V_q \subset \mathbb{R}^n$ of p and q and a unique function $h : V_q \to U_p$ such that

$$F^{-1}(0) \cap (U_p, V_q) = \text{graph}(h) = \{(y, h(y)) : y \in V_q\}$$

and h is C^r. By construction, we have $F(h(y), y) = 0$ for all $y \in V_q$. Therefore $0 = f(h(y)) - y \iff f(h(y)) = y$, for all $y \in V_q$, so $f \circ h = \text{id}_{V_q}$.

Now we run the same argument in reverse. Define : $G : U_q \times V_p \to \mathbb{R}^n$ by $G(x, y) = x - h(y)$. Of course, $h(q) = p$ because $(p, q) = F^{-1}(0) \cap (U_p \times V_q) = \text{graph}(h)$, so we have $G(p, q) = p - h(q) = 0$. Then the Jacobian matrix of G at (p, q) is

$$(DG)_{(p,q)} = \begin{bmatrix} \boldsymbol{I}_m & -(Dh)_q \end{bmatrix}.$$

Note that since $p = h(q)$,

$$f \circ h = \text{id}_{V_q} \quad \implies \quad (Df)_{h(q)} \circ (Dh)_q = \text{Id}_{\mathbb{R}^n} \quad \iff \quad (Df)_p \circ (Dh)_q = \text{Id}_{\mathbb{R}^n},$$

where we applied the Chain Rule. So $(Dh)_q$ is invertible. Therefore by the Implicit Function Theorem, there exists an open neighborhood $U'_p \times V'_q \subset U_p \times V_q$ of (p, q) and a unique function $g : U'_p \to V'_q$ such that

$$G^{-1}(0) \cap (U'_p \times V'_q) = \text{graph}(g) = \{(x, g(x)) : x \in U'_p\}.$$

Again, we have $G(x, g(x)) = 0$ for all $x \in U'_p$, so $h(g(x)) = x$. Therefore $0 = x - h(g(x)) \iff h \circ g = \text{id}_{U'_p}$.

Note that for points $x \in U'_p$, $g(x) = f(x)$ (i.e., g is the restriction of f to U'_p). To see why, note that for $x \in U'_p$, we have $g(x) \in V'_q \subset V_q$. Since $f \circ h = id_{V_q}$, we have $g(x) = f(h(g(x))) = f(x)$ (where we used that $h \circ g = id_{U'_p}$).

Finally, let $U := U'_p \subset \mathbb{R}^n$ and let $V := h^{-1}(U) \subset V'_q \subset \mathbb{R}^n$. Observe that f maps U into V since for any $x \in U'_p$, $f(x) = g(x) = h^{-1}(x)$. Also, h maps V into U because for any $v \in h^{-1}(U)$ we have $h(v) = x$ for some $x \in U$. Therefore f is a C^r diffeomorphism from U to V.

$$\square$$

1.7.5   Tensors (Homeworks 11 - 14 in Math 425b, Chapter 16 of Lang [2005])

Definition 1.7.23 (Direct sums; definition from Homework 11, 425b). Let S be any set. Assume that for every $s \in S$ we have a vector space V_s over $\mathbb{R}$. Define

$$\bigoplus_{s \in S} V_s,$$

the **direct sum** of the vector spaces V_s to be the set of functions $f : S \to \bigcup_{s' \in S} V_{s'}$ such that $f(s) \in V_s$ for all $s \in S$, and such that $f(s) = 0$ for all but finitely many $s \in S$. (note that the union is a "disjoint union;" the vector spaces $V_{s'}$ are not given as subsets of a larger set *a priori*.)

For $f, g \in \oplus_{s \in S} V_s$, define $f + g$ by $(f + g)(s) = f(s) + g(s)$. For $c \in \mathbb{R}$, define cf by $(cf)(s) = c(f(s))$.

Definition 1.7.24 (Direct sums; definition from Section 1.7 of Lang [2005], p. 36 of textbook (p. 51 of pdf)). Let $\{V_s\}_{s \in S}$ be a family of abelian groups (recall Definition ??). Define

$$\bigoplus_{s \in S} V_s$$

to be the subset of the direct product $\prod V_s$ consisting of all families $(x_s)_{s \in S}$ with $x_s \in V_s$ such that $x_s = 0$ for all but a finite number of indices s.

For $v_i \in V_{s_i}$, $i \in [n]$, $\bigoplus_{s \in S} V_s$ is the vector space of all linear combinations $c_1 v_1 + \ldots + c_n v_n$ (where $s_1 \neq \ldots \neq s_n$). The linear combinations are finite sums since $f(s) = 0$ for all but finitely many $s \in S$.

Definition 1.7.25. Let S be any set. Assume that for every $s \in S$ we have a vector space V_s over $\mathbb{R}$. Define the **free vector space** on S to be

$$F(S) := \bigoplus_{s \in S} \mathbb{R};$$

the set of functions $f : S \to \mathbb{R}$ such that $f(s) \in \mathbb{R}$ for all $s \in S$, and such that $f(s) = 0$ for all but finitely many $s \in S$. That is, it is a direct sum where we take each vector space V_s to be $\mathbb{R}$. An **inclusion map** is defined to be the function $i : S \to F(S)$ sending $s \in S$ to the function $i(s) : S \to \mathbb{R}$ defined by

$$i(s)(s') = \begin{cases} 1, & s' = s, \\ 0, & \text{otherwise.} \end{cases}$$

$F(S)$ is the vector space of all linear combinations of elements of S. The elements of S form a basis for $F(S)$ by definition. Equivalently, $F(S)$ is a way to construct a vector space having a given set S as a basis.

Proposition 137 (Homework 11 problem 4)**.** (a) $\bigoplus_{s \in S} V_s$ is closed under the above addition and scalar multiplication properties.

(b) Suppose S has two elements; let $S = \{s, t\}$. Then $V_s \oplus V_t$ is isomorphic to the vector space of ordered pairs (v, w) where $v \in V_s$, $w \in V_t$ (i.e., the Cartesian product $V_s \times V_t$), and addition and scalar multiplication are given by $(v, w) + (v', w') = (v + v', w + w')$ and $c(v, w) = (cv, cw)$.

(c) Let V be any vector space and let $g : S \to V$ be any function. Then there exists a unique linear transformation $T : F(S) \to V$ such that $g = T \circ i$. Equivalently, the following diagram commutes:

$$F(S) \xrightarrow{\ T\ } V$$
$$i \uparrow \qquad \nearrow g$$
$$S$$

Proof. (a) We will check that sums and scalar multiples preserve the quality of vanishing at all but finitely many s.

For sums, observe that for $f, g \in \bigoplus_{s \in S} V_s$ we have $(f + g)(s) = f(s) + g(s)$. By definition $f(s)$ and $g(s)$ both equal 0 for all but finitely many $s \in S$; therefore the same is true for $f(s) + g(s)$ and $(f + g)(s)$.

For scalar multiplication, since for $f \in \bigoplus_{s \in S} V_s$ we have $(cf)(s) = cf(s)$ and $f(s)$ equals 0 for all but finitely many $s \in S$, the same is true for $cf(s)$ for any scalar c and therefore for $(cf)(s)$.

(b) For $f \in V_s \oplus V_t$, let $f(s) = v$ and $f(t) = w$ and define the linear map to $V_s \times V_t$ by $f \mapsto (f(s), f(t)) = (v, w)$. Observe that this map is linear because $cf \mapsto ((cf)(s), (cf)(t)) = (c(f(s)), c(f(t))) = c(f(s), f(t)) = c(v, w)$ and for g satisfying $g(s) = v'$ and $g(t) = w'$ we have $f + g \mapsto ((f(s), f(t)) + (g(s), g(t)) = (v, w) + (v', w') = (v + v', w + w')$. Note also that all of these mappings are bijective.

(c) For $f \in F(S)$, define $T(f)$ to be the element of V given by $\sum_{s \in S} f(s)g(s)$ (recall that $f(s) \in \mathbb{R}$ and $g(s) \in V$). Observe that since $f(s) = 0$ for all but finitely many s, $\sum_{s \in S} f(s)g(s)$ is finite. We see that T is linear since

$$T(cf) = \sum_{s \in S} (cf)(s)g(s) = \sum_{s \in S} cf(s)g(s) = c \sum_{s \in S} f(s)g(s) = cT(f)$$

and for another $f' \in F(S)$

$$T(f + f') = \sum_{s \in S} (f + f')(s)g(s) = \sum_{s \in S} [f(s) + f'(s)]g(s) = \sum_{s \in S} f(s)g(s) + \sum_{s \in S} f'(s)g(s) = T(f) + T(f').$$

Then the diagram commutes because for any $s \in S$

$$(T \circ i)(s) = \sum_{s' \in S} i(s)(s')g(s) = \sum_{s' \in S, s' \neq s} 0 \cdot g(s) + 1 \cdot g(s) = g(s).$$

Now we will show uniqueness. Suppose T and T' are two linear maps such that the diagram commutes, so we have $T(i(s)) = g(s) = T'(i(s))$. Let $f \in F(S)$. Observe that

$$f(s) = \sum_{s' \in S} i(s)(s')f(s) = \sum_{s' \in S, f(s') \neq 0} i(s)(s')f(s),$$

which is a finite sum by the requirement that $f(s) = 0$ for all but finitely many $s \in S$. Now (using the linearity of T and T')

$$
\begin{aligned}
T(f(s)) &= T\left(\sum_{s' \in S, f(s') \neq 0} i(s)(s')f(s) \right) \\
&= \sum_{s' \in S, f(s') \neq 0} T\left(i(s)(s')f(s) \right) \\
&= \sum_{s' \in S, f(s') \neq 0} f(s)T\left(i(s)(s') \right) \\
&= \sum_{s' \in S, f(s') \neq 0} f(s)g(s) \\
&= \sum_{s' \in S, f(s') \neq 0} f(s)T'\left(i(s)(s') \right) \\
&= \sum_{s' \in S, f(s') \neq 0} T'\left(i(s)(s')f(s) \right) \\
&= T'\left(\sum_{s' \in S, f(s') \neq 0} i(s)(s')f(s) \right) \\
&= T'(f(s)).
\end{aligned}
$$

Therefore $T(f) = T'(f)$, showing uniqueness.

$\square$

Part (c) of Proposition 137 shows that any function g from a set S to a vector space V can be extended linearly in a unique way to a linear transformation defined on $F(S) := \oplus_{s \in S} \mathbb{R}$, the vector space of formal linear combinations of elements of S.

Definition 1.7.26 (Quotient Vector Space). Let V be a vector space and let W be a subspace of V. Define an equivalence relation on V by stating that $v_1 \sim v_2$ if $v_1 - v_2 \in W$. Let V/W denote the set of equivalence classes under this relation (see Definition 1.3.2). The space V/W is called a **quotient vector space**.

If $v \in V$, let $[v]$ denote its equivalence class. For two elements α, β of V/W, define $\alpha + \beta$ to be $[v_1 + v_2]$ where $\alpha = [v_1]$ and $\beta = [v_2]$. Also, for $c \in \mathbb{R}$, define $c\alpha$ to be $[cv]$ where $\alpha = [v]$.

Proposition 138. The quotient vector space V/W satisfies the vector space axioms under the above definitions of addition and scalar multiplication. Also, if V has dimension n and W has dimension m, then V/W has dimension $n - m$.

Proposition 139 (Math 425b Homework 12 problem). $\alpha + \beta$ and $c\alpha$ as defined in Definition 1.7.26 are well-defined (i.e., independent of the choice of v_1 and v_2 representing α and β).

Proof. Let $v_1, v_1', v_2, v_2' \in V$ and let $c \in \mathbb{R}$. Suppose that $v_1 \sim v_1'$ and $v_2 \sim v_2'$; that is, $v_1 - v_1' \in W$ and $v_2 - v_2' \in W$. Since W is a subspace of V (and therefore a vector space), it is closed under addition, so $v_1 - v_1' + v_2 - v_2' \in W$. It also satisfies associativity of addition, so $(v_1 + v_2) - (v_1' + v_2') \in W$. Therefore $(v_1 + v_2) \sim (v_1' + v_2')$.

Next, observe that for any $w \in W$, we have $cw \in W$. Therefore $c(v_1 - v_1') \in W$, which means $cv_1 - cv_1' \in W$ (and $cv_1 \sim cv_1'$) by distributivity of scalar multiplication with respect to vector addition. Since these statements would be true for any $\tilde{v}_1, \tilde{v}_1' \in \alpha$, $\tilde{v}_2, \tilde{v}_2' \in \beta$, or $c \in \mathbb{R}$, this shows that $\alpha + \beta$ and $c\alpha$ are well-defined.

$\square$

Proposition 140 (Math 425b Homework 12 problem). Let V be a vector space and let W be a subspace of V. Let Z be another vector space and let $f : V \to Z$ be a linear map such that $f(w) = 0$ for all $w \in W$. Let $p : V \to V/W$ denote the linear map sending $v \to [v]$. Then there exists a unique linear map $g : V/W \to Z$ such that $f = g \circ p$.

Proof. First we will show existence. Define $g([v]) := f(v)$. To see that this is well-defined, consider another $v' \in V$ such that $v - v' \in W$; that is, $v \sim v'$. Observe that $f(v) - f(v') = f(v - v') = 0$ (by assumption that $f(w) = 0$ for all $w \in W$), and since $[v] = [v']$, $g([v]) - g([v']) = g([v]) - g([v]) = 0$. Now consider an arbitrary $v' \in V$ such that $v - v' \notin W$. Note that $f(v + v') = f(v) + f(v')$, and from Exercise 1, we have $[v] + [v'] = [v + v']$, so $g([v + v']) = g([v]) + g([v'])$. Finally, for any $c \in \mathbb{R}$ we have $f(cv) = cf(v)$, and since from Exercise 1 we have $[cv] = c[v]$, we have $g([cv]) = g(c[v]) = cg([v])$ by linearity of g.

Now we will show uniqueness. Note that any $\alpha \in V/W$ is equal to $[v] = p(v)$ for some $v \in V$. Consider two maps g and $\tilde{g}$ that make the diagram commute; that is, $g([v']) = f(v')$ and $\tilde{g}([v']) = f(v')$ for any $v' \in V.$. Then for any $\alpha \in V/W$, we have $g(\alpha) = g([v]) = f(v) = \tilde{g}([v]) = \tilde{g}(\alpha)$.

$\square$

Let V and W be vector spaces over $\mathbb{R}$. Consider the free vector space $F(V \times W)$ (see Definition 1.7.25). It is the set of functions $f : V \times W \to \mathbb{R}^2$. $F(V \times W)$ is the vector space of all linear combinations of elements of $V \times W$. The elements of $V \times W$ form a basis for $F(V \times W)$ by definition.

Given an element $(v, w) \in V \times W$, we view (v, w) as an element of $F(V \times W)$ via the inclusion map $i : V \times W \to F(V \times W)$. Any element of $F(V \times W)$ is a finite linear combination of such elements (v, w).

Note that $F(V \times W)$ disregards the vector space structures on V and W, and just treats $V \times W$ as the set of ordered pairs (v, w) where $v \in V$ and $w \in W$. For example, if $v \neq 0 \in V$, then $(v, 0)$ and $(2v, 0)$ are linearly independent in $F(V \times W)$, even though v and $2v$ are not linearly independent in V.

Definition 1.7.27 (Tensor product). Let S be the subset of $F(V \times W)$ consisting of the following elements:

- $(v, w) + (v', w) - (v + v', w)$ for all $v, v' \in V$ and $w \in W$.

- $(v, w) + (v', w) - (v, w + w')$ for all $v \in V$ and $w, w' \in W$.

- $c(v, w) - (cv, w)$ for all $v \in V$, $w \in W$, and $c \in \mathbb{R}$.

- $c(v, w) - (v, cw)$ for all $v \in V$, $w \in W$, and $c \in \mathbb{R}$.

Define

$$V \otimes W := \frac{F(V \times W)}{\operatorname{span}(S)}.$$

That is, for $f_1 : V \times W \to \mathbb{R}^2 \in F(V \times W)$ and $f_2 : V \times W \to \mathbb{R}^2 \in F(V \times W)$, for an equivalence relation defined as $f_1 \sim f_2$ if $f_1 - f_2 \in \operatorname{span}(S)$, $V \otimes W$ is the quotient vector space consisting of the set of equivalence classes under this relation. In particular, if we view (v, w) as an element of $F(V \times W)$ via the inclusion map $i : V \times W \to F(V \times W)$ (recall Definition 1.7.25), we can characterize the equivalence relation by $(v_1, w_1) \sim (v_2, w_2)$ if $(v_1, w_1) - (v_2, w_2) \in \operatorname{span}(S)$. Then $V \otimes W$ is the quotient vector space consisting of the set of equivalence classes under this relation.

Given an element (v, w) of $F(V \times W)$, its class $[(v, w)]$ in the quotient space $V \otimes W$ will be denoted by $v \otimes w$. Note that not every element of $V \otimes W$ can be written as $v \otimes w$ for some $v \in V$ and $w \in W$ (since not every function $f : V \times W \to \mathbb{R}^2$ can be expressed as the inclusion map of a single $(v, w) \in V \times W$). Elements like $v \otimes w$ are called **pure tensors**. In general, an element of $V \otimes W$ is a linear combination of pure tensors.

See also Section 4.1 of Spivak [1971] (p. 88 of pdf, p. 75 of book).

Example 1.7.6. The tensor product of zero copies of $\mathbb{R}^n$ is defined to be $\mathbb{R}$. (This is relevant later in the dimensions of spaces of alternating multilinear functionals.)

Definition 1.7.28 (Bilinear maps; also in Section 5.2 of Pugh [2015], p. 297 of pdf, p. 287 of book). Let V, W, Z be vector spaces over $\mathbb{R}$. A map $f : V \times W \to Z$ is said to be **bilinear** if

- $f(v + cv', w) = f(v, w) + cf(v', w)$ for all $v, v' \in V$, $w \in W$, and $c \in \mathbb{R}$.

- $f(v, w + cw') = f(v, w) + cf(v, w')$ for all $v \in V$, $w, w' \in W$, and $c \in \mathbb{R}$.

For a generalization of this concept, see multilinear maps (Definition 1.7.33).

Proposition 141 (Math 425b Homework 12 Problem 3)**.** The map $\pi : V \times W \to V \otimes W$ sending (v, w) to $v \otimes w$ is bilinear.

Proof. We will show that $\pi(v + cv', w) = \pi(v, w) + c\pi(v', w)$. Let $c \in \mathbb{R}$, $v, v' \in V$, $w \in W$. Note that since $cv' \in V$, we have

$$
\begin{aligned}
(v + cv', w) - (v, w) - c(v', w) &= (v + cv', w) - (v, w) - c(v', w) - (cv', w) + (cv', w) \\
&= -\underbrace{(v, w) + (cv', w) - (v + cv', w)}_{\in S} - \underbrace{c(v', w) - (cv', w)}_{\in S}.
\end{aligned}
$$

Therefore $(v + cv', w) - (v, w) - c(v', w)$ is in the span of S. That means it is in the equivalence class $[0]$, so $0 = \pi\left((v + cv', w) - (v, w) - c(v', w)\right) = \pi(v + cv', w) - \pi(v, w) - c\pi(v', w) \iff \pi(v + cv', w) = \pi(v, w) + c\pi(v', w)$. The other equation is similar.

$\square$

Tensor products have a "universal property:" roughly speaking, a bilinear map from $V \times W$ to Z is the same as a linear map from $V \otimes W$ to Z.

Proposition 142 (Math 425b Homework 12 Problem 4)**.** Let V, W, Z be vector spaces over $\mathbb{R}$. Let $f : V \times W \to Z$ be a bilinear map. Let $\pi : V \times W \to V \otimes W$ be defined by $\pi(v, w) := v \otimes w$. Then there exists a unique linear map $g : V \otimes W \to Z$ such that $f = g \circ \pi$.

Using this universal property, we have $V \otimes W \cong W \otimes V$ for any vector spaces V and W. Other standard properties also follow, like $(V \otimes W) \otimes Z \cong V \otimes (W \otimes Z)$. One can also deduce relations like $V \otimes (W_1 \oplus W_2) \cong (V \otimes W_1) \oplus (V \otimes W_2)$; i.e., tensor products distributed over direct sums (up to isomorphism). The vector space $\mathbb{R}$ acts as a "unit" for the tensor product operations: for any V, we have $V \otimes \mathbb{R} \cong V$.

It follows that if V has dimension n, so $V \cong \mathbb{R}^n \cong \underbrace{\mathbb{R} \oplus \ldots \oplus \mathbb{R}}_{n \text{ terms}}$, and W has dimension m, then

$$
V \otimes W \cong \underbrace{(\mathbb{R} \oplus \ldots \oplus \mathbb{R})}_{n \text{ terms}} \otimes \underbrace{(\mathbb{R} \oplus \ldots \oplus \mathbb{R})}_{m \text{ terms}},
$$

which expands out to nm copies of $\mathbb{R}$. Thus, $V \otimes W$ has dimension nm. Concretely, if $\{e_i\}$ forms a basis for V and $\{f_j\}$ forms a basis for W, then $\{e_i \otimes f_j\}$ forms a basis for $V \otimes W$ (see also Theorem 4-1 in Spivak [1971], p. 89 of pdf, p. 76 of book).

Mathematicians often define a vector to be an element of a vector space. Correspondingly, one can define a tensor as follows.

Definition 1.7.29 (**Tensors**)**.** A **tensor of rank** k is an element of the vector space $V_1 \otimes \cdots \otimes V_k$ for some $k \geq 1$ (i.e., a tensor is a vector in the space $V_1 \otimes \cdots \otimes V_k$).

Along with tensor products, another essential operation on vector spaces is the notion of the dual space (recall Definition 1.7.6).

Definition 1.7.30 (Basis vectors for dual space; 425b Homework 13). If $\{e_i\}$ is a basis for V, define the linear functionals $e_i^* : V \to \mathbb{R} \in V^*$ by

$$e_i^*(e_j) := \delta_{i,j} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

where V^* is the dual space of V (recall Definition 1.7.6).(Following the up/down notation mentioned below, one often write e^i instead of e_i^*.) In Proposition 143, we show that these vectors are a basis for V^*.

Remark 15. In some contexts it is common to present tensors as being higher-dimensional analogues of vectors and matrices. If a vector is a one-dimensional array of numbers and a matrix is a two-dimensional array, a tensor of rank k should be a k-dimensional array. One can connect this view with tensors with Definition 1.7.29 as follows: given bases $\{e_{i,j}\}_j$ for each V_i, one gets a basis for $V_1 \otimes \cdots \otimes V_k$ by taking tensor products of basis vectors as defined in Definition 1.7.27 (see Theorem 4-1 in Spivak [1971], p. 89 of pdf, p. 76 of book). An arbitrary element of $V_1 \otimes \cdots \otimes V_k$ can be expanded as a linear combination of basis vectors; one has a unique coefficient $c_{j_1,\ldots,j_k}$ on each basis vector $e_{1,j_1} \otimes \cdots \otimes e_{k,j_k}$. One organizes these coefficients into an array of the appropriate dimensionality:

- For $k = 1$, the coefficients c_j form a one-dimensional array (vector),

- For $k = 2$, the coefficients c_{j_1,j_2} form a two-dimensional array (matrix),

- For $k = 3$, the coefficients c_{j_1,j_2,j_3} form a three-dimensional array (rank-3 tensor), etc.

Now that we have dual vector spaces, we often prefer to view a matrix as representing a linear transformation between vector spaces $T : V \to W$ and then interpret this transformation as an element of $W \otimes V^*$ (still a rank-2 tensor, but involving a dual on one of the factors).

Definition 1.7.31 (Covariant tensors). Let $V_1, \ldots, V_k$ be fixed vector spaces. An element x of $V_1 \otimes \cdots \otimes V_k$ is called a **covariant tensor**. We write its coefficient on a basis vector $e_{1,j_1} \otimes \cdots \otimes e_{kj_k}$ as $x^{j_1,\ldots,j_k}$ (i.e., we use "up indices").

Definition 1.7.32 (Contravariant tensors). Let $V_1, \ldots, V_k$ be fixed vector spaces. An element x of $V_1^* \otimes \cdots \otimes V_k^*$ is called a **contravariant tensor**. We write its coefficient on the dual basis vector $e^{1,j_1} \otimes \cdots \otimes e^{kj_k}$ as $x_{j_1,\ldots,j_k}$ (i.e., we use "down indices").

In between these extreme cases, we have various types of **mixed tensors** whose coefficients have both up and down indices.

Remark 16. The up/down notation is part of "Einstein notation," developed by Albert Einstein for use in relativity. Another part of the convention is to write only the coefficients and note the basis vectors they're multiplied by when specifying a tensor. For example,

- A vector $x = \sum_i x^i e_i$ in V gets written as x^i,

- A dual vector $x = \sum_i x_i e^i \in V^*$ gets written as x_i,

- A rank-2 covariant tensor $\sum_{i,j} x^{i,j} e_i \otimes f_j$ in $V \otimes W$ gets written as $x^{i,j}$ where $\{f_j\}$ is the given basis for W,

- A rank-2 mixed tensor $\sum_{i,j} x^i_j e_i \otimes f^j$ in $W \otimes V^*$ gets written as x^i_j,

- A rank-2 contravariant tensor $\sum_{i,j} x_{i,j} e^i \otimes f^j$ in $V^* \otimes W^*$ gets written as $x_{i,j}$.

Often one will consider certain sums of these coefficients; the convention is that one sums over repeated indices (one must be up and the other down) and omits the sum symbol. For example, let a^i_j represent an element of $W \otimes V^*$ (corresponding to a linear transformation T from V to W; in fact, in the assumed bases, a^i_j is the usual matrix entry of T in row i and column j). Mathematically, we would write a^i_j as $\sum_{i,j} a^i_j e_i \otimes f^j$. Let v^i represent a vector in V; mathematically, we would write v^i as $\sum_i v^i e_i$. Applying the transformation $a - j^i$ to the vector v^i gives $\sum_{i,j} a^i_j v^j f_i$ (a vector in W), and the convention is to write this vector as simply $a^i_j v^j$ (the sum over j is implicit, as is the sum over i with the "invisible" basis vectors f_i). You know this quantity represents a vector since it has one "free' index i and this index is up.

Proposition 143 (**Math 425b Homework 13 Problem 1; similar to Theorem 4-1 in Spivak [1971], p. 89 of pdf, p. 76 of book, also Proposition 11.1 in Lee [2012]**). Let V be a finite-dimensional vector space with basis $\{e_i\}$. Define the linear functionals $e^*_i \in V^*$ by

$$e^*_i(e_j) := \delta_{i,j} = \begin{cases} 1, & i = j, \\ 0, & i \neq j \end{cases}$$

as in Definition 1.7.30. The linear functionals $\{e^*_i\}$ form a basis for V^*. Further, V^* has the same dimension as V.

Also, if V and W are finite-dimensional with bases $\{e_i\}$ and $\{f_j\}$, then the elements $e_i \otimes f_j$ form a basis for $V \otimes W$; in particular, the dimension of $V \otimes W$ is the product of the dimensions of V and W.

Proof. To show that $\{e^*_i\}$ forms a basis for V^*, we will show that the $\{e^*_i\}$ are linearly independent and that they span V^*. Let $n := |\{e^*_i\}| = \dim(V)$. Since $\{e_i\}$ is a basis for V, for every $v \in V$ there exists a unique $(c_1, \ldots, c_n) \in \mathbb{R}^n$ such that $v = \sum_{i=1}^n c_i e_i$. Suppose for some $(c^*_1, \ldots, c^*_n) \in \mathbb{R}^n$ we have $\sum_{j=1}^n c^*_j e^*_j(v) = 0$ for all $v \in V$. Then for any $v \in V$

$$0 = \sum_{j=1}^n c^*_j e^*_j(v) = \sum_{j=1}^n c^*_j e^*_j \left(\sum_{i=1}^n c_i e_i \right) = \sum_{j=1}^n c^*_j \sum_{i=1}^n c_i e^*_j(e_i) = \sum_{i=1}^n c^*_i c_i e^*_i(e_i) = \sum_{i=1}^n c^*_i c_i$$

(where the third equality follows from the fact that e^*_j is linear and the fourth equality follows from $e^*_j(e_i) = 0$ for all $j \neq i$), which only holds for all $v \in V$ if $(c^*_1, \ldots, c^*_n) = (0, \ldots, 0)$. Therefore the $\{e^*_i\}$ are linearly independent. Next we will show that the $\{e^*_i\}$ span V^*. Let $\phi : V \to \mathbb{R}$ be an arbitrary linear functional. We will show in particular that $\phi(v) = \sum_{j=1}^n c^*_j e^*_j(v)$ for some $(c^*_1, \ldots, c^*_n) \in \mathbb{R}^n$ for any $v = \sum_{i=1}^n c_i e_i \in V$. We have

$$\phi(v) = \phi \left(\sum_{j=1}^n c_j e_j \right) = \sum_{j=1}^n c_j \phi(e_j) = \sum_{j=1}^n \phi(e_j) c_j e^*_j(e_j) = \sum_{j=1}^n \phi(e_j) \sum_{i=1}^n c_i e^*_j(e_i) = \sum_{j=1}^n \phi(e_j) e^*_j \left(\sum_{i=1}^n c_i e_i \right)$$

$$= \sum_{j=1}^n \phi(e_j) \cdot e^*_j(v) = \sum_{j=1}^n c^*_j \cdot e^*_j(v)$$

where $(c_1^*, \dots, c_n^*) = (\phi(e_1), \dots, \phi(e_n))$.

Next we will show that if V and W are finite-dimensional with bases $\{e_i\}$ and $\{f_j\}$, then the elements $e_i \otimes f_j$ form a basis for $V \otimes W$; in particular, the dimension of $V \otimes W$ is the product of dimensions of V and W. We can write V as $\mathbb{R} \oplus \cdots \oplus \mathbb{R}$, with n copies of $\mathbb{R}$. Similarly, we can write W as $\mathbb{R} \oplus \cdots \oplus \mathbb{R}$, with $\dim(W)$ copies of $\mathbb{R}$.

We will use the fact that if V_1, V_2, W are vector spaces and $T_1 : V_1 \to W, T_2, V_2 \to W$ are linear maps, there exists a unique linear map $T : (V_1 \oplus V_2) \to W$ such that $T((v_1, 0)) = T_1(v_1)$ and $T((0, v_2)) = T_2(v_2)$.

$\hfill\square$

Proposition 144 (**Math 425b Homework 13 Problem 2**)**.** Let V and W be vector spaces over $\mathbb{R}$. Define a map $T : V^* \times W^* \to (V \otimes W)^*$ that sends $(\phi, \psi) \in V^* \times W^*$ to $(v \otimes w) \mapsto \phi(v)\psi(w) \in (V \otimes W)^*$ (recall Definition 1.7.6 for the definition of the dual spaces V^* and W^*). This map T is well-defined and bilinear (recall Definition 1.7.28), and thus induces a linear map $V^* \otimes W^* \to (V \otimes W)^*$. Further, if V and W are finite-dimensional, this induced map is an isomorphism.

Remark 17. The left side $(V \otimes W)^*$ of the isomorphism in Proposition 144 can be identified as the space of bilinear maps from $V \times W$ to $\mathbb{R}$. Proposition 144 shows that this space of bilinear maps can also be described as $V^* \otimes W^*$, when V and W are finite-dimensional. More generally, if all spaces V_i are finite-dimensional we can identify $V_1^* \otimes \cdots \otimes V_n^*$ with the space of multlinear maps from $V_1 \times \cdots \times V_n$ into $\mathbb{R}$.

Even more generally, if all spaces V_i and W_j are finite-dimensional, we can identify $V_1^* \otimes \cdots \otimes V_n^* \otimes W_1 \otimes \cdots \otimes W_m$ with the space of multlinear maps from $V_1 \times \cdots \times V_n$ to $W_1 \otimes \cdots \otimes W_n$. A special case is when $n = m = 1$; we can identify $V^* \otimes W$ (or $W \otimes V$) with the space of linear maps from V to W.

1.7.6   Differential Forms (Section 5.8 of Pugh [2015]; Chapter 10 of Rudin [1976])

Definition 1.7.33 (**Multilinear functionals; Section 5.9 of Pugh [2015]**)**.** A map $\beta : \underbrace{\mathbb{R}^n \times \dots \times \mathbb{R}^n}_{k \text{ times}} \to \mathbb{R}$ which is linear in each vector variable separately is a k-**multilinear functional.** That is, for $i \in [k]$ we have

$$\beta(x_1, \dots, x_i + cx_i', \dots, x_k) = \beta(x_1, \dots, x_i \dots, x_k) + c\beta(x_1, \dots, x_i', \dots, x_k).$$

A bilinear map is a 2-multilinear map; see Definition 1.7.28.

Definition 1.7.34 (**Symmetric groups and permutations**)**.** We call the set of all permutations of $[k]$ (of size $k!$) the **symmetric group on k** and denote it by S_k. We refer to its elements as $\sigma \in S_k$. The number of transpositions modulo 2 in any factorization of σ into transpositions is well-defined and denoted by $\text{sgn}(\sigma)$.

See Section **??**.

Definition 1.7.35 (**Alternating multilinear functionals; Section 5.9 of Pugh [2015], Section 4.1 of Spivak [1971] (p. 91 of pdf, p. 78 of book)**)**.** A k-multilinear functional $\beta : \mathbb{R}^n \times \dots \times \mathbb{R}^n \to \mathbb{R}$ is **alternating** if for each permutation $\sigma \in S_k$ we have

$$\beta(v_1, \dots, v_k) = \text{sgn}(\sigma)\beta(v_{\sigma(1)}, \dots, v_{\sigma(k)}).$$

The set of alternating k-linear forms is a vector space. In particular, we denote by $\mathrm{Alt}_k(\mathbb{R}^n, \mathbb{R})$ the vector space of multilinear maps $\alpha : \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{k \text{ times}} \to \mathbb{R}$ with

$$\alpha(v_1, \ldots, v_{i+1}, v_i, \ldots, v_k) = -\alpha(v_1, \ldots, v_i, v_{i+1}, \ldots, v_k)$$

for $1 \leq i \leq k$. (In Spivak [1971], this space is denoted by $\Lambda^k(V)$.)

Example 1.7.7. The set of 1-multilinear functionals is the set of linear functionals from $\mathbb{R}^n \to \mathbb{R}$.

Example 1.7.8. The set of 2-multilinear functionals is the set of bilinear functionals from $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfying

$$\beta(v, w) = -\beta(w, v)$$

for all $v, w \in \mathbb{R}^n$.

Example 1.7.9. The set of 0-multilinear functionals is $\mathbb{R}$.

Definition 1.7.36 (**Differential k-forms; from Math 425B Homeworks 13 and 14**). A **differential k-form** on $U \subset \mathbb{R}^n$ is a function $\alpha : U \to \mathrm{Alt}_k(\mathbb{R}^n, \mathbb{R})$.

From Remark 17, $\mathrm{Alt}_k(\mathbb{R}^n, \mathbb{R})$ can be viewed as a subspace of $\underbrace{(\mathbb{R}^n)^* \otimes \cdots \otimes (\mathbb{R}^n)^*}_{k \text{ times}}$ (recall Definition 1.7.27 for the definition of the tensor product $\otimes$, and recall Definition 1.7.6 for the definition of the dual space $(\mathbb{R}^n)^*$). So a differential k-form is a particular type of rank-k contravariant tensor field on U (namely one that satisfies the "alternating property" at each point). While this perspective is not strictly speaking necessary to define differential k-forms, it is very useful when defining wedge products (see Definition 1.7.40 and 1.7.50).

C^r regularity for differential k-forms can be defined by looking at their coordinates in any basis for the finite-dimensional vector space $\mathrm{Alt}_k(\mathbb{R}^n, \mathbb{R})$ (independently of the choice of basis). That is, we say that α is of class C^r if its coordinates in the standard basis of $(\mathbb{R}^n)^* \otimes \cdots \otimes (\mathbb{R}^n)^*$ (equivalently, any basis for this vector space) are C^r functions from U to $\mathbb{R}$.

That is, a differential k-form on U is a function α from U to $\underbrace{(\mathbb{R}^n)^* \otimes \cdots \otimes (\mathbb{R}^n)^*}_{k \text{ times}}$ which is alternating in the following sense: viewing $\alpha(p) \in (\mathbb{R}^n)^* \otimes \cdots \otimes (\mathbb{R}^n)^*$ as an element of $(\mathbb{R}^n \otimes \cdots \otimes \mathbb{R}^n)^*$ **by the below Propositions [which?]** (i.e., a multilinear map from $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ to $\mathbb{R}$), the map $\alpha(p)$ is alternating in the sense that for all k-tuples $(v_1, \ldots, v_k)$ of vectors in $\mathbb{R}^n$, we have

$$\alpha(p)(v_1 \otimes \cdots \otimes v_i \otimes v_{i+1} \otimes \cdots \otimes v_k) = -\alpha(p)(v_1 \otimes \cdots \otimes v_{i+1} \otimes v_i \otimes \cdots \otimes v_k)$$

for any i with $1 \leq i \leq k - 1$.

Example 1.7.10.)

Let $U \subset \mathbb{R}^n$ be open. Since $\mathrm{Alt}_0(\mathbb{R}^n, \mathbb{R}) = \mathbb{R}$ (see Example 1.7.9), a **differential 0-form** on U is a function $f : U \to \mathbb{R}$.

Example 1.7.11. Let $U \subset \mathbb{R}^n$ be open. A **differential 1-form** on U is a function $\alpha : U \to (\mathbb{R}^n)^*$. Given any basis $\{\phi_1, \ldots, \phi_n\}$ for $(\mathbb{R}^n)^*$, we can write $\alpha(p) = a_1(p)\phi_1 + \ldots + a_n(p)\phi_n$, where $a_1, \ldots, a_n$ are functions from U to $\mathbb{R}$. We say that α is of class C^r ($1 \le r \le \infty$) if all the functions a_i are of class C^r. Since changes of basis on $\mathbb{R}^n$ are C^∞ (even linear) functions, this notion is independent of the choice of basis $\{\phi_1, \ldots, \phi_n\}$.

Since $\mathrm{Alt}_1(\mathbb{R}^n, \mathbb{R})$ is the set of linear functionals from $\mathbb{R}^n \to \mathbb{R}$ (see Example 1.7.7), we can also say that a differential 1-form α, evaluated at a point $p \in U$, gives a linear functional $\alpha(p)$ from $\mathbb{R}^n$ to $\mathbb{R}$.

Also, any 1-form α on $[a, b]$ is $f(t)dt$ for some function f.

Given a basis for $\mathbb{R}^n$ (in particular, the standard basis), one can identify differential 1-forms on U with **vector fields** on U (see Definition 1.7.16); i.e., functions from U to $\mathbb{R}^n$ (rather than $(\mathbb{R}^n)^*$) (see Remark 18).

Definition 1.7.37 (**Differential 1-forms; from Pugh [2015] Section 5.8, p. 337 of pdf, p. 327 of book**). A **differential 1-form** is a function that sends paths to real numbers and which can be expressed as a path integral (for example, $f\ dx + g\ dy$ is a 1-form).

Example 1.7.12 (**Example from Pugh [2015] Section 5.8, p. 337 of pdf, p. 327 of book**). Consider a path integral as its defined in calculus:

$$\int_C (f\ dx + g\ dy) = \int_0^1 f(x(t), y(t)) \frac{\mathrm{d}x(t)}{\mathrm{d}t} dt + \int_0^1 g(x(t), y(t)) \frac{\mathrm{d}y(t)}{\mathrm{d}t} dt, \tag{1.34}$$

where f and g are smooth real-valued functions of (x, y) and C is a smooth path parameterized by $(x(t), y(t))$ as t varies on $[0, 1]$. Taking the dual approach of differential forms, we will think of the integral as a number that depends on the path C. What property of C does the differential 1-form $f\ dx + g\ dy$ measure?

Consider the case $f(x, y) = 1$ and $g(x, y) = 0$. Then the path integral (1.34) is

$$\int_C dx = \int_0^1 \frac{\mathrm{d}x(t)}{\mathrm{d}t} dt = x(1) - x(0),$$

the "net x-variation" of the path C. In functional notation, we can write $dx : C \mapsto x(1) - x(0)$, so dx assigns to each path C its net x-variation. Similarly, dy assigns to each path its net y-variation.

For a general $f\ dx$, the function f "weights" x-variation. If the path C passes through a region in which f is large, its x-variation is magnified accordingly, and the integral $\int_C f\ dx$ reflects the net f-weighted x-variation of C. Similarly, $g\ dy$ assigns a path to its net g-weighted variation, and the 1-form $f\ dx + g\ dy$ assigns to C the sum of the two variations. See Figure 1.7.

Example 1.7.13. Note that $\mathrm{Alt}_2(\mathbb{R}^n, \mathbb{R})$ is the set of bilinear functionals from $\beta : \mathbb{R}^2 \to \mathbb{R}$ satisfying $\beta(v, w) = -\beta(w, v)$ for all $(v, w) \in \mathbb{R}^2$ (see Example 1.7.8). Therefore a differential 2-form α, evaluated at a point $p \in U$, gives a bilinear map $\alpha(p)$ from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ such that $\alpha(p)(v, w) = -\alpha(p)(w, v)$ for all $v, w \in \mathbb{R}^n$.

There are four fundamental operations on differential forms that we need to understand.

1. **Exterior derivative:** if α is a k-form (say C^∞), then we have a $(k + 1)$-form $d\alpha$. This operation will generalize the gradient (Remark 18), curl, and divergence operations (Proposition 152). See Def-

Figure 121 suggests why $\int_C y\,dx$ is positive and $\int_{C'} y\,dx$ is negative: The weight factor is positive on C and negative on C'. On the other hand, if the weight factor is the constant c then both integrals are $c(q-p)$.
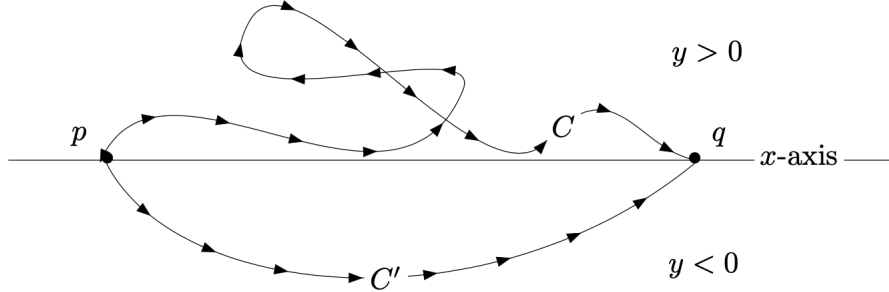
Figure 121 C and C' are paths from p to q where p and q lie on the x-axis. The integrals $\int_C y\,dx$ and $\int_{C'} y\,dx$ express the net y-weighted x-variation along C and C'.

Figure 1.7: Figure 121 from Pugh [2015], illustrating 1-forms.

inition 1.7.38 for the exterior derivative of a 0-form (also called a differential) and Definition 1.7.41 for the exterior derivative of a k-form.

2. **Wedge product:** if α is a k-form and β is an ℓ-form, then we have a $(k+\ell)$-form $\alpha \wedge \beta$ (this operation will generalize the cross product of vectors in $\mathbb{R}^3$). See Definition 1.7.40 and Definition 1.7.50.

3. **Pullback:** if α is a k-form on U and $F : V \to U$ is smooth where $V \subset \mathbb{R}^m$ is open (for some m), then we have a k-form $F^*(\alpha)$ on V. (See also Definition 1.7.39 below for the definition of pullbacks of 1-forms and Definition 1.7.53 for the general definition.)

4. **Integration:** if α is a k-form on a k-dimensional cube $[0,1]^k \subset \mathbb{R}^k$ (it's okay that this isn't an open set although α should at least be right-continuous), we have a real number $\int_{[0,1]^k} \alpha$. (We can also integrate on more general rectangular sets than just $[0,1]^k$.) Combined with pullbacks, this operation will let us generalize line integrals and surface integrals. (See also Example 1.7.12.)

Example 1.7.14. Integration of zero-forms on $\mathbb{R}^0$ is just evaluation at the unique point $0 \in \mathbb{R}^0$.

Also, any 1-form α on $[a,b]$ is $f(t)dt$ for some function f, and if f is smooth or just Riemann integrable, we can just define $\int_{[a,b]} \alpha := \int_a^b f(t)dt$ as usual.

Now we will consider less trivial operations than just 0-forms and 1-forms. We will start by studying the exterior derivative df of a zero-form f (also known as the differential of f), closely related to the gradient of f.

Definition 1.7.38 (**Exterior derivative of a 0-form (differential)**)**.** Let $U \subset \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ be a smooth function (0-form). Define a differential 1-form df by the equation

$$df(p) = (Df)_p.$$

(In other words, df is just the Jacobian Df (see Theorem 124), which at a point p gives a linear map from $\mathbb{R}^n$ to $\mathbb{R}$.) Basically the transpose of the gradient. See also Definition 1.7.41 for the exterior derivative of a k-form.

Proposition 145 (**Math 425b Homework 13 Problem 3**). Let $\{e_1, \ldots, e_n\}$ be the standard basis vectors of $\mathbb{R}^n$ and let $\{e_1^*, \ldots, e_n^*\}$ be their dual basis vectors (see Definition 1.7.30). Let f be as above; prove that

$$df = \frac{\partial f}{\partial x^1} e_1^* + \ldots + \frac{\partial f}{\partial x^n} e_n^*.$$

(we use "up indices" x^i rather than "down indices" x_i to match the physics conventions; a physicist may write df as $\frac{\partial f}{\partial x^i}$ or even as $\partial_i f$, which is a 1-form since the i index is down, because up indices in a denominator are down. See Remark 16.)

Proof. It is enough to show that $df(p)(e_i) = \frac{\partial f}{\partial x^i}(p)$ for each i and each $p \in U$. Since for each $p \in U$ we have

$$(DF)_p = \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(p) & \cdots & \frac{\partial F_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1}(p) & \cdots & \frac{\partial F_m}{\partial x_n}(p) \end{bmatrix},$$

it holds that

$$(DF)_p e_i = \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(p) & \cdots & \frac{\partial F_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1}(p) & \cdots & \frac{\partial F_m}{\partial x_n}(p) \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \sum_{j=1}^n \frac{\partial F_i}{\partial x_j}(p) = \frac{\partial f}{\partial x^i}(p)$$

for each i. Therefore $df(p)(e_i) = \frac{\partial f}{\partial x^i}(p)$ for each i and each $p \in U$. $\qquad\square$

Remark 18 (**From 425b Homework 13 Problem 4; Theorem 4-7 in** Spivak [1971] (**p. 102 of pdf, p. 89 of book**)). Given a basis for $\mathbb{R}^n$ (in particular, the standard basis), one can identify differential 1-forms on U with **vector fields** on U (see Definition 1.7.16); i.e., functions from U to $\mathbb{R}^n$ (rather than $(\mathbb{R}^n)^*$). The vector field associated to the 1-form df is the **gradient** of f, denoted ∇f:

$$\nabla f = \frac{\partial f}{\partial x^1} e_1 + \ldots + \frac{\partial f}{\partial x^n} e_n.$$

In general, df is a bit more natural of an object than ∇f, since it doesn't require any choice of basis to define.

One can also identify 1-forms and vector fields by picking an inner product for $\mathbb{R}^n$, rather than a basis. On a general smooth manifold M (see Definition 1.7.10), the 1-form df for a function $f : M \to \mathbb{R}$ is always defined, but the gradient ∇f requires a choice of "Riemannian metric" on M (inner product on each tangent space; see Definition 1.7.14). In Einstein notation (see Remark 16), one sometimes writes $\partial^i f$ for

the gradient of f, as opposed to the 1-form $\partial_i f$. The Riemannian metric itself can be viewed as a function on M with values in $(\mathbb{R}^n)^* \otimes (\mathbb{R}^n)^*$ (recall Definition 1.7.27 for the definition of the tensor product $\otimes$), since an inner product on $\mathbb{R}^n$ is a bilinear map from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$. Thus, **as mentioned in a remark towards the end of Homework 12 (not in these notes yet)**, the metric is determined locally by functions g_{ij} on M for $1 \leq i, j \leq \dim(M)$. The gradient of f is determined uniquely by the formula $\partial_i f = g_{ij} \partial^j f$ (note the implicit sum over j); for the usual metric on $\mathbb{R}^n$, $g_{ij} = \delta_{ij}$.

Proposition 146 (**Homework 13 Problem 4**). Recall that for $1 \leq i \leq n$, we have a coordinate projection function $x^i : U \to \mathbb{R}$ (sending a point in U to its ith coordinate), and thus a 1-form dx^i. For all $p \in U$, we have $dx^i(p) = e_i^*$, where $\{e_i\}$ is the standard basis of $\mathbb{R}^n$.

It follows that if $f : U \to \mathbb{R}$ is a smooth function as above, then we can write

$$df = \frac{\partial f}{\partial x^1} dx^1 + \ldots + \frac{\partial f}{\partial x^n} dx^n,$$

a natural-looking formula.

Proof. By Proposition 145 we have that for all $p \in U$,

$$dx^i(p) = \sum_{j=1}^{n} \frac{\partial x^i}{\partial x^j} e_j^* = \frac{\partial x^i}{\partial x^i} e_i^* = e_i^*$$

where $\{e_i\}$ is the standard basis of $\mathbb{R}^n$.

$\square$

Definition 1.7.39 (**Pullbacks of 1-forms**). Let $V \subset \mathbb{R}^m$ be open and let α be a differential 1-form on V (a function from V to $(\mathbb{R}^m)^*$). Let $U \subset \mathbb{R}^n$ be open and let $F : U \to V$ be a smooth function. The **pullback** $F^*(\alpha)$ is the differential 1-form on U (that is, a function from U to $(\mathbb{R}^n)^*$, so a function from U to the space of functionals with domain $\mathbb{R}^n$) defined at a point $p \in U$ by

$$(F^*(\alpha))_p(v) := \alpha_{F(p)}((DF)_p(v))$$

for $v \in \mathbb{R}^n$. (Note that $(DF)_p(v) \in \mathbb{R}^m$, so it makes sense to evaluate $\alpha_{F(p)}$ on the vector $(DF)_p(v)$.) See also Definition 1.7.53 for the general definition of pullbacks.

Example 1.7.15. The pullback of a zero-form f by a function F is just defined to be $f \circ F$.

The idea is that to pull back a differential 1-form by F, you push forward the corresponding "input vector" v by DF; the same idea is used to define pullbacks of k-forms (see Definition 1.7.53).

Proposition 147 (**Math 425b Homework 13 Problem 5**). (a) If $f : V \to \mathbb{R}$ is a smooth function and $F : U \to V$ is smooth, then

$$F * (df) = d(f \circ F) = d(F^*(f)).$$

(That is, pullbacks commute with exterior derivatives acting on zero-forms.)

(b) If $\alpha = f_1 \alpha_1 + \ldots + f_N \alpha_N$, then

$$F^*(\alpha) = (f_1 \circ F)F^*(\alpha_1) + \ldots + (f_n \circ F)F^*(\alpha_N).$$

(c) If $\alpha = f_1 dx^1 + \ldots + f_m dx^m$ and $\boldsymbol{r} : [a, b] \to V$ is a smooth pair with components $r_1, \ldots, r_m$, then

$$\boldsymbol{r}^*(\alpha) = f_1(\boldsymbol{r}(t))r_1'(t)dt + \ldots + f_m(\boldsymbol{r}(t))r_m'(t)dt.$$

Start of Homework 14: Now we will begin working with k-forms with $k > 1$. An important goal will be understanding the standard identifications

- 0-forms on $\mathbb{R}^3$ correspond to functions on $\mathbb{R}^3$ (see Example 1.7.10);

- 1-forms on $\mathbb{R}^3$ correspond to vector fields (see Definition 1.7.16) on $\mathbb{R}^3$ (see Example 1.7.11);

- 2-forms on $\mathbb{R}^3$ correspond to vector fields on $\mathbb{R}^3$; and

- 3-forms on $\mathbb{R}^3$ correspond to functions on $\mathbb{R}^3$.

All k-forms on $\mathbb{R}^3$ are zero if $k > 3$.

Remark 19. Implicitly, these identifications use the standard inner product/Riemannian metric on $\mathbb{R}^3$, as with the identification between df and ∇f discussed in Remark 18.

Note that when working only with functions and vector fields, it may not be immediately clear whether a vector field $\boldsymbol{F}$ came from a 1-form or a 2-form (or a legitimate vector field); similarly, a given function f might represent a 3-form instead of a 0-form.

Given these identifications, we will see that

- The gradient operator ∇ taking functions to vector fields (see Definition 1.7.16) becomes identified with the exterior derivative d taking 0-forms to 1-forms (see Remark 18).

- The curl operation curl (or $\nabla \times$), taking vector fields to vector fields, becomes identified with the exterior derivative d taking 1-forms to 2-forms (Proposition 152).

- The divergence operator div (or $\nabla \cdot$), taking vector fields to functions, becomes identified with the exterior derivative d taking 2-forms to 3-forms (Proposition 152).

The relations $\text{curl} \circ \nabla = 0$ and $\text{div} \circ \text{curl} = 0$ are special cases of the fundamental property $d^2 = 0$ for the exterior derivative.

Recall Definition 1.7.36. The identifications we want will become more plausible once we know that

- $\dim \text{Alt}_0(\mathbb{R}^3, \mathbb{R}) = 1$, (recall Example 1.7.6)

- $\dim \text{Alt}_1(\mathbb{R}^3, \mathbb{R}) = 3$ (recall Proposition 145),

- $\dim \text{Alt}_2(\mathbb{R}^3, \mathbb{R}) = 3$ (Proposition 148), and

- $\dim \mathrm{Alt}_3(\mathbb{R}^3, \mathbb{R}) = 1$ (Proposition 150).

In general, $\dim \mathrm{Alt}_k(\mathbb{R}^n, \mathbb{R}) = \binom{n}{k}$ (and equals 0 unless $0 \le k \le n$). The case of Alt_0 is tautological since the tensor product of 0 copies of $\mathbb{R}^n$ is defined to be $\mathbb{R}$ (see Example 1.7.10). The case of Alt_1 just says that $(\mathbb{R}^n)^*$ has dimension n, which follows from **a proposition from Homework 13 [which?]** and Example 1.7.11.

These dimension counts will follow from the fact that the wedge products $\{dx^{i_1} \wedge \cdots \wedge dx^{i_k} : 1 \le i_1 \le \cdots \le i_k \le n\}$ form a basis for $\mathrm{Alt}_k(\mathbb{R}^n, \mathbb{R})$, but for this to make sense, we first need to define wedge products.

Definition 1.7.40 (Abstract definition of wedge products). If $\phi, \psi \in (\mathbb{R}^n)^*$, define an alternating bilinear map (an element of $\mathrm{Alt}_2(\mathbb{R}^n, \mathbb{R})$) $\phi \wedge \psi$ from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ by

$$(\phi \wedge \psi)(v, w) := \frac{1}{2}(\phi(v)\psi(w) - \phi(w)\psi(v)).$$

If α and β are 1-forms on U (that is, α and β are functions from U to $(\mathbb{R}^n)^*$, see Example 1.7.11), then $\alpha \wedge \beta$ is a 2-form on U defined by $(\alpha \wedge \beta)(p) := \alpha(p) \wedge \beta(p)$ (wedge products of differential forms are defined "pointwise;" see also Example 1.7.13).

Note that we have $\psi \wedge \phi = -\phi \wedge \psi$ (so $\phi \wedge \phi = 0$), and that wedge products are bilinear:

$$(c_1\phi_1 + c_2\phi_2) \wedge \psi = c_1\phi_1 \wedge \psi + c_2\phi_2 \wedge \psi,$$

and similarly in the second slot.

More generally, if α is a k-form and β is an ℓ-form, then $\alpha \wedge \beta$ is a $(k + \ell)$-form.

See also the differently formulated definition in Section 4.1 of Spivak [1971] (p. 92 of pdf, p. 79 of book).

Example 1.7.16. The wedge product of a zero-form f with any k-form is always defined as ordinary scalar multiplication by the value of f at each point.

Proposition 148 (425b Homework 14 Problem 1). $\{dx \wedge dy, dx \wedge dz, dy \wedge dz\}$ is a basis for the vector space of alternating bilinear maps from $\mathbb{R}^3 \times \mathbb{R}^3$ to $\mathbb{R}$.

(Strictly speaking, dx, dy, and dz are differential 1-forms defined on $\mathbb{R}^3$, so we should evaluate them at a point $p \in \mathbb{R}^3$, but their values are independent of p and it is standard to just write dx, dy, and dz.)

Proof. From Proposition 146, we have $dx(p) = e_1^*, dy(p) = d_2^*$, and $dz(p) = e_3^*$ for any $p \in \mathbb{R}^3$, so we want to show that $\{e_1^* \wedge e_2^*, e_1^* \wedge e_3^*, e_2^* \wedge e_3^*\}$ is a basis for $\mathrm{Alt}_2(\mathbb{R}^n, \mathbb{R})$. By definition, since each $e_i^* \wedge e_j^* \in \mathrm{Alt}_2(\mathbb{R}^n, \mathbb{R})$, they are all bilinear and alternating. We want to show that this set of three maps is linearly independent and spans the space of alternating bilinear maps.

Recall Definition 1.7.30. Let $v, w \in \mathbb{R}^n$ with $v = \sum_{i=1}^n c_{vi}e_i$ and $w = \sum_{i=1}^n c_{wi}e_i$, where e_i are the standard basis vectors for $\mathbb{R}^n$. Then

$$e_j^*(v) = \sum_{i=1}^n c_{vi}e_j^*(e_i) = \sum_{i=1}^n c_{vi}\delta_{ij} = c_{vj} \qquad \forall j \in [n]$$

and similarly $e_j^*(w) = c_{wj}$. So for $i \neq j \in [n]$ we have

$$(e_i^* \wedge e_j^*)(v, w) = \frac{1}{2}(e_i^*(v)e_j^*(w) - e_i^*(w)e_j^*(v)) = \frac{1}{2}(c_{vi}c_{wj} - c_{wi}c_{vj}). \tag{1.35}$$

First we will show linear independence. $\{e_1^* \wedge e_2^*, e_1^* \wedge e_3^*, e_2^* \wedge e_3^*\}$ are independent if

$$c_1(e_1^* \wedge e_2^*) + c_2(e_1^* \wedge e_3^*) + c_3(e_2^* \wedge e_3^*) = 0 \quad \Longleftrightarrow \quad c_1 = c_2 = c_3 = 0.$$

By (1.35), we have

$$c_1(e_1^* \wedge e_2^*)(v, w) + c_2(e_1^* \wedge e_3^*)(v, w) + c_3(e_2^* \wedge e_3^*)(v, w) = 0$$
$$\Longleftrightarrow \quad c_1 \cdot \frac{1}{2}(c_{v1}c_{w2} - c_{w1}c_{v2}) + c_2 \cdot \frac{1}{2}(c_{v1}c_{w3} - c_{w1}c_{v3}) + c_3 \cdot \frac{1}{2}(c_{v2}c_{w3} - c_{w2}c_{v3}) = 0$$
$$\Longleftrightarrow \quad c_1 = c_2 = c_3 = 0,$$

since none of the terms cancel. Next we will show that $\{e_1^* \wedge e_2^*, e_1^* \wedge e_3^*, e_2^* \wedge e_3^*\}$ spans the space of alternating bilinear maps $\mathrm{Alt}_2(\mathbb{R}^n, \mathbb{R})$. Consider an arbitrary map $\Phi \in \mathrm{Alt}_2(\mathbb{R}^n, \mathbb{R})$. For $(e_i, e_j) \in \mathbb{R}^n \times \mathbb{R}^n$ with $i < j$, we have $\Phi(e_1, e_2) =$,

$$\square$$

In multivariable calculus, one often studies the cross product of vectors in $\mathbb{R}^3$. Viewed in terms of the dual space $(\mathbb{R}^3)^*$ instead of $\mathbb{R}^3$, the below proposition shows that the cross product becomes a special case of the wedge product, taking in two elements of $(\mathbb{R}^3)^* = \mathrm{Alt}_1(\mathbb{R}^3, \mathbb{R})$ and producing an element of the three-dimensional space $\mathrm{Alt}_2(\mathbb{R}^3, \mathbb{R})$.

Proposition 149 (**425b Homework 14 Problem 2**). For an element $a = a_1 e_1^* + a_2 e_2^* + a_3 e_3^*$ of $(\mathbb{R}^3)* = \mathrm{Alt}_1(\mathbb{R}^3, \mathbb{R})$, define $\Phi(\alpha) \in \mathbb{R}^3$ to be the vector (a_1, a_2, a_3). Similarly, for an element $\alpha = a_{12}e_1^* \wedge e_2^* + a_{13}e_1^* \wedge e_3^* + a_{23}e_2^* \wedge e_3^*$ of $\mathrm{Alt}_2(\mathbb{R}^3, \mathbb{R})$, define $\Phi(\alpha) \in \mathbb{R}^3$ to be $(a_{23}, -a_{13}, a_{12})$.

Then for $\alpha, \beta \in (\mathbb{R}^3)^*$, we have

$$\Phi(\alpha \wedge \beta) = \Phi(\alpha) \times \Phi(\beta),$$

where $\times$ denotes the usual cross product of vectors in $\mathbb{R}^3$.

Remark 20. Seeing that $(a_{23}, -a_{13}, a_{12})$ is the natural vector in $\mathbb{R}^3$ to define a given element of $\mathrm{Alt}_2(\mathbb{R}^3, \mathbb{R})$ involves looking at the Hodge star operator (covered later in these notes). Note that the minus sign disappears if you use the basis vector $e_3^* \wedge e_1^*$ rather than $e_1^* \wedge e_3^*$ in the basis for $\mathrm{Alt}_2(\mathbb{R}^3, \mathbb{R})$.

Proof. Expressing α as $a_1 e_1^* + a_2 e_2^* + a_3 e_3^*$ and $\beta = b_1 e_1^* + b_2 e_2^* + b_3 e_3^*$, we have

$$
\begin{aligned}
\alpha \wedge \beta =& (a_1 e_1^* + a_2 e_2^* + a_3 e_3^*) \wedge (b_1 e_1^* + b_2 e_2^* + b_3 e_3^*) \\
=& a_1 e_1^* \wedge (b_1 e_1^* + b_2 e_2^* + b_3 e_3^*) + a_2 e_2^* \wedge (b_1 e_1^* + b_2 e_2^* + b_3 e_3^*) + a_3 e_3^* \wedge (b_1 e_1^* + b_2 e_2^* + b_3 e_3^*) \\
=& a_1 b_1 e_1^* \wedge e_1^* + a_1 b_2 e_1^* \wedge e_2^* + a_1 b_3 e_1^* \wedge e_3^* + a_2 b_1 e_2^* \wedge e_1^* + a_2 b_2 e_2^* \wedge e_2^* + a_2 b_3 e_2^* \wedge e_3^* \\
& + a_3 b_1 e_3^* \wedge e_1^* + a_3 b_2 e_3^* \wedge e_2^* + a_3 b_3 e_3^* \wedge e_3^* \\
=& a_1 b_2 e_1^* \wedge e_2^* + a_1 b_3 e_1^* \wedge e_3^* - a_2 b_1 e_1^* \wedge e_2^* + a_2 b_3 e_2^* \wedge e_3^* - a_3 b_1 e_1^* \wedge e_3^* - a_3 b_2 e_2^* \wedge e_3^* \\
=& (a_1 b_2 - a_2 b_1) e_1^* \wedge e_2^* + (a_1 b_3 - a_3 b_1) e_1^* \wedge e_3^* + (a_2 b_3 - a_3 b_2) e_2^* \wedge e_3^*.
\end{aligned}
$$

Now we have $\alpha \wedge \beta$ in terms of the basis $\{e_1^* \wedge e_2^*, e_1^* \wedge e_3^*, e_2^* \wedge e_3^*.\}$ for $\mathrm{Alt}_2(\mathbb{R}^3, \mathbb{R})$. Then

$$
\begin{aligned}
\Phi(\alpha \wedge \beta) &= \Phi\left((a_1 b_2 - a_2 b_1) e_1^* \wedge e_2^* + (a_1 b_3 - a_3 b_1) e_1^* \wedge e_3^* + (a_2 b_3 - a_3 b_2) e_2^* \wedge e_3^*\right) \\
&= (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1) \\
&= (a_2 b_3 - a_3 b_2) e_1 + (a_3 b_1 - a_1 b_3) e_2 + (a_1 b_2 - a_2 b_1) e_3.
\end{aligned}
$$

On the other hand,

$$
\Phi(\alpha) = (a_1, a_2, a_3) = a_1 e_1 + a_2 e_2 + a_3 e_3, \qquad \Phi(\beta) = (b_1, b_2, b_3) = b_1 e_1 + b_2 e_2 + b_3 e_3.
$$

By computing the cross product as in multivariable calculus, we have

$$
\Phi(\alpha) \times \Phi(\beta) = \begin{vmatrix} e_1 & e_2 & e_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = (a_2 b_3 - a_3 b_2) e_1 - (a_1 b_3 - a_3 b_1) e_2 + (a_1 b_2 - a_2 b_1) e_3 = \Phi(\alpha \wedge \beta).
$$

$$\square$$

More generally, the abstract definition of wedge products (Definitions 1.7.40 and 1.7.50) will imply that if $\phi_1, \ldots, \phi_k$ are in $(\mathbb{R}^n)^*$, then

$$
\phi_1 \wedge \cdots \wedge \phi_k = \frac{1}{k!} \left(\sum_{\sigma \in S_k} (-1)^{\mathrm{sgn}(\sigma)} \phi_{\sigma(1)} \otimes \cdots \otimes \phi_{\sigma(k)} \right), \tag{1.36}
$$

where S_k and $\mathrm{sgn}(\sigma)$ are as defined in Definition 1.7.34 (and recall Definition 1.7.27 for the definition of the tensor product $\otimes$). Definition 1.7.40 (see also Definition 1.7.50) also implies basic properties like associativity of wedge products, and makes it clear that wedge products of elements of $(\mathbb{R}^n)^* = \mathrm{Alt}_1(\mathbb{R}^n, \mathbb{R})$ form a spanning set for $\mathrm{Alt}_k(\mathbb{R}^n, \mathbb{R})$.

Thus, (1.36) suffices for computing all wedge products. In particular, for $\phi_1, \phi_2, \phi_3 \in (\mathbb{R}^3)^*$, we have

$$
\phi_1 \wedge \phi_2 \wedge \phi_3 = \frac{1}{6} \left(\phi_1 \otimes \phi_2 \otimes \phi_3 - \phi_1 \otimes \phi_3 \otimes \phi_2 + \phi_2 \otimes \phi_3 \otimes \phi_1 - \phi_2 \otimes \phi_1 \otimes \phi_3 + \phi_3 \otimes \phi_1 \otimes \phi_2 - \phi_3 \otimes \phi_2 \otimes \phi_1 \right).
$$
$$\tag{1.37}$$

Proposition 150 (**425b Homework 14 Problem 3**). $\{dx \wedge dy \wedge dz\}$ is a basis for $\text{Alt}_3(\mathbb{R}^n, \mathbb{R})$.

To recap, we now have the following facts:

- A 0-form on $\mathbb{R}^3$ is a function f on $\mathbb{R}^3$ (Example 1.7.10).

- A 1-form on $\mathbb{R}^3$ can be written uniquely as $fdx + gdy + hdz$ where f, g, h are functions on $\mathbb{R}^3$.

- A 2-form on $\mathbb{R}^3$ can be written uniquely as $f\ dx \wedge dy + g\ ds \wedge dz + h\ dy \wedge dz$ where f, g, h are functions on $\mathbb{R}^3$ (Proposition 148).

- A 3-form of $\mathbb{R}^3$ can be written uniquely as $f\ dx \wedge dy \wedge dz$ where f is a function on $\mathbb{R}^3$ (Proposition 150).

It turns out that in general, a differential k-form α on $\mathbb{R}^n$ can be written uniquely as

$$\alpha = \sum_{1 \leq i_1 \leq \ldots \leq i_k \leq n} f_{i_1, \ldots, i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}.$$

(This is Theorem 4-5 in Spivak [1971], p. 94 of pdf, p. 81 of book.) This expression is especially convenient when defining the exterior derivative.

Definition 1.7.41 (**Exterior derivative of a k-form**). Let

$$\alpha = \sum_{1 \leq i_1 \leq \ldots \leq i_k \leq n} f_{i_1, \ldots, i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}$$

be a differential k-form on an open subset U of $\mathbb{R}^n$. Its **exterior derivative** $d\alpha$ is the $(k+1)$-form on U defined by

$$d\alpha = \sum_{1 \leq i_1 \leq \ldots \leq i_k \leq n} (df_{i_1, \ldots, i_k} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k}.$$

Thus, if $\alpha = f$ is a 0-form, then $d\alpha = df$ as defined in Definition 1.7.38 for the exterior derivative of a 0-form (also called a differential).

A crucial fact about d is that $d \circ d = 0$, i.e., $d(d(\alpha)) = 0$ for any k-form α. One can prove this in general; here we'll just prove it for forms on $\mathbb{R}^3$.

Proposition 151 (**425b Homework 14 Problem 4**). $d \circ d = 0$ for d acting on

(a) 0-forms on $\mathbb{R}^3$ and

(b) 1-forms on $\mathbb{R}^3$.

(The equation is automatic for 2-forms and 3-forms on $\mathbb{R}^3$.)

Proof. We will use the following result:

Corollary 151.1 (**Corollary 5.17 of** Pugh [2015]). Corresponding mixed second partial derivatives of a second-differentiable function are equal:

$$\frac{\partial^2 f_k(p)}{\partial x_i \partial x_j} = \frac{\partial^2 f_k(p)}{\partial x_j \partial x_i}.$$

(a) From the previous problem set, we have

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz.$$

Then

$$
\begin{aligned}
d(df) =& d\left(\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz\right)\\
=& d\left(\frac{\partial f}{\partial x}\right) \wedge dx + d\left(\frac{\partial f}{\partial y}\right) \wedge dy + d\left(\frac{\partial f}{\partial z}\right) \wedge dz\\
=& \left(\frac{\partial^2 f}{\partial x \partial x}dx + \frac{\partial^2 f}{\partial y \partial x}dy + \frac{\partial^2 f}{\partial z \partial x}dz\right) \wedge dx + \left(\frac{\partial^2 f}{\partial x \partial y}dx + \frac{\partial^2 f}{\partial y \partial y}dy + \frac{\partial^2 f}{\partial z \partial y}dz\right) \wedge dy\\
& + \left(\frac{\partial^2 f}{\partial x \partial z}dx + \frac{\partial^2 f}{\partial y \partial z}dy + \frac{\partial^2 f}{\partial z \partial z}dz\right) \wedge dz\\
=& \frac{\partial^2 f}{\partial y \partial x}dy \wedge dx + \frac{\partial^2 f}{\partial z \partial x}dz \wedge dx + \frac{\partial^2 f}{\partial x \partial y}dx \wedge dy + \frac{\partial^2 f}{\partial z \partial y}dz \wedge dy\\
& + \frac{\partial^2 f}{\partial x \partial z}dx \wedge dz + \frac{\partial^2 f}{\partial y \partial z}dy \wedge dz\\
=& -\frac{\partial^2 f}{\partial x \partial y}dx \wedge dy - \frac{\partial^2 f}{\partial x \partial z}dx \wedge dz + \frac{\partial^2 f}{\partial x \partial y}dx \wedge dy - \frac{\partial^2 f}{\partial y \partial z}dy \wedge dz\\
& + \frac{\partial^2 f}{\partial x \partial z}dx \wedge dz + \frac{\partial^2 f}{\partial y \partial z}dy \wedge dz\\
=& \ 0.
\end{aligned}
$$

(by Proposition 146)

where we applied Corollary 5.17 of Pugh [2015]

(b) For a 1-form $\alpha = fdx + gdy + hdz$, we have $d(\alpha) = d(fdx + gdy + hdz) = dfdx + dgdy + dhdz$, so

$$
\begin{aligned}
d(df(\alpha)) =& d\left(dfdx + dgdy + dhdz\right)\\
=& d\Bigg(\left[\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz\right]dx + \left[\frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy + \frac{\partial g}{\partial z}dz\right]dy\\
& + \left[\frac{\partial h}{\partial x}dx + \frac{\partial h}{\partial y}dy + \frac{\partial h}{\partial z}dz\right]dz\Bigg)\\
=& d\left(\left[\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz\right]\right)dx \wedge dx + d\left(\left[\frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy + \frac{\partial g}{\partial z}dz\right]\right)dy \wedge dy\\
& + d\left(\left[\frac{\partial h}{\partial x}dx + \frac{\partial h}{\partial y}dy + \frac{\partial h}{\partial z}dz\right]\right)dz \wedge dz\\
=& \ 0.
\end{aligned}
$$

□

Recapping again, we can interpret k-forms on $\mathbb{R}^3$ in terms of familiar vector calculus quantities as follows, compatibly with Proposition 149.

Proposition 152 (**425b Homework 14 Problem 5**). (a) Let α be a 1-form on $\mathbb{R}^3$. Let $\boldsymbol{V} = V_1 e_1 + v_2 e_2 + V_3 e_3$ be the corresponding vector field (see Definition 1.7.16). Then the 2-form $d\alpha$ corresponds to the vector field $\operatorname{curl}(V)$.

(b) Let β be a 2-form on $\mathbb{R}^3$. Let $\boldsymbol{W} = W_1 e_1 + W_2 e_2 + W_3 e_3$ be the corresponding vector field. Then the 3-form $d\beta$ corresponds to the function $\operatorname{div}(W)$.

Proposition 153 (**Math 425b Final Exam problem**). Let $\boldsymbol{F} = (F_1, F_2, F_3)^\top$ be a vector field on $\mathbb{R}^3$. Let ω be the 2-form corresponding to the vector field $\boldsymbol{F}$ in the usual way. Then $d\omega = \left(\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} \right) dx \wedge dy \wedge dz$, which corresponds to the divergence of $\boldsymbol{F}$.

Proof. This vector field corresponds with the 2-form $\omega = F_1 dy \wedge dz + F_2 dz \wedge dx + F_3 dx \wedge dy = F_1 dy \wedge dz - F_2 dx \wedge dz + F_3 dx \wedge dy$. We have

$$
\begin{aligned}
d\omega &= d\left(F_1 dy \wedge dz + F_2 dz \wedge dx + F_3 dx \wedge dy \right) \\
&= dF_1 \wedge dy \wedge dz + dF_2 \wedge dz \wedge dx + dF_3 \wedge dx \wedge dy \\
&= \left(\frac{\partial F_1}{\partial x} dx + \frac{\partial F_1}{\partial y} dy + \frac{\partial F_1}{\partial z} dz \right) \wedge dy \wedge dz + \left(\frac{\partial F_2}{\partial x} dx + \frac{\partial F_2}{\partial y} dy + \frac{\partial F_2}{\partial z} dz \right) \wedge dz \wedge dx \\
&\quad + \left(\frac{\partial F_3}{\partial x} dx + \frac{\partial F_3}{\partial y} dy + \frac{\partial F_3}{\partial z} dz \right) \wedge dx \wedge dy \\
&= \frac{\partial F_1}{\partial x} dx \wedge dy \wedge dz + \frac{\partial F_1}{\partial y} dy \wedge dy \wedge dz + \frac{\partial F_1}{\partial z} dz \wedge dy \wedge dz + \frac{\partial F_2}{\partial x} dx \wedge dz \wedge dx + \frac{\partial F_2}{\partial y} dy \wedge dz \wedge dx \\
&\quad + \frac{\partial F_2}{\partial z} dz \wedge dz \wedge dx + \frac{\partial F_3}{\partial x} dx \wedge dx \wedge dy + \frac{\partial F_3}{\partial y} dy \wedge dx \wedge dy + \frac{\partial F_3}{\partial z} dz \wedge dx \wedge dy \\
&= \frac{\partial F_1}{\partial x} dx \wedge dy \wedge dz + 0 - \frac{\partial F_1}{\partial z} dz \wedge dz \wedge dy - \frac{\partial F_2}{\partial x} dx \wedge dx \wedge dz + \frac{\partial F_2}{\partial y} dx \wedge dy \wedge dz \\
&\quad + 0 + 0 - \frac{\partial F_3}{\partial y} dy \wedge dy \wedge dx + \frac{\partial F_3}{\partial z} dx \wedge dy \wedge dz \\
&= \left(\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} \right) dx \wedge dy \wedge dz.
\end{aligned}
$$

By Proposition 152, if ω is a 2-form on $\mathbb{R}^3$ and $\boldsymbol{F}$ is the corresponding vector field, then $d\omega$ corresponds to $\operatorname{div}(\boldsymbol{F})$. This makes sense because

$$
\operatorname{div}(\boldsymbol{F}) = \nabla \cdot \boldsymbol{F} = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z},
$$

which corresponds to the 3-form $\left(\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} \right) dx \wedge dy \wedge dz$.

□

Homework 15

One abstract way to define the tangent space $(T\mathbb{R}^n)_p$ at a point $p \in \mathbb{R}^n$ (see Definition 1.7.14) is as a set of equivalence classes of differentiable or smooth curves γ with $\gamma(0) = p$, where the equivalence relation is $\gamma_1 \sim \gamma_2 \iff \gamma_1'(0) = \gamma_2'(0)$. (On a manifold you'd require that this holds in some, or equivalently any, coordinate chart; see Definition 1.7.13.) It's clear that for $\mathbb{R}^n$ we can think of an equivalence class $[\gamma]$ as being uniquely determined by the velocity vector $\gamma'(0)$, so that the set of equivalences classes can be identified with $\mathbb{R}^n$, and we can pass the vector-space structure of $\mathbb{R}^n$ to this set of equivalence classes (making it an n-dimensional vector space).

We will consider the cotangent space $(T^*\mathbb{R}^n)_p$. It will consist of the sorts of things that give linear functionals on tangent vectors—a reasonable thing to think of for this is "functions." If f is a smooth function, then the directional derivative of f with respect to tangent vectors v should give a linear functional on the tangent space (abstractly, this functional is $v \mapsto (Df)_p(v)$.

So we would like to define $(T^*\mathbb{R}^n)_p$ as a vector space of functions. The first issue that arises is that spaces of functions are typically infinite-dimensional. One reason why this is is that functions have values at points far away from p, but these values should be irrelevant for directional derivatives, so to work toward a finite-dimensional cotangent space, we will use the following concept.

Definition 1.7.42. Let $p \in \mathbb{R}^n$ and consider the set $\mathcal{F}$ of pairs (U, f) where U is an open neighborhood of p and $f : U \to \mathbb{R}$ is a smooth function. Define an equivalence relation on $\mathcal{F}$ by $(U_1 f_1) \sim (U_2, f_2)$ if there exists $U \subset U_1 \cap U_2$ with $p \in U$ such that f_1 and f_2 are equal on all points of U (i.e., they agree when restricted to U). The set of equivalence classes $F/\sim$ is denoted by $C^\infty(\mathbb{R}^n)_p$, the "stalk at p of the sheaf of smooth functions on $\mathbb{R}^n$." Elements of $C^\infty(\mathbb{R}^n)_p$, i.e. equivalence classes $[U, f]$, are called **germs of smooth functions at** p.

Definition 1.7.43 (**Tangent covectors on manifolds; p. 275 of** Lee [**2012**]**, p. 290 of pdf**)**.** Let M be a smooth manifold with or without boundary. For each $p \in M$, we define the **cotangent space at** p, denoted by T_p^*M or $(T^*M)_p$, to be the dual space to T_pM: $T_p^*M = (T_pM)^*$, or $(T^*M)_p = ((TM)_p)^*$. Elements of T_p^*M are called **tangent covectors at** p, or just **covectors at** p.

$$\vdots$$

Definition 1.7.44. A **vector bundle**/M is a vector space V_p for each $p \in M$, "ranging smoothly." TM is the **tangent bundle** of M.

Definition 1.7.45. If $V : \{V_p : p \in M\}$ is a vector bundle/M, then a **section** of V is a map $M \xrightarrow{\sigma} \bigcup_{p \in M} V_p$ such that for all $p \in M$, $\sigma(p) \in V_p$.

Thus, a vector field on M (see Definition 1.7.16) is a section of the tangent bundle.

Similarly, we have another vector bundle on M: the value at the point p is the dual vector space $(TM)_p)^*$, the **cotangent space** of M at p. This bundle is called the **cotangent bundle** of M, written T^*M. A differential 1-form on M is a **section of the cotangent bundle**. (A differential k-form on M is a section of the kth exterior power of the cotangent bundle, $\wedge^k T^*M$.)

Note: standard vector calculus textbooks (e.g. Stewart [2015]) turns all 1-forms on $\mathbb{R}^n$ into vector fields (implicitly using standard inner product/Riemanian metric on $\mathbb{R}^n$; see Remark 18). This is why ∇f is familiar while df is not.

We will also see: 2-forms, 3-forms on $\mathbb{R}^3$ are turned into vector fields and functions, respectively.

Now we will consider line integrals and integrals of 1-forms on curves and 1-manifolds (see Definition 1.7.10). Stewart [2015] discusses 3 types of line integrals. We can understand all three in terms of integrals of 1-forms on curves C. But, we need to consider C as its own manifold that we can integrate things on. If M is a manifold of dimension n, what can I integrate on M? Answer: a **top-degree differential form** (similar to saying "a function," but when $M \neq \mathbb{R}^n$, this distinction is important). Case we care about now:

- $M = [a, b] \subset \mathbb{R}$: "1-manifold with boundary." Let $t \in M$. Any 1-form on M is $\alpha = f(t)dt$ for some smooth function f. Can define $\int_{[a,b]} \alpha := \int_a^b \alpha = \int_a^b f(t)dt$.

  Note that differential forms language gives meaning to $f(t)$ and dt individually. $f(t)$ is a 0-form, dt is a 1-form, can multiply together. (However, you can't divide differential forms—$\frac{dy}{dx}$ can't be understood in terms of differential forms.)

- C: **curve** in $\mathbb{R}^2$ (or $\mathbb{R}^3$, $\mathbb{R}^n$, etc.). e.g. C is a circle centered at the origin with radius 1. If we have a function f on C, we don't know how to integrate it (*a priori*). But if we have a 1-form α on C, we should be able to integrate it.

  How to define this? Pick a parameterization of C, $\boldsymbol{r} = [a,b] \mapsto C \subset \mathbb{R}^n$ (want $\mathbb{R}$ to be nice enough: $\boldsymbol{r}'(t) \neq 0$ for all t, $\boldsymbol{r}$ is surjective, $\boldsymbol{r}$ is injective except possibly at finitely many points). Then define

  $$\int_C \alpha := \int_{[a,b]} \boldsymbol{r}^*(\alpha). \tag{1.38}$$

  Can show it's independent of $\boldsymbol{r}$, as long as the orientation (the direction we travel on the curve) matches the orientation of $\boldsymbol{r}$ ("orientation compatible"). This allows us to compute $\int_C \alpha$.

If α is a 1-form on $\mathbb{R}^n$, $\alpha = \alpha_1 dx^1 + \ldots + \alpha_n dx^n$. Can view α as giving a 1-form on C (pullback of α via inclusion $C \mapsto \mathbb{R}^n$; recall Definitions 1.7.39 and 1.7.53). Then we can compute $\int_C \alpha$ be picking $\boldsymbol{r}$ and using (1.38); this will be concrete and will recover line integrals of vector fields (see Definition 1.7.16).

Definition 1.7.46 (From Stewart [2015]).   1. **Line integral of a scalar field/function on C with respect to arc length:** Let $f : \mathbb{R}^n \to \mathbb{R}$ (or $C \to \mathbb{R}$) be a smooth function. The line integral $\int_C f \ ds = \int_C f(\boldsymbol{r}) \ ds$ is defined by

$$\int_C f \ ds := \int_a^b f(\boldsymbol{r}(t)) \|\boldsymbol{r}'(t)\| dt.$$

(This definition is independent of the parameterization $\boldsymbol{r}$, as long as orientations are respected.) If $\gamma(a) = \gamma(b)$ (C is closed), we can write $\oint_C f \ ds$.

2. **Line integral of a scalar field/function on C with respect to dx:** Let $f : \mathbb{R}^n \to \mathbb{R}$ (or $C \to \mathbb{R}$) be a smooth function. The line integral $\int_C f \ dx = \int_C f(\boldsymbol{r}) \ dx$ is defined by

$$\int_C f \ dx := \int_a^b f(\boldsymbol{r}(t)) r_i'(t) \ dt,$$

where $\boldsymbol{r} = (r_1, \ldots, r_n)^\top$. For example, let $P = P(x, y)$ be a function from $\mathbb{R}^2$ to $\mathbb{R}$. We can take

$$\int_C P(x, y) \ dx = \int_C P \ dx = \int_a^b P(r_1(t), r_2(t)) r_1'(t) dt$$

and $\int_C P(x, y) \ dy = \int_C P \ dy$, etc. (See Example 1.7.12.)

3. **Line integral of a vector field (see Definition 1.7.16) on** C**:** Let $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$ (or $C \to \mathbb{R}^n$) be a vector field on C. The line integral $\int_C \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r} = \int_C \boldsymbol{F} \cdot d\boldsymbol{r}$ is defined by

$$\int_C \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r} := \int_a^b \boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) \ dt,$$

Proposition 154. Let P be a function on $\mathbb{R}^n$, so that $P \ dx$ is a 1-form on $\mathbb{R}^n$. Let $C, \boldsymbol{r}$ be as above. Then the definitions of $\int_C P dx^i$ as in Definitions 1.7.46 and [**in the above differential 1-form sense**] agree.

Proof. Note $P dx^i$ is a 1-form on $\mathbb{R}^n$, pulled back to C. So starting from the 1-form formulation, we have

$$\int_C P dx^i = \int_{[a,b]} \boldsymbol{r}^*(P dx^i) \qquad \text{(using (1.38))}$$

$$\text{(by Proposition 147(b))} \quad = \int_{[a,b]} (P \circ \boldsymbol{r}) \boldsymbol{r}^*(dx^i)$$

$$\text{(by Proposition 147(a))} \quad = \int_{[a,b]} (P \circ \boldsymbol{r}) d(\boldsymbol{r}^*(x_i))$$

$$= \int_{[a,b]} (P \circ \boldsymbol{r}) d(r_i)$$

$$= \int_{[a,b]} (P \circ \boldsymbol{r}) r_i'(t) dt \qquad \text{(at a point } t \in [a,b], \text{ by HW 13)},$$

which is the calculus definition of an integral in Definition 1.7.46 part (2). (See also Example 1.7.12.)

$$\square$$

Now we will address Definition 1.7.46 part (3). Write $\boldsymbol{F} = (F_1, \ldots, F_n)^\top$ (a vector field (see Definition 1.7.16) on $\mathbb{R}^n$ and let $\alpha := F_1 dx^1 + \ldots + F_n dx^n$, a 1-form on $\mathbb{R}^n$. These have the usual correspondence from the homework.

Proposition 155. The definition $\int_C \alpha$ in 1-forms matches the Definition 1.7.46 part (3) for $\int_C \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r}$.

Proof. Starting from the 1-form definition, we have

$$\int_C \alpha = \int_{[a,b]} \boldsymbol{r}^*(\alpha) \qquad \text{(using (1.38))}$$

$$= \int_{[a,b]} \boldsymbol{r}^*(F_1 dx^1 + \ldots + F_n dx^n)$$

$$\text{(by Proposition 147(c))} \quad = \int_{[a,b]} F_1(\boldsymbol{r}(t)) r_1'(t) dt + \ldots + \int_{[a,b]} F_n(\boldsymbol{r}(t)) r_n'(t) dt$$

$$= \int_{[a,b]} \left(F_1(\boldsymbol{r}(t)) r_1'(t) + \ldots + \int_{[a,b]} F_n(\boldsymbol{r}(t)) r_n'(t) \right) dt$$

$$= \int_C \boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) dt$$

$$= \int_C \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r}.$$

$\square$

For example, if $\boldsymbol{F}(x,y) = (P(x,y), Q(x,y))^\top$, then $\int_C \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r} = \int_C (P(x,y)dx + Q(x,y)dy)$, which is a 1-form integral of type 2 from Definition 1.7.46. Stewart [2015] gives a similar computation as the relationship between Type 2/3 integrals.

4/24 notes

Finally, we will address Type 1 integrals from Definition 1.7.46. The general principle is that on a curve $C \subset \mathbb{R}^n$, parameterized by a smooth function $\boldsymbol{r}$ (so $\boldsymbol{r} : [a,b] \to C \subset \mathbb{R}^n$), there exists a unique 1-form α on C (a function from C to $(\mathbb{R}^n)^*$) with

$$\boldsymbol{r}^*(\alpha) = \|\boldsymbol{r}'(t)\| dt.$$

α is typically called ds, the "arc length 1-form on C." That is, ds is the unique 1-form on C such that

$$\boldsymbol{r}^*(ds) = \|\boldsymbol{r}'(t)\| dt. \tag{1.39}$$

Recall from Definition 1.7.39 that at a point $t \in [a,b]$ and for $v \in \mathbb{R}$, we have that the pullback $\boldsymbol{r}^*(ds)$ is the differential 1-form on $[a,b]$ (that is, a function from $[a,b]$ to $(\mathbb{R})^*$, so a function from $[a,b]$ to the space of functionals with domain $\mathbb{R}$) defined at a point $t \in [a,b]$ by

$$(\boldsymbol{r}^*(ds))_t(v) := ds_{\boldsymbol{r}(t)}((D\boldsymbol{r})_t(v))$$

for $v \in \mathbb{R}$. (Note that $(D\boldsymbol{r})_t(v) \in \mathbb{R}^n$, so it makes sense to evaluate $ds_{\boldsymbol{r}(t)}$ on the vector $(D\boldsymbol{r})_t(v)$.)

Warning: this is not d of something; ds is historical notation. In particular, if $\boldsymbol{r}$ is injective, define $\boldsymbol{r} : [a,b] \to C \subset \mathbb{R}^n$. Let s be the unique function on C such that $s(\boldsymbol{r}(t)) = \int_a^t \|\boldsymbol{r}'(u)\| du$. Then ds is d applied to s (the exterior derivative of s). However, if $\boldsymbol{r}$ is not injective (e.g. $\boldsymbol{r}(a) = \boldsymbol{r}(b)$ for a closed curve), then $s(\boldsymbol{r}(a)) = 0$ while $s(\boldsymbol{r}(b)) \neq 0$ (this is the arc length of C). This is impossible if $\boldsymbol{r}(a) = \boldsymbol{r}(b)$. So in this case, ds is not d of s. (not important aside: ds is a generator for the deRham cohomology group $H'(C)$ (see Remarks on homework).)

Given a type (1) integral $\int_C f ds$, we have a 1-form $f ds$ on C. We expect the classical type (1) integral is the integral of this 1-form.

Proposition 156. The definition of $\int_C f ds$ as defined from the above function from f on C agrees with the definition from Definition 1.7.46 part (1).

Proof. Starting from the 1-form definition,

$$\int_C f ds := \int_{[a,b]} \boldsymbol{r}^*(f \ ds) \qquad \text{(using (1.38))}$$

$$\text{(by Proposition 147(b))} \quad = \int_{[a,b]} (f \circ \boldsymbol{r}) \boldsymbol{r}^*(ds)$$

$$= \int_a^b f(\boldsymbol{r}(t)) \boldsymbol{r}^*(ds)$$

$$\text{(using (1.39))} \quad = \int_a^b f(\boldsymbol{r}(t)) \|\boldsymbol{r}'(t)\| dt,$$

which matches Definition 1.7.46 part (1).

$\square$

Remark 21. Interesting fact about 1-forms on C: any 1-form on C is of the form $\alpha = f \ ds$; i.e., α is some function times arc-length 1-form ds. i.e. integrals of type (1) are just as general as integrals of type (3). In particular, any 1-form α on C (corresponding to a vector field $\boldsymbol{F}$ on C) (see Definition 1.7.16) can be written as $\alpha = (\boldsymbol{F} \cdot \boldsymbol{T}) ds$. Therefore classical line integrals are integrals of 1-forms on curves. (We will see later that the same is true for surface integrals and 2-forms.)

If we have a vector field $\boldsymbol{F}$ on C and we want to interpret $\int_C \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r}$ as a type (1) integral, we can write

$$\int_C \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r} = \int_a^b \boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) dt = \int_a^b \boldsymbol{F}(\boldsymbol{r}(t)) \cdot \underbrace{\frac{\boldsymbol{r}'(t)}{\|\boldsymbol{r}'(t)\|}}_{\text{unit tangent vector}} \|\boldsymbol{r}'(t)\| dt$$

Let $\boldsymbol{T}(\boldsymbol{r}(t)) := \frac{\boldsymbol{r}'(t)}{\|\boldsymbol{r}'(t)\|}$ where $\boldsymbol{T}$ is the unique vector field on C defined this way. (This really makes most sense if $\boldsymbol{r}$ is injective.) Then we can write this as

$$\int_C \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r} = \int_a^b (\boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{T}(\boldsymbol{r}(t))) \|\boldsymbol{r}'(t)\| dt = \int_C (\boldsymbol{F} \cdot \boldsymbol{T}) ds$$

where $\boldsymbol{F} \cdot \boldsymbol{T}$ is a scalar field on C obtained from the original $\boldsymbol{F}$ (vector field on C; see Definition 1.7.16) by the dot product with $\boldsymbol{T}$ (a fixed vector field on C). (See Proposition 157.) Thus, $\int_C (\boldsymbol{F} \cdot \boldsymbol{T}) ds$ is another standard notation for $\int_C \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r} = \int_C \alpha$. This shows that in terms of forms, any 1-form α on C (corresponding to a vector field $\boldsymbol{F}$ on C) can be written as $\alpha = (\boldsymbol{F} \cdot \boldsymbol{T}) ds$.

Proposition 157 (Math 425b Final Exam problem). Let C be a curve in $\mathbb{R}^n$ parameterized by $\boldsymbol{r} : [a,b] \to C \subset \mathbb{R}^n$. Assume that $\boldsymbol{r}$ is smooth when viewed as a map from $[a,b]$ to $\mathbb{R}^n$ (i.e. $i \circ \boldsymbol{r}$ is smooth where $i : C \to \mathbb{R}^n$ is the inclusion map), that $\boldsymbol{r}$ is bijective, and that $\boldsymbol{r}'(t) \neq 0$ for all $t \in [a,b]$.

Let α be a 1-form on $\mathbb{R}^n$, corresponding to a vector field $\boldsymbol{F}$ on $\mathbb{R}^n$ under the usual correspondence. Then

$$\int_C \alpha = \int_C (\boldsymbol{F} \cdot \boldsymbol{T}) ds.$$

Proof. ds is the unique 1-form on C such that $\boldsymbol{r}^*(ds) = \|\boldsymbol{r}'(t)\| dt$. The unit tangent vector function can be defined on C in terms of $\boldsymbol{r}$ by $\boldsymbol{T}(\boldsymbol{r}(t)) := \frac{\boldsymbol{r}'(t)}{\|\boldsymbol{r}'(t)\|}$ (note that this definition is okay since by assumption $\boldsymbol{r}'(t) \neq 0$ for all $t \in [a,b]$). Let

$$\boldsymbol{F} = \begin{bmatrix} F_1 \\ \vdots \\ F_n \end{bmatrix}.$$

The 1-form α on $\mathbb{R}^n$ corresponding to $\boldsymbol{F}$ in the usual way can be written as

$$\alpha = F_1 dx_1 + \ldots + F_n dx_n.$$

Then we have

$$
\begin{aligned}
\int_C \alpha &= \int_{[a,b]} \boldsymbol{r}^*(\alpha) \qquad \text{(using (1.38))} \\
&= \int_{[a,b]} \boldsymbol{r}^* \left(\sum_{i=1}^n F_i dx_i \right) \\
\text{(by Proposition 147(c))} \quad &= \int_{[a,b]} \sum_{i=1}^n F_i \left(\boldsymbol{r}(t) \right) r_i'(t) dt \\
&= \int_a^b \boldsymbol{F}(\boldsymbol{r}) \cdot \boldsymbol{r}'(t) dt \\
&= \int_a^b \left(\boldsymbol{F}(\boldsymbol{r}(t)) \cdot \frac{\boldsymbol{r}'(t)}{\|\boldsymbol{r}'(t)\|} \right) \|\boldsymbol{r}'(t)\| dt \\
&= \int_a^b \left(\boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{T}(\boldsymbol{r}(t)) \right) \|\boldsymbol{r}'(t)\| dt \\
&= \int_a^b \left(\boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{T}(\boldsymbol{r}(t)) \right) \boldsymbol{r}^*(ds) \\
\text{(by Proposition 147(a); permissable because } \boldsymbol{r} \text{ is injective)} \quad &= \int_a^b \left(\boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{T}(\boldsymbol{r}(t)) \right) d(s \circ \boldsymbol{r}) \\
&= \int_C (\boldsymbol{F} \cdot \boldsymbol{T}) ds.
\end{aligned}
$$

$\square$

Note: In Section 5.8 of Pugh [2015], differential 1-forms are defined by how they pair with curves C (or $\boldsymbol{r}$) to give an integral $\int_C \alpha$. I.e., if α is a differential 1-form in the sense described in these notes, α gives a functional on the set of parameterized curves $\boldsymbol{r} : [0,1] \to \mathbb{R}^n$ by $\Phi_\alpha(\boldsymbol{r} : [a,b] \to \mathbb{R}^n) := \int_C \alpha$, $C = \text{image}(\boldsymbol{r})$. Pugh [2015] defines α in terms of Φ_α; this is non-standard, but follows Rudin [1976] for pedagogical reasons (to hide the algebra). We are doing the algebra-based approach. In this viewpoint, a differential form doesn't mean anything at a point. In ours, it does—it's an alternating multilinear map (see Definition 1.7.33).

Example 1.7.17 (**Line integrals in complex analysis**). We often encounter integrals like $\int_C f(z)dz$ where C is a curve in the complex plane. We can interpret this in the following way. First, consider that $\mathbb{C} \cong \mathbb{R}^2$. We can write $z = x + iy$, where z is the identify function on $\mathbb{C}$, x is the first coordinate in $\mathbb{R}^2$, and y is the second coordinate in $\mathbb{R}^2$. Now we can take d of these functions on $\mathbb{R}^2$, and get the 1-form

$dz = dx + idy$ on $\mathbb{R}^2$. (It's ok to do complex-valued differential 1-forms $\text{Alt}_k(\mathbb{R}^n, \mathbb{C})$ (with $\mathbb{R}$ multilinear) or $\text{Alt}_k(\mathbb{C}^n, \mathbb{C})$ (with $\mathbb{C}$ multilinear). Then

$$\int_C f(z)dx = \int_C f(x,y)dx + i \int_C f(x,y)dy$$

(and usually we write $f = u + iv$ and expand further). Now this integral is in the form of the integrals we've worked with in these notes.

Some terminology: if f is a holomorphic function, then $f(z)dz$ is a holomorphic 1-form while $f(z)d\overline{z} = f(z)(dx - idy)$ is not a holomorphic 1-form.

find definition of smooth manifold

in inclusion map notes:

Summary: if $\boldsymbol{F}$ is a vector field (see Definition 1.7.16) on a curve $C \subset \mathbb{R}^n$, the corresponding 1-form on C is

$$\underbrace{(\boldsymbol{F} \cdot \boldsymbol{T})}_{\text{function on } C} \underbrace{ds}_{\text{arc-length 1-form of } C} .$$

This explains why $\int_C (\boldsymbol{F} \cdot \boldsymbol{T})ds$ occurs so often in vector calculus (e.g., line integrals).

Definition 1.7.47 (**Closed differential forms; p. 92 of** Spivak [**1971**]**, p. 105 of pdf**). A differential form α is called **closed** if $d\alpha = 0$.

Definition 1.7.48 (**Exact differential forms; p. 92 of** Spivak [**1971**]**, p. 105 of pdf**). A differential form α is called **exact** if $\alpha = df$ for some function (0-form) f.

Remark 22. Theorem 4-10 in Spivak [1971] shows that every exact form is closed.

4/27 notes: FTC for line integrals, wedge products, multiple integrals, surface integrals, etc.

Theorem 158 (**Fundamental Theorem of Calculus for line integrals**). Let C be a curve in $\mathbb{R}^n$ (the image of $\boldsymbol{r} : [a,b] \to \mathbb{R}^n$ where $\boldsymbol{r}'(t) \neq 0$ and $\boldsymbol{r}$ is injective except maybe at the endpoints). Let $\boldsymbol{f} : C \to \mathbb{R}$ be a smooth function (∇f: vector field on C). Then

$$\int_C (\nabla f \cdot \boldsymbol{T})ds = f(\boldsymbol{r}(b)) - f(\boldsymbol{r}(a)).$$

(In particular, this means $\oint (\nabla f \cdot \boldsymbol{T})ds = 0$, where $\oint$ denotes that the endpoints are equal so the curve is closed.)

Proof.

$$\int_C (\nabla f \cdot \boldsymbol{T}) ds = \text{(by result from 4/24)} \int_C df$$

$$:= \int_{[a,b]} \boldsymbol{r}^*(df)$$

$$= \int_{[a,b]} d(f \circ \boldsymbol{r})$$

$$= \int_a^b \frac{\mathrm{d} f \circ \boldsymbol{r}}{\mathrm{d} t} dt$$

$$\text{(by FTC)} \quad = f(\boldsymbol{r}(b)) - f(\boldsymbol{r}(a)).$$

$\square$

We can write this in forms language as

$$\int_C df = \int_{\partial C} f$$

(where ∂C means "boundary of C"). In fact, we can generalize this.

Theorem 159 (Generalized Stokes' Theorem). If M is a smooth n-manifold (see Definition 1.7.10) with boundary[4] and ω is an $(n-1)$-form on M, then

$$\int_M d\omega = \int_{\underbrace{\partial M}_{\text{boundary of } M}} \underbrace{\omega}_{\text{really } i^*(\omega)}$$

(where $i : \partial M \to M$ is the inclusion operator).

Remark 23. This theorem implies Green's Theorem, Stokes' Theorem, the Divergence Theorem, and more. Our goal for the remainder of class will be to see how this generalizes the classical theorems.

First, one remark about 1-forms.

Remark 24. Consider a first order ODE $y'(x) = F(x, y(x))$ (this is not autonomous). A method for solving this: write $F(x, y) = \frac{-P(x,y)}{Q(x,y)}$ for some functions $P(x, y)$ and $Q(x, y)$. (Note that there is not a unique way of doing this—exist several valid choices of P and Q.) The ODE is

$$\frac{\mathrm{d} y}{\mathrm{d} x} = \frac{-P(x, y)}{Q(x, y)},$$

i.e., $P(x, y) + Q(x, y)\frac{\mathrm{d} y}{\mathrm{d} x} = 0$. Now we "multiply by dx" (not meaningful in our formulation with differential forms):

$$\underbrace{P(x, y) dx + Q(x, y) dy}_{\text{looks like a 1-form}} = 0. \tag{1.40}$$

---

[4]we will not precisely define this term here; treat this as an intuitive notion for now

Now we will try to find a function $f(x,y)$ with $df = P(x,y)dx + Q(x,y)dy$. (Recall a 1-form α is called **exact** if $\alpha = df$; see Definition 1.7.48.) If we can find f, then $\frac{\partial f}{\partial x} = P$, $\frac{\partial f}{\partial y} = Q$, so the ODE is $\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x} = 0$, or $\begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} 1 \\ \frac{dy}{dx} \end{bmatrix} = 0$. By the Chain Rule (Theorem 122) we can write this as

$$= \frac{d}{dx}(f(x,y(x))) = (DF)_{(x,y(x))}\left(\begin{bmatrix} 1 \\ y'(x) \end{bmatrix}\right) = 0.$$

Then we can integrate, which yields $f(x,y(x)) = c$ for some $c \in \mathbb{R}$. Then (try to) solve for $y(x)$ as a function of x (by the Implicit Function Theorem, this is usually possible).

How can we express this in the language of differential forms? The equation (1.40) is suspect given the un-allowed transformations we did. Should mean P and Q are the zero function, which is not what we want. The missing ingredient is pullbacks (Definitions 1.7.39 and 1.7.53).

Proposition 160. A function $y = y(x)$ solves the ODE $y' = \frac{-P(x,y)}{Q(x,y)}$ if and only if $\gamma^*(Pdx + Qdy) = 0$, where $\gamma(t) := (t, y(t))$.

Proof.

$$\gamma^*(P(x,y)dx + Q(x,y)dy) = P(t,y(t))d(x \circ \gamma) + Q(t,y(t))d(y \circ \gamma)$$
$$= P(t,y(t))d(t) + Q(t,y(t))d(y(t))$$
$$= P(t,y(t))dt + Q(t,y(t))y'(t)dt$$
$$= (P(t,y(t)) + Q(t,y(t))y'(t))\, dt$$
$$= 0 \qquad \Longleftrightarrow \quad \text{coefficient of } dt \text{ equals } 0 \text{ for all } t,$$

which is true if and only if $y(t)$ solves the ODE.

$\square$

To summarize: we can view any first order ODE $y' = F(x,y)$ as an equation $\gamma^*(\alpha) = 0$ for some non-unique choice of 1-form α on $\mathbb{R}^2$. So an equation for an unknown function of y is equivalent to an equation for an unknown path γ of the form $\gamma(t) = (t, y(t))$. Saying "the ODE is exact" is bad, since α is not unique. Should say α is exact (see Definition 1.7.48). Indeed:

Theorem 161. If $F : \mathbb{R}^2 \to \mathbb{R}$ is smooth, there exist smooth functions $P(x,y), Q(x,y) : \mathbb{R}^2 \to \mathbb{R}$ with $F(x,y) = -\frac{P(x,y)}{Q(x,y)}$ and $Pdx + Qdy$ is an exact 1-form (see Definition 1.7.48) on $\mathbb{R}^2$.

Equivalently: if $P(x,y)dx + Q(x,y)dy$ is a 1-form on $\mathbb{R}^2$ such that P, Q don't both vanish at (x,y), then there exists a smooth functoin $\mu(x,y) : \mathbb{R}^2 \to \mathbb{R}$ such that $\mu Pdx + \mu Qdy$ is exact. (note: $-\frac{\mu P}{\mu Q} = -\frac{P}{Q}$.)

Definition 1.7.49. A function μ like this (that we multiply a 1-form by to get an exact 1-form) is called an **integrating factor** for the 1-form $Pdx + Qdy$.

So the theorem is saying any nonvanishing 1-form α on $\mathbb{R}^2$ has an integrating factor. (It turns out that on $\mathbb{R}^n$ you need $\alpha^n d\alpha = 0$; see the Frobenius theorem in differential geometry.) So you can solve an arbitrary first-order ODE $y' = F(x,y)$ by finding an integrating factor, then applying the above method. Finding an integrating factor is hard in general: if $\alpha = (P(x)y - Q(x))dx + dy$, then can take $e^{\int P(x)dx}$ as an integrating factor (compare with ODE books), but other than tricks in special cases like this, the general problem is hard.

Remark 25. We can often say that given $\alpha = Pdx + Qdy$, we can test if it's exact by taking $d\alpha = \left(\frac{\partial Q}{\partial xZ} - \frac{\partial P}{\partial y}\right) dx \wedge dy$. If this is 0, α is closed (see Definition 1.7.47) and (usually) exact. This statement depends on the domain U that α is defined on—all closed 1-forms on U are exact if and only if (by definition of H') $H'(U) = 0$, where $H'(U)$ is the de Rhan cohomology group of U: $H'(U) = \frac{\text{closed 1 forms on } U}{\text{exact 1-forms on } U}$.

4/29 notes

Definition 1.7.50 (In-class general definition of wedge products). If V is a finite-dimensional vector space over $\mathbb{R}$ (or $\mathbb{C}$ or any field), the **tensor algebra** of V, denoted by $T^{\cdot}(V)$, is defined to be

$$\bigoplus_{k \geq 0} \underbrace{V \oplus \ldots \oplus V}_{k \text{ factors}} = \mathbb{R} \oplus V \oplus (V \oplus V) \oplus \ldots.$$

(The summand $V \oplus \ldots \oplus V$ with k factors is called $T^k V$, the k**th tensor power** of V. We have a bilinear map $T^k V \times T^\ell V \to T^{k+\ell} V$ (from the linear map $\text{id} : T^k V \otimes T^\ell V \to T^{k+\ell} V$) giving us a multiplication operator on $T^{\cdot} V$ (so we have a ring (see Definition **??** or $\mathbb{R}$-algebra structure) (and recall Definition 1.7.27 for the definition of the tensor product $\otimes$).

For example,

$$\underbrace{(V_1 \otimes V_2)}_{\in T^2(V)} \underbrace{(V_3 \otimes V_4 \otimes V_5)}_{\in T^3(V)} := \underbrace{V_1 \otimes V_2 \otimes V_3 \otimes V_4 \otimes V_5}_{\in T^5(V)}$$

Definition 1.7.51. The **exterior algebra** $\wedge^{\cdot} V$ is defined to be $(T^{\cdot}V)/(x \otimes x : x \in V)$, where $V = T^{-1}(v)$. Regarding $(x \otimes x : x \in V)$: R ring (see Definition **??**: $I \subset R$ is **ideal** if it's closed under addition and if $r \in R$; see Definition **??**, $s \in I$ then $rs \in I$. (Recall Definition 1.7.27 for the definition of the tensor product $\otimes$.) Exercise: R/I is also a ring defined like V/W for vector spaces.

If $S \subset R$ is any subset, then (S) is defined to be the ideal generated by S ("ring version of span," see Definition **??**). It equals $\text{span}_{\mathbb{R}}\{rs : r \in R, s \in S\}$. (This is an ideal of R.)

Exercise 8.   1. $\wedge^{\cdot}(V) = \bigoplus_{k \geq 0} \wedge^k V$, where $\wedge^k V$ is the image of $T^k V$ in quotient $\wedge^{\cdot} V = TV/(x \otimes x)$. ($T^k V$ is the set of equivalence classes of ??? $\sum_i v_{i1} \otimes \ldots \otimes v_{ik}$ (k factors))

Multiplication sends $\wedge^k V \times \wedge^\ell V \to \wedge^{k+\ell} V$.

Abstractly: $T^{\cdot} V$ is a **graded** ring (the degree-k part is $T^k V$). $(x \otimes x : x \in V)$ [**note: not sure if the symbol between xs in this part is $\otimes$ or $\oplus$**] is a **homogeneous ideal** of $T^{\cdot} V$. We have a "quadratic" ideal.

Quotient of a graded ring b homogeneous ideal is itself a graded ring.

   2.

$$\{v \in V^{\otimes k} = \text{swap}_i(v) = -v\} \to \wedge^k V$$
$$v + 1 \otimes \ldots \otimes v_k \to \frac{1}{k!}(v_n \otimes \ldots \otimes v_k)$$

(where $\text{swap}_i(v)$ sends $v_1 \otimes \ldots \otimes v_i \otimes v_{i+1} \otimes \ldots \otimes v_k$ to $v_1 \otimes \ldots \otimes v_{i+1} \otimes v_i \otimes \ldots \otimes v_k$. Exercise: this is a well-defined map from $V^{\otimes k}$ to itself.)

[$V^{\otimes k}$ **is "V tensor k"**] This is really a map $V^{\otimes k} \to \wedge^k V$, $v + 1 \otimes \ldots \otimes v_k \to \frac{1}{k!}(v_1 \otimes \ldots \otimes v_k)$, and we can restrict to $\{v \in V^{\otimes k} : \text{swap}_i(v) = -v \forall i\}$, a vector subspace of $V^{\otimes k}$.

The exercise is: The natural map from above is an isomorphism with inverse

$$[v_1 \otimes \ldots \otimes v_k] \mapsto \sum_{\sigma \in S_k} \text{sign}(\sigma) v_{\sigma(1)} \otimes \ldots \otimes v_{\sigma(k)}.$$

(To check this is well-defined: could define as $\wedge^{\cdot} V$ by defining on $T^{\cdot} V$ and checking the ideal $(x \otimes x : x \in V)$ gets sent to 0.

Note: the second map looks like what appeared on the homework, but without the $1/k!$ factor. The homework conventions are non-standard. Should use above conventions instead.

Notation: write $v_1 \wedge \ldots \wedge v_k$ for the equivalence class of $v_1 \otimes \ldots \otimes v_k$ in the quotient $\wedge^k V$. If $\alpha \in \wedge^k V$, $\beta \in \wedge^{\ell} V$, then write $\alpha \wedge \beta$ for their product in $\wedge^{k+\ell} V$.

We can apply this to differential forms. Take $V = (\mathbb{R}^n)^*$. Then $\text{Alt}_k(\mathbb{R}^n, \mathbb{R})$ can be viewed as a subspace of $V^{\otimes k}$. (From the homework: $\text{Alt}_k(\mathbb{R}^n, \mathbb{R}) \subset (\mathbb{R}^n \otimes \ldots \otimes \mathbb{R}^n)^* = (\mathbb{R}^n)^* \otimes \ldots \otimes (\mathbb{R}^n)^* = V \otimes \ldots \otimes V = V^{\otimes k}$.) Specifically, it's $\{\psi \in (\mathbb{R}^n)^* \otimes \ldots \otimes (\mathbb{R}^n)^* : \text{swap}_i(\psi) = -\psi \quad \forall i\}$.

The previous exercise gives us an isomorphism between $\text{Alt}_k(\mathbb{R}^n, \mathbb{R})$ and $\wedge^k(\mathbb{R}^n)^*$. Corollary: a differential k-form on $U \subset \mathbb{R}^n$ (U open) is a function from $U \to \wedge^k((\mathbb{R}^n)^*)$.

Definition 1.7.52. If α is a k-form on U and β is an ℓ-form on U, define the $(k + \ell)$-form $\alpha \wedge \beta$ on U by

$$(\alpha \wedge \beta)_p := \underbrace{\alpha_p}_{\wedge^k((\mathbb{R}^n)^*)} \wedge \underbrace{\beta_p}_{\wedge^{\ell}((\mathbb{R}^n)^*)}$$

The result is in $\wedge^{k+\ell}((\mathbb{R}^n)^*)$ as desired.

Upshot: wedge products are trivial to define in the "quotient description:" sum over permutations appearing when relating to the subspace description.

Corollary 161.1. If $\alpha \in \text{Alt}_k(\mathbb{R}^n, \mathbb{R})$ and $\beta \in \text{Alt}_{\ell}(\mathbb{R}^n, \mathbb{R})$, then $\alpha \wedge \beta \in \text{Alt}_{k+\ell}(\mathbb{R}^n, \mathbb{R})$ sends $v_1, \ldots, v_{k+\ell}$ to

$$\frac{1}{k! \ell} \sum_{\sigma \in S_{k+\ell}} \text{sign}(\sigma) \alpha(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) \beta(v_{\sigma(1)}, \ldots, v_{\sigma(k)})$$

Proof. exercise: important to use lecture conventions over homework conventions, especially about factorial placement.

$\square$

Certain things follow immediately from the quotient setup.

- Wedge products are associative (since ring multiplication is associative).

- $\text{Alt}_k(\mathbb{R}^n, \mathbb{R})$ (or $\wedge^k((\mathbb{R}^n)^*)$) is spanned by wedge products of things in $\text{Alt}_1(\mathbb{R}^n, \mathbb{R}) = \wedge^1((\mathbb{R}^n)^*) = (\mathbb{R}^n)^*$. (Why? $T^k((\mathbb{R}^n)^*)$ is spanned by pure k-tensors of ??? of $(\mathbb{R}^n)^*$, and in general, for vector

spaces $W \subset V$, a spanning set for V gives a spanning set for V/W. Thus, $\text{Alt}_k(\mathbb{R}^n, \mathbb{R})$ is spanned by $\{dx_{i_1} \wedge \ldots \wedge dx_{i_k} : 1 \le i_1 \le n, \ldots, 1 \le i_k \le n\}$.

Exercise: can use properties of wedge products to reduce to smaller spanning set: $\{dx_{i_1} \wedge \ldots \wedge dx_{i_n} : 1 \le i_1 < \ldots < i_k \le n\}$. (basic properties like: $dx_i \wedge dx_j = -dx_j \wedge dx_i$, since $(dx_i + dx_j) \wedge (dx_i + dx_j) = 0$. in general: $\alpha \wedge \beta = (-1)^{k\ell} \beta \wedge \alpha$ if $\alpha \in \wedge^k V, \beta \in \wedge^\ell V$. e.g. $dx \wedge dy$, $dx \wedge dz$, $dy \wedge dz$ form a basis for $\text{Alt}_2(\mathbb{R}^3, \mathbb{R})$ as in homework, $dx \wedge dy \wedge dz$ is a basis for $\text{Alt}_3(\mathbb{R}^3, \mathbb{R})$.

5/1 Notes covering pullbacks and 1-forms in general (have so far only done for 1-forms). For pullbacks: use $\text{Alt}_k(\mathbb{R}^n, \mathbb{R})$ instead of $\wedge^k((\mathbb{R}^n)^*)$.

Definition 1.7.53 (Pullbacks). Let $U \subset \mathbb{R}^n$ be open, $V \subset \mathbb{R}^m$ be open, and let $F : U \to V$ be smooth. Let α be a k-form on V ($\alpha : V \to \text{Alt}_k(\mathbb{R}^m, \mathbb{R})$). Define a k-form $F^*\alpha$ on U ($F^*\alpha : U \to \text{Alt}_k(\mathbb{R}^n, \mathbb{R})$) by

$$(F^*(\alpha))_p(\underbrace{v_1}_{\text{vector in } \mathbb{R}^n}, \ldots, v_k) := \alpha_{f(p)}(\underbrace{(DF)_p(v_1)}_{\text{``pushforward vectors'' in } \mathbb{R}^m}, \ldots, (DF)_p(v_k)).$$

One can check that this is multilinear and alternating in the inputs $v_1, \ldots, v_k$. One can also check that is is smooth assuming α is smooth.

Pullbacks and wedge products: we have the formula $F^*(\alpha \wedge \beta) = F^*(\alpha) \wedge F^*(\beta)$ if α is a k-form and β is an ℓ-form (can check using concrete $\frac{1}{k!\ell!}\sum \ldots$ definition for $\alpha \wedge \beta$).

Exterior derivatives and wedge products: we have $F^*(d\alpha) = d(F^*\alpha)$ (can check using chain rule, Theorem 122).

Pullbacks and integrals: below.

Also: wedge products and exterior derivatives: we have the formula $d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^k \alpha \wedge (d\beta)$ if α is a k-form. You can check that the sign shows up; intuition: there is a general pattern where d has "degree 1" in some sense, α has degree k. The general sign rule is that when commuting an object of degree k with an object of degree ℓ, the sign is multiplied by $(-1)^{k\ell}$ ("super sign rule"—see within Supersymmetry and Morse theory).

Now we will discuss integrals of n-forms on n-manifolds M. (We need multiple integrals, which is Pugh [2015] section 5.7. Or, just learning Lebesgue theory envelops what we need to know about multiple integrals.) The idea is to use partitions of unity to reduce to the case of n-forms on $\mathbb{R}^n$ (n-forms are completely supported, so 0 outside a compact set), then use your favorite multi-vcr **???** integration theory in $\mathbb{R}^n$.

As before: if $\boldsymbol{r}$ is a parameterization of a surface S and α is a 2-form on S, then $\int_S \alpha := \int_{\text{domain of } \boldsymbol{r}} \boldsymbol{r}^*\alpha$.

Now we will discuss Riemann integration of compactly supported functions on $\mathbb{R}^n$. The book discusses functions on $[a_1, b_1] \times \ldots \times [a_n, b_n] = Q$ (a hyperprism). Any compactly supported function on $\mathbb{R}^n$ is zero outside some cube Q. If we can define $\int_Q f dx_1 \ldots dx_n$, then we can define $\int_{\mathbb{R}} f dx_1 \ldots dx_n$ by picking any Q such that $f = 0$ outside Q, then setting $\int_{\mathbb{R}^n} f dx_1 \ldots dx_n := \int_Q f dx_1 \ldots dx - n$. (Can check: this is independent of the choice of Q.)

Pugh [2015] discusses Riemann integrability for functions $f : Q \to \mathbb{R}$. There is a mesh definition, a Darboux-style definition, and a limits of nets definition; all three are equivalent. Lebesgue's criterion for Riemann integrability (f is continuous except on a null set) still holds.

New: Fubini's Theorm (save real versions for Lebesgue theory (see Theorem 180 below), where it is properly covered; we will do a very simple version that covers our needs.).

Theorem 162 (**Fubini's Theorem, as stated in 425b**)**.** Let $f : Q \to \mathbb{R}$ be continuous. Then f is Riemann integrable on Q and $\int_Q f dx_1 \ldots dx_n$ (the classical Riemann integral) equals

$$\int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} \ldots \int_{a_n}^{b_n} f dx_n \ldots dx_2 \right) dx_1,$$

the "iterated integral." Further, the order is irrelevant.

Theorem 163 (**Fubini's Theorem**)**.** Let $h : \mathbb{R}^2 \to \mathbb{R}$ be a continuous function such that $\int \int_{\mathbb{R}^2} |h(x,y)| dx dy < \infty$. Then

$$\int \int_{\mathbb{R}^2} h(x,y) dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x,y) dx \right) dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x,y) dy \right) dx$$

Note: the measure theoretic version of Fubini's Theorem requires less than continuity, only integrability.

Finally, the change of variables formula. This is a multivariable analogue of u-substitution.

Theorem 164. Let $\phi : \mathbb{R}^n \to \mathbb{R}^n$ be a C^1 diffeomorphism and let $f : \mathbb{R}^n \to \mathbb{R}$ be a compactly supported Riemann integrable function. Then

$$\int_{\mathbb{R}^n} f dx_1 \ldots dx_n = \int_{\mathbb{R}^n} \sim \overbrace{(f \circ \phi)}^{\text{also compactly supported}} \underbrace{\lfloor \det(J\phi) \rfloor}_{\text{"Jacobian determinant"}} dx_1 \ldots dx_n.$$

Proof. Not easy, takes several pages in Pugh [2015].

$\square$

As above: use this to define $\int_M \omega$, where ω is an n-form on M. (ω: also a common name for differential forms.)

Proposition 165 (**Integrals and pullbacks**)**.** If M, N are smooth n-manifolds and $f : M \to N$ is a diffeomorphism, and ω is an n-form on N, then

$$\int_N \omega = \int_M f^* \omega.$$

Exercise: is this correctly stated? subtle question of orientations; gloss over for now.

Sketch of proof. In chart, $w = g dy_1 \wedge \ldots \wedge dy_n$ (y_i are the coordinates on N). This implies

$$f^* \omega = (g \circ f) d(\underbrace{y_1 \circ f}_{f_1}) \wedge \ldots \wedge d(\underbrace{y_n \circ f}_{f_n})$$

where f_i are the "coordinates" of f. We can write this as

$\square$

Integrals and wedge products: $\int_M (\alpha \wedge \beta)$ vs. $\int_? \alpha \wedge \int_? \beta$: can't say much here based on what we've covered so far.

Theorem 166 (**Stokes' Theorem**)**.** Let ω be a $n-1$-form on M with boundary ∂M. Then $\int_M d\omega = \int_{\partial M} \omega$.

On Monday (final lecture), we will see how this generalizes Green's Theorem, Stokes' Theorem, and the Divergence Theorem.

5/4/20

Definition 1.7.54 (orientation; rough definition)**.** An **orientation** on a smooth n-manifold M is an equivalence class of smooth atlases $\mathcal{A}$ such that $\det J\phi > 0$ for all transition maps ϕ, modulo $A \sim A'$ if their union still satisfies this positive Jacobian determinant property.

The definition of $\int_M \omega$ requires choice of orientation on M; the other choice gives $-\int_M \omega$.

Definition 1.7.55. $f : M \to N$ (smooth) is **orientation-preserving** if $\det(Jf) > 0$ in some (equivalently: any) coordinate charts compatible with the orientation.

Theorem 167. If $f : M \to N$ is an orientation-preserving diffeomorphism and ω is an n-form on N, then

$$\int_N \omega = \int_M f^* \omega.$$

Sketch of proof. same as before (?)

$\square$

For $C \in \mathbb{R}^n$, orientation is encoded by the unit tangent vector field $\boldsymbol{T}$ ($\boldsymbol{r}$ is an "oriented parameterization" if $\boldsymbol{r}'(t)/\|\boldsymbol{r}'(t)\| = \boldsymbol{T}(\boldsymbol{r}(t))$ (as opposed to $-\boldsymbol{T}(\boldsymbol{r}(t))$)).

For $S \in \mathbb{R}^3$ parameterized by $\boldsymbol{r}(u, v)$, orientation on S is encoded by the unit normal vector $\hat{n}$ (vector-valued ($\in \mathbb{R}^3$) function on S). ($\boldsymbol{r}$ is an **oriented parameterization** is

$$\frac{\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}}{\|\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}\|} = \hat{n} \qquad \forall u, v \qquad (\text{not } -\hat{n})$$

Thus, parameterization of $C \subset \mathbb{R}^n$, or $S \subset \mathbb{R}^3$, gives an orientation.

important thing: this theorem about pullbacks and integrals is about orientations preserving integrals.

Next we will proceed to Stokes' and Divergence theorems. We will start with Green's Theorem.

Theorem 168 (**Green's Theorem (somewhat imprecise statement) (stated by Green in 1828, proven by Riemann in 1851)**)**.** Let $D \subset \mathbb{R}^2$ be a "region" bounded by a (say smooth) curve C. D is

given the "usual orientation," and C has an induced orientation[5]—C is parameterized counterclockwise. A classical line integral of the form $\oint_C \underbrace{Pdx + Qdy}_{\text{1-form on } D}$ (if the 1-form is defined on the whole bulk) is equal to

$$\int\int_D \underbrace{\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)}\, dxdy$$

(The underlined 1-form can be extended by zero outside D; get a compactly supported function on $\mathbb{R}^n$ (not continuous; ok if D is "nice."))

Proof. Let $\omega := Pdx + Qdy$, a 1-form on D. Then

$$d\omega = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dxdy,$$

so

$$\oint_C \omega = \text{(by Generalized Stokes' Theorem)} \int_D d\omega = \int\int_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy = \int\int_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dxdy.$$

$\square$

For classical Stokes' and divergence theorems, we need **surface integrals/fluxes** (applied terminology).

Let $S \subset \mathbb{R}^3$ be a surface parameterized by $\boldsymbol{r} : K \to S \subset \mathbb{R}^3$ where K is a compact subset of $\mathbb{R}^2$ ("nice enough," not a Cantor set). (Classically you write $\boldsymbol{r} : K \to \mathbb{R}^3$, which we would call $i \circ \boldsymbol{r}$; we'll sometimes be ambiguous about i, like for curves.)

We want an "area form" (2-form) dA on S, analogous to ds on a curve C.

Definition 1.7.56. The **vector area function** of $(S, \boldsymbol{r})$ is a function $\boldsymbol{A} : K \to \mathbb{R}^3$ defined by

$$\boldsymbol{A}(u, v) := \frac{\partial \boldsymbol{r}}{\partial u}(u, v) \times \frac{\partial \boldsymbol{r}}{\partial v}(u, v).$$

Note that $\boldsymbol{A}(u, v)$ is normal to S and its length is proportional to both $\frac{\partial \boldsymbol{r}}{\partial u}$ and $\frac{\partial \boldsymbol{r}}{\partial v}$.

The **unit normal vector function** $\hat{n} : K \to \mathbb{R}^3$ is defined by

$$\hat{n}(u, v) := \frac{\boldsymbol{A}(u, v)}{\|\boldsymbol{A}(u, v)\|}.$$

Note that $\|\boldsymbol{A}(u, v)\| \neq 0$ since $\frac{\partial \boldsymbol{r}}{\partial u}$ and $\frac{\partial \boldsymbol{r}}{\partial v}$ are linearly independent for all (u, v) (this should be assumed as part of "niceness" for $\boldsymbol{r}$).

---

[5]In general, if M is an oriented smooth manifold with boundary, then you get an induced "boundary orientation" on ∂M, so $\int_M d\omega = \int_{\partial M} \omega$ makes sense.

We'll usually view $\hat{n}$ as a function on S, defined by

$$\hat{n}(\boldsymbol{r}(u,v)) := \frac{\boldsymbol{A}(u,v)}{\|\boldsymbol{A}(u,v)\|}.$$

Proposition 169. There exists a unique 2-form dA on S such that

$$\boldsymbol{r}^*(dA) = \|\boldsymbol{A}\|du \wedge dv = \left\|\frac{\partial \boldsymbol{r}}{\partial u}(u,v) \times \frac{\partial \boldsymbol{r}}{\partial v}(u,v)\right\| du \wedge dv.$$

Proof. Basically true by definition; can define things by what they look like in a parameterization. If not considering different $\boldsymbol{r}$, then don't need to check independence of $\boldsymbol{r}$, etc.

$\square$

Definition 1.7.57 (**Classical**)**.** If f is a function on $S \subset \mathbb{R}^3$, then

$$\int\int_S f dA := \int\int_K f(\boldsymbol{r}(u,v))\left\|\frac{\partial \boldsymbol{r}}{\partial u}(u,v) \times \frac{\partial \boldsymbol{r}}{\partial v}(u,v)\right\| dudv.$$

Proposition 170. The classical definition of $\int\int_S f DA$ is equal to $\overbrace{\int_S}^{\text{typical notation for forms}} \underbrace{f dA}_{\text{2-form}}$.

Proof.

$$\int_S \underbrace{fdA}_{\text{2-form}} := \int_K \boldsymbol{r}^*(fdA) = \int_K f(\boldsymbol{r})\left\|\frac{\partial \boldsymbol{r}}{\partial u}(u,v) \times \frac{\partial \boldsymbol{r}}{\partial v}(u,v)\right\| du \wedge dv = \int\int_K f(\boldsymbol{r}(u,v))\left\|\frac{\partial \boldsymbol{r}}{\partial u}(u,v) \times \frac{\partial \boldsymbol{r}}{\partial v}(u,v)\right\| dudv.$$

$\square$

This is the first kind of surface integral defined in Stewart [2015]. Stewart [2015] Uses these to define surface integrals of vector fields on S.

Definition 1.7.58 (**Classical**)**.** If $\boldsymbol{F}$ is a vector field on S (here and in analogous situations, defined to be a function from S to $\mathbb{R}^3$; not necessarily a tangent vector field to S), then

$$\int\int_S \boldsymbol{F} \cdot d\boldsymbol{A} := \int\int_S (\boldsymbol{F} \cdot \hat{n}) dA$$

(Both sides of the equation are equally common notation, but the right side indicates more clearly how the integral is defined; $f = \boldsymbol{F} \cdot \hat{n}$ is a function on S and we know $\int\int_S f dA$. The left side is the surface integral of $\boldsymbol{F}$ on S, or the flux of $\boldsymbol{F}$ across S.)

This is analogous to $\int_C (\boldsymbol{F} \cdot \boldsymbol{T}) ds$. Why $\hat{n}$ here and not $\boldsymbol{T}$? Why $\boldsymbol{T}$ for curves? Ansewr: curves in $\mathbb{R}^n$ are a 1-form α on C, so a 1-form α on $\mathbb{R}^n$, which can be viewed as a vector field $\boldsymbol{F}$ on $\mathbb{R}^n$, which we can restrict to a vector field $\boldsymbol{F}$ on C. Then $\alpha = (\boldsymbol{F} \cdot \boldsymbol{T}) ds$, so this is the most natural thing to integrate on C.

For surfaces: let α be a 2-form on $S \subset \mathbb{R}^3$. We can extend this to a 2-form α on $\mathbb{R}^3$, which corresponds in the usual way to a vector field $\boldsymbol{F}$ on $\mathbb{R}^3$. Then we can restrict it to S to get a vector field $\boldsymbol{F}$ on S.

Proposition 171. $\alpha = (\boldsymbol{F} \cdot \hat{n})dA.$

Proof. Write

$$\boldsymbol{F} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix};$$

corresponds to $\alpha = F_3 dx \wedge dy - F_2 dx \wedge dz + F_1 dy \wedge dz$ as in the homework. We want to show

$$\boldsymbol{r}^*(\alpha) \tag{1.41}$$

is equal to

$$\boldsymbol{r}^*((\boldsymbol{F} \cdot \hat{n})dA). \tag{1.42}$$

Let

$$\boldsymbol{r}(u,v) := \begin{bmatrix} r_1(u,v) \\ r_2(u,v) \\ r_3(u,v) \end{bmatrix}.$$

Then (1.41) is

$$\boldsymbol{r}^*(F_3 dx \wedge dy - F_2 dx \wedge dz + F_1 dy \wedge dz)$$
$$=(F_3 \cdot \boldsymbol{r})dr_1 \wedge dr_2 - (F_2 \cdot \boldsymbol{r})dr_1 \wedge dr_3 + (F_1 \cdot \boldsymbol{r})dr_2 \wedge dr_2$$
$$=(F_3 \circ \boldsymbol{r})\det(J_{\boldsymbol{r}_{12}}) - \dots$$

Note that

Meanwhile, (1.42) equals

$$\boldsymbol{r}^*((\boldsymbol{F} \cdot \hat{n})dA) \tag{1.43}$$
$$=((\boldsymbol{F} \circ \boldsymbol{r})(\hat{n} \circ \boldsymbol{r}))\|\boldsymbol{A}(u,v)\|du \wedge dv \tag{1.44}$$
$$=((\boldsymbol{F} \circ \boldsymbol{r}) \cdot \frac{\boldsymbol{A}(u,v)}{\|\boldsymbol{A}(u,v)\|})\|\boldsymbol{A}(u,v)\|du \wedge dv \tag{1.45}$$
$$=((\boldsymbol{F}(\boldsymbol{r}(u,v)) \cdot \boldsymbol{A}(u,v))du \wedge dv \tag{1.46}$$

Note that

$$\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v} = \hat{i}\begin{vmatrix} \frac{\partial r_2}{\partial u} & \frac{\partial r_2}{\partial v} \\ \frac{\partial r_3}{\partial v} & \frac{\partial r_3}{\partial v} \end{vmatrix} - \hat{j}\det(J_{\boldsymbol{r}_{13}}) + \hat{k}\det(J_{\boldsymbol{r}_{12}}).$$

Then

$$\boldsymbol{F}(\boldsymbol{r}(u,v)) \cdot \boldsymbol{A}(u,v) = F_1(\boldsymbol{r}*,v)) \det(J_{\boldsymbol{r}_{2,3}}) - F_2(\boldsymbol{r}(u,v)) \det(J_{\boldsymbol{r}_{13}}) + F_1(\boldsymbol{r}(u,v) \det J_{\boldsymbol{r}_{12}}.$$

But that's what we got for (1.41).

$\square$

Corollary 171.1. $\int_S \underbrace{\alpha}_{2-form} = \underbrace{\int \int_S (\boldsymbol{F} \cdot \hat{n}) dA}_{classical}$, where $\boldsymbol{F}$, α correspond as above.

Now we can state

Theorem 172 (Stokes' Theorem). If $S \subset \mathbb{R}^3$ is an oriented surface with boundary C and $\boldsymbol{F}$ is a vector field on S, then

$$\oint_C (\boldsymbol{F} \cdot \boldsymbol{T}) ds = \int \int_S ((\nabla \times \boldsymbol{F}) \cdot \hat{n}) dA.$$

(where $\nabla \times \boldsymbol{F} = \mathrm{curl}(F)$).

Proof. Extend $\boldsymbol{F}$ to $\mathbb{R}^3$; write $F = \begin{bmatrix} F_1 & F_2 & F_3 \end{bmatrix}^\top$ with corresponding 1-form $\alpha := F_1 dx + F_2 dy + F_3 dz$. Then

$$\oint_C (\boldsymbol{F} \cdot \boldsymbol{T}) ds = \int_C \alpha = \text{(by Generalized Stokes' Theorem)} \int_S d\alpha = \text{(by above + homework)} \int \int_S ((\nabla \times \boldsymbol{F}) \cdot \hat{n}) dA.$$

$\square$

Theorem 173 (Gauss Divergence Theorem). If S is the boundary of $\Omega \in \mathbb{R}^3$ and $\boldsymbol{F}$ is a vector field on Ω, then

$$\oiint_S (\boldsymbol{F} \cdot \hat{n}) dA = \iiint_\Omega (\nabla \cdot \boldsymbol{F}) dx dy dz$$

where $\nabla \cdot \boldsymbol{F}) = \mathrm{divergence}(\boldsymbol{F})$.

Proof. Write $\boldsymbol{F} = \begin{bmatrix} F_1 & F_2 & f_3 \end{bmatrix}^\top$; corresponding 2-form on $\mathbb{R}^3$ is $\alpha = F_3 dx \wedge dy - F_2 dx \wedge dz + F_1 dy \wedge dz$. Then

---

$$\oiint_S (\boldsymbol{F} \cdot \hat{n}) dA$$

$$= \int_S \alpha$$

$$= (\text{by Stokes' Theorem}) \int_\Omega d\alpha$$

$$= \text{by Homework} \int_\Omega (\nabla \cdot \boldsymbol{F}) dx \wedge dy \wedge dz$$

$$= \iiint_\Omega (\nabla \cdot \boldsymbol{F}) dx dy dz.$$

(the homework property is that a function corresponding to a 3-form $d\alpha$ is the divergence $\nabla \cdot \boldsymbol{F}$)

$\square$

1.8   Gateaux Derivatives (Section 1.6 of Koroljuk et al. [1994], Section 6.2 of Serfling [1980])

1.9   Preliminaries to Measure Theory (Chapter 0 of Folland [2013])

A primary goal of this course is integration in as general a context as possible. In particular, want to avoid assumptions of continuity, because it is not always a natural assumption.

Example 1.9.1. Define

$$L^2[0,1] = \left\{ f : [0,1] \to \mathbb{R} : \int_0^1 f^2(x) dx < \infty \right\}.$$

Distance in $L^2[0,1]$:

$$d_2(f,g) = \left[\int_0^1 (f(x) - g(x))^2 dx \right]^{1/2}.$$

$f_n \to f$ in L^2 means $d_2(f_n, f) \to 0$. Suppose we restrict to continuous functions:

$$L_C^2[0,1] = \left\{ \text{all continuous functions in } L^2[0,1] \right\}.$$

We may then have sequences $\{f_n\}$ in $L_C^2[0,1]$ which converge

1.10 Measures (Chapter 1 of Folland [2013])

Definition 1.10.1 (**p. 25 of** Folland [2013]). Let $(X, \mathcal{M}, \mu)$ be a measure space. If $\mu(X) < \infty$ (which implies that $\mu(E) < \infty$ for all $E \in \mathcal{M}$), μ is called **finite**. If $X = \cup_{j=1}^{\infty} E_j$ where $E_j \in \mathcal{M}$ and $\mu(E_j) < \infty$ for all j, μ is called σ-**finite**. More generally, if $E = \cup_{j=1}^{\infty} E_j$ where $E_j \in \mathcal{M}$ and $\mu(E_j) < \infty$ for all j, the set E is said to be σ-**finite** for μ. If for each $E \in \mathcal{M}$ with $\mu(E) = \infty$ there exists $F \in \mathcal{M}$ with $F \subset E$ and $0 < \mu(F) < \infty$, μ is called **semifinite**.

Theorem 174 (**Theorem 1.8 in** Folland [2013]). Let $(X, \mathcal{M}, \mu)$ be a measure space.

(a) (**Montonicity.**) If $E, F \in \mathcal{M}$ and $E \subset F$, then $\mu(E) \leq \mu(F)$.

(b) (**Subadditivty.**) If $\{E_j\}_{j=1}^{\infty} \subset \mathcal{M}$, then $\mu\left(\cup_{j=1}^{\infty} E_j\right) \leq \sum_{j=1}^{\infty} \mu(E_j)$.

(c) (**Continuity from below.**) If $\{E_j\}_{j=1}^{\infty} \subset \mathcal{M}$ and $E_1 \subset E_2 \subset \ldots$, then $\mu\left(\cup_{j=1}^{\infty} E_j\right) = \lim_{j \to \infty} \mu(E_j)$.

(d) (**Continuity from above.**) If $\{E_j\}_{j=1}^{\infty} \subset \mathcal{M}$, $\mu(E_1) < \infty$, and $E_1 \supset E_2 \supset \ldots$, then $\mu\left(\cap_{j=1}^{\infty} E_j\right) = \lim_{j \to \infty} \mu(E_j)$.

1.11 Integration (Chapter 2 of Folland [2013])

Definition 1.11.1. Given a measure space $(X, \mathcal{M}, \mu)$,

$$L^+ := \text{ the space of all measurable functions from } X \text{ to } [0, \infty).$$

Definition 1.11.2. Given a measure space $(X, \mathcal{M}, \mu)$, denote by L^1 the space of all integrable functions on X; that is, the set of all functions such that $\int |f| < \infty$.

Theorem 175 (**Monotone Convergence Theorem; Theorem 2.14 in** Folland [2013]). If $\{f_n\}$ is a sequence in L^+ such that $f_j \leq f_{j+1}$ for all j and $f = \lim_{n \to \infty} f_n(= \sup_n f_n)$, then

$$\left(\int \lim_{n \to \infty} f_n = \right) \int f = \lim_{n \to \infty} \int f_n.$$

Theorem 176 (**Fatou's Lemma; Theorem 2.18 in** Folland [2013]). If $\{f_n\}$ is any sequence in L^+, then

$$\int (\liminf f_n) \leq \liminf \int f_n.$$

Theorem 177 (**Dominated Convergence Theorem; Theorem 2.24 in** Folland [2013]). Let $f_n \in L^1$ and $f_n \to f$ almost everywhere. Suppose there exists $g \in L^1$ with $|f_n| \leq g$ for all n. Then $f \in L^1$ and $\int f = \lim_n \int f_n$.

Remark 26. See also Theorem **??** for an application of this theorem to probability.

Remark 27. Let $\{f_n\}$ be measurable functions on $(X, \mathcal{M}, \mu)$, $\overline{\mathcal{M}}$ be the completion of $\mathcal{M}$, and $f_n \to f$ almost everywhere. Then we know $\{x : \{f_n(x)\} \text{ converges}\} \in \mathcal{M}$ by the assumption that each f_n is measurable. However, since the convergence is merely almost everywhere, we only know $\{x : f_n(x) \to f(x)\} \in \overline{\mathcal{M}}$ because we don't know about f on null sets.

Note that $g(x) = \lim_n f_n(x)$ is defined for $x \in E$, and g is measurable from $(X, \mathcal{M})$. $g = f$ almost everywhere, but $\{g \neq f\}$ may only be in $\overline{\mathcal{M}}$.

There is a null set $N \in \mathcal{M}$ with $\{x : f_n(x) \nrightarrow f(x)\} \subset N$. If we change f to 0 on N then still $f_n \to f$ almost everywhere, but now f is measurable for $\mathcal{M}$.

Proof. Change f to 0 on a null set so that f is measurable (per Remark 27). Since $|f_n| \leq g$ for all n, we have $|f| \leq g$, so $f \in L^1$. We can handle the real and imaginary parts separately for complex-valued function, so assume f is real-valued. It is sufficient to show

$$\liminf_n \int f_n \geq \int f \geq \limsup_n \int f_n. \tag{1.47}$$

Note that

$$|f_n| \leq g$$
$$\iff \quad -g \leq f_n \leq g$$
$$\iff \quad g + f_n \geq 0, \qquad g - f_n \geq 0,$$

so $f_n + g \in L^+$ and $g - f_n \in L^+$. Therefore by Fatou's Lemma,

$$\liminf_n \int (f_n + g) \geq \int \liminf_n \{f_n + g\}$$
$$= \int (f + g)$$
$$\iff \quad \liminf_n \int f_n \geq \int f,$$

where we subtracted $\int g$. This shows the left inequality in (1.47) holds. Similarly,

$$-\limsup_n \int (f_n - g) = \liminf_n \int (g - f_n) \geq \int \liminf_n \{g - f_n\}$$
$$= \int (g - f)$$
$$\iff \quad \liminf_n \int (-f_n) \geq \int (-f)$$
$$\iff \quad \limsup_n \int f_n \leq \int f,$$

which proves the right inequality in (1.47).

$\square$

Consider the point mass measure

$$\delta_a(E) = \begin{cases} 1, & a \in E, \\ 0, & \text{otherwise.} \end{cases}$$

What is $\int f d\delta_a$? First consider a simple $\phi = \sum_{j=1}^n c_j \chi_{E_j}$:

$$\int \phi \, d\delta_a = \sum_{i=1}^n c_j \delta_a(E_j) = \sum_{i=1}^n \sum_{j:a \in E_j} c_j = \phi(a).$$

For $f \geq 0$, taking limits of ϕ_n up to f shows

$$\int f \, d\delta_a = f(a).$$

For a general $f = f^+ - f^-$ we can show the same thing. The counting measure is defined by

$$\mu(E) = \sum_{x \in E} \mathbb{1}(x \in \mathbb{N}) = \sum_{n=1}^{\infty} \delta_n(E) = |\mathbb{N} \cap E|,$$

where $\mathbb{1}(\cdot) = 1$ if $\cdot$ holds and 0 otherwise. For simplicity, we can take $\mathcal{M} = \mathcal{P}(\mathbb{N})$.

Proposition 178. For $\phi : \mathbb{N} \to \mathbb{R}$ a simple sequence with only finitely many term values (that is, the set of numbers that appears in the sequence is finite),

$$\int \phi \, d\mu = \int \phi \, d\left(\sum_{n=1}^{\infty} \delta_n\right) = \sum_{n=1}^{\infty} \int \phi \, d\delta_n$$

where μ is the counting measure

Proof. Take $\mathcal{M} = \mathcal{P}(\mathbb{N})$. Since $\phi = \sum_{j=1}^{n} c_j \chi_{E_j}$ (and recall that $E_j \in \mathcal{P}(\mathbb{N})$),

$$\int \phi \, d\mu = \sum_{j=1}^{n} c_j \mu(E_j)$$
$$= \sum_{j=1}^{n} c_j \sum_{n'=1}^{\infty} \delta_{n'}(E_j)$$
$$= \sum_{j=1}^{n} c_j |E_j|$$
$$\stackrel{(*)}{=} \sum_{n=1}^{\infty} \sum_{j:n \in E_j} c_j$$
$$= \sum_{n=1}^{\infty} \sum_{j=1}^{n} c_j \delta_n(E_j)$$
$$= \sum_{n=1}^{\infty} \int \phi \, d\delta_n$$

where $(*)$ follows because c_j appears $|E_j|$ times in the sum on the right.

$\square$

Proposition 179. Let $u : \mathbb{N} \to \mathbb{R}_+$ be a positive sequence. Then

$$\int u \, d\mu = \sum_{n=1}^{\infty} u(n)$$

for the measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$, where μ is the counting measure.

Proof. Express μ as

$$\mu = \sum_{j=1}^{\infty} \delta_j.$$

Define

$$u_n(k) = \begin{cases} u(k), & k \le n, \\ 0, & k > n \end{cases} = \sum_{k=1}^{n} u(k)\chi_{\{k\}}.$$

Note that $u_n(k)$ is a simple function for all $n \in \mathbb{N}$. Therefore

$$\int u_n \ d\mu = \sum_{k=1}^{n} u(k)\mu(\{k\}) = \sum_{k=1}^{n} u(k).$$

Since $0 \le u_n \le u$ and $\lim_{n \to \infty} u_n = u$, by the Monotone Convergence Theorem we have

$$\int_{\mathbb{N}} u \ d\mu = \int_{\mathbb{N}} \lim_{n \to \infty} u_n \ d\mu = \lim_{n \to \infty} \int_{\mathbb{N}} u_n \ d\mu = \lim_{n \to \infty} \sum_{k=1}^{n} u(k) = \sum_{k=1}^{\infty} u(k).$$

$\square$

Corollary 179.1. Let $u_1, \dots, u_n$ be a finite positive sequence. Then

$$\int u \ d\mu = \sum_{i=1}^{n} u_i$$

for the measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$, where μ is the counting measure.

Proof. Apply Proposition 179 to the sequence

$$u(i) = \begin{cases} u_i, & i \le n, \\ 0, & i > n. \end{cases}$$

$\square$

Corollary 179.2. Let $u : \mathbb{N} \to \mathbb{R}$ be a sequence. Suppose

$$\sum_{n=1}^{\infty} |u(n)| < \infty.$$

Then

$$\int u \ d\mu = \sum_{n=1}^{\infty} u(n)$$

for the measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$, where μ is the counting measure.

Proof. We have

$$\int u \ d\mu = \int u^+ \ d\mu - \int u^- \ d\mu.$$

By Proposition 179,

$$\int u^+ \ d\mu = \sum_{n=1}^{\infty} u^+(n), \qquad \int u^- \ d\mu = \sum_{n=1}^{\infty} u^-(n).$$

By assumption,

$$\sum_{n=1}^{\infty} |u(n)| < \infty$$

$$\Longleftrightarrow \quad \sum_{n=1}^{\infty} (u^+(n) + u^-(n)) < \infty$$

$$\Longleftrightarrow \quad \sum_{n=1}^{\infty} u^+(n) + \sum_{n=1}^{\infty} u^-(n) < \infty$$

$$\Longleftrightarrow \quad \sum_{n=1}^{\infty} u^+(n) < \infty, \qquad \sum_{n=1}^{\infty} u^-(n) < \infty$$

since these are both nonnegative sequences. Therefore

$$\int u \, d\mu = \int u^+ \, d\mu - \int u^- \, d\mu$$

$$= \sum_{n=1}^{\infty} u^+(n) - \sum_{n=1}^{\infty} u^-(n)$$

$$= \sum_{n=1}^{\infty} \left(u^+(n) - u^-(n) \right)$$

$$= \sum_{n=1}^{\infty} u(n).$$

$\square$

Definition 1.11.3 ((p. 65 of Folland [2013])). Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be measure spaces. If $E \subset X \times Y$, for $x \in X$ and $y \in Y$ we define the **x-section** E_x and the **y-section** E^y of E by

$$E_x := \{y \in Y : (x, y) \in E\}, \qquad E^y := \{x \in X : (x, y) \in E\}.$$

Also, if f is a function on $X \times Y$ we define the **x-section** f_x and the **y-section** f^y of f by

$$f_x(y) = f^y(x) = f(x, y).$$

Thus, for example, $(\chi_E)_x = \chi_{E_x}$ and $(\chi_E)^y = \chi_{E^y}$.

Theorem 180 (**The Fubini-Tonelli Theorem; Theorem 2.37 in** Folland [2013]). Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces.

(a) (Tonelli.) If $f \in L^+(X \times Y)$, then the functions $g(x) = \int f_x \, d\nu$ and $h(y) = \int f^y \, d\mu$ are in $L^+(X)$ and $L^+(Y)$ respectively, and

$$\int f \, d(\mu \times \nu) = \int \left[\int f(x, y) \, d\nu(y) \right] d\mu(x) = \int \left[\int f(x, y) \, d\mu(x) \right] d\nu(y). \qquad (1.48)$$

(b) (Fubini.) If $f \in L^1(\mu \times \nu)$, then $f_x \in L^1(\nu)$ for almost every $x \in X$, $f^y \in L^1(\mu)$ for almost every $y \in Y$, the almost everywhere-defined functions $g(x) = \int f_x \, d\nu$ and $h(y) = \int f^y \, d\mu$ are in $L^1(\mu)$ and $L^1(\nu)$, respectively, and (1.48) holds.

Remark 28. Recall Theorem 162 gave a simpler, non-measure-theoretic version of Fubini's Theorem.

1.12 Signed Measures and Differentiation (Chapter 3 of Folland [2013])

Theorem 181 (**The Lebesgue-Radon-Nikodym Theorem; Theorem 3.8 in** Folland [2013]). Let ν be a σ-finite signed measure and μ a σ-finite positive measure on $(X, \mathcal{M})$. There exist unique σ-finite signed measures λ, ρ on $(X, \mathcal{M})$ such that

$$\lambda \perp \mu, \quad \rho \ll \mu, \quad \text{and} \quad \nu = \lambda + \rho.$$

Moreover, there is an extended μ-integrable function $f : X \to \mathbb{R}$ such that $d\rho = f\,d\mu$ (that is, $\rho(E) = \int_E f\,d\mu$), and any two such functions are equal μ-almost everywhere.

Definition 1.12.1. The decomposition $\nu = \lambda + \rho$ where $\lambda \perp \mu$ and $\rho \ll \mu$ is called the **Lebesgue decomposition** of ν with respect to μ. In the case where $\nu \ll \mu$, Theorem 181 says that $d\nu = f\,d\mu$ for some f. This result is usually known as the **Radon-Nikodym Theorem**, and f is called the **Radon-Nikodym derivative** of ν with respect to μ. We denote it by $d\nu/d\mu$:

$$d\nu = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}d\mu \qquad \left(\iff \nu(E) = \int_E \frac{\mathrm{d}\nu}{\mathrm{d}\mu}d\mu \right).$$

(Strictly speaking, $d\nu/d\mu$ should be construed as the class of functions equal to f μ-almost everywhere.)

Theorem 182 (**Theorem 3.22 in** Folland [2013]). Let ν be a regular signed or complex Borel measure on $\mathbb{R}^n$, and let $d\nu = d\lambda + f\,dm$ be its Lebesgue-Radon-Nikodym representation. Then for m-almost every $x \in \mathbb{R}^n$,

$$\lim_{r \to 0} \frac{\nu(E_r)}{m(E_r)} = f(x)$$

for every family $\{E_r\}_{r>0}$ that shrinks nicely to x.

1.13 Elements of Functional Analysis (Chapter 5 of Folland [2013])

1.14 Problems from Practice Math GRE Subject Tests

38. Let A and B be nonempty subsets of $\mathbb{R}$ and let $f : A \to B$ be a function. If $C \subseteq A$ and $D \subseteq B$, which of the following must be true?

(A) $C \subseteq f^{-1}\Big(f(C)\Big)$

(B) $D \subseteq f\Big(f^{-1}(D)\Big)$

(C) $f^{-1}\Big(f(C)\Big) \subseteq C$

Solution 38. (A) Neither of the equalities should hold – these are in fact nonsense statements, as one side lies in A and the other in B. To unravel the remaining two sets,

$$f^{-1}(f(C)) = \{x \in A : f(x) \in f(C)\}, \quad f(f^{-1}(D)) = f(\{y \in A : f(y) \in D\})$$

Clearly the second set must always be contained in D, but not the other way around. Similarly the first set certainly contains all $c \in C$ (as $f(c) \in f(C)$) but not the other way around.

47. The function $f : \mathbb{R} \to \mathbb{R}$ is defined as follows.

$$f(x) = \begin{cases} 3x^2 & \text{if } x \in \mathbb{Q} \\ -5x^2 & \text{if } x \notin \mathbb{Q} \end{cases}$$

Which of the following is true?

(A) f is discontinuous at all $x \in \mathbb{R}$.

(B) f is continuous only at $x = 0$ and differentiable only at $x = 0$.

(C) f is continuous only at $x = 0$ and nondifferentiable at all $x \in \mathbb{R}$.

(D) f is continuous at all $x \in \mathbb{Q}$ and nondifferentiable at all $x \in \mathbb{R}$.

(E) f is continuous at all $x \notin \mathbb{Q}$ and nondifferentiable at all $x \in \mathbb{R}$.

Solution 47. (B) A classic kind of problem. We are clearly continuous and differentiable at 0. Anywhere else, near a rational number there is an irrational number and vice versa. Therefore there can be no continuity anywhere but at 0, and hence no differentiability either.

57. For each positive integer n, let x_n be a real number in the open interval $\left(0, \dfrac{1}{n}\right)$. Which of the following statements must be true?

I. $\lim\limits_{n \to \infty} x_n = 0$

II. If f is a continuous real-valued function defined on $(0, 1)$, then $\{f(x_n)\}_{n=1}^{\infty}$ is a Cauchy sequence.

III. If g is a uniformly continuous real-valued function defined on $(0, 1)$, then $\lim\limits_{n \to \infty} g(x_n)$ exists.

(A) I only   (B) I and II only   (C) I and III only   (D) II and III only   (E) I, II, and III

Solution 57. (C) I is true, since $\lim_{n \to \infty} x_n$ must be bounded between 0 and $\lim_{n \to \infty} 1/n = 0$. Unfortunately, x_n does not converge inside $(0, 1)$. There is no reason therefore that $f(x_n)$ should be a convergent sequence – suppose that $f(x) = 1/x$, so that $f(x_n)$ is certainly not Cauchy. However, if g is uniformly continuous, then g extends to a continuous function on $[0, 1]$. Now x_n is a convergent sequence, so $\lim_{n \to \infty} g(x_n) = g(\lim_{n \to \infty} x_n) = g(0)$ exists.

60. A real-valued function f defined on $\mathbb{R}$ has the following property.

> For every positive number ϵ, there exists a positive number δ such that
> $$|f(x) - f(1)| \geq \epsilon \text{ whenever } |x - 1| \geq \delta.$$

This property is <u>equivalent</u> to which of the following statements about f?

(A) f is continuous at $x = 1$.

(B) f is discontinuous at $x = 1$.

(C) f is unbounded.

(D) $\displaystyle\lim_{|x| \to \infty} |f(x)| = \infty$

(E) $\displaystyle\int_0^\infty |f(x)| dx = \infty$

Solution 60. (D) While it looks like this is the opposite of continuity, that should read 'there exists $\varepsilon > 0$'. What the statement says is that we not only get arbitrarily far away from $f(1)$, but we must for all x sufficiently far away from 1. So as $|x|$ gets very large, so does $|f(x)|$.

63. For any nonempty sets A and B of real numbers, let $A \cdot B$ be the set defined by

$$A \cdot B = \{xy : x \in A \text{ and } y \in B\}.$$

If A and B are nonempty bounded sets of real numbers and if $\sup(A) > \sup(B)$, then $\sup(A \cdot B) =$

(A) $\sup(A) \sup(B)$

(B) $\sup(A) \inf(B)$

(C) $\max\{\sup(A) \sup(B), \ \inf(A) \inf(B)\}$

(D) $\max\{\sup(A) \sup(B), \ \sup(A) \inf(B)\}$

(E) $\max\{\sup(A) \sup(B), \ \inf(A) \sup(B), \ \inf(A) \inf(B)\}$

Solution 63. (E) The supremum is either going to be the product of the two largest positive numbers in A and B or the product of the two smallest negative numbers in A and B. That means we should look for $\sup \cdot \sup$ or $\inf \cdot \inf$. However, it might be the case that B contains only negative numbers and A contains only positive numbers. Then the largest value in $A \cdot B$ will be attained by the smallest positive element of A and the largest negative element of B, giving us our third option: $\inf A \cdot \sup B$.

Bibliography

G. Folland. *Real Analysis: Modern Techniques and Their Applications.* Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, 2013. ISBN 9781118626399. URL https://books.google.com/books?id=wI4fAwAAQBAJ.

V. Koroljuk, V. Korolyuk, I. Borovskikh, Y. Borovskich, I. Borovskikh, I. Borovskikh, and J. Borovskich. *Theory of U-Statistics.* Mathematics and Its Applications. Springer Netherlands, 1994. ISBN 9780792326083. URL https://books.google.com/books?id=bp2DgkDxMdMC.

S. Lang. *Algebra.* Graduate Texts in Mathematics. Springer New York, 2005. ISBN 9780387953854. URL https://books.google.com/books?id=Fge-BwqhqIYC.

J. Lee. *Introduction to Smooth Manifolds.* Graduate Texts in Mathematics. Springer New York, 2012. ISBN 9781441999825. URL https://books.google.com/books?id=xygVcKGPsNwC.

C. Pugh. *Real Mathematical Analysis.* Undergraduate Texts in Mathematics. Springer International Publishing, 2015. ISBN 9783319177717. URL https://books.google.com/books?id=2NVJCgAAQBAJ.

W. Rudin. *Principles of Mathematical Analysis.* International series in pure and applied mathematics. McGraw-Hill, 1976. ISBN 9780070856134. URL https://books.google.com/books?id=kwqzPAAACAAJ.

R. Serfling. *Approximation theorems of mathematical statistics.* Wiley series in probability and mathematical statistics : Probability and mathematical statistics. Wiley, New York, NY [u.a.], [nachdr.] edition, 1980. ISBN 0471024031. URL http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+024353353&sourceid=fbw_bibsonomy.

M. Spivak. *Calculus On Manifolds: A Modern Approach To Classical Theorems Of Advanced Calculus.* Mathematics monograph series. Avalon Publishing, 1971. ISBN 9780813346120. URL https://books.google.com/books?id=POIJJJcCyUkC.

J. Stewart. *Calculus.* Cengage Learning, 2015. ISBN 9781305480513. URL https://books.google.com/books?id=spiaBAAAQBAJ.

T. Tao. *An Epsilon of Room, I: Real Analysis.* An Epsilon of Room. American Mathematical Society, 2010. ISBN 9780821852781. URL https://books.google.com/books?id=royDAwAAQBAJ.