MATH 208 Lecture Notes

Chenjia Lin

December 12, 2021

Course Introduction

The goal of this course is to develop the foundations for understanding integration on manifolds. Roughly speaking, a manifold is a space that locally resembles $\mathbb{R}^n$, and simple examples include the surface of spheres and torii. Before we study integration, we will cover differentiation on manifolds and make sense of what dx means while studying differential forms.

Contents

1 Review of Topology

In this lecture, we review definitions and propositions regarding metric spaces and point-set topology.

Definition 1.1. Let X be a set. A *topology* $\mathcal{T}$ on X is a family of subsets of X such that

1. $\varnothing$, X belong to $\mathcal{T}$,
2. the union of elements in any subcollection of $\mathcal{T}$ belongs in $\mathcal{T}$,
3. and the intersection of elements in any finite subcollection of $\mathcal{T}$ belongs in $\mathcal{T}$.

The elements of $\mathcal{T}$ are known as the *open subsets of* X, and we say X is a topological space.

Example 1.2. If (X, ρ) is a metric space, then the *metric topology* is the set of all open sets in X, where a set $U \subseteq X$ is open if for all $a \in U$, there exists $\varepsilon > 0$ such that $B(a, \varepsilon) \subseteq U$.

Definition 1.3. A function $f : X \to Y$ between topological spaces is *continuous* if for all open $V \subseteq Y$, its preimage $f^{-1}(V) \subseteq X$ is open as well.

Example 1.4. In a metric space (X, ρ), continuity can also be formulated using the ε-δ definition. The preimage and ε-δ turn out to be equivalent.

Definition 1.5. A subset $C \subseteq X$ is *closed* if $X \setminus C$ is open.

Definition 1.6. A sequence $(x_n) \subseteq X$ *converges* to $x \in X$ if for all open sets U containing x, there exists $N \in \mathbb{N}$ such that $x_n \in U$ for all $n \geq N$.

Remark. Note that a sequence does not always converge to a single point! For instance, consider the *indiscrete topology* on $X := \{a, b\}$, where $\mathcal{T} = \{X, \varnothing\}$. Note that X is an open set containing a and b, and the sequence $\{a, b, a, b, \dots\}$ converges to both a and b. If we were working with a metric topology, then limits are necessarily unique.

Definition 1.7. A point $x \in X$ is a *limit point* of a subset $A \subseteq X$ if $U \cap (A \setminus \{x\}) \neq \varnothing$ for all open sets U containing x.

Proposition 1.8. A set $C \subseteq X$ is closed if and only if C contains all of its limit points.

Definition 1.9. Let $A \subseteq X$. We define

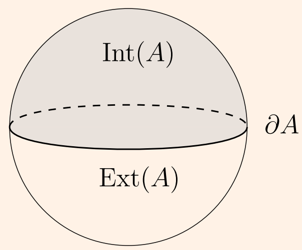
1. the *closure* $\overline{A}$ of A to be the smallest closed set containing A,
2. the *interior* $\operatorname{int}(A)$ of A the largest open set in A,
3. the *exterior* $\operatorname{ext}(A)$ of A the complement of $\overline{A}$,
4. and the *boundary* ∂A of A the set $\overline{A} \setminus \operatorname{int}(A)$.

Remark. Note that because $\overline{A}$ is closed, $\text{ext}(A) = X \setminus \overline{A}$ must be open. The boundary ∂A is equivalently the intersection $\overline{A} \cap \text{ext}(A) = \overline{A} \cap (X \setminus \overline{A})$.

Proposition 1.10. For any subset $A \subseteq X$, we have

$$X = \text{int}(A) \sqcup \partial A \sqcup \text{ext}(A).$$

Example 1.11. Consider $X = \mathbb{S}^n$ and the set A to be the upper hemisphere (including the equator). We have that the interior of A is the upper hemisphere (excluding the equator), the exterior the bottom hemisphere (excluding equator), and the boundary the equator.



2 Review of Topology (continued)

For this lecture, assume X is a topological space. We will review notions of compactness and connectedness from topology.

Definition 2.1. An *open cover* of $A \subseteq X$ is a collection $\mathcal{C} := \{U_\lambda\}_{\lambda \in \Lambda}$ such that

$$A \subseteq \bigcup_{\lambda \in \Lambda} U_\lambda.$$

Definition 2.2. The *subspace topology* on $A \subseteq X$ is the set

$$\mathcal{T}_A := \{U \cap A \mid U \subseteq X \text{ is open}\}.$$

We can verify that the subspace topology on $A \subseteq X$ does indeed satisfy the axioms defining a topology. Note that if $\mathcal{C} = \{U_\lambda\}_{\lambda \in \Lambda}$ is an open cover of A (in terms of $\mathcal{T}$), then $\mathcal{C}' := \{U_\lambda \cap A\}_{\lambda \in \Lambda}$ is an open cover of A (in terms of $\mathcal{T}_A$).

Example 2.3. Consider the subspace topology on $[0, 1] \subseteq \mathbb{R}$. Note that $[0, 1/2)$ is open in $[0, 1]$ because $[0, 1/2)$ is the intersection of $[0, 1]$ and the open interval $(-1/2, 1/2)$. However, $[0, 1/2)$ is clearly not open in $\mathbb{R}$.

Definition 2.4. Let X be a topological space. We say X is

1. *compact* if every open cover of X admits a finite subcover,
2. *sequentially compact* if every sequence in X has a finite subcover.

In topology, these two notions are different. However, if the topology on X is induced by a metric, then the two notions are equivalent.

Theorem 2.5. Let (X,ρ) be a metric space. Then the following are equivalent:

1. X is compact.
2. X is sequentially compact.
3. X is complete and totally bounded.

Theorem 2.6 (Heine-Borel)**.** A subset $K \subseteq \mathbb{R}^n$ is compact if and only if it is closed and bounded

Theorem 2.7 (Bolzano-Weierstrass)**.** Any bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Note that it follows from Definition 2.4 that a subset $A \subseteq X$ is compact if every open cover of A admits a finite subcover and sequentially compact if every sequence in A converges in A. Recall that sequences in topological spaces can have multiple limits, and for sequential compactness, it suffices for a sequence to have at least one limit point in A.

The following is something we did not touch on last quarter:

Definition 2.8. A topological space X is *connected* if it cannot be expressed as a union of two disjoint, non-empty open subsets.

Example 2.9.

1. The set $A = (0,1) \cup (2,3)$ is not connected because $(0,1)$ and $(2,3)$ are open and disjoint.
2. The set of rationals $\mathbb{Q} \subseteq \mathbb{R}$ is not connected because $\mathbb{Q} = ((-\infty, \sqrt{2}) \cap \mathbb{Q}) \cup ((\sqrt{2}), \infty) \cap \mathbb{Q})$.

Theorem 2.10. A subset $A \subseteq \mathbb{R}$ is connected if and only if A is an interval.

Proof. First, suppose A is not an interval. Then there exists $a, b \in A$ and $x \in \mathbb{R} \setminus A$ such that $a < x < b$. Notice that $A = ((-\infty, x) \cap A) \cup ((x, \infty) \cap A)$, which means A is not connected. Thus, if A is connected, then A must be an interval.

Now suppose A is an interval. We proceed by contradiction, so suppose A is not connected. As such, there exist disjoint, nonempty open subsets $A \cap U$ and $A \cap V$ such that $(A \cap U) \cup (A \cap V) = A$. Since $A \cap U$ and $A \cap V$ are disjoint, let $a \in A \cap U$ and $b \in A \cap V$. Without loss of generality, let $a < b$. Now let $S := \{x \in \mathbb{R} \mid a < x < b \text{ and } (a, x] \subseteq A \cap U\}$. Note that S is nonempty; if U is open, then there exists $\varepsilon > 0$ so that $(a - \varepsilon, a + \varepsilon) \subseteq U$ and $a + \varepsilon < b$. Since A is an interval, we know $a + \varepsilon \in A$ and $(a, a + \varepsilon) \subseteq A \cap U$, as well as $a + \varepsilon/2 \in S$. Note that S is bounded above by b, so by the least-upper-bound property of $\mathbb{R}$, we know $c := \sup(S)$ exists.

We now show that c belongs to neither $A \cap U$ nor $A \cap V$. So suppose $c \in A \cap U$. Because U is open and A is an interval, there exists $\varepsilon > 0$ such that $c + \varepsilon < b$ and $(c, c + \varepsilon) \subseteq U$. Notice that $(a, c + \varepsilon/2] \in A \cap U$, which contradicts our definition of c as a supremum. Similarly, if $c \in A \cap V$, then there exists $\varepsilon > 0$ such that $(c - \varepsilon, c) \subseteq V$. By definition of c, we know $(a, c - \varepsilon/2] \subseteq A \cap U$, but this contradicts our assumption that U and V are disjoint. And so, we have struck a contradiction. □

Proposition 2.11. Let $f : X \to Y$ be continuous and X connected. Then $f(X)$ is connected.

Proof. Suppose instead that $f(X)$ is not connected. Then there exist disjoint, nonempty open sets U and V in $f(X)$ such that $f(X) = U \cup V$. By continuity of f, we know $f^{-1}(U \cup V) = f^{-1}(U) \cup f^{-1}(V)$ must be open. Moreover, note that $X = f^{-1}(U) \cup f^{-1}(V)$, which implies X is not connected. □

Corollary 2.12. Let $f : [a, b] \to \mathbb{R}$ be continuous. If $f(a) < x < f(b)$, then there exists $c \in [a, b]$ so that $f(c) = x$.

Proof. Since $[a, b]$ is connected, its image $f([a, b])$ is also connected. By the previous proposition, we know $f([a, b])$ is an interval and $x \in f([a, b])$. Hence, there exists $c \in [a, b]$ so that $f(c) = x$. □

3 Introduction to Differentiation in $\mathbb{R}^n$

Last time, we defined connectedness for topological spaces and showed that sets in $\mathbb{R}$ are connected if and only if they are intervals. We now define a simpler notion of path-connectedness.

Definition 3.1. A topological space X is *path-connected* if for all $x, y \in X$, there exists a continuous function $\gamma : [0, 1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = y$.

In the homework, we will show that path-connectedness always implies connectedness. The converse is not always true, but for a connected open subset A of $\mathbb{R}^n$, it is true that A is also path-connected.

We will now move on to differentiation in $\mathbb{R}^n$. Our general setup will consist of an open subset $U \subseteq \mathbb{R}^n$ and a function $f : U \to \mathbb{R}^m$. In single-variable calculus, the derivative of f at $a \in U$ is the "best linear approximation" to f at a. We would find that the derivative is a scalar—the slope of the line tangent to a point on the graph of f—but in multivariable calculus, the derivative is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Definition 3.2. Let $U \subseteq \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^m$ a function. We say $f : U \to \mathbb{R}^m$ is *differentiable* at $a \in U$ if there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ so that

$$\lim_{u \to 0} \frac{f(a + u) - f(a) - Au}{\|u\|} = 0.$$

for all $u \in \mathbb{R}^n$. If such a map A exists, then A is known as the *total derivative* of f at a, and we denote $Df(a) := A$.

Note that the limit $u \to 0$ is the same as $\|u\| \to 0$. And as always, it is wise to check that mathematical objects are unique:

Proposition 3.3. If A and B are total derivatives of $f : U \to \mathbb{R}^m$, then $A = B$.

Proof. As given, we have

$$\lim_{u \to 0} \frac{f(a + u) - f(a) - Au}{\|u\|} = 0,$$
$$\lim_{u \to 0} \frac{f(a + u) - f(a) - Bu}{\|u\|} = 0,$$

which imply

$$\lim_{u \to 0} \frac{f(a + u) - f(a) - Au}{\|u\|} - \lim_{u \to 0} \frac{f(a + u) - f(a) - Bu}{\|u\|} = \lim_{u \to 0} \frac{Bu - Au}{\|u\|} = 0.$$

Note that there exists $v \in \mathbb{R}^n$ so that $\|v\| = 1$ and $tv = u$ for some $t > 0$. And so, we have

$$\lim_{u \to 0} \frac{Bu - Au}{\|u\|} = \lim_{u \to 0} \frac{B(tv) - A(tv)}{c\|v\|} = \lim_{u \to 0} B(tv) - A(tv) = 0,$$

which implies $A(v) = B(v)$ and $A(u) = B(u)$. □

Unlike single-variable calculus, our definition of the multivarialbe derivative does not tell us what the linear map looks like.

Example 3.4. Let $A \in M_{m\times n}$ and consider $f : \mathbb{R}^n \to \mathbb{R}^m$ where $f(x) = Ax$. For $a \in \mathbb{R}^n$, note that $Df = f$, as

$$\lim_{u\to 0} \frac{f(a+u) - f(a) - Au}{\|u\|} = \lim_{u\to 0} \frac{f(a+u) - f(a) - f(u)}{\|u\|} = 0$$

for all $u \in \mathbb{R}^n$.

Theorem 3.5. If $f : U \to \mathbb{R}^m$ is differentiable at $a \in U$, then f is continuous at a.

Proof. If f is differentiable at $a \in U$, then there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{u\to 0} \frac{f(a+u) - f(a) - Au}{\|u\|} = 0.$$

for all $u \in \mathbb{R}^n$. Notice that

$$f(a+u) = f(a) + Au + \|u\| \left(\frac{f(a+u) - f(a) - Au}{\|u\|} \right),$$

which implies

$$\begin{aligned}\lim_{u\to 0} f(a+u) &= \lim_{u\to 0} \left(f(a) + Au + \|u\| \left(\frac{f(a+u) - f(a) - Au}{\|u\|} \right) \right),\\ &= f(a) + 0 + 0 \cdot 0,\\ &= f(a).\end{aligned}$$

And so, f is continuous at a. □

Note that $f : U \to \mathbb{R}^m$ is given by an m-tuple of functions $f = (f_1, f_2, \dots, f_m)$, where $f_i : U \to \mathbb{R}$.

Proposition 3.6. A function $f : U \to \mathbb{R}^m$ is differentiable at $a \in U$ if and only if $f_i : U \to \mathbb{R}$ is differentiable at a for $1 \le i \le m$. Moreover, we have

$$Df(a) = (Df_1(a), Df_2(a), \dots, Df_m(a)).$$

Proof. Note that

$$\lim_{u\to 0} \frac{f(a+u) - f(a) - Au}{\|u\|} = 0$$

is true if and only if

$$\lim_{u\to 0} \frac{\pi_i \circ f(a+u) - \pi_i \circ f(a) - \pi_i(Au)}{\|u\|} = 0,$$

where $\pi_i : \mathbb{R}^m \to \mathbb{R}$ for $1 \le i \le m$ returns the ith component of a point in $\mathbb{R}^m$. Furthermore, the assertion above is true if and only if $\pi_i \circ f = f_i$ is differentiable, with derivative $\pi_i(Au)$. Observe that

$$Au = (\pi_1(Au), \pi_2(Au), \dots, \pi_m(Au)),$$

so if $A = Df(a)$ and $\pi_i \circ A = Df_i(a)$, then we see that

$$Df(a) = (Df_1(a), Df_2(a), \dots, Df_m(a)).$$ □

Remark. To clarify, $Df(a)$ consists of the m linear maps $Df_i(a)$ for $1 \leq i \leq m$ and it acts on $u \in \mathbb{R}^n$ in the following way

$$Df(a)u = (Df_1(a)u, Df_2(a)u, \ldots, Df_m(a)u)$$

to produce a vector in $\mathbb{R}^m$.

Definition 3.7. The *partial derivative* of $f : U \to \mathbb{R}^m$ with respect to x_i at $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ is defined as

$$\frac{\partial f}{\partial x_i}(a) := \lim_{h \to 0} \frac{f(a_1, \ldots, a_{i-1}, a_i + h, a_{i+1}, \ldots, a_n) - f(a)}{h},$$

given that the limit exists. In this context, $h \in \mathbb{R}$.

We'd like to know when a multivariable function is differentiable, and this leads to two questions:

1. If all partial derivatives of a function exist, then does the total derivative also exist?
2. Likewise, would the function be continuous?

Unfortunately, the answer to both questions is a "no."

Example 3.8. Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) = \begin{cases} \dfrac{x^2 y}{x^4 + y^2} & \text{if } (x,y) \neq 0, \\ 0 & \text{if } (x,y) = 0. \end{cases}$$

All partial derivatives of f exist, but f is not continuous at $(0,0)$ and does not have a total derivative.

Definition 3.9. Let $v \in \mathbb{R}^n$. The *directional derivative* of f at a in the direction $v \in \mathbb{R}^n$ is given by

$$\frac{\partial f}{\partial v}(a) = \lim_{h \to 0} \frac{f(a + hv) - f(a)}{h},$$

where $h \in \mathbb{R}$.

Note that the function in our counterexample has all directional derivatives as well. Thus, having directional derivatives in all directions is not enough to guarantee the existence of the total derivative. Fortunately, we will see that if all the partial derivatives are *continuous*, then total derivative exists. Conversely, if the total derivatves exist, then all partial derivatives are continuous!

Proposition 3.10. If $f : U \to \mathbb{R}^m$ is differentiable at $a \in U$, then the directional derivative of f at a exist for all directions $v \in \mathbb{R}^n$ and

$$\frac{\partial f}{\partial v}(a) = Df(a)(v).$$

Proof. Let $u := hv$ for $h \in \mathbb{R}$, and by the definition of the total derivative, we have

$$\begin{aligned} 0 = \lim_{u \to 0} \frac{f(a+u) - f(a) - Df(a)(u)}{\|u\|} &= \lim_{h \to 0} \frac{f(a+hv) - f(a) - Df(a)(hv)}{h}, \\ &= \lim_{h \to 0} \frac{f(a+hv) - f(a) - h \cdot Df(a)(v)}{h}, \\ &= \lim_{h \to 0} \frac{f(a+hv) - f(a)}{h} - Df(a)(v), \end{aligned}$$

which implies $\dfrac{\partial f}{\partial v}(a) = Df(a)(v)$. □

4 Differentiability and the Chain Rule

Last time, we considered functions $f : U \to \mathbb{R}^m$ on an open subset $U \subseteq \mathbb{R}^n$ and defined the derivative $Df(a)$ of f at $a \in U$. Since the map $Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, it has an $m \times n$ matrix representation. We defined partial and directional derivatives, and we showed that if $Df(a)$ exists, then all directional derivatives exist and $D_v f(a) = Df(a)(v)$.

If $\{e_1, e_2, \ldots, e_n\}$ is the standard basis of $\mathbb{R}^n$ and $f = (f_1, f_2, \ldots, f_m)$, then

$$\begin{aligned} Df(a)(e_i) &= (Df_1(a)(e_i), Df_2(a)(e_i), \ldots, Df_m(a)(e_i)), \\ &= \left(\frac{\partial f_1}{\partial x_i}(a), \frac{\partial f_2}{\partial x_i}(a), \ldots, \frac{\partial f_m}{\partial x_i}(a) \right). \end{aligned}$$

As such, the matrix representation of $Df(a)$ is given by

$$Df(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}$$

This matrix is known as the *Jacobian matrix* of f at a. If f is differentiable, then we can describe the total derivative using its Jacobian matrix. However, we still need to know when the total derivative exists, and in today's lecture, we prove the following theorem.

Theorem 4.1. Let $U \subseteq \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^m$ a function. If all partial derivatives of f exist and are continuous on U, then f is differentiable on U.

Proof. Let $f = (f_1, f_2, \ldots, f_m)$, where $f_i : U \to \mathbb{R}$ for $1 \leq i \leq m$. Without loss of generality, it suffices to consider the case $m = 1$. Let $a = (a_1, \ldots, a_n) \in U$ and $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$, and define

$$v_j := (a_1, a_2, \ldots, a_{j-1}, \underbrace{a_j + u_j}_{j\text{th term}}, a_{j+1} + u_{j+1}, \ldots, a_n + u_n)$$

for $1 \leq j \leq n$ and $v_{n+1} := 0$. We have $f(v_1) = f(a+u)$ and $f(v_{n+1}) = f(a)$, and notice that

$$\begin{aligned} f(a+u) - f(a) &= (f(v_1) - f(v_2)) + (f(v_2) - f(v_3)) + \cdots + (f(v_{n-1}) - f(v_n)) + (f(v_n) - f(v_{n+1})), \\ &= \sum_{j=1}^{n} (f(v_j) - f(v_{j+1})) \end{aligned}$$

By the Mean Value Theorem, there exists $c_j \in (a_j, a_j + h_j)$ such that

$$\frac{\partial f}{\partial x_j}(a_1, \ldots, a_{j-1}, c_j, a_{j+1} + u_{j+1}, \ldots, a_n + u_n) = \frac{f(v_j) - f(v_{j+1})}{u_j},$$

where

$$\frac{f(v_j) - f(v_{j+1})}{u_j} = \frac{f(a_1, \ldots, a_j + u_j, \ldots, a_n + u_n) - f(a_1, \ldots, a_j, a_{j+1} + u_{j+1}, \ldots, a_n + u_n)}{u_j}.$$

We now have

$$f(a+u) - f(a) = \sum_{j=1}^{n} u_j \cdot \frac{\partial f}{\partial x_j}(a_1, \ldots, a_{j-1}, c_j, a_{j+1} + u_{j+1}, \ldots, a_n + u_n).$$

Let $\pi_j : \mathbb{R}^n \to \mathbb{R}^n$ describe the projection $(x_1, \ldots, x_n) \mapsto (0, \ldots, 0, x_j, 0, \ldots, 0)$, and if J_f is the Jacobian matrix of f at a, then

$$J_f(a)(u) = \sum_{j=1}^{n} (J_f(a) \circ \pi_j)(u) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a) \cdot u_i.$$

Now observe that $\dfrac{f(a+u) - f(a) - J_f(a)(u)}{\|u\|}$ is equal to

$$\begin{aligned} &\sum_{j=1}^{n} \frac{1}{\|u\|} \left(u_j \cdot \frac{\partial f}{\partial x_j}(a_1, \ldots, a_{j-1}, c_j, a_{j+1} + u_{j+1}, \ldots, a_n + u_n) - (J_f(a) \circ \pi_j)(u) \right), \\ &= \sum_{j=1}^{n} \frac{1}{\|u\|} \left(u_j \cdot \frac{\partial f}{\partial x_j}(a_1, \ldots, a_{j-1}, c_j, a_{j+1} + u_{j+1}, \ldots, a_n + u_n) - \frac{\partial f}{\partial x_j}(a) \cdot u_j \right), \\ &= \sum_{j=1}^{n} \frac{u_j}{\|u_j\|} \left(\frac{\partial f}{\partial x_j}(a_1, \ldots, a_{j-1}, c_j, a_{j+1} + u_{j+1}, \ldots, a_n + u_n) - \frac{\partial f}{\partial x_j}(a) \right). \end{aligned}$$

Note that $|u_i / \|u_i\|| \leq 1$, and because the partial derivatives of f are continuous, we have

$$\lim_{u \to 0} \left(\frac{\partial f}{\partial x_j}(a_1, \ldots, a_{j-1}, c_j, a_{j+1} + u_{j+1}, \ldots, a_n + u_n) - \frac{\partial f}{\partial x_j}(a) \right) = \frac{\partial f}{\partial x_j}(a) - \frac{\partial f}{\partial x_j}(a) = 0,$$

which implies

$$\lim_{u \to 0} \frac{f(a+u) - f(a) - J_f(a)(u)}{\|u\|} = 0.$$

Therefore, we have shown that $J_f(a)$ is the derivative of f at $a \in U$. $\square$

Remark.

- The converse is not necessarily true; a function can be differentiable but not all of the partial derivatives are continuous.
- It sufficed for us to consider f simply as a functional because showing

$$\lim_{u\to 0} \frac{f(a+u)-f(a)-J_f(a)(u)}{\|u\|} = 0$$

is equivalent to showing

$$\lim_{u\to 0} \pi_i \left(\frac{f(a+u)-f(a)-J_f(a)(u)}{\|u\|}\right) = 0$$

for $1 \le i \le m$.

To end this lecture, we prove the chain rule for multivariable derivatives:

Theorem 4.2 (Chain Rule)**.** Let $f : U \to \mathbb{R}^n$ and $g : V \to \mathbb{R}^p$ be functions on open sets $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$, respectively. If f is differentiable at $a \in U$ and g is differentiable at $b := f(a) \in V$, then $g \circ f : U \to \mathbb{R}^p$ is differentiable at a and $D(g \circ f)(a) = Dg(b) \circ Df(a)$.

Proof. Let $A := Df(a)$ and $B := Dg(b)$. Let $\tilde{f} : \mathbb{R}^n \to \mathbb{R}^m$ be the function $\tilde{f}(u) = f(a+u) - f(a) - Au$, and if f is differentiable, then

$$\lim_{u\to 0} \frac{\tilde{f}(u)}{\|u\|} = 0.$$

Observe that

$$\begin{aligned}
(g \circ f)(a+h) &= g(f(a+h)), \\
&= g\left(f(a) + Au + \tilde{f}(u)\right), \\
&= g(f(a)) + B\left(Au + \tilde{f}(u)\right) + \tilde{g}\left(Au + \tilde{f}(u)\right), \\
&= g(f(a)) + BAu + B(\tilde{f}(u)) + \tilde{g}\left(Au + \tilde{f}(u)\right).
\end{aligned}$$

We want to show

$$\lim_{u\to 0} \frac{B(\tilde{f}(u)) + \tilde{g}(Au + \tilde{f}(u))}{\|u\|} = 0.$$

Note that $\lim_{u\to 0} \frac{\tilde{f}(u)}{\|u\|} = 0$, and because B is linear—therefore continuous—we have

$$\lim_{u\to 0} \frac{B(\tilde{f}(u))}{\|u\|} = B\left(\lim_{u\to 0} \frac{\tilde{f}(u)}{\|u\|}\right) = B(0) = 0.$$

Now observe that

$$\lim_{u\to 0} \frac{\tilde{g}(Au + \tilde{f}(u))}{\|u\|} = \lim_{u\to 0} \left(\frac{\tilde{g}(Au + \tilde{f}(u))}{\|Au + \tilde{f}(u)\|} \cdot \frac{\|Au + \tilde{f}(u)\|}{\|u\|}\right).$$

By the Triangle Inequality,

$$\frac{\|Au + \tilde{f}(u)\|}{\|u\|} \le \frac{\|Au\|}{\|u\|} + \frac{\|\tilde{f}(u)\|}{\|u\|} \le \|A\| + \frac{\|\tilde{f}(u)\|}{\|u\|},$$

and because f is differentiable, we have

$$\lim_{u\to 0}\frac{\|Au+\tilde{f}(u)\|}{\|u\|}\leq \lim_{u\to 0}\left(\|A\|+\frac{\|\tilde{f}(u)\|}{\|u\|}\right)=\|A\|+0=\|A\|.$$

Since g is differentiable, we have $\lim_{u\to 0}\frac{\tilde{g}(Au+\tilde{f}(u))}{\|Au+\tilde{f}(u)\|}=0$, which implies

$$\begin{aligned}\lim_{u\to 0}\left(\frac{\tilde{g}(Au+\tilde{f}(u))}{\|Au+\tilde{f}(u)\|}\cdot\frac{\|Au+\tilde{f}(u)\|}{\|u\|}\right) &\leq \left(\lim_{u\to 0}\frac{\tilde{g}(Au+\tilde{f}(u))}{\|Au+\tilde{f}(u)\|}\right)\cdot\left(\lim_{u\to 0}\frac{\|Au+\tilde{f}(u)\|}{\|u\|}\right),\\ &\leq 0\cdot\|A\|,\\ &= 0.\end{aligned}$$

Thus, we have shown that $g\circ f$ is differnetiable at $a\in U$, and $D(g\circ f)(a)=Dg(b)\circ Df(a)$. $\square$

5 Applications of the Chain Rule and Higher-Ordered Derivatives

Last time, we characterized differentiability and proved the multivariable Chain Rule. We will examine some applications of the Chain Rule and discuss higher-ordered derivatives.

Theorem 5.1. Let $U\subseteq\mathbb{R}^n$ be an open subset.

1. If $f,g:U\to\mathbb{R}^m$ are differentiable at $a\in U$, then $f+g$ is differentiable at a and
$$D(f+g)(a)=Df(a)+Dg(a).$$
2. If $f,g:U\to\mathbb{R}$ are differentiable at $a\in U$, then $f\cdot g$ is differentiable and
$$D(f\cdot g)(a)=Df(a)\cdot g(a)+Dg(a)\cdot f(a)$$
3. If $f:U\to\mathbb{R}$ is differentiable at $a\in U$ and $f(a)\neq 0$, then $1/f$ is differentiable at a and
$$D(1/f)(a)=-\frac{1}{f(a)^2}Df(a)$$
.

Proof.

1. Define functions $F:U\to\mathbb{R}^{2m}$ and $G:\mathbb{R}^{2m}\to\mathbb{R}^m$ by $F(x)=(f(x),g(x))$ and $G(v,w)=v+w$. Note that F is differentiable at a because each of its components f and g are differentiable, and G is differentiable because addition is linear. We have $DF(a)=(Df(a),Dg(a))$ and $DG(v,w)=G$. The function $f+g$ is the composition $G\circ F:U\times U\to\mathbb{R}^m$, and the Chain Rule tells us $G\circ F$ is differentiable and
$$\begin{aligned}D(G\circ F)(a)&=DG(F(a))\circ DF(a),\\ &=DG(F(a))\circ(Df(a),Dg(a)),\\ &=G\circ(Df(a),Dg(a))\\ &=Df(a)+Dg(a).\end{aligned}$$
2. Once again, define $F:U\to\mathbb{R}^{2m}$ by $F(x)=(f(x),g(x))$ and $G:\mathbb{R}^2\to\mathbb{R}$ by $G(u,v)=u\cdot v$.

Note that G is differentiable, where

$$DG(u,v) = \begin{pmatrix} \frac{\partial G}{\partial u}(u,v) & \frac{\partial G}{\partial v}(u,v) \end{pmatrix} = \begin{pmatrix} v & u \end{pmatrix}.$$

And so, observe that

$$\begin{aligned} D(G \circ F)(a) &= DG(F(a)) \circ DF(a), \\ &= \begin{pmatrix} g(a) & f(a) \end{pmatrix} \circ \left(Df(a), Dg(a),\right), \\ &= g(a) \cdot Df(a) + f(a) \cdot Dg(a). \end{aligned}$$

Note that $g(a)$ and $f(a)$ are scalars, so $g(a) \cdot Df(a) + f(a) \cdot Dg(a) = Df(a) \cdot g(a) + Dg(a) \cdot f(a)$.

3. Once more, define $F : U \to \mathbb{R}^{2m}$ by $F(x) = (\mathrm{Id}_n(x), f(x))$ and $G : \mathbb{R}^2 \to \mathbb{R}$ by $G(u,v) = 1/v$. We have that G is differentiable, where

$$DG(u,v) = \begin{pmatrix} \frac{\partial G}{\partial u}(u,v) & \frac{\partial G}{\partial v}(u,v) \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{v^2} \end{pmatrix},$$

and as such, we have

$$\begin{aligned} D(G \circ F)(a) &= DF(F(a)) \circ DF(a), \\ &= \begin{pmatrix} 0 & -\frac{1}{f(a)^2} \end{pmatrix} \circ (D(\mathrm{Id}_n)(a), Df(a)), \\ &= -\frac{1}{f(a)^2} \cdot Df(a). \end{aligned}$$

□

Theorem 5.2 (Mean Value Theorem). If $f : U \to \mathbb{R}$ is differentiable on U and U contains the line segment joining $[a, a+h]$ for $a \in U$ and $h \in \mathbb{R}^n$, then there exists $t_0 \in (0,1)$ such that

$$f(a+h) - f(a) = Df(a + t_0 h)(h).$$

Proof. Define $\phi : \mathbb{R} \to \mathbb{R}$ as the function $\phi(t) := f(a+th)$. By the Mean Value Theorem in $\mathbb{R}$, there exists $t_0 \in (0,1)$ such that

$$\phi'(t_0) = \frac{\phi(1) - \phi(0)}{1 - 0} = f(a+h) - f(a),$$

and by the Chain Rule $\phi'(t_0) = Df(a + t_0 h)(h)$. □

Now let's turn our discussion to higher-ordered derivatives. So far, we have that for a differentiable function $f : U \to \mathbb{R}^m$, the total derivative is the map $Df : U \to \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) \cong M_{m \times n}$ that takes in $a \in U$ and returns a linear transformation. We can further differentiate Df, where

$$D^2 f = D(Df) : U \to \mathrm{Hom}(\mathbb{R}^n, \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m))$$

takes a point $a \in U$ and returns a linear map from $\mathbb{R}^n$ to the space of $m \times n$ matrices. The matrix representation of $D^2 f(a)$ would be a "three-dimensional" array instead of the two-dimensional Jacobian matrix.

Definition 5.3. A function $f : U \to \mathbb{R}^m$ is of *class* C^k if the derivatives $Df, D^2 f, \ldots, D^k f$ all exist and are continuous on U. A function is of *class* C^∞ is called *smooth*.

To make sense of the second multivariable derivative, we will "derive" the matrix representation of $D^2f(a) = D(Df)(a)$ for a functional $f : U \to \mathbb{R}$. To begin, note that $Df : U \to \mathrm{Hom}(\mathbb{R}^n, \mathbb{R})$ takes a point $a \in U$ and returns the map $Df(a) : \mathbb{R}^n \to \mathbb{R}$, whose matrix representation is given by

$$Df(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) & \frac{\partial f}{\partial x_2}(a) & \cdots & \frac{\partial f}{\partial x_n}(a) \end{pmatrix}.$$

Following this setup, $D(Df) : U \to \mathrm{Hom}(\mathbb{R}^n, \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}))$ will also take a point $a \in U$, but this time it will return a map $D(Df)(a) : \mathbb{R}^n \to \mathrm{Hom}(\mathbb{R}^n, \mathbb{R})$, whose matrix representation is given by

$$D(Df)(a) = \begin{pmatrix} \frac{\partial (Df)}{\partial x_1}(a) & \frac{\partial (Df)}{\partial x_2}(a) & \cdots & \frac{\partial (Df)}{\partial x_n}(a). \end{pmatrix}$$

By the definition of partial derivatives, we have

$$\begin{aligned} \frac{\partial (Df)}{\partial x_i}(a) &= \lim_{h \to 0} \frac{Df(a + he_i) - Df(a)}{h}, \\ &= \lim_{h \to 0} \frac{1}{h} \begin{pmatrix} \frac{\partial f}{\partial x_1}(a + he_i) - \frac{\partial f}{\partial x_1}(a) & \frac{\partial f}{\partial x_2}(a + he_i) - \frac{\partial f}{\partial x_2}(a) & \cdots & \frac{\partial f}{\partial x_n}(a + he_i) - \frac{\partial f}{\partial x_n}(a). \end{pmatrix} \end{aligned}$$

If $\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(a)$ exists for $1 \leq j \leq n$, then

$$\frac{\partial (Df)}{\partial x_i}(a) = \left(\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_1}\right)(a) \quad \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_2}\right)(a) \quad \cdots \quad \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_n}\right)(a) \right).$$

Treating the n-tuple above as a vector, we have

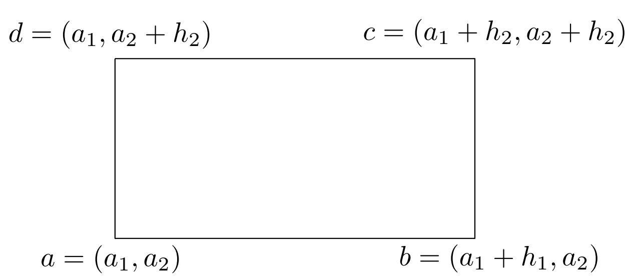
$$D(Df)(a) = \begin{pmatrix} \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right)(a) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right)(a) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_1}\right)(a) \\ \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right)(a) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right)(a) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_2}\right)(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_n}\right)(a) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_n}\right)(a) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_n}\right)(a) \end{pmatrix}$$

The matrix above is also known as the *Hessian*. If f is of class C^k where $k \geq 2$, then we will see that partial derivatives can be interchanged.

Theorem 5.4. Let $f : U \to \mathbb{R}^m$ be a function of class C^2. Then for $1 \leq i, j \leq n$, we have

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(a) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(a).$$

Before diving into our proof, let's get some intuition for the theorem above. Without loss of generality, assume $n = 2$ and $m = 1$ because only two variables are being interchanged and it suffices to consider functionals. Let $a \in U$ and $h \in \mathbb{R}^2$, and consider the $R \subseteq \mathbb{R}^2$ depicted below:



Roughly speaking, h_1 and h_2 are small quantities, and

$$\frac{\partial f}{\partial x_1}(a) \approx \frac{f(b)-f(a)}{h_1}, \qquad \frac{\partial f}{\partial x_2}(a) \approx \frac{f(d)-f(a)}{h_2}.$$

Consequently, this means

$$\begin{aligned}
\frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right)(a) &\approx \frac{1}{h_1}\left(\frac{f(c)-f(b)}{h_2} - \frac{f(d)-f(a)}{h_2}\right) = \frac{1}{h_1 h_2}\left(f(c)-f(b)-f(d)+f(a)\right), \\
\frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right)(a) &\approx \frac{1}{h_2}\left(\frac{f(c)-f(d)}{h_1} - \frac{f(b)-f(a)}{h_1}\right) = \frac{1}{h_1 h_2}\left(f(c)-f(d)-f(b)+f(a)\right),
\end{aligned}$$

which roughly shows

$$\frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right)(a) = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right)(a).$$

For our actual proof, we will apply the Mean Value Theorem to "replicate" the approximations that we are making above.

Proof. Define $\lambda : \mathbb{R}^2 \to \mathbb{R}$ by

$$\lambda(h_1, h_2) := f(a) - f(b) + f(c) - f(d).$$

We will show that there exist $p, q \in R$ such that

$$\lambda(h_1, h_2) = \frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_1}(p) \cdot h_1 h_2 = \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_2}(q) \cdot h_1 h_2$$

and apply the continuity of partial derivatives to obtain our result. Note that it suffices to obtain one of the equalities above because of symmetry.

First, consider $\phi : \mathbb{R} \to \mathbb{R}$ defined by

$$\phi(s) = f(s, a_2 + h_2) - f(s, a_2).$$

Then observe that $\lambda(h_1, h_2) = \phi(a_1+h_1) - \phi(a_1)$, and by the Theorem 5.2, there exists $s_0 \in (a_1, a_1+h_1)$ such that

$$\lambda(h_1, h_2) = \phi'(s_0) \cdot h_1 = \underbrace{\left(\frac{\partial f}{\partial x_1}(s_0, a_2 + h_2) - \frac{\partial f}{\partial x_1}(s_0, s_2)\right)}_{\text{by Chain Rule}} \cdot h_1.$$

Applying Theorem 5.2 again, there exists $t_0 \in (a_2, a_2 + h_2)$ such that

$$\frac{\partial f}{\partial x_1}(s_0, a_2 + h_2) - \frac{\partial f}{\partial x_1}(s_0, a_2) = \left(\frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_1}(s_0, t_0)\right) \cdot h_2.$$

And so, we have

$$\lambda(h_1, h_2) = \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_1}(s_0, t_0) \cdot h_1 h_2,$$

and we can take $p = (s_0, t_0) \in R$. We repeat this process to get $q \in R$, and we have

$$\frac{\lambda(h_1, h_2)}{h_1 h_2} = \frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_1}(p) = \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_2}(q).$$

Because the second partial derivatives are continuous, we have

$$\lim_{p \to a} \frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_1}(p) = \frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_1}(a) = \frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_1}(a) = \lim_{p \to a} \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_2}(q).$$

□

Remark. The proof technique to take away here is that the Mean Value Theorem is a way to formalize heuristics/intuition that we may have.

By this point, we have some sense of what higher-derivatives may look like. Note that if we differentiated $D^2 f$ from our example once more, we would end up with a three-dimensional array of iterated partial derivative functions, which we denote by

$$D^3 f = \left(\frac{\partial^3 f}{\partial x_{a_3} \partial x_{a_2} \partial x_{a_1}} \right)_{(a_1, a_2, a_3)}$$

for all sequences (a_1, a_2, a_3) where $1 \leq a_i \leq 3$. In general, higher-ordered total derivatives are simply arrays of iterated partial derivatives of a function. So if $f : \mathbb{R}^n \to \mathbb{R}^m$ is a C^k function with $f = (f_1, f_2, \dots, f_m)$, then $D^k f$ is the array

$$D^k f = \left(\frac{\partial^k f_i}{\partial x_{a_k} \cdots \partial x_{a_1}} \right)_{\substack{(a_1, \dots, a_k) \\ 1 \leq i \leq m}}.$$

6 Inverse Function Theorem

So far, we have seen that the derivative is roughly a linear approximation of a function at a point. This week, we will explore the implications of differentiability in describing the local behaviors of functions. In this lecture, we introduce the statement of the Inverse Function Theorem and some of the tools we will use. Roughly speaking, if a function's derivative is invertible (non-vanishing), then we can expect the function to be invertible within a sufficiently small neighborhood.

Theorem 6.1 (Inverse Function Theorem)**.** Let $f : U \to \mathbb{R}^n$ be a C^k function defined on an open set $U \subseteq \mathbb{R}^n$. If $a \in U$ such that $Df(a) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, then there exist open neighbohoods $V \subseteq U$ of a and $W \subseteq \mathbb{R}^n$ of $b := f(a)$ and a map $g : W \to V$ such that

- $f(V) = W$ and $g(W) = V$,
- $(g \circ f)(x) = x$ and $(f \circ g)(y) = y$ for $x \in V$ and $y \in W$.

Moreover, $Dg(y) = Df(g(y))^{-1}$ for all $y \in W$.

Part of the proof will be showing that within a sufficiently small neighborhood $W \subseteq \mathbb{R}^n$ of y, there exists $x \in V$ such that $f(x) = y$. This problem is simple if $f : U \to \mathbb{R}^n$ were linear, as $f(x) = y$ if and only if $x = a + Df(a)^{-1}(y - b)$ (resembles slope-intercept form). However, this is not necessarily true if f were not linear. Nonetheless, we can use the expression to get a sequence of points that approximate the preimage of y, namely $x_0 := a + Df(a)^{-1}(y - b)$ and

$$x_n := x_{n-1} + Df(a)^{-1}(f(x_{n-1}) - b)$$

for $n \geq 1$. The sequence that we obtain should converge to a fixed point $F_y(x) = x + Df(a)^{-1}(f(x) - y)$, and note that $f(x) = y$ if and only if $F_y(x) = x$. This leads us to an application of the Contraction Mapping Theorem, which encapsulates our algorithm.

Theorem 6.2 (Contraction Mapping Theorem)**.** Let (X, d) be a complete metric space and $T : X \to X$ is a function such that there exists $r < 1$ such that

$$d(T(x), T(y)) \leq r \cdot d(x, y)$$

for all $x, y \in X$. Then T has a unique fixed point. More generally, if Λ is another metric space and

$\{T_\lambda\}_{\lambda \in \Lambda}$ is a continuous family of contraction mappings with scale factor r—in other words,

$$\lim_{\lambda \to \lambda_0} \sup_{x \in X} d(T_\lambda(x), T_{\lambda_0}(x)) = 0$$

—then the fixed points of T_λ belong in Λ. Equivalently, if we consider $f : \Lambda \to X$ to be the function that maps $\lambda \in \Lambda$ to the fixed point of $T_\lambda : X \to X$, then T_λ is continuous.

In order to invoke the Contraction Mapping Theorem, we must show that $f : U \to \mathbb{R}^n$ satisfies the theorem's hypotheses, which requires the following result:

Theorem 6.3. Let $f : U \to \mathbb{R}^m$ be a differentiable function from a convex open set $U \subseteq \mathbb{R}^n$ such that $\|Df(x)\| < h$ for all $x \in U$. Then $\|f(x) - f(y)\| \leq h\|x - y\|$ for all $x, y \in U$.

Proof. Let $x, y \in U$ and $v := \dfrac{f(x) - f(y)}{\|f(x) - f(y)\|}$. Define the linear map $g : \mathbb{R}^m \to \mathbb{R}$ by

$$g(y) = v \cdot y = \sum_{i=1}^{m} v_i y_i.$$

By the definition of v, we have $g(f(x) - f(y)) = \|f(x) - f(y)\|$, the Cauchy-Schwarz Inequality implies $\|g\| = 1$. Now consider $g \circ f : U \to \mathbb{R}$; the Chain Rule tells us $g \circ f$ is differentiable, and the Mean Value Theorem tells us there exists $u \in [x, y]$ such that

$$(g \circ f)(x) - (g \circ f)(y) = D(g \circ f)(u)(x - y).$$

Because g is linear,

$$(g \circ f)(x) - (g \circ f)(y) = g(f(x - y)) = g(f(x) - f(y)) = \|f(x) - f(y)\|.$$

By the Chain Rule, we have $D(g \circ f)(u) = g \circ Df(u)$, so

$$\|f(x) - f(y)\| = (g \circ Df(u))(x - y).$$

Since $\|g\| = 1$, we have

$$\|g \circ Df(u)\| \leq \|Df(u)\| \leq h,$$

which shows $\|f(x) - f(y)\| \leq h\|x - y\|$. $\square$

7 Inverse Function Theorem (continued)

In this lecture, we prove the Inverse Function Theorem.

Theorem. Let $f : U \to \mathbb{R}^n$ be a C^k function defined on an open set $U \subseteq \mathbb{R}^n$. If $a \in U$ such that $Df(a) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, then there exist open neighbohoods $V \subseteq U$ of a and $W \subseteq \mathbb{R}^n$ of $b := f(a)$ and a map $g : W \to V$ such that

- $f(V) = W$ and $g(W) = V$,
- $(g \circ f)(x) = x$ and $(f \circ g)(y) = y$ for $x \in V$ and $y \in W$.

Moreover, $Dg(y) = Df(g(y))^{-1}$ for all $y \in W$.

Proof. Let $y \in \mathbb{R}^n$, and define $F_y(x) := x + A^{-1}(f(x) - y)$, where $A = Df(a)$. If $x \in U$ is a fixed point of F_y, then we have $f(x) = y$. Now observe that

$$DF_y(x) = \mathrm{Id}(x) + A^{-1}Df(x),$$

and note that $DF_y(a) = 0$ and $DF_y(x)$ does not depend on y. And so, for $0 < \varepsilon < 1$ there exists $R > 0$ such that $\|DF_y(x)\| < \varepsilon$ for all $y \in \mathbb{R}^n$ and $x \in B(a, R)$. We want to restrict y so that F_y maps $B(a, R)$ to $B(a, R)$ and is a contraction mapping with factor ε. If $\|DF_y(z)\| < \varepsilon$ for all $z \in B(a, R)$, then for $x, x' \in B_R(a)$ the Mean Value Theorem implies that

$$\|F_y(x) - F_y(x')\| < \varepsilon\|x - x'\|.$$

Observe that

$$\begin{aligned}\|F_y(x) - a\| &\leq \|F_y(x) - F_y(a)\| + \|F_y(a) - a\|,\\ &< \varepsilon\|x - a\| + \|F_y(a) - a\|,\\ &< \varepsilon R + \|F_y(a) - a\|.\end{aligned}$$

We know $F_y(a)$ is a continuous function of y and $F_{f(a)}(a) = a$. And so, if $y = f(a)$, then $\|F_y(a) - a\| = 0$. By the continuity of F_y, there exists a neighborhood W of $f(a)$ such that for $y \in W$, $\|F_y(a) - a\| \leq R(1 - \varepsilon)$. For $y \in W$, we have

$$\|F_y(x) - a\| < \varepsilon R + R(1 - \varepsilon) = R.$$

Thus, we have that F_y maps $B(a, R)$ into itself. Because $\|F_y(x) - F_y(x')\| < \varepsilon\|x - x'\|$ for $x, x' \in B_R(a)$, then the Contraction Mapping Theorem tells us that there exists a unique fixed point of F_y in U. Let $g(y)$ denote this unique fixed point, and by construction $f(g(y)) = y$.

Now define $V := f^{-1}(W) \cap B(a, R)$, which is open because W is open and f is continuous. Note that g maps from W to V and $f(V) \subseteq W$. Moreover, we have $g \circ f(x) = x$ and $f \circ g(y) = y$. We now want to show that g is continuous, differentiable, and of class C^k. To see that g is continuous, note that the Contraction Mapping Theorem tells us that it suffices to show

$$\lim_{y \to y_0} \sup_{x \in V} \|F_y(x) - F_{y_0}(x)\| = 0.$$

for $y_0 \in W$. Note that if $x \in B(a, R)$, then

$$\begin{aligned}\|F_y(x) - F_{y_0}(x)\| &= \|A^{-1}(y) - A^{-1}(y_0)\|,\\ &= \|A^{-1}(y - y_0)\|,\\ &\leq \|A^{-1}\|\|y - y_0\|.\end{aligned}$$

Thus, we have

$$\lim_{y \to y_0} \sup_{x \in V} \|F_y(x) - F_{y_0}(x)\| \leq \lim_{u \to y_0} \|A^{-1}\|\|y - y_0\| = 0,$$

which implies g is continuous. To see that g is differentiable, note that $f \circ g = \mathrm{Id}$ implies $D(f \circ g)(y) = Df(g(y)) \circ Dg(y) = \mathrm{Id}(y)$ and $Dg(y) = Df(g(y))^{-1}$ by the Chain Rule. We want to verify that the derivative of g is indeed $Df(g(y))^{-1}$, namely

$$\lim_{y \to y_0} \frac{g(y) - g(y_0) - Df(g(y_0))^{-1}(y - y_0)}{\|y - y_0\|} = 0.$$

for $y_0 \in W$. Let $x_0 = g(y_0)$ and $x = g(y)$. By continuity of g, we have

$$\begin{aligned}&= \lim_{x \to x_0} \frac{x - x_0 - Df(x_0)^{-1}(f(x) - f(x_0))}{\|f(x) - f(x_0)\|},\\ &= \lim_{x \to x_0} \frac{-Df(x_0)^{-1}\left(f(x) - f(x_0) - Df(x_0)(x - x_0)\right)}{\|x - x_0\|} \cdot \frac{\|x - x_0\|}{\|f(x) - f(x_0)\|}.\end{aligned}$$

Since f is differentiable, we have

$$= \lim_{x \to x_0} \frac{f(x) - f(x_0) - Df(x_0)(x - x_0)}{\|x - x_0\|} = 0.$$

And now, it suffices to show that $\dfrac{\|x - x_0\|}{\|f(x) - f(x_0)\|}$ is bounded, namely that there exists $c > 0$ so that

$$0 < c < \liminf_{x \to x_0} \frac{\|f(x) - f(x_0)\|}{\|x - x_0\|}.$$

By the Reverse Triangle Inequality, notice that

$$\begin{aligned} \liminf_{x \to x_0} \frac{\|f(x) - f(x_0)\|}{\|x - x_0\|} &\geq \liminf_{x \to x_0} \left| \frac{\|Df(x_0)(x - x_0)\|}{\|x - x_0\|} - \frac{\|f(x) - f(x_0) - Df(x_0)(x - x_0)\|}{\|x - x_0\|} \right|, \\ &\geq \liminf_{x \to x_0} \frac{\|Df(x_0)(x - x_0)\|}{\|x - x_0\|}. \end{aligned}$$

If $Df(x_0)$ is invertible, then $\|Df(x_0)\| > 0$. In particular, this means that there exists $c > 0$ such that

$$\|Df(x_0)(y)\| \geq c\|y\|$$

for $y \neq 0$. To clarify, c is the minimum norm of $Df(x_0)$ applied to a unit vector, and such a quantity exists because the unit sphere is compact. And so,

$$\liminf_{x \to x_0} \frac{\|D(f(x_0)(x - x_0)\|}{\|x - x_0\|} > 0,$$

which implies

$$\lim_{y \to y_0} \frac{g(y) - g(y_0) - Df(g(y_0))^{-1}(y - y_0)}{\|y - y_0\|} = 0.$$

Note that $Dg : U \to \mathrm{GL}_n \subset \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^n)$ maps $y \in U$ to $\mathrm{inv} \circ Df \circ g$. The inverse map inv is smooth and Df is of class C^{k-1}. And because g is continuous, we know Dg is continuous, and therefore of class C^1. Consequently, g is of class C^2 and arguing inductively, we have that g is of class C^k. $\square$

8 Implicit Function Theorem

In this lecture, we will use the Inverse Function Theorem to prove the Implicit Function Theorem, which describes when the level set of a function is locally parametrizable based on properties of function's derivative. Very often in the next few lectures we will associate the parametrization of a set with the set being the *graph* of some function.

Definition 8.1. If $f : U \to \mathbb{R}^m$ is a function defined on an open set $U \subseteq \mathbb{R}^n$, then the *graph* of f is the set

$$\Gamma(f) := \{(x, f(x)) : x \in U\} \subseteq \mathbb{R}^n \times \mathbb{R}^m.$$

Example 8.2. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is a C^k function and consider the function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ defined by $F(x, y) = y - f(x)$. The level set $F^{-1}(0)$ is defined as

$$F^{-1}(0) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : F(x, y) = 0\}.$$

If $(x, y) \in F^{-1}(0)$, then $F(x, y) = y - f(x) = 0$. Note that $F(x, y) = 0$ if and only if $y = f(x)$, and this implies $F^{-1}(0)$ can also be defined by

$$F^{-1}(0) = \{(x, f(x)) : x \in \mathbb{R}^n\}.$$

Notice that $F^{-1}(0)$ here is precisely the graph of f, and more importantly, the y-coordinates of points in $F^{-1}(0)$ are *parametrized* by the x-coordinates.

In our example above, we started with a function $f : \mathbb{R}^n \to \mathbb{R}^m$ and constructed a function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ whereby its level set $F^{-1}(0)$ can be parametrized in terms of $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^m$. We can flip this situation around and ask when the level set $F^{-1}(0)$ of a function $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$ is parametrizable. This is the essense of the Implicit Function Theorem.

Theorem 8.3 (Implicit Function Theorem). Let $F : U \to \mathbb{R}^m$ be a C^k function defined on an open set $U \subseteq \mathbb{R}^n \times \mathbb{R}^m$. If $a = (a_1, a_2) \in U$ such that $F(a) = 0$ and $DF(a) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is surjective, then there exists an open neighborhood $V \subseteq \mathbb{R}^n$ of $a_1 \in \mathbb{R}^n$ and a C^k function $f : V \to \mathbb{R}^m$ such that $(x, f(x)) \in U$ and $F(x, f(x)) = 0$ for all $x \in V$. Moreover, if V is connected, then f is unique.

Remark. Given $DF(a) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, we have

$$DF(a) = \begin{pmatrix} D_xF(a) & D_yF(a) \end{pmatrix},$$

where $D_xF(a) : \mathbb{R}^n \to \mathbb{R}^m$ and $D_yF(a) : \mathbb{R}^m \to \mathbb{R}^m$ are block matrices that constitute the first n and last m columns of $DF(a)$, respectively. Note that if $DF(a) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, then a rearrangement of coordinates will allow $D_yF(a)$ to be an isomorphism between $\mathbb{R}^m$ and itself. In other words, $D_yF(a)$ is invertible. This is a convenient assumption that we will make in our later proofs.

Before we dive into a proof of the Implicit Function Theorem, let us consider the simpler case involving linear transformations. Suppose that $T : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is linear and surjective. We want to parametrize the level set $T^{-1}(0)$, which is actually $\ker T$. Let $(x, y) \in T^{-1}(0)$, and observe that

$$T(x, y) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_{11} & b_{12} & \cdots & b_{1m} \\ a_{21} & a_{22} & \cdots & a_{2n} & b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_{m1} & b_{m2} & \cdots & b_{mm} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

yields the following system of equations:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + b_{11}y_1 + b_{12}y_2 + \cdots + b_{1m}y_m = 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + b_{21}y_1 + b_{22}y_2 + \cdots + b_{2m}y_m = 0, \\ \qquad\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + b_{m1}y_1 + b_{m2}y_2 + \cdots + b_{mm}y_m = 0. \end{cases}$$

There are m equations in $n + m$ variables. Note that the system above is equivalent to

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = -(b_{11}y_1 + b_{12}y_2 + \cdots + b_{1m}y_m), \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = -(b_{21}y_1 + b_{22}y_2 + \cdots + b_{2m}y_m), \\ \qquad\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = -(b_{m1}y_1 + b_{m2}y_2 + \cdots + b_{mm}y_m), \end{cases}$$

which we can reformulate as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = - \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mm} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

Letting $A := (a_{ij})$ and $B := (b_{ij})$, we have that $Ax = -By$. Note that because T is surjective, we can assume B is invertible after a change of basis. And so, we have $y = -B^{-1}Ax$, which parametrizes y in terms of x for $(x, y) \in \ker T$. The key assumption that allowed us to parametrize $\ker T$ is that T is surjective. This hopefully illustrates how $DF(a) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ might be a useful assumption in our proof of the general Implicit Function Theorem. Unlike the linear case, however, we need to somehow construct the parametrizing function for $F^{-1}(0)$. For that purpose, we will use the Inverse Function Theorem.

Proof. Let $G : U \to \mathbb{R}^n \times \mathbb{R}^m$ be the map $G(x, y) = (x, F(x, y))$. Note that G is a C^k function and

$$DG = \begin{pmatrix} DG_1 & DG_2 \end{pmatrix} = \begin{pmatrix} \dfrac{\partial G_1}{\partial x} & \dfrac{\partial G_1}{\partial y} \\ \dfrac{\partial G_2}{\partial x} & \dfrac{\partial G_2}{\partial y} \end{pmatrix} = \begin{pmatrix} \mathrm{Id}_n & 0 \\ D_xF & D_yF \end{pmatrix}.$$

If $a \in U$ and $D_yF(a)$ is invertible, then $DG(a)$ is invertible, whereby

$$DG(a)^{-1} = \begin{pmatrix} \mathrm{Id}_n & 0 \\ -D_yF(a)^{-1} \circ D_xF(a) & D_yF(a)^{-1} \end{pmatrix}.$$

By the Inverse Function Theorem, there exists an open neighborhood $W_1 \times W_2$ of $a \in U$ such that the restriction $G|_{W_1 \times W_2}$ is invertible and its inverse is differentiable. Now let $Z := G(W_1, W_2) \subseteq \mathbb{R}^n \times \mathbb{R}^m$, and note that Z is a neighborhood of $(a_1, 0)$ and there is an C^k inverse function $H : Z \to W_1 \times W_2$ such that $H(x, F(x, y)) = (x, y)$. And so, $H(x, y) = (x, h(x, y))$ for some C^k function $h : Z \to W_2$.

Now consider an open neighborhood $V \subseteq \mathbb{R}^n$ of a_1 such that $V \times \{0\} \subseteq Z$. Then for $x \in V$, we have $H(x, 0) = (x, h(x, 0))$ and $(x, 0) = G(x, h(x, 0)) = (x, F(x, h(x, 0)))$, which implies $F(x, h(x, 0)) = 0$. And so, define $f : V \to \mathbb{R}^m$ by $f(x) = h(x, 0)$. □

9 Implicit Function Theorem and Manifolds

Last class, we proved the existence of a parametrizing function stated in the Implicit Function Theorem.

Theorem (Implicit Function Theorem)**.** Let $F : U \to \mathbb{R}^m$ be a C^k function defined on an open set $U \subseteq \mathbb{R}^n \times \mathbb{R}^m$. If $a = (a_1, a_2) \in U$ such that $F(a) = 0$ and $DF(a) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is surjective, then there exists an open neighborhood $V \subseteq \mathbb{R}^n$ of $a_1 \in \mathbb{R}^n$ and a C^k function $f : V \to \mathbb{R}^m$ such that $(x, f(x)) \in U$ and $F(x, f(x)) = 0$ for all $x \in V$. Moreover, if V is connected, then f is unique.

We complete our proof of the thereom today by showing f is unique if V is connected.

Proof. Suppose $f_1, f_2 : V \to \mathbb{R}^m$ satisfy the conclusion of the Implicit Function Theorem. We have $f_1(a_1) = f_2(a_1) = a_2$ and $F(x, f_1(x)) = F(x, f_2(x)) = 0$ for all $x \in V$. Now define

$$A := \{x \in V : f_1(x) = f_2(x)\},$$

and note that A is nonempty because $a_1 \in A$. Moreover, A is closed because f_1 and f_2 are continuous, and this implies $V \setminus A$ is open. Notice that $V = A \cup (V \setminus A)$, and if we show that A is also open, then connectedness of V tells us $V \setminus A$ must be empty.

Let $b_1 \in A$ and $b_2 = f_1(b_1) = f_2(b_1)$. Define $G : U \to \mathbb{R}^{n+m}$ by $G(x_1, x_2) = (x_1, F(x_1, x_2))$ and recall that from last class that G satisfies the hypothesis of the Inverse Function Theorem at (b_1, b_2). And so, there exists an open neighborhood $W_1 \times W_2$ of (b_1, b_2) such that $G : W_1 \times W_2 \to G(W_1 \times W_2)$ is invertible with a C^k inverse and $G(W_1 \times W_2) \subseteq \mathbb{R}^{n+m}$ is open. Note that for $(x_1, 0) \in G(W_1 \times W_2)$, there exists a unique $x_2 \in W_2$ such that $G(x_1, x_2) = (x_1, 0)$. We have

$$F(x_1, x_2) = F(x_1, f_1(x_1)) = F(x_1, f_2(x_1)) = 0,$$

and it follows from the uniqueness of x_2 that $x_2 = f_1(x_1) = f_2(x_1)$. Because $G(W_1 \times W_2)$ is open, there exists an open neighborhood $B \times \{0\} \subseteq G(W_1 \times W_2)$ of $(x_1, 0)$ where $f_1(x_1) = f_2(x_1)$, and this shows $B \subseteq A$ is thus A is open. $\square$

Example 9.1. Consider $F : \mathbb{R}^2 \to \mathbb{R}$ defined by $F(x, y) = x^2 + y^2 - 1$. Then

$$F^{-1}(0) = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}.$$

Our goal is to parametrize the y-coordinate of points in $F^{-1}(0)$ in terms of x (within an appropriate neighborhood of x). Now let $a = (a_1, a_2) \in F^{-1}(0)$, and we have $a_1^2 + a_2^2 = 1$. To apply the Implicit Function Theorem, we need $Df(a) : \mathbb{R}^2 \to \mathbb{R}$ to be surjective, or equivalently that $D_y F(a) = 2a_2$ is invertible. This is true for all $a \in F^{-1}(0)$ except $(1, 0)$ and $(-1, 0)$.

Example 9.2. Let's consider an example where the parametrizing function is not unique if the neighborhood of the coordinates we're parametrizing by is not connected. Define $F : \mathbb{R}^2 \to \mathbb{R}$ by $F(x, y) = x - y^2$, which yields

$$F^{-1}(0) = \{(x, y) \in \mathbb{R}^2 : y^2 = x\}.$$

Suppose y is a function of x within the neighborhood $V = (0, 2) \cup (3, 4)$ of $(1, 1)$. Note that there are two ways to parametrize the y-coordinate of points in $F^{-1}(0)$, namely with $f_1(x) = \sqrt{x}$ for all $x \in V$ and

$$f_2(x) = \begin{cases} \sqrt{x} & x \in (0, 2) \\ -\sqrt{x} & x \in (3, 4) \end{cases},$$

And indeed, we have $F(x, f_1(x)) = F(x, f_2(x)) = 0$ for all $x \in V$. This example demonstrates the importance of a connected domain to enable a unique parametrization.

Our proof of the Implicit Function Theorem concludes our discussion of derivatives and their implications about the local behaviors of functions. We will continue to use these techniques when we discuss manifolds in depth. For the rest of the lecture, we introduce some basic definitions related to manifolds.

Definition 9.3. Let $W_1, W_2 \subseteq \mathbb{R}^n$. We say a function $f : W_1 \to W_2$ is a C^k *embedding* if f is an injective C^k function where $Df(x) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible for all $x \in W_1$.

Definition 9.4. Let M be a subset of $\mathbb{R}^n$. We say M is an *embedded m-dimensional C^k manifold* if for all $x \in M$, there exists a neighborhood $U \subseteq \mathbb{R}^n$ of x and a smooth function $F : U \to \mathbb{R}^{n-m}$ such that

1. $M \cap U = F^{-1}(0)$,
2. $DF(y) : \mathbb{R}^n \to \mathbb{R}^{n-m}$ is surjective for all $y \in U$.

For $x \in M$, the *tangent space* T_xM at x is the vector space given by

$$T_xM := \ker(DF(x))$$

for any choice of $f : U \to \mathbb{R}^n$.

Remark. The function $F : U \to \mathbb{R}^{n-m}$ guaranteed by each $x \in M$ is not necessarily. This is not really an issue, and in fact, we can view this non-uniqueness as a special feature of manifolds.
Because $DF(x) : \mathbb{R}^n \to \mathbb{R}^{n-m}$ is surjective, the Rank-Nullity Theorem tells us that

$$\dim(\ker DF(x)) = n - (n-m) = m$$

Note that for an m-dimensional manifold, the tangent space also has dimension m.

Our "official" definition of an (embedded) m-manifold may seem somewhat unintuitive. No worries, the following are equivalent formulations of manifolds:

Proposition 9.5. Let $M \subseteq \mathbb{R}^n$. For $x \in M$, the following are equivalent:

1. There exists an open neighborhood $\widehat{U} \subseteq \mathbb{R}^n$ of 0 and a C^k embedding $\varphi : \widehat{U} \to \mathbb{R}^n$ such that $\varphi(0) = x$ and $\varphi(\widehat{U} \cap \mathbb{R}^m) = M \cap \varphi(\widehat{U})$. To clarify, we identify $\mathbb{R}^m \cong \mathbb{R}^m \times \{0\}^{n-m}$.

2. There exists a neighborhood $U \subseteq \mathbb{R}^n$ of x and a C^k function $F : U \to \mathbb{R}^{n-m}$ such that $DF(y)$ is surjective for all $y \in U$ and $M \cap U = F^{-1}(0)$. In other words, M is a manifold.

3. There exist neighborhoods $U_1 \subseteq \mathbb{R}^m$ of $(x_1, \cdots, x_m)$ and $U_2 \subseteq \mathbb{R}^{n-m}$ of $(x_{m+1}, \ldots, x_n)$ and a C^k function $g : U_1 \to U_2$ such that $A \cap (U_1 \times U_2)$ contains the graph of g.

4. There exist open neighborhoods $U \subseteq \mathbb{R}^n$ of x and $V \subseteq \mathbb{R}^m$ of 0 and a C^k function $h : V \to U$ such that $h(0) = x$, $Dh(y)$ is injective for all $y \in U$, and $h(V) = M \cap U$.

Example 9.6.

1. $\mathbb{R}^n$ is an n-manifold.

2. An open subset $U \subseteq \mathbb{R}^n$ is also an n-manifold.

3. The n-sphere $S^n := \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : x_1^2 + \cdots + x_{n+1}^2 = 1\}$ is an n-manifold.

4. If $M \subseteq \mathbb{R}^n$ is an m-manifold, then the *tangent space of M* is the set

$$TM := \{(x, y) : x \in M, y \in T_xM\} \subseteq \mathbb{R}^n \times \mathbb{R}^n.$$

   We can check that TM is a $2m$-manifold.

5. The graph of $f(x) = |x|$ is not a manifold because the function is not differentiable at $(0, 0)$.

10 Introduction to Riemann Integration in $\mathbb{R}^n$

In this lecture, we introduce the Riemann integral on rectangles in $\mathbb{R}^n$.

Definition 10.1. Let $Q = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be a rectangle in $\mathbb{R}^n$. The *volume* of Q is given by

$$\mathrm{vol}(Q) = \prod_{i=1}^{n} (b_i - a_i).$$

The *width* of Q is defined as $w(Q) = \max(b_i - a_i)$. We refer to any of the intervals $[a_i, b_i]$ as a *component interval* of Q.

Definition 10.2. A *partition* P of a closed interval $[a, b]$ is a finite collection of points in $[a, b]$ such that $a, b \in P$. Arranging the elements of P in increasing order, we have

$$a = t_0 < t_1 < \cdots < t_k = b.$$

The intervals $[t_i, t_{i+1}]$ for $0 \leq i \leq k-1$ are known as *subintervals determined by* P. More generally, a partition $P = (P_1, \ldots, P_n)$ of a rectangle Q is a an n-tuple of partitions P_i for each component interval of Q. A *subrectangle* determined by P is a product of subintervals determined by each P_i.

Example 10.3. Consider the rectangle $[0, 1] \times [0, 1]$ in $\mathbb{R}^n$ and the partition $P = (P_1, P_2)$ where $P_1 = \{0, \frac{1}{2}, 1\}$ and $P_2 = \{0, \frac{1}{3}, \frac{2}{3}, 1\}$. A subrectangle determined by P would be $[\frac{1}{2}, 1] \times [\frac{1}{3}, \frac{2}{3}]$.

Definition 10.4. Let $f : Q \to \mathbb{R}$ be a bounded function defined on a rectangle Q and P a partition of Q. If R is a subrectangle determined by P, then we define

$$m_R(f) := \inf_{x \in R} f(x), \qquad M_R(f) := \sup_{x \in R} f(x).$$

The *lower sum of f determined by P* as

$$L(f, P) := \sum_R m_R(f) \cdot \mathrm{vol}(R),$$

and *the upper sum of f determined by P* as

$$U(f, P) := \sum_R M_R(f) \cdot \mathrm{vol}(R).$$

Definition 10.5. If $P = (P_1, \ldots, P_n)$ and $P' = (P'_1, \ldots, P'_n)$ are partititions of Q, then we say P' is a *refinement* of P if $P_i \subseteq P'_i$ for all $1 \leq i \leq n$.

Note that given partitions P and P' of Q, we can always form a partition that refines them by taking the union of the points in P and P'. This leads us to the following result:

Lemma 10.6. Suppose P is a partition of Q and P' refines P. If $f : Q \to \mathbb{R}$ is a bounded function on Q, then $L(f, P) \leq L(f, P')$ and $U(f, P') \leq U(f, P)$.

Proof. By definition of $L(f, P)$, we have

$$L(f, P) = \sum_R m_R(f) \cdot \mathrm{vol}(R).$$

If P' refines P, then note that P' partitions each rectangle R determined by P into finitely many subrectangles $R'_1, \dots, R'_n$ determined by P'. Moreover, note that $m_R(f) \le m_{R'_i}(f)$, which implies

$$m_R(f) \cdot \mathrm{vol}(R) = \sum_{R'_i \subset R} m_R(f) \cdot \mathrm{vol}(R'_i) \le \sum_{R'_i \subset R} m_{R'_i}(f) \cdot \mathrm{vol}(R'_i)$$

and thus

$$L(f, P) = \sum_R m_R(f) \cdot \mathrm{vol}(R) \le \sum_R \sum_{R'_i \subset R} m_{R'_i}(f) \cdot \mathrm{vol}(R'_i) = \sum_{R'} m_{R'}(f) \cdot \mathrm{vol}(R'_i) = L(f, P').$$

By a similar approach, we also have $U(f, P') \le U(f, P)$. $\square$

Definition 10.7. Let $f : Q \to \mathbb{R}$ be a bounded function. Then the *lower integral* of f over Q is defined as

$$\underline{\int_Q} f := \sup_P L(f, P),$$

and the *upper integral* as

$$\overline{\int_Q} f = \inf_P U(f, P).$$

We say f is *integrable* if $\overline{\int_Q} f = \underline{\int_Q} f$.

Remark. Note that the upper and lower integrals exist because

$$\sup_P L(f, P) \le \sup\{f(x) : x \in Q\} \cdot \mathrm{vol}(Q),$$
$$\inf_P U(f, P) \ge \inf\{f(x) : x \in Q\} \cdot \mathrm{vol}(Q).$$

Example 10.8. Let $c \in \mathbb{R}$ and consider the constant function $f : Q \to \mathbb{R}$ defined by $f(x) = c$. Notice that $L(f, P) = U(f, P) = c \cdot \mathrm{vol}(Q)$ for any partition P of Q, which implies

$$\overline{\int_Q} f = \underline{\int_Q} f$$

and f is integrable.

Lemma 10.9. Suppose $f : Q \to \mathbb{R}$ is a bounded function on a rectangle $Q \subseteq \mathbb{R}^n$, and P and P' are partitions of Q. Then

$$L(f, P) \le U(f, P').$$

Proof. Let P'' be the union of points in P and P'. We have that P'' refines P and P', so by the previous lemma, we have

$$L(f, P) \le L(f, P''), \qquad U(f, P'') \le U(f, P').$$

Note that $L(f, P'') \le U(f, P'')$, so we have $L(f, P) \le L(f, P'') \le U(f, P'') \le U(f, P')$. $\square$

The definition of the integral is quite cumbersome if we want to determine which functions $f : Q \to \mathbb{R}$ are (Riemann) integrable. In the next few lectures, we will explore an alternate criterion for integrability. In the meantime, we will prove the following statement:

Theorem 10.10. Suppose $f : Q \to \mathbb{R}$ is bounded. Then f is integrable if and only if for all $\varepsilon > 0$, there exists a partition P of Q such that $U(f, P) - L(f, P) < \varepsilon$.

Proof. First, suppose that for all $\varepsilon > 0$, there exists a partition P such that $U(f, P) - L(f, P) < \varepsilon$. Observe that

$$\overline{\int_Q} f \leq U(f, P), \qquad \underline{\int_Q} f \geq U(f, P),$$

and if $U(f, P) - L(f, P) < \varepsilon$, then we have $\overline{\int_Q} f - \underline{\int_Q} f < \varepsilon$ and thus $\overline{\int_Q} f = \underline{\int_Q} f$.

Now suppose instead that f is integrable. Then

$$\underline{\int_Q} f = \sup_P L(f, P) = \inf_P U(f, P) = \overline{\int_Q} f,$$

which means there exist partitions P and P' of Q such that

$$I - L(f, P) < \frac{\varepsilon}{2}, \qquad U(f, P') - I < \frac{\varepsilon}{2}.$$

Now let P'' be a refinement of P and P'. Then $I - L(f, P'') < \frac{\varepsilon}{2}$ and $U(f, P'') - I < \frac{\varepsilon}{2}$, which implies

$$U(f, P'') - L(f, P'') = (U(f, P'') - I) + (I - L(f, P'')) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} < \varepsilon. \qquad \square$$

11 Integration and Measure Zero

In this lecture, we begin to formulate an alternative criterion for Riemann integrability. Intuitively, we can imagine that a function is integrable if there aren't too many "holes" on the domain. In other words, we want the set of discontinuities in the domain to be "small enough" or perhaps "negligible." We characterize such sets as having *measure zero*.

Definition 11.1. A subset $A \subseteq \mathbb{R}^n$ has *measure zero* if for all $\varepsilon > 0$, there exists a countable cover of A by $\{Q_i\}_{i=1}^{\infty}$ such that

$$\sum_{i=1}^{\infty} \mathrm{vol}(Q_i) < \varepsilon.$$

Proposition 11.2. If $\{A_i\}_{i=1}^{\infty}$ is a collection of sets in $\mathbb{R}^n$ with measure zero, then their union

$$A := \bigcup_{i=1}^{\infty} A_i$$

has measure zero.

Proof. Let $\varepsilon > 0$. For each $i \geq 1$, let $\{Q_{ij}\}_{j=1}^{\infty}$ be a countable collection of rectangles covering A_i such that

$$\sum_{j=1}^{\infty} \operatorname{vol}(Q_{ij}) < \frac{\varepsilon}{2^i}.$$

Note that $\bigcup_{i,j} Q_{ij}$ covers A, so

$$\sum_{i,j} \operatorname{vol}(Q_{ij}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \operatorname{vol}(Q_{ij}) < \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon.$$

Thus, A has measure zero. □

Proposition 11.3. For $m < n$, the linear subspace $\mathbb{R}^m \times \{0\}^{n-m} \subseteq \mathbb{R}^n$ has measure zero.

Proof. Consider the subspace $Z := \mathbb{Q}^m \times \{0\}^{n-m} \subseteq \mathbb{R}^m \times \{0\}^{n-m}$. Note that Z is countable, so let $Z = \{z_1, z_2, \dots\}$. Suppose that $z_i = (x_{i1}, \dots, x_{im}, 0, 0, \dots, 0)$, and consider rectangle

$$R_i = \left[x_{i1} - \frac{1}{2}, x_{i1} + \frac{1}{2}\right] \times \cdots \times \left[x_{im} - \frac{1}{2}, x_{im} + \frac{1}{2}\right] \times \left[-\frac{\varepsilon^{1/(n-m)}}{2^{(i+1)/(n-m)}}, \frac{\varepsilon^{1/(n-m)}}{2^{(i+1)/(n-m)}}\right]^{n-m},$$

which has volume $\operatorname{vol}(R_i) = \dfrac{\varepsilon}{2^i}$. Note that $\{R_i\}_{i=1}^{\infty}$ covers $\mathbb{R}^m \times \{0\}^{n-m}$, and observe that

$$\sum_{i=1}^{\infty} \operatorname{vol}(R_i) = \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon.$$

Thus, $\mathbb{R}^m \times \{0\}^{n-m}$ has measure zero. □

Remark. We were able to shrink the volume of each R_i because the final $n - m$ coordinates were all zero. And so, we require $m < n$.

Example 11.4. The following are examples of measure zero sets:

- A point in $\mathbb{R}^n$ has measure zero.
- Any countable subset of $\mathbb{R}$—such as $\mathbb{Q}$ or $\mathbb{Z}$—has measure zero.

Proposition 11.5. A set $A \subseteq \mathbb{R}^n$ has measure zero if and only if there exists a countable collection $\{Q_i\}_{i=1}^{\infty}$ of rectangles such that $A \subset \displaystyle\bigcup_{i=1}^{\infty} \operatorname{int}(Q_i)$ and $\displaystyle\sum_{i=1}^{\infty} \operatorname{vol}(Q_i) < \varepsilon$.

Proof. Suppose that A has measure zero. Then there exists a countable collection $\{Q_i\}_{i=1}^{\infty}$ of rectangles such that $A \subset \bigcup_{i=1}^{\infty} Q_i$ and $\sum_{i=1}^{\infty} \operatorname{vol}(Q_i) < \varepsilon/2$. For each Q_i, let Q_i' be a rectangle such that $Q_i \subset \operatorname{int}(Q_i')$ and $\operatorname{vol}(Q_i') < 2 \cdot \operatorname{vol}(Q_i)$. Then $A \subseteq \bigcup_{i=1}^{\infty} Q_i \subseteq \bigcup_{i=1}^{\infty} \operatorname{int}(Q_i')$ and

$$\sum_{i=1}^{\infty} \operatorname{vol}(Q_i') < 2 \cdot \sum_{i=1}^{\infty} \operatorname{vol}(Q_i) < \varepsilon.$$

The converse follows immediately. □

Proposition 11.6. Let $Q \subseteq \mathbb{R}^n$ be a rectangle and $\{Q_i\}_{i=1}^n$ a set of rectangles covering Q. Then

$$\mathrm{vol}(Q) \leq \sum_{i=1}^{n} \mathrm{vol}(Q_i)$$

Proof. Let Q' be a rectangle containing $\bigcup_{i=1}^n Q_i$. Using the component intervals of Q_i and Q, we construct a partition P of Q' whereby Q and each Q_i are unions of rectangles determined by P. In other words, P partitions Q and each Q_i. As such, we have

$$\mathrm{vol}(Q) = \sum_{R \subseteq Q} \mathrm{vol}(R), \qquad \mathrm{vol}(Q_i) = \sum_{R \subseteq Q_i} \mathrm{vol}(R),$$

where R is determined by P. Note that if $R \subseteq Q$, then R is a subset of some Q_i. And so, we have

$$\sum_{i=1}^{n} \mathrm{vol}(Q_i) = \sum_{R \subseteq Q_i} \mathrm{vol}(R) \geq \sum_{R \subseteq Q} \mathrm{vol}(R) = \mathrm{vol}(Q).$$

□

Corollary 11.7. If Q is a rectangle in $\mathbb{R}^n$, then Q does not have measure zero.

Proof. Suppose instead that $Q \subseteq \mathbb{R}^n$ has measure zero. Then there exists a countable collection $\{Q_i\}_{i=1}^\infty$ such that $Q \subseteq \bigcup_{i=1}^\infty \mathrm{int}(Q_i)$ and

$$\sum_{i=1}^{\infty} \mathrm{vol}(Q_i) < \frac{\mathrm{vol}(Q)}{2}.$$

Because Q is compact (closed and bounded), a finite subcover $\{Q_i\}_{i=1}^n$ of Q exists. In other words, we have $Q \subseteq \bigcup_{i=1}^n \mathrm{int}(Q_i)$, which implies

$$\mathrm{vol}(Q) \leq \mathrm{vol}(Q_1) + \cdots + \mathrm{vol}(Q_n) < \frac{\mathrm{vol}(Q)}{2}$$

and we arrive at a contradiction. □

12 Integration and Measure Zero (continued)

In this lecture, we will prove *Lebesgue's Criterion for Riemann Integrability*:

Theorem 12.1. Let $f : Q \to \mathbb{R}$ be a bounded function on a rectangle $Q \subseteq \mathbb{R}^n$. Let $D \subseteq Q$ be the set of points where f is not continuous. Then f is integrable if and only if D has measure zero.

Recall that

- the countable union of measure zero sets has measure zero,
- the subset of a measure zero set is measure zero,
- and if $Q \subseteq \mathbb{R}^n$ is a rectangle, then the boundary of Q has measure zero (each "face" of the rectangle has a lower dimension than the entire rectangle).

We will formulate continuity in a different way that will be convenient for proving our theorem.

Definition 12.2. Let $f : Q \to \mathbb{R}$ be a bounded function and $a \in Q$. For any $\delta > 0$, let

$$A_\delta(a) := \{f(x) : x \in Q, \|x - a\| < \delta\},$$

and define $M_\delta(f; a) = \sup A_\delta(a)$ and $m_\delta(f; a) = \inf A_\delta(a)$. We define the *oscillation* by

$$\operatorname{osc}(f; a) = \inf_{\delta > 0} \left(M_\delta(f; a) - m_\delta(f; a)\right).$$

In other words, $A_\delta(a)$ is the image $f(B(a, \delta))$, and the oscillation roughly measures how "wide" the image is. We can imagine that f is continuous at $a \in Q$ if $f(B(a, \delta))$ shrinks to a single point when $\delta \to 0$, and this is reflected by $\operatorname{osc}(f; a) = 0$. This intuition is indeed true, and we have the following:

Proposition 12.3. A function $f : Q \to \mathbb{R}$ is continuous at $a \in Q$ if and only if $\operatorname{osc}(f; a) = 0$.

Proof. If $f : Q \to \mathbb{R}$ is continuous at $a \in Q$, then for all $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon/2$ for all $x \in Q$ where $\|x - a\| < \delta$. And so, we have

$$f(a) - \frac{\varepsilon}{2} \le m_\delta(f : a) \le M_\delta(f; a) \le f(a) + \frac{\varepsilon}{2},$$

which implies $M_\delta(f; a) - m_\delta(f; a) \le \varepsilon$ and thus $\operatorname{osc}(f; a) = 0$.

If $\operatorname{osc}(f; a) = 0$, then for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$M_\delta(f; a) - m_\delta(f; a) < \varepsilon.$$

If $x \in Q$ and $\|x - a\| < \delta$, then $f(x)$ and $f(a)$ lie between $M_\delta(f; a)$ and $m_\delta(f; a)$. In other words, $|f(x) - f(a)| < \varepsilon$, which shows f is continuous at a. $\square$

Using this fact, we prove Theorem 12.1.

Proof. First, suppose $f : Q \to \mathbb{R}$ is integrable. We want to show that the set of discontinuities $D \subseteq Q$ has measure zero, so define

$$D_m := \left\{d \in D : \operatorname{osc}(f; d) \ge \frac{1}{m}\right\}$$

for $m \in \mathbb{N}$. If d is a point of discontinuity, then $\operatorname{osc}(f; d) > 0$. Now observe that

$$D = \bigcup_{m=1}^{\infty} D_m,$$

and note that it suffices to show D_m has measure zero. Let $m \in \mathbb{N}$ and $\varepsilon > 0$. If f is integrable, then there exists a partition P of Q such that $U(f, P) - L(f, P) < \varepsilon/m$. Let $D_m' \subset D_m$ denote the set of points which lie on the boundary of some subrectangle determined by P, and let $D_m'' := D_m \setminus D_m'$. To clarify, the boundary of a rectangle is its "faces," which are measure zero sets according to Proposition 11.3. Because D_m' is a subset of the boundaries of subrectangles determined by P, we have that D_m' has measure zero. And so, it remains to show that D_m'' has measure zero.

Let $R_1, \ldots, R_k$ be subrectangles of Q determined by P such that $D_m'' \subseteq \bigcup_{i=1}^k R_i$. Notice that each R_i contains some $a \in D_m''$ in its interior, and because $\operatorname{int}(R_i)$ is open, there exists $\delta > 0$ such that $B(a, \delta) \subseteq R_i$. Now observe that

$$\frac{1}{m} \le \operatorname{osc}(f; a) \le M_\delta(f; a) - m_\delta(f; a) \le M_{R_i}(f) - m_{R_i}(f),$$

and in addition

$$\sum_{i=1}^{k}(M_{R_i}(f)-m_{R_i}(f))\cdot \mathrm{vol}(R_i)\le \underbrace{\sum_{R}(M_R(f)-m_R(f))\cdot \mathrm{vol}(R)}_{U(f,P)-L(f,P)}<\frac{\varepsilon}{m}.$$

As such, we have

$$\sum_{i=1}^{k}\frac{1}{m}\mathrm{vol}(R_i)\le \sum_{i=1}^{k}(M_{R_i}(f)-m_{R_i}(f))<\frac{\varepsilon}{m},$$

and it follows that

$$\sum_{i=1}^{k}\mathrm{vol}(R_i)<\varepsilon.$$

And so, if $D''_m\subseteq\bigcup_{i=1}^{k}R_i$, then D''_m has measure zero.

Now suppose instead that D has measure zero. Given $\varepsilon>0$, we want to find a partititon P such that $U(f,P)-L(f,P)<\varepsilon$. If D has measure zero, then for all $\varepsilon'>0$, there exists a countable collection of rectangles $\{Q_i\}_{i=1}^{\infty}$ such that $\sum_{i=1}^{\infty}\mathrm{vol}(Q_i)<\varepsilon'$ and $D\subseteq\bigcup_{i=1}^{\infty}\mathrm{int}(Q_i)$. If $a\in Q\setminus D$—in other words, f is continuous at a—then there exists a rectangle $Q_a\subseteq Q$ such that $a\in\mathrm{int}(Q_a)$ and $|f(x)-f(a)|<\varepsilon'$ for all $x\in Q_a\cap Q$. Now observe that

$$\{\mathrm{int}(Q_i)\}_{i=1}^{\infty}\cup\{\mathrm{int}(Q_a)\}_{a\in Q\setminus D}$$

is an open cover of Q. Because Q is compact, there exists a finite subcover

$$\{\mathrm{int}(Q_i)\}_{i=1}^{N}\cup\{\mathrm{int}(Q_{a_j})\}_{j=1}^{M}.$$

Define $R_i:=Q_i\cap Q$ for $i=1,\dots,N$ and $R'_j:=Q'_{a_j}\cap Q$ for $j=1,\dots,M$. Let P be the partition of Q given by the component intervals of R_i and R'_j. Notice that each R_i and R'_j is a union of subrectangles determined by P, and each subrectangle determined by P lies in some R_i or R'_j.

Let $\mathcal{S}$ be the family of subrectangles determined by P that is a subset of some R_i, and let $\mathcal{S}'$ be the family of subrectangles determined by P that lie in some R'_j. We have

$$\sum_{S\in\mathcal{S}}(M_S(f)-m_S(f))\cdot\mathrm{vol}(S)\le\sum_{S\in\mathcal{S}}M\cdot\mathrm{vol}(S),$$

where $M=\sup\limits_{x\in Q}f(x)-\inf\limits_{x\in Q}f(x)$. Notice that

$$\sum_{S\in\mathbb{S}}\mathrm{vol}(S)\le\sum_{i=1}^{\infty}\mathrm{vol}(Q_i)<\varepsilon',$$

so we have

$$\sum_{S\in\mathbb{S}}(M_S(f)-m_S(f))\cdot\mathrm{vol}(S)\le\sum_{S\in\mathcal{S}}M\cdot\mathrm{vol}(S)<\varepsilon' M.$$

Now observe that if $x,y\in Q_a$, then $|f(x)-f(y)|\le|f(x)-f(a)|+|f(y)-f(a)|<2\varepsilon'$. And so, we have $M_s(f)-m_S(f)<2\varepsilon'$ and

$$\sum_{S\in\mathcal{S}'}(M_S(f)-m_S(f))\cdot\mathrm{vol}(S)<2\varepsilon'\cdot\sum_{S\in\mathcal{S}'}\mathrm{vol}(S)$$

Note that

$$\sum_{S \in \mathbb{S}'} \operatorname{vol}(S) \leq \operatorname{vol}(Q),$$

and altogether, we have

$$\sum_{S \in \mathcal{S}} (M_S(f) - m_S(f)) \cdot \operatorname{vol}(S) + \sum_{S \in \mathcal{S}'} (M_S(f) - m_S(f)) \cdot \operatorname{vol}(S) \leq \varepsilon' M + 2\varepsilon' \cdot \sum_{S \in \mathcal{S}'} \operatorname{vol}(S),$$

$$U(f, P) - L(f, P) \leq \varepsilon' M + 2\varepsilon' \cdot \operatorname{vol}(Q).$$

Setting $\varepsilon' := \dfrac{\varepsilon}{M + 2 \cdot \operatorname{vol}(Q)}$, we conclude that f is indeed integrable. □

13 Applications of Lebesgue's Criterion for Riemann Integrability

In this lecture, we will apply our result from the previous lecture to nonnegative functions.

Theorem 13.1. Suppose $f : Q \to \mathbb{R}$ is an integrable function on a rectangle $Q \subseteq \mathbb{R}^n$.

- If f is nonzero on a measure zero set, then $\int_Q f = 0$.
- If f is nonnegative and $\int_Q f = 0$, then f is zero except possibly on a measure zero set.

Proof. Let $E := \{x \in Q : f(x) \neq 0\}$, and suppose that E has measure zero. Let P be a partition of Q, and note that a subsectange $R \subset Q$ determined by P does not have measure zero (Corollary 11.7). If E has measure zero, then $R \not\subseteq E$. Because R does not have measure zero, there exists $x \in R$ such that $f(x) = 0$. Observe then that $m_R(f) \leq 0 \leq M_R(f)$, and consequently

$$\underline{\int_Q} f \leq 0 \leq \overline{\int_Q} f.$$

If f is integrable, then we have that $\underline{\int_Q} f = \overline{\int_Q} f = 0$.

Now suppose that f is nonnegative and $\int_Q f = 0$. If f is integrable, then the set of discontinuities $D \subset Q$ has measure zero. Once again, let $E := \{x \in Q : f(x) \neq 0\}$. We want to show E has measure zero and to do so, we show $E \subseteq D$. Suppose instead that there exists $x \in Q$ such that $f(x) \neq 0$ and f is continuous at x. Then let $\varepsilon = f(x) > 0$ and by continuity of f, there exists $\delta > 0$ such that $f(x) - f(y) < \varepsilon/2$ for all $y \in Q$ such that $\|x - y\| < \delta$. Now let P be a partition of Q such that the distance between any two points in a subrectangle determined by P is less than δ. If R is a subrectangle determined by P containing x and $y \in R$, then we have $f(x) - f(y) < \varepsilon/2$, which implies $\varepsilon/2 < f(y)$ and thus $m_R(f) \geq \varepsilon/2$. For every other subrectangle R', we have $m_R(f) \geq 0$ because f is nonnegative, and as such $L(f, P) \geq \frac{\varepsilon}{2} \cdot \operatorname{vol}(R) > 0$. However, this contradicts the fact that $L(f, P) \leq \int_Q f = 0$. □

Example 13.2. Let $f : Q \to \mathbb{R}$ be a function on a rectangle $Q \subseteq R$ defined by

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

Note that f is not integrable since it's discontinuous everywhere. However, $f(x) = 0$ except for a set of measure zero.

Lebesgue's Criterion is a great way for us to determine whether a function on rectangles is integrable. However, we won't stop here. We want to develop a notion of integration on arbitrary sets in $\mathbb{R}^n$ and on manifolds. Moreover, Lebesgue's Criterion does not really tell us *how* we can integrate functions in higher dimensions. For the latter half of the course, we will generalize the notion of integration to various contexts and extend the Fundamental Theorem of Calculus from single-variable calculus.

Theorem (Fundamental Theorem of Calculus).

1. Suppose $f : [a, b] \to \mathbb{R}$ is continuous. Let

$$F(x) = \int_a^x f$$

Then F is a C^1 function on $[a, b]$ and $F'(x) = f(x)$ on (a, b).

2. If $f : [a, b] \to \mathbb{R}$ is continuous and $g : [a, b] \to \mathbb{R}$ is continuous with $g'(x) = f(x)$ on (a, b), then

$$\int_a^b f = g(b) - g(a)$$

For the rest of this lecture, we will prove Fubini's Theorem, which tells us that we can integrate over a region simply by integrating over the intervals that comprise the region in a step-by-step manner. In other words, for a continuous function $f : [a, b] \times [c, d] \to \mathbb{R}$, we can imagine that

$$\int_{[a,b]\times[c,d]} f = \int_c^d g(y),$$

where $g(y) := \int_a^b f(x, y)$ (integrate with respect to x). Although seemingly intuitive and perhaps obvious, such a maneuver involves the interchanging of limits and can be problematic at times.

Theorem 13.3 (Fubini's Theorem). Suppose $f : Q \to \mathbb{R}$ is a bounded function defined on a rectangle $Q := A \times B \subseteq \mathbb{R}^{n+m}$. For $x \in Q$, define $g(x) := \underline{\int_{y\in B}} f(x, y)$ and $h(x) := \overline{\int_{y\in B}} f(x, y)$. If f is integrable on Q, then g and h are both integrable on A and

$$\int_Q f = \int_A g = \int_A h.$$

Proof. Let $P = (P_1, P_2)$ be a partition of Q where P_1 and P_2 partition A and B, respectively. Note that every subrectangle of Q determined by P is given by $R_1 \times R_2$, where R_1 and R_2 are subrectangles determined by P_1 and P_2, respectively.

First, we will show that $L(f, P) \le L(g, P_1)$. Let $R = R_1 \times R_2$ be a subrectangle determined by P. Then for any $x_0 \in R_1$, we have $m_R(f) \le m_{R_2}(f(x_0, y))$ and

$$\begin{aligned}\sum_{R_2} m_{R_1\times R_2}(f) \cdot \mathrm{vol}(R_2) &\le \sum_{R_2} m_{R_2}(f(x_0, y)) \cdot \mathrm{vol}(R_2),\\ &= L(f(x_0, y), P_2),\\ &\le g(x_0).\end{aligned}$$

We have $\sum_{R_2} m_{R_1\times R_2}(f)(R_2) \le g(x_0)$ for all $x_0 \in R_1$, which means

$$\sum_{R_2} m_{R_1\times R_2}(f) \cdot \mathrm{vol}(R_2) \le m_{R_1}(g).$$

And so, observe that

$$\sum_{R_1}\sum_{R_2} m_{R_1\times R_2}(f)\cdot \mathrm{vol}(R_2)\cdot \mathrm{vol}(R_1) \le \sum_{R_1} m_{R_1}(g)\cdot \mathrm{vol}(R_1),$$

$$\sum_{R\subseteq Q} m_R(f)\cdot \mathrm{vol}(R) \le L(g,P_1),$$

$$L(f,P) \le L(g,P_1).$$

By the same argument we likewise have $U(h,P_1)\le U(f,P)$, which implies

$$L(f,P)\le L(g,P_1)\le L(h,P_1)\le U(h,P_1)\le U(f,P).$$

Similarly, we have

$$L(f,P)\le L(g,P_1)\le U(g,P_1)\le U(h,P_1)\le U(f,P).$$

If f is integrable, then for all $\varepsilon>0$, there exists a partition P such that $U(f,P)-L(f,P)<\varepsilon$. By our inequalities above, we have $U(h,P_1)-L(h,P_1)<\varepsilon$ and $U(g,P_1)-L(g,P_1)<\varepsilon$. And so, g and h are integrable and we find that $\int_Q f=\int_A g=\int_A h$. □

14 Integration Beyond Rectangles

In this lecture, we start extending integration on rectangles to more general sets. We will then prove some properties that allow us to compute integrals.

Definition 14.1. Let $S\subseteq\mathbb{R}^n$ be a bounded subset and $f:S\to\mathbb{R}$ a bounded function. If $Q\subseteq\mathbb{R}^n$ is a rectange containing S, then define $f_S:Q\to\mathbb{R}$ by

$$f_S(x):=\begin{cases} f(x) & \text{if } x\in S,\\ 0 & \text{if } x\notin S.\end{cases}$$

We define the *integral on S* as

$$\int_S f := \int_Q f_S,$$

given that $\int_Q f_S$ exists.

Note that f_S is not necessarily continuous on Q. We will show in the following lemma that $\int_S f$ does not depend on our choice of a rectangle Q.

Lemma 14.2. Suppose Q and Q' are rectangles in $\mathbb{R}^n$ and $f:\mathbb{R}^n\to\mathbb{R}$ be a bounded function such that $f(x)=0$ for $x\notin Q\cap Q'$. Then f is integrable over Q if and only if it is integrable over Q' and

$$\int_Q f=\int_{Q'} f.$$

Proof. First, note that it suffices to consider without loss of generality that $Q\subseteq Q'$ since we can apply our lemma to $Q\cap Q'$ and get $\int_Q f=\int_{Q'} f$. Now define $E:=\{x\in Q : f \text{ is not continuous on } Q\}$ and $F:=\{x\in Q' : f \text{ not continuous on } Q'\}$. Because $Q\subseteq Q'$ and $f(x)=0$ for $x\in Q\setminus Q'$, we have $E\subseteq F$ and $F\subseteq E\cup\partial Q$. Note that ∂Q has measure zero, so E has measure zero if and only if F has measure zero. Therefore, f is integrable on Q if and only if it is integrable on Q'.

Now let P' be a partition of Q' such that $Q \subseteq Q'$ is a union of subrectangles determined by P'. As such, P' also determines a partition of Q. Notice that if R is a subrectangle determined by P', then either $R \subseteq Q$ or $R \not\subseteq Q$, whereby the latter implies $m_R(f) \leq 0 \leq M_R(f)$. And so, we have $L(f, P') \leq L(f, P) \leq \int_Q f$, which implies $\int_{Q'} f \leq \int_Q f$. Similarly, we have $\int_Q f \leq \int_{Q'} f$ and thus

$$\int_{Q'} f = \int_Q f. \qquad \square$$

Lemma 14.3. Suppose $S \subseteq \mathbb{R}^n$ and $f, g : S \to \mathbb{R}$ are bounded functions. Let $F(x) = \max\{f(x), g(x)\}$ and $G(x) = \min\{f(x), g(x)\}$.

1. If f and g are continuous at x_0, then so are F and G.
2. If f and g are integrable on S, then so are F and G.

Proof.

1. Note that max and min are continuous functions, so if f and g are continuous, then it follows that F and G are continuous.
2. Let E and E' be sets discontinuities of f and g, respectively. If f and g are integrable on S, then E and E' have measure zero, which implies $E \cup E'$ has measure zero. If F or G is not continuous at $x \in S$, then either $x \in E$ or $x \in E'$. As such, F and G are integrable. $\square$

Theorem 14.4. Let $S \subseteq \mathbb{R}^n$ be a bounded region and $f, g : S \to \mathbb{R}$ functions on S.

1. If f and g are integrable on S and $a, b \in \mathbb{R}$, then $af + bg$ is also integrable on S and
$$\int_S (af + bg) = a \int_S f + b \int_S g.$$
2. If f and g are integrable on S and $f(x) \leq g(x)$ for all $x \in S$, then
$$\int_S f \leq \int_S g.$$
3. If $T \subseteq S$ and f is nonnegative and integrable on both S and T, then
$$\int_T f \leq \int_S f.$$
4. If $S = S_1 \cup S_2$ and f is integrable over S_1 and S_2, then f is integrable over S and $S_1 \cap S_2$ and
$$\int_S f = \int_{S_1} f + \int_{S_2} f - \int_{S_1 \cap S_2} f.$$

Proof.

1. Notice that $(af + bg)_S = af_S + bg_S$, so for the sake of simplicity assume that S is also a rectangle. Since f and g are integrable on S, we know that their respective sets of discontinuities D and E have measure zero. Note that the union $D \cup E$ is also the set of discontinuities of $af + bg$ and $D \cup E$ has measure zero. As such, we have that $af + bg$ is integrable. To show linearity of the

integrals, we will proceed by the following steps:

- *Step 1:* If $a > 0$, then $\int_S af = a\int_S f$.
Let P be a partition of S and R is a subrectangle determined by P. Observe that $m_R(af) = a \cdot m_R(f)$ and $M_R(af) = a \cdot M_R(f)$ by properties of the supremum and infimum, which imply $U(af, P) = a \cdot U(f, P)$ and $L(af, P) = a \cdot L(f, P)$. And so, we have
$$\underline{\int_S} af = \sup_P L(af, P) = \sup_P (a \cdot L(f, P)) = a \cdot \sup_P L(f, P) = a\underline{\int_S} f,$$
$$\overline{\int_S} af = \inf_P U(af, P) = \inf_P (a \cdot U(f, P)) = a \cdot \inf_P U(f, P) = a\overline{\int_S} f,$$
which implies
$$\int_S af = a\int_S f.$$
- *Step 2:* $\int_S (f+g) = \int_S f + \int_S g$
If P is a partition of S, then observe that
$$L(f, P) + L(g, P) \le L(f+g, P) \le U(f+g, P) \le U(f, P) + U(g, P).$$
Notice that
$$L(f, P) + L(g, P) \le \int_S f + \int_S g \le U(f, P) + U(g, P),$$
$$L(f+g, P) \le \int_S (f+g) \le U(f+g, P),$$
which implies
$$\int_S (f+g) = \int_S f + \int_S g$$
when we consider the limits of $L(f+g, P)$, $U(f+g, P)$, $L(f, P)$, $U(f, P)$, $L(g, p)$, and $U(g, P)$.
- *Step 3:* $\int_S (-f) = -\int_S f$
If P is a partition of S, then observe that $L(-f, P) = -U(-f, P)$ and $U(-f, P) = -L(f, P)$ by properties of the supremum and infimum. And so, we find that
$$\int_Q -f = -\int_Q f.$$

2. Once again, it suffices to assume S is a rectangle. If P is a partition of Q, then observe that $L(f, P) \le L(g, P)$ and $U(f, P) \le U(g, P)$. Considering the limiting cases, we have
$$\int_Q f \le \int_Q g.$$
3. We have $f_T \le f_S$, which implies $\int_T f \le \int_S f$.
4. First, suppose f is nonnegative and Q a rectangle containing S. Because f is nonnegative, observe that
$$f_S(x) = \max(f_{S_1}(x), f_{S_2}(x)),$$
$$f_{S_1\cap S_2}(x) = \underbrace{\min(f_{S_1}(x), f_{S_2}(x))}_{0 \text{ outside } S_1\cap S_2}$$

for all $x \in Q$, and we have $f_S = f_{S_1} + f_{S_2} - f_{S_1 \cap S_2}$. By linearity of integrals, we have

$$\int_S f = \int_{S_1} f + \int_{S_2} f - \int_{S_1 \cap S_2} f.$$

Now for the general case, let

$$\begin{aligned} f_+(x) &= \max(f(x), 0) \\ f_-(x) &= \max(-f(x), 0), \end{aligned}$$

and note that $f = f_+ - f_-$ and f_+ and f_- are nonnegative. By Lemma 14.2, f_+ and f_- are integrable on S_1, S_2, S, and $S_1 \cap S_2$. Because f_+ and f_- are nonnegative, we have from our first case that

$$\begin{aligned} f_{+,S} &= f_{+,S_1} + f_{+,S_2} - f_{+,S_1 \cap S_2}, \\ f_{-,S} &= f_{-,S_1} + f_{-,S_2} - f_{-,S_1 \cap S_2}, \end{aligned}$$

which yields

$$\begin{aligned} \int_S f_+ &= \int_{S_1} f_+ + \int_{S_2} f_+ - \int_{S_1 \cap S_2} f_+, \\ \int_S f_- &= \int_{S_1} f_- + \int_{S_2} f_- - \int_{S_1 \cap S_2} f_-. \end{aligned}$$

By linearity of integrals, we have

$$\begin{aligned} \int_S (f_+ - f_-) &= \int_{S_1} (f_+ - f_-) + \int_{S_2} (f_+ - f_-) - \int_{S_1 \cap S_2} (f_+ - f_-), \\ \int_S f &= \int_{S_1} f + \int_{S_2} f - \int_{S_1 \cap S_2} f. \end{aligned}$$

□

15 Integrability Beyond Rectangles

In the last lecture, we extended our notion of integrability to arbitrary subsets of $\mathbb{R}^n$. In this lecture, we will explore when a function $f : S \to \mathbb{R}$ on some subset $S \subseteq \mathbb{R}^n$ is integrable. To begin, consider the following lemma:

Lemma 15.1. Let $f : S \to \mathbb{R}$ be a continuous function on a bounded subset $S \subseteq \mathbb{R}^n$. The function $f_S : \mathbb{R}^n \to \mathbb{R}$ is continuous at $x_0 \in \mathbb{R}^n$ if and only if either $x_0 \notin \partial S$ or $\lim\limits_{x \to x_0} f(x) = 0$ for $x \in S$.

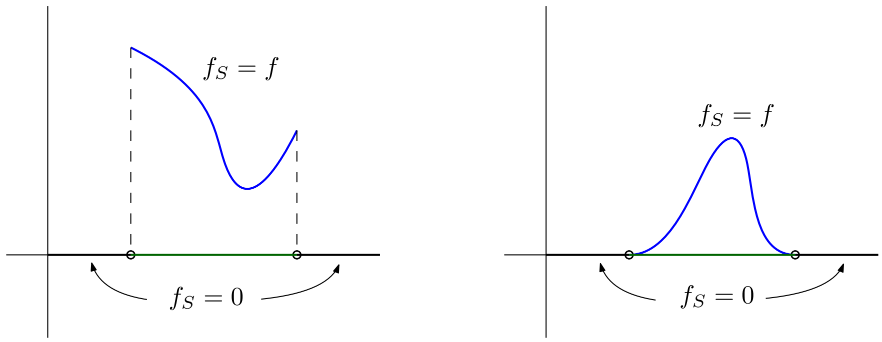
Proof. As a refresher, recall that $f_S : \mathbb{R}^n \to \mathbb{R}$ is the function defined by

$$f_S(x) = \begin{cases} f(x) & \text{if } x \in S, \\ 0 & \text{if } x \notin S. \end{cases}$$

First, note that $\mathbb{R}^n = \operatorname{int}(S) \cup \partial S \cup \operatorname{ext}(S)$, where $\operatorname{int}(S)$ and $\operatorname{ext}(S)$ are open. Since f is continuous on S, $f_S|_{\operatorname{int}(S)}$ is continuous. By definition of f_S, we have that $f_S|_{\operatorname{ext}(S)} \equiv 0$, which is continuous as well. And so, if $x_0 \in \operatorname{int}(S) \cup \operatorname{ext}(S)$, then f_S is continuous at x_0.

Now suppose $x_0 \in \partial S$. If f_S is continuous at x_0, then for all $\varepsilon > 0$, there exists $\delta > 0$ such that $|f_S(x_0) - f_S(y)| < \varepsilon$ for all $y \in B(x_0, \delta)$. Because $B(x_0, \delta)$ is a neighborhood of x_0 and $x_0 \in \partial S$, there exists $y \in B(x_0, \delta) \cap \operatorname{ext}(S)$. Notice then that $f_S(y) = 0$, so we have $|f_S(x_0) - f_S(y)| = |f_S(x_0)| < \varepsilon$, and for this to be true, we must have $f_S(x_0) = 0$. And so, if f_S is continuous at x_0, then we must have $\lim_{x\to x_0} f(x) = 0$ for $x \in S$.

We now prove the converse, so suppose $\lim_{x\to x_0} f(x) = 0$. Then for $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x)| < \varepsilon$ for all $x \in B(x_0, \delta) \cap S$. Consequently, we have $|f_S(x)| < \varepsilon$ for arbitrary $x \in B(x_0, \delta)$, which implies $\lim_{x\to x_0} f_S(x) = 0$. To see that f_S is continuous at x_0, we neec $f_S(x_0) = 0$. If $x_0 \in S$, then the continuity of f implies $f(x_0) = 0$ and thus $f_S(x_0) = 0$. If instead $x_0 \notin S$, then $f_S(x_0) = 0$ by definition of f_S. And so, we have that $\lim_{x\to x_0} f_S(x_0) = 0$ and $f_S(x_0) = 0$, which implies f_S is continuous at x_0. □



Intuitively, Lemma 15.1 tells us that for a continuous map $f : S \to \mathbb{R}$, its extension f_S is continuous everywhere in $\mathbb{R}^n$ if and only if f converges to 0 (from the interior of S) at the boundary of S. The left image above is the case where f_S is not continuous everywhere on $\mathbb{R}$ because of the jump at the boundary of the graph of f, whereas the right image above is the case where f tapers down to 0 on its boundary.

We can use Lemma 15.1 to also deduce some facts about integrability, namely that a continuous function $f : S \to \mathbb{R}$ is integrable as long as the jumps between the interior and exterior of S are negligible.

Corollary 15.2. If $f : S \to \mathbb{R}$ is a bounded continuous function on a bounded subset $S \subseteq \mathbb{R}^n$, then f is integrable over S if and only if

$$D_\partial := \left\{ x_0 \in \partial S : \lim_{x\to x_0} f(x) \neq 0 \right\}$$

has measure zero.

Proof. If f is integrable, then $\int_S f = \int_Q f_S$, where Q is some rectangle that contains S. If f_S is integrable on Q and D is the set of discontinuities of f_S in Q, then D has measure zero. By Lemma 15.1, D_∂ corresponds to the set of discontinuities of f_S on ∂S. Because $D_\partial \subseteq D$, we find that D_∂ has measure zero as well. Conversely, if D_∂ has measure zero, then the set of discontinuities of f_S on Q has measure zero, which implies f_S is integrable on Q. □

The boundary of S either makes or breaks integrability of a continuous function, and if the jumps at the boundary are not too significant, it happens that the integral of $f : S \to \mathbb{R}$ over S is equal to the integral over the interior.

Corollary 15.3. Let $f : S \to \mathbb{R}$ be a bounded function on a bounded subset $S \subseteq \mathbb{R}^n$, and define $A := \operatorname{int}(S)$. If f is integrable on S, then f is integrable on A and $\int_A f = \int_S f$.

Proof. First, observe that if $x_0 \in S \setminus A$, then $x_0 \in \partial S$. If f_S is continuous at x_0, then $f_S(x_0) = 0$. Because $x_0 \notin A$, we have $f_A(x_0) = 0$ by definition of f_A. Thus, we have $f_S(x_0) = f_A(x_0) = 0$. Notice that if $\lim_{x\to x_0} f(x) = 0$ for $x \in S$, then $\lim_{x\to x_0} f(x) = 0$ for $x \in A$. We have that

$f_A(x_0) = \lim_{x \to x_0} f(x)$ for $x \in S$, which means f_A is continuous at x_0. And so, we find that the set of discontinuities of f_A is a subset of the set of discontinuities of f_S. Thus, if f_S is integrable, then f_A is intergrable as well. $\square$

Definition 15.4. Let $S \subseteq \mathbb{R}^n$ be a bounded subset. We say that S is *rectifiable* if the constant function $\mathbf{1} : S \to \mathbb{R}$ defined as $\mathbf{1}(x) = 1$ is integrable over S. And so,

$$\mathrm{vol}(S) = \int_S \mathbf{1}.$$

Proposition 15.5. If $S \subseteq \mathbb{R}^n$ is a bounded subset, then S is rectifiable if and only if ∂S has measure zero.

Proof. Note that $\mathbf{1} : S \to \mathbb{R}$ is a bounded function on S and for $x_0 \in \partial S$, we have $\lim\limits_{x \to x_0} f(x) = 1$ for $x \in S$, which implies

$$D_\partial = \left\{ x_0 \in \partial S : \lim_{x \to x_0} f(x) \neq 0 \right\} = \partial S.$$

By Corollary 15.2, we have that $\mathbf{1}$ is integrable over S—in other words, rectifiable—if and only if ∂S has measure zero. $\square$

Theorem 15.6.

1. If S is rectifiable, then $\mathrm{vol}(S) \geq 0$.
2. If S_1 and S_2 are rectifiable and $S_1 \subseteq S_2$, then $\mathrm{vol}(S_1) \leq \mathrm{vol}(S_2)$.
3. If S_1 and S_2 are rectifiable, then $S_1 \cap S_2$ and $S_1 \cup S_2$ are rectifiable. Moreover,
$$\mathrm{vol}(S_1 \cup S_2) = \mathrm{vol}(S_1) + \mathrm{vol}(S_2) - \mathrm{vol}(S_1 \cap S_2).$$
4. If S is rectifiable and $\mathrm{vol}(S)$, then $\mathrm{vol}(S) = 0$ if and only if S has measure zero.
5. If S is rectifiable, then $\mathrm{int}(S)$ is rectifiable as well. Moreover,
$$\mathrm{vol}(S) = \mathrm{vol}(\mathrm{int}(S)).$$
6. If S is rectifiable and $f : S \to \mathbb{R}$ is a bounded continuous function, then f is integrable on S.

Proof. Note that statements (1) through (3) follow from Theorem 14.4, statement (4) follows from the defintion of integrals and measure zero, and statements (5) and (6) follow from the preceding corollaries and proposition. $\square$

We showed that a set is rectifiable if and only if its boundary has measure zero. One might think that boundedness is a sufficient condition for rectifiability, but this is not always true. To demonstrate this, consider the following example.

Example 15.7. Consider $T := \mathbb{Q} \cap (0,1)$. Note that T is countable, so suppose $T = \{q_1, q_1, \dots\}$. Fix $a \in (0,1)$, and for $q_i \in T$, define $I_i := (a_i, b_i)$ such that $q_i \in I_i$ and $b_i - a_i \leq a/2^i$. Now let $A = \bigcup_{i=1}^\infty I_i$, and we have that A is open. We will show that A does not have measure zero.

Observe that $[0,1] = A \cup \partial A$ and A is covered by intervals whose lengths sum to a. If we cover ∂A

with intervals with total length ε, then $[0, 1]$ is covered by countably many intervals with total length $a + \varepsilon$. For sufficiently small $\varepsilon > 0$, we have $a + \varepsilon < 1$. However, we now arrive at a contradiction.

16 Simple Regions and Constructing Rectifiable Sets

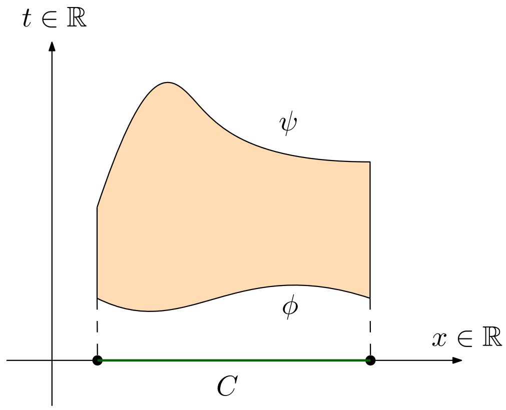
In this lecture, we will consider a specific type of rectifiable sets known as *simple regions*, and we will extend Fubini's Theorem on rectangles to such regions.

Definition 16.1. Let C be a compact, rectifiable set in $\mathbb{R}^{n-1}$ and $\phi, \psi : C \to \mathbb{R}$ continuous functions such that $\phi(x) \leq \psi(x)$ for all $x \in C$. We define the set

$$S := \{(x,t) : x \in C \text{ and } \phi(x) \leq t \leq \psi(x)\} \subseteq \mathbb{R}^{n-1} \times \mathbb{R}$$

as a *simple region* in $\mathbb{R}^n$.

A simple region describes the region between the graphs of two functions ϕ and ψ, as shown in the diagram below.



Lemma 16.2. Any simple region is compact and rectifiable.

Proof. First, we show that a simple region S is compact by showing S is closed and bounded. Note that S is bounded because C is compact, which implies ϕ and ψ achieve extrema on C. To see that S is closed, consider the map $H : C \times \mathbb{R} \to \mathbb{R}^2$ defined by $H(x,y) = (t - \phi(x), \psi(x) - t)$. Notice that $(x,t) \in S$ if and only if $H(x,t) \geq 0$. Moreover, H is continuous and so $S = H^{-1}(\mathbb{R}{\geq}\, 0 \times \mathbb{R}_{\geq 0})$. Note that $\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ is closed in $\mathbb{R}^2$, so S is closed in $C \times \mathbb{R}$. Since $C \times \mathbb{R}$ is closed in $\mathbb{R}^{n-1} \times \mathbb{R}$, we have that S is also closed in $\mathbb{R}^{n-1} \times \mathbb{R}$. Thus, we find that S is compact.

To show that S is rectifiable, we show that ∂S has measure zero. Define

$$G_\phi := \{(x, \phi(x)) : x \in C\}, \quad G_\psi := \{(x, \psi(x)) : x \in C\}, \quad D := \{(x,t) : x \in \partial C \text{ and } (x,t) \in S\},$$

which are the graphs of ϕ and ψ and the "edge" of our set in $\mathbb{R}^n$. We will that show $\partial S \subseteq G_\phi \cup G_\psi \cup D$, and if $G_\phi \cup G_\psi \cup D$ has measure zero, then ∂S has measure zero as well. Notice that because S is closed, we have $\partial S = S \setminus \operatorname{int}(S)$. Moreover, note that $\partial S \subseteq G_\phi \cup G_\psi \cup D$ is equivalent to

$$S \setminus (G_\phi \cup G_\psi \cup D) \subseteq S \setminus \partial S = \operatorname{int}(S),$$

and so it suffices to show that $S \setminus (G_\phi \cup G_\psi \cup D)$ is open. Now define $H' : \operatorname{int}(C) \times \mathbb{R} \to \mathbb{R}^2$ by

$$H'(x,t) := (t - \phi(x), \psi(x) - t),$$

and observe that

$$(H')^{-1}(\mathbb{R}_{>0} \times \mathbb{R}_{>0}) = S \setminus (G_\phi \cup G_\psi \cup D)$$

is open in $\operatorname{int}(C) \times \mathbb{R}$. Since $\operatorname{int}(C) \times \mathbb{R}$ is open in $\mathbb{R}^n$, we see that $S \setminus (G_\phi \cup G_\psi \cup D) \subseteq \mathbb{R}^{n+1}$ is also open in $\mathbb{R}^n$.

And now, we will show that D, G_ϕ, and G_ψ all have measure zero. Given that C is compact, we know $\phi : C \to \mathbb{R}$ is uniformly continuous and there exists a rectangle $Q \subseteq \mathbb{R}^{n-1}$ containing C. And so, for $\varepsilon > 0$ there exists $\delta > 0$ such that $|\phi(x) - \phi(y)| < \varepsilon$ for all $x, y \in R \cap C$, where R is a square with width$(R) < \delta$. Now let P be a partition of Q such that mesh$(P) < \delta$, and for each subrectangle R determined by P, let $x_R \in R$. Since $|\phi(x_R) - \phi(x)| < \varepsilon$ for all $x \in R$, we have that $G_\phi \cap (R \times R)$ is contained in $R \times [\phi(x_R) - \varepsilon, \phi(x_R) + \varepsilon]$. Now notice that

$$\sum_{R \subseteq Q} \text{vol}(R \times [\phi(x_R) - \varepsilon, \phi(x_R) + \varepsilon]) = \sum_{R \subseteq Q} 2\varepsilon \text{vol}(R) = 2\varepsilon \text{vol}(Q),$$

and because we can allow ε to be arbitrarily small, we see that G_ϕ has measure zero. Note that we have shown a general fact that the graph of a uniformly continuous function has measure zero. And likewise, we have that G_ψ also has measure zero.

Now it remains to show that $D = (\partial C \times \mathbb{R}) \cap S$ has measure zero. Notice that $C \subseteq \mathbb{R}^{n-1}$ is rectifiable, so ∂C has measure zero. Since ϕ and ψ are uniformly continuous and C is compact, there exists $M > 0$ such that $-M \leq \phi(x) \leq \psi(x) \leq M$ for all $x \in C$. Notice then that $(\partial C \times \mathbb{R}) \cap S$ is a subset of $\partial C \times [-M, M]$, and it suffices to show $\partial C \times [-M, M]$ has measure zero. Since ∂C has measure zero, for $\varepsilon > 0$ there exists a cover $\{Q_i\}_{i=1}^\infty$ of ∂C such that $\sum_{i=1}^\infty \text{vol}(Q_i) < \varepsilon/2M$. Observe that $Q_i \times [-M, M]$ covers Q_i and $\text{vol}(Q_i \times [-M, M]) = 2M \text{vol}(Q_i)$, and we have

$$\sum_{i=1}^\infty \text{vol}(Q_i \times [-M, M]) = \sum_{i=1}^\infty 2M \text{vol}(Q_i) < \varepsilon.$$

Note that $\partial C \times [-M, M] \subseteq \bigcup_{i=1}^\infty Q_i \times [-M, M]$, so we have shown that $\partial C \times [-M, M]$ has measure zero. □

We can imagine that simple regions aren't too different from rectangles, and in fact, we can compute the value of an integral over a simple region using a generalized version of Fubini's Theorem.

Theorem 16.3 (Fubini's Theorem on Simple Regions)**.** Let $S := \{(x, t) : x \in C \text{ and } \phi(x) \leq t \leq \psi(x)\}$ be a simple region in $\mathbb{R}^{n-1} \times \mathbb{R}$ and $f : S \to \mathbb{R}$ a continuous function. Then f is integrable over S and

$$\int_S f = \int_{x \in C} \int_{\phi(x)}^{\psi(x)} f(x, t).$$

Proof. Let $Q \times [-M, M] \subseteq \mathbb{R}^{n-1} \times \mathbb{R}$ be a rectangle containing S, where $M > 0$ such that

$$-M \leq \phi(x) \leq \psi(x) \leq M.$$

For a fixed $x_0 \in C$, we have

$$f_S(x_0, t) = \begin{cases} f(x_0, t) & \text{if } (x_0, t) \in S, \\ 0 & \text{if } (x_0, t) \notin S. \end{cases}$$

Note that f_S is a continuous function of t except at the boundaries $t = \phi(x_0)$ and $t = \psi(x_0)$. And so, f_S is integrable on $[-M, M]$, and by Fubini's Theorem, we have

$$\int_S f := \int_{Q \times [-M, M]} f_S = \int_Q \int_{-M}^M f(x, t) = \int_Q \int_{\phi(x)}^{\psi(x)} f(x, t).$$ □

Proposition 16.4. Suppose $f : U \to \mathbb{R}^n$ is a C^1 function such that $Df(x)$ is invertible for all $x \in U$. If $C \subseteq U$ is compact and rectifiable, then $f(C)$ is also compact and rectifiable.

Proof. To show that $f(C)$ is rectifiable, we will show that $\partial f(C)$ has measure zero. By the Inverse Function Theorem, f is an open map, which means $f(\text{int}(C)) \subseteq \text{int}(f(C))$. Because C is compact and f is continuous, $f(C)$ is compact and therefore closed. As such, we have $\partial f(C) \subseteq f(C)$ and $\partial f(C) \subseteq f(\partial C)$. If C is rectifiable, then ∂C has measure zero. Consequently, $f(\partial C)$ has measure zero as well (see homework), which means $\partial f(C)$ has measure. □

17 Determinants and Volume

In the last few lectures, we extended integration on rectangles to arbitrary sets in $\mathbb{R}^n$. We conclude our discussion of integration on Euclidean space in the next few lectures by studying how volumes change under a function. We will prove a *change of variables* formula that generalizes the familiar u-substitution in single-variable calculus to the context of $\mathbb{R}^n$. To ease our way into this final result, we begin by studying how volumes transform under *linear* maps.

Theorem 17.1. Let $S \subseteq \mathbb{R}^n$ be a rectifiable subset. If $g : \mathbb{R}^n \to \mathbb{R}^n$ an invertible linear transformation, then

$$\text{vol}(g(S)) = \text{vol}(S) \cdot |\det(g)|.$$

A careful characterization of the determinant will be provided in the appendix, but for now we will define the determinant of a linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ to be the *signed volume* of the image of the unit cube, and we will accept the following properties about determinants:

1. The determinant of the identity matrix is 1.
2. The determinant is multilinear with respect to the rows and columns. In other words,

$$\det\begin{pmatrix} \text{---}\; A_1 \;\text{---} \\ \vdots \\ \text{---}\; A_i + \lambda A_i' \;\text{---} \\ \vdots \\ \text{---}\; A_n \;\text{---} \end{pmatrix} = \det\begin{pmatrix} \text{---}\; A_1 \;\text{---} \\ \vdots \\ \text{---}\; A_i \;\text{---}, \\ \vdots \\ \text{---}\; A_n \;\text{---} \end{pmatrix} + \lambda \cdot \det\begin{pmatrix} \text{---}\; A_1 \;\text{---} \\ \vdots \\ \text{---}\; A_i' \;\text{---}, \\ \vdots \\ \text{---}\; A_n \;\text{---} \end{pmatrix},$$

$$\det\begin{pmatrix} | & & | & & | \\ A_1 & \cdots & A_i + \lambda A_i' & \cdots & A_n \\ | & & | & & | \end{pmatrix} = \det\begin{pmatrix} | & & | & & | \\ A_1 & \cdots & A_i & \cdots & A_n \\ | & & | & & | \end{pmatrix} + \lambda \cdot \det\begin{pmatrix} | & & | & & | \\ A_1 & \cdots & A_i' & \cdots & A_n \\ | & & | & & | \end{pmatrix}.$$

3. The determinant is antisymmetric; swapping two rows multiplies the determinant by -1.
4. The determinant does not change if a row A_i is replaced by the sum $A_i + A_j$.
5. If $A, B \in M_{n\times n}(\mathbb{R})$, then $\det(A \circ B) = \det(A) \cdot \det(B)$.
6. A matrix $A \in M_{n\times n}(\mathbb{R})$ is invertible if and only if $\det(A) \neq 0$.

In addition to the determinant properties listed above, we define the *elementary linear transformations.*

Definition 17.2. The following are elementary linear transformations:

1. $E_1^c(x_1,\dots,x_n) := (x_1,\dots,cx_i,x_{i+1},\dots,x_n)$ for some $c \neq 0$.
2. $E_2^{ij}(x_1,\dots,x_n) = (x_1,\dots,x_{i-1},x_j,x_{i+1},\dots,x_{j-1},x_i,x_{j+1},\dots,x_n)$ (swapping)
3. $E_3^{ij}(x_1,\dots,x_n) = (x_1,\dots,x_{i-1},x_i+x_j,x_{i+1},\dots,x_n)$ (shearing)

An important fact we will use is that a linear map is invertible if and only if it can expressed as a composition of elementary linear transformations applied to the identity map. With these facts in mind, we now prove Theorem 17.1.

Proof. If $g : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, then g can be expressed as a composition of the elementary linear transformations. Since the determinant of a product is the product of the determinants, note that it suffices to show Theorem 17.1 for only the elementary linear transformations.

Suppose that $g = E_1^c$ where $c \neq 0$. Then we have $\det(g) = \det(E_1^c) = c$. To compute $\mathrm{vol}(g(S)) = \int_{g(S)} \mathbf{1}_{g(S)}$, let Q be a rectangle containing $g(S)$ and P a partition of Q. Notice that $g^{-1}(Q)$ is also a rectangle, where the ith components of Q are scaled by $1/c$, and $g^{-1}(P)$ is a partition of $g^{-1}(Q)$. Now observe that

$$\begin{aligned} L(\mathbf{1}_{g(S)},P) &= \sum_R m_R(\mathbf{1}_{g(S)}) \cdot \mathrm{vol}(R), \\ &= \sum_{g^{-1}(R)} m_{g^{-1}(R)}(\mathbf{1}_S) \cdot |c| \cdot \mathrm{vol}(g^{-1}(R)), \\ &= |c| \cdot L(\mathbf{1}_S, g^{-1}(P)), \end{aligned}$$

and likewise, we have $U(\mathbf{1}_{g(S)},P) = |c| \cdot U(\mathbf{1}_S, g^{-1}(P))$. As such, we have that

$$\mathrm{vol}(g(S)) = \int_Q \mathbf{1}_{g(S)} = |c| \cdot \int_{g^{-1}(Q)} \mathbf{1}_S = |c| \cdot \mathrm{vol}(S) = |\det(g)| \cdot \mathrm{vol}(S).$$

If $g = E_2^{ij}$, then we have $\det(g) = \det(E_2^{ij}) = -1$. Proceeding in the same way as before, we find that $\mathrm{vol}(g(S)) = \mathrm{vol}(S) = |\det(g)| \cdot \mathrm{vol}(S)$.

Now suppose $g = E_3^{ij}$, where we assume without loss of generality that $i = 1$ and $j = 2$. Let Q be a rectangle containing $g(S)$, and if $Q = \prod_{i=1}^n [a_i,b_i]$, then observe that $g^{-1}(Q)$ is the simple region

$$g^{-1}(Q) = \{(x,t) : x \in C \text{ and } \phi(x) \leq t \leq \psi(x)\} \subseteq \mathbb{R}^{n-1} \times \mathbb{R}$$

where $C = [a_2,b_2] \times \cdots \times [a_n,b_n] \subseteq \mathbb{R}^{n-1}$ and $\phi, \psi : \mathbb{R}^{n-1} \to \mathbb{R}$ are defined by

$$\phi(x_2,\dots,x_n) = a_1 - x_2, \qquad \psi(x_2,\dots,x_n) = b_1 - x_2.$$

By Fubini's Theorem over simple regions, we have

$$\int_{g^{-1}(Q)} \mathbf{1}_S = \int_{x \in C} \int_{\phi(x)}^{\psi(x)} \mathbf{1}_S = \int_{x_2=a_2}^{b_2} \cdots \int_{x_n=a_n}^{b_n} \int_{x_1=a_1-x_2}^{b_1-x_2} \mathbf{1}_S.$$

Observe that

$$\int_{x_1=a_1-x_2}^{b_1-x_2} \mathbf{1}_S = \int_{x_1=a_1}^{x_1=b_1} \mathbf{1}_S(x_1+x_2,\dots,x_n) = \int_{a_1}^{b_1} \mathbf{1}_{g(S)}$$

which implies

$$\int_{g^{-1}(Q)} \mathbf{1}_S = \int_{x_2=a_2}^{b_2} \cdots \int_{x_n=a_n}^{b_n} \int_{x_1=a_1}^{b_1} \mathbf{1}_S = \int_Q \mathbf{1}_{g(S)} = \mathrm{vol}(g(S)).$$

□

18 Change of Variables

Last time, we showed that a linear map $g : \mathbb{R}^n \to \mathbb{R}^n$ scales the volume of a rectifiable set $S \subseteq \mathbb{R}^n$ by a factor of $|\det(g)|$. In this lecture, we will begin proving the change of variables formula for integrals, which extends the familiar u-substitution from single-variable calculus to $\mathbb{R}^n$. Before we state our main result, consider the following definitions:

Definition 18.1. Let $U, V \subseteq \mathbb{R}^n$ be open subsets. A C^k *diffeomorphism* from U to V is a bijective C^k function $g : U \to V$ such that $g^{-1} : V \to U$ is also a C^k function.

Remark. If $U \subseteq \mathbb{R}^n$ is open and $g : U \to \mathbb{R}^n$ is an injective C^k function such that $Dg(x)$ is invertible for all $x \in U$, then g is a C^k diffeomorphism between U and $g(U)$ by the Inverse Function Theorem.

Definition 18.2. Let $g : U \to \mathbb{R}^n$ be differentiable on $U \subseteq \mathbb{R}^n$. The *Jacobian determinant* of g is the function $Jg : U \to \mathbb{R}$ defined by

$$Jg(x) := |\det(Dg(x))|.$$

Theorem 18.3 (Change of Variables)**.** Let $U, V \subseteq \mathbb{R}^n$ be open subsets and $g : U \to V$ a C^k diffeomorphism. If $S \subseteq \mathbb{R}^n$ is a rectifiable set such that $\overline{S} \subseteq V$ and $f : S \to \mathbb{R}$ is integrable, then

$$\int_S f = \int_{g^{-1}(S)} (f \circ g) \cdot Jg.$$

Proof. We will organize our proof by the following steps:

<u>*Step 1:*</u> If $A \in M_{n\times n}(\mathbb{R})$, then define

$$\|A\|' := \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|,$$

which bounds the norm of each row vector of A. Observe that

$$\|Av\|_{\sup} \le \|A\|' \cdot \|v\|_{\sup},$$

where $\|v\|_{\sup} = \max_i |v_i|$. Now let $C \subseteq U$ be a square with width$(C) = \ell$ centered at $x_0 \in C$, and define

$$\lambda = \max_{x\in C} \|Dg(x)\|'.$$

We have that $\|x_0 - y\| \le \ell/2$ for all $y \in C$, and note that $g(C)$ can be inscribed in a square with volume $\lambda^n \cdot \mathrm{vol}(C)$; if $g = (g_1, g_2, \ldots, g_n)$, then observe by the Mean Value Theorem that for $y \in C$, there exists $0 < h < 1$ such that

$$g_i(x_0) - g_i(y) = Dg_i(x_0 + h(y - x_0))(x_0 + h(y - x_0)).$$

And so, we have

$$|g_i(x_0) - g_i(y)| \le \|Dg(x_0 + h(y - x_0))\|' \cdot \|x_0 + h(y - x_0)\|_{\sup}, \le \lambda \cdot \frac{\ell}{2}.$$

which means $g(C)$ is contained in a square centered at x_0 with side length $\lambda \cdot \ell$, hence volume $\lambda^n \cdot \mathrm{vol}(C)$. Now note that if $A : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, then $g = A \circ A^{-1} \circ g$ and we have

$$\mathrm{vol}(g(C)) = |\det A| \cdot \mathrm{vol}(A^{-1} g(C))$$

by Theorem 17.1. Consequently, we see that

$$\mathrm{vol}(g(C)) = |\det A| \cdot \mathrm{vol}(A^{-1}g(C)) \leq |\det A| \left(\max_{x \in C} \|A^{-1} \circ Dg(x)\|' \right)^n \cdot \mathrm{vol}(C).$$

To summarize, we produced a bound on the volume of $g(C)$ by inscribing $g(C)$ within a square that is a scalar multiple of $C \subseteq U$.

Step 2: Following the previous step, we now want to select a linear map A to obtain a tighter bound on $\mathrm{vol}(g(C))$. Consider the function $F : C \times C \to \mathbb{R}$ given by $F(x, y) = \|Dg(x)^{-1} \circ Dg(y)\|'$; because g is a C^1 function on a compact set C, we know $Dg : U \to \mathrm{Hom}(U, \mathbb{R}^n)$ is uniformly continuous. Then for all $\varepsilon > 0$, there exists $\delta_1 > 0$ such that $\|Dg(x)^{-1} \circ Dg(y)\|' < 1 + \varepsilon$ for all $x, y \in C$ such that $\|x - y\|_{\sup} < \delta_1$. We are considering the bound $1 + \varepsilon$ because $\|Dg(x)^{-1} \circ Dg(x)\|' = \|\mathrm{Id}_n\|' = 1$.

Now note that $Jg(x) = |\det(Dg(x))|$ is continuous, hence integrable on C, and as such, there exists $\delta_2 > 0$ such that $U(Jg, P) - L(Jg, P) < \varepsilon$ for all partitions P of C where $\mathrm{mesh}(P) < \delta_2$. Let $\delta = \min(\delta_1, \delta_2)$ and P' a partition of C with $\mathrm{mesh}(P') < \delta$, and for each subrectangle C_i determined by P', let $x_i \in C_i$. Now observe that

$$C = \left(\bigsqcup_i \mathrm{int}(C_i) \right) \sqcup B,$$

where $B := \bigcup_i \partial C_i$. Because g is injective, we have

$$g(C) = \left(\bigsqcup_i g(\mathrm{int}(C_i)) \right) \sqcup g\left(B\right).$$

Note that $\bigcup_i \partial C_i$ has measure zero, so $g\left(\bigcup_i \partial C_i\right)$ also has measure zero. Now notice that

$$\mathbf{1}_{g(C)} = \mathbf{1}_B + \sum_i \mathbf{1}_{g(C_i)},$$

and by the addition of integrals over disjoint regions, we have

$$\begin{aligned}
\int_{g(C)} \mathbf{1}_{g(C)} &= \int_B \mathbf{1}_B + \sum_i \int_{g(C_i)} \mathbf{1}_{g(C_i)}, \\
\mathrm{vol}(g(C)) &= \mathrm{vol}(B) + \sum_i \mathrm{vol}(g(C_i)), \\
&= \sum_i \mathrm{vol}(g(C_i)).
\end{aligned}$$

And so, observe that

$$\begin{aligned}
\mathrm{vol}(g(C)) &= \sum_{C_i} \mathrm{vol}(g(C_i)), \\
&\leq \sum_{C_i} |\det(Dg(x_i))| \cdot \left(\max_{y \in C} \|Dg(x_i)^{-1} \circ Dg(y)\| \right)^n \cdot \mathrm{vol}(C_i), \\
&\leq \sum_{C_i} Jg(x_i) \cdot (1 + \varepsilon)^n \cdot \mathrm{vol}(C_i), \\
&\leq (1 + \varepsilon)^n \cdot U(Jg, P'), \\
&\leq (1 + \varepsilon)^n \cdot \int_C Jg,
\end{aligned}$$

which implies

$$\mathrm{vol}(g(C)) \leq \int_C Jg.$$

This conclude the second step in our proof. $\square$

Corollary 18.4. If $g : U \to V$ is a C^k diffeomorphism and $T \subseteq \mathbb{R}^n$ is rectifiable subset such that $\overline{T} \subseteq U$, then $\mathrm{vol}(g(T)) = \int_T J_g$.

19 Change of Variables and an Introduction to Differential Forms

At the end of the last lecture, we proved two steps of the proof for the change of variables theorem for multidimensional integrals. In summary, we took a square $C \in U$ and showed that

$$\mathrm{vol}(g(c)) \leq \int_C Jg,$$

where $Jg : U \to \mathbb{R}$ is the Jacobian determinant of g on U. We will now "cover" $g^{-1}(S)$ using squares to evaluate $\int_{g^{-1}(S)}(f \circ g) \cdot J_g$ using our result on squares.

Theorem. Let $U, V \subseteq \mathbb{R}^n$ be open subsets and $g : U \to V$ a C^k diffeomorphism. If $S \subseteq \mathbb{R}^n$ is a rectifiable set such that $\overline{S} \subseteq V$ and $f : S \to \mathbb{R}$ is integrable, then

$$\int_S f = \int_{g^{-1}(S)} (f \circ g) \cdot Jg.$$

Proof. We will now prove the theorem.

Step 3: As introduced in the proof of Theorem 14.4, let

$$f_+(x) := \max(f(x), 0),$$
$$f_-(x) := \min(f(x), 0),$$

whereby $f = f_+ - f_-$ and f_+ and f_- are non-negative integrable functions on S. By linearity of the integral, it suffices to prove the theorem for f_+ and f_- individually. Hence, it also suffices to assume that f is non-negative.

We begin by showing $\int_S f \leq \int_{g^{-1}(S)}(f \circ g)J_g$, so let C be a square containing S (not necessarily contained in V) and P a partition of C by squares. Now notice that

$$L(f, P) = \sum_{C_i \subseteq C} m_{C_i}(f_S) \cdot \mathrm{vol}(C_i) = \sum_{C_i \subseteq S} m_{C_i}(f_S) \cdot \mathrm{vol}(C_i),$$

as $m_{C_i}(f_S) = 0$ if C_i contains a point not in S—even if C_i intersects S. We have

$$\sum_{C_i \subseteq S} m_{C_i}(f_S) \cdot \mathrm{vol}(C_i) \leq \sum_{C_i \subseteq S} m_{C_i}(f_S) \cdot \int_{g^{-1}(C_i)} Jg.$$

Note that for $x \in g^{-1}(C_i)$, we have $m_{C_i}(f_S) \leq (f \circ g)(x)$, which implies

$$\begin{aligned} L(f_S, P) &\leq \sum_{C_i \subseteq S} \int_{g^{-1}(C_i)} (f \circ g) \cdot J_g, \\ &= \sum_{C_i \subseteq S} \int_{g^{-1}(\mathrm{int}(C_i))} (f_S \circ g) \cdot J_g, \\ &\leq \int_{g^{-1}(S)} (f \circ g) \cdot J_g. \end{aligned}$$

Since $\mathrm{mesh}(P)$ can be made arbitrarily small, $L(f_S, P)$ converges to $\int_S f$ and we see that

$$\int_S f \leq \int_{g^{-1}(S)} (f \circ g) \cdot Jg.$$

To obtain the reverse inequality, define $h := g^{-1}$ and $S' := g^{-1}(S)$ and $f' := (f \circ g) \cdot Jg$. We have that $h^{-1}(S') = g(g^{-1}(S)) = S$. and applying the inequality above to the inverse function yields

$$\int_{g^{-1}(S)} (f \circ g) \cdot Jg = \int_{S'} f' \leq \int_{h^{-1}(S')} (f' \circ h) \cdot Jh.$$

Note that for $x \in S$, we have

$$(f' \circ h)(x) = f'(h(x)) = f(g(h(x))) \cdot Jg(h(x)) = f(x) \cdot Jg(h(x)),$$

and the Chain Rule applied to $g \circ h = \mathrm{Id}_n$ yields $D(g \circ h)(x) = Dg(h(x)) \circ Dh(x) = \mathrm{Id}_n$. As such, we have $Jg(h(x)) \cdot Jh(x) = 1$. Notice that $(f' \circ h) \cdot Jh = f$, so we have

$$\int_{g^{-1}(S)} (f \circ g) \cdot Jg \leq \int_S f$$

and therefore

$$\int_S f = \int_{g^{-1}(S)} (f \circ g) \cdot Jg. \qquad \square$$

This concludes our proof of the change of variables formula for integrals in $\mathbb{R}^n$. The single-variable version is the well-known u-substitution formula, which is stated as follows:

Theorem. Suppose that $f : [a, b] \to \mathbb{R}$ is integrable and $g : [c, d] \to \mathbb{R}$ a C^1 function such that $g'(x) \neq 0$ for all $x \in [c, d]$ and $\{g(c), g(d)\} = \{a, b\}$. Then

$$\int_a^b f(x)dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(t)) \cdot g'(t)dt.$$

Since $g : [c, d] \to \mathbb{R}$ is a function between an interval and $\mathbb{R}$, the Jacobian matrix of g simply consists of its first derivative g'. However, it appears that absolute value in our multivariable result is absent in the single-variable case. To explain this descrepancy, note the following facts:

- If $g'(t) > 0$, then $g(c) = a$ and $g(d) = b$. As such, we have

$$\int_a^b f(x)dx = \int_c^d f(g(t)) \cdot g'(t)dt.$$

- If $g'(t) < 0$, then $g(c) = b$ and $g(d) = a$, which implies

$$\int_a^b f(x)dx = \int_d^c f(g(t)) \cdot g'(t)dt = -\int_c^d f(g(t)) \cdot g'(t)dt.$$

As we see above, the sign of g' alters the "direction" in which we integrate. A notion of direction is encoded by the upper and lower bounds of integration, but notice that this feature is absent in our current definition of integrals. Our next step in the course is to reintroduce some notion of signs and *orientation* through *differential forms*. Differential forms are defined using exterior products, which provides a framework for defining the determinant of a map intrinsically.

Definition 19.1. Let $U \subseteq \mathbb{R}^n$ be an open subset. A *differential k-form* on U is a map

$$\phi : U \to \bigwedge^k (\mathbb{R}^n)^*$$

We say ϕ is a C^r-form if it is a C^r function.

Note that the dual space $(\mathbb{R}^n)^*$ is an n-dimensional vector space with basis $\{e^1, \dots, e^n\}$, where $e^i : \mathbb{R}^n \to \mathbb{R}$ is a functional defined by

$$e^i(e_j) := \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

As such, the basis of $\bigwedge^k (\mathbb{R}^n)^*$ is given by

$$B := \{e^{i_1} \wedge e^{i_2} \wedge \cdots \wedge e^{i_k} : 1 \leq i_1 < i_2 < \cdots < i_k \leq n\},$$

which consists of $\binom{n}{k}$ many pure wedges. Given an open subset $U \subseteq \mathbb{R}^n$, we define the family of *constant* 1-forms $dx_1, \dots, dx_n : U \to \bigwedge^1 (\mathbb{R}^n)^*$ where

$$dx_i(y) = e^i$$

for all $y \in U$ and $1 \leq i \leq n$. We will later see that "d" is actually an operation on differential k-forms, but for now we will accept $dx_i(y) = e^i$ as a definition. Anyways, note that each dx_i is smooth, and because B is a basis of $(\mathbb{R}^n)^*$, we see that every k-form $\phi : U \to \bigwedge^k (\mathbb{R}^n)^*$ can be expressed as

$$\phi = \sum_{i_1 < \cdots < i_k} \phi_{i_1, \dots, i_k} \, dx_{i_1} \wedge \cdots \wedge dx_{i_k},$$

where each $\phi_{i_1, \dots, i_k} : U \to \mathbb{R}$ is a C^k function.

Example 19.2. Suppose that $f : U \to \mathbb{R}$ be a C^1 function. Recall that $Df : U \to \operatorname{Hom}(\mathbb{R}^n, \mathbb{R})$ is a function that takes points in U and returns linear functionals on $\mathbb{R}^n$. More concretely, for $x \in U$, the output $Df(x)$ is a linear map with matrix representation

$$Df(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) & \frac{\partial f}{\partial x_2}(x) & \cdots & \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

with respect to the standard basis vectors. Notice also that $\operatorname{Hom}(\mathbb{R}^n, \mathbb{R})$ is isomorphic to $\bigwedge^1 (\mathbb{R}^n)^*$, and with this identification in mind, we can rewrite $Df(x)$ as

$$Df(x) = \frac{\partial f}{\partial x_1}(x) \cdot e^1 + \frac{\partial f}{\partial x_2}(x) \cdot e^2 + \cdots + \frac{\partial f}{\partial x_n}(x) \cdot e^n.$$

For now, we will define the 1-form $df : U \to \bigwedge^1 (\mathbb{R}^n)^*$ as

$$df(x) = \frac{\partial f}{\partial x_1}(x) \cdot e^1 + \frac{\partial f}{\partial x_2}(x) \cdot e^2 + \cdots + \frac{\partial f}{\partial x_n}(x) \cdot e^n.$$

Remark. For the sake of completeness, we define 0-forms to be smooth functions into $\mathbb{R}$.

20 Differential Forms and Pullbacks

We ended the last lecture with a definition of differential forms and considered an example of a 1-form on $\mathbb{R}^n$. In this lecture, we introduce the *pullback* for differential forms.

Definition 20.1. Suppose $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ are open subsets and $f : U \to V$ is differentiable. If ϕ is a k-form on V, then the k-form $f^*(\phi) : U \to \bigwedge^k (\mathbb{R}^m)^*$ defined by

$$f^*(\phi)(x) := \left(\bigwedge^k Df(x)^* \right) (\phi(f(x)))$$

is known as the *pullback* of ϕ along f.

To clarify, $Df(x)^* : (\mathbb{R}^m)^* \to (\mathbb{R}^n)^*$ is the adjoint of $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$, where

$$Df(x)^*(T) := T \circ Df(x)$$

for $T \in (\mathbb{R}^m)^*$. On top of that, $\bigwedge^k Df(x)^* : \bigwedge^k (\mathbb{R}^m)^* \to \bigwedge^k (\mathbb{R}^n)^*$ defines a multilinear map where

$$\begin{aligned}\bigwedge^k Df(x)^*(T_1 \wedge \cdots \wedge T_k) &= Df(x)^*(T_1) \wedge \cdots \wedge Df(x)^*(T_k),\\ &= (T_1 \circ Df(x)) \wedge \cdots \wedge (T_k \circ Df(x))\end{aligned}$$

for $T_1 \wedge \cdots \wedge T_k \in \bigwedge^k (\mathbb{R}^n)^*$. Now observe that if $f = (f_1, \dots, f_m)$ and

$$\phi(y) = \sum_{i_1 < \cdots < i_k} \phi_{i_1,\dots,i_k}(y)\ dy_{i_1} \wedge \cdots \wedge dy_{i_k}$$

then the pullback of ϕ along f is given by

$$\begin{aligned}\left(\bigwedge^k Df(x)^*\right)(\phi(f(x))) &= \bigwedge^k Df(x)^* \left(\sum_{i_1 < \cdots < i_k} \phi_{i_1,\dots,i_k}(f(x))\ dy_{i_1} \wedge \cdots \wedge dy_{i_k} \right),\\ &= \sum_{i_1 < \cdots < i_k} \phi_{i_1,\dots,i_k}(f(x)) \Big(\bigwedge^k Df(x)^* (\underbrace{dy_{i_1} \wedge \cdots \wedge dy_{i_k}}_{e^{i_1} \wedge \cdots \wedge e^{i_k}}) \Big),\\ &= \sum_{i_1 < \cdots < i_k} \phi_{i_1,\dots,i_k}(f(x)) \Big(Df(x)^*(e^{i_1}) \wedge \cdots \wedge Df(x)^*(e^{i_k}) \Big),\\ &= \sum_{i_1 < \cdots < i_k} \phi_{i_1,\dots,i_k}(f(x)) \Big(e^{i_1} \circ Df(x) \wedge \cdots \wedge e^{i_k} \circ Df(x) \Big),\\ f^*(\phi)(x) &= \sum_{i_1 < \cdots < i_k} \phi_{i_1,\dots,i_k}(f(x))\ df_{i_1}(x) \wedge \cdots \wedge df_{i_k}(x).\end{aligned}$$

Example 20.2. Recall that if $f : U \to \mathbb{R}$ is a C^1 function, then the 1-form $df : U \to \bigwedge^1 \mathbb{R}$ is given by

$$df(x) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x) \cdot dx_i.$$

Let $dt : U \to \bigwedge^1 \mathbb{R}^*$ be the constant 1-form defined by $dt(x) = e^1$, and note that df is also the pullback of dt along f, namely

$$f^*(dt)(x) = \bigwedge^1 Df(x)^*(dt) = dt \circ Df(x) = e^1 \circ Df(x) = df(x).$$

Example 20.3. Suppose $U, V \subseteq \mathbb{R}^n$ are open subsets and $f : U \to V$ is differentiable. Let

$$\phi = g(y_1, \dots, y_n)\ dy_1 \wedge \cdots \wedge dy_n$$

be an n-form on V, and note that every n-form is a multiple of $dy_1 \wedge \cdots \wedge dy_n$ because $dy_1 \wedge \cdots \wedge dy_n$ is the sole basis wedge of $\bigwedge^n (\mathbb{R}^n)^*$. Now observe that

$$\begin{aligned}f^*(\phi)(x) &= g(f(x))\ df_1(x) \wedge \cdots \wedge df_n(x),\\ &= g(f(x)) \cdot \left(\sum_{i=1}^n \frac{\partial f_1}{dx_i}(x) \cdot dx_i \right) \wedge \cdots \wedge \left(\sum_{i=1}^n \frac{\partial f_n}{dx_n}(x) \cdot dx_i \right),\end{aligned}$$

$$
\begin{aligned}
&= g(f(x)) \cdot \sum_{\sigma \in S_n} \frac{\partial f_1}{\partial x_{\sigma(1)}}(x) \cdots \frac{\partial f_n}{\partial x_{\sigma(n)}}(x) \cdot dx_{\sigma(1)} \wedge \cdots \wedge dx_{\sigma(n)}, \\
&= g(f(x)) \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \frac{\partial f_1}{\partial x_{\sigma(1)}}(x) \cdots \frac{\partial f_n}{\partial x_{\sigma(n)}}(x) \cdot dx_1 \wedge \cdots \wedge dx_n, \\
&= g(f(x)) \cdot \det(Df(x)) \cdot dx_1 \wedge \cdots \wedge dx_n.
\end{aligned}
$$

To clarify, note that we implicitly used the fact that $\det(T) = \det(T^*)$ for linear maps $T : V \to V$ while expanding $df_i(x)$. Nonetheless, we see that

$$f^*(g\ dy_1 \wedge \cdots \wedge dy_n)(x) = g(f(x)) \cdot \det(Df(x)) \cdot dx_1 \wedge \cdots \wedge dx_n.$$

Definition 20.4. Suppose $U \subseteq \mathbb{R}^n$ is an open subset and ϕ is an n-form on U defined by

$$\phi = g \cdot dx_1 \wedge \cdots \wedge dx_n.$$

We define the *support of* ϕ to be the set

$$\operatorname{supp}(\phi) := \overline{\{x \in U : \phi(x) \neq 0\}}.$$

We say ϕ has *compact support* on U if the $\operatorname{supp}(\phi)$ is compact. We define the *integral* of ϕ as

$$\int_U \phi := \int_U g$$

if it exists.

Remark. Now is probably a good time to address what dx means when we write

$$\int_a^b f(x)\ dx$$

like we do in single-variable calculus; in our theory of integration so far, we have (intentionally) avoided writing dx after the functions that we wish to integrate. With what we now know about differential forms, we can see that $f(x)\ dx$ is a 1-form—whereby f is an integrable function and $dx : U \to \bigwedge^1 \mathbb{R}^*$ is the constant 1-form defined by $dx(y) = e^1$ for all $y \in U$.

Definition 20.5. Let $U, V \subseteq \mathbb{R}^n$ be open subsets and $f : U \to V$ a C^k diffeomorphism. We say f is *orientation preserving* if $\det(Df(x)) > 0$ for all $x \in U$.

Based on our definition of the pullback, we can reframe the change of variables formula from the previous lecture in the context of differential n-forms.

Theorem 20.6 (Change of Variables)**.** If $f : U \to V$ is an orientation preserving C^k diffeomorphism and $\phi : V \to \bigwedge^n (\mathbb{R}^n)^*$ an integrable n-form on V, then

$$\int_U f^* \circ \phi = \int_V \phi.$$

21 Differential Forms and Pullbacks (continued)

In this lecture, we will continue our discussion of the pullbacks of differential forms and formally reintroduce manifolds. From here on out, we will denote the vector space of k-forms on U by $\mathcal{A}^k(U)$. And so, for a C^r function $f : U \to V$, the pullback is the linear map $f^* : \mathcal{A}^k(V) \to \mathcal{A}^k(U)$ defined by

$$f^*(\phi)(x) = \left(\bigwedge^k Df(x)^*\right)(\phi(f(x)))$$

for $\phi \in A^k(V)$. Now consider the following maneuvers:

1. For a vector space V, we have a linear map $\mathcal{M} : \bigwedge^k V \otimes \bigwedge^\ell V \to \bigwedge^{k+\ell} V$ defined by

   $$\mathcal{M}\left((v_1 \wedge \cdots \wedge v_k) \otimes (w_1 \wedge \cdots \wedge w_k)\right) = v_1 \wedge \cdots \wedge v_k \wedge w_1 \wedge \cdots \wedge w_\ell.$$

   As such, if ϕ is a k-form and ψ a ℓ-form, then we can create a $(k+\ell)$-form $\phi \wedge \psi$ defined by

   $$(\phi \wedge \psi)(x) := \phi(x) \wedge \psi(x).$$

   By definition of the exterior product, we have $\phi \wedge \psi = (-1)^{k\ell}\psi \wedge \phi$.

2. Let ϕ be a differentiable k-form on an open set $U \subseteq \mathbb{R}^n$ given by

   $$\phi = \sum_{i_1 < \cdots < i_k} \phi_{i_1,\cdots,i_k} \underbrace{dx_{i_1} \wedge \cdots \wedge dx_{i_k}}_{\text{basis wedge in } \bigwedge^k V},$$

   where each $\phi_{i_1,\cdots,i_k}$ is differentiable. We define the operation $d : \mathcal{A}^k(U) \to \mathcal{A}^{k+1}(U)$ by

   $$d\phi := \sum_{i_1 < \cdots < i_k} d\phi_{i_1,\cdots,i_k} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k},$$

   To clarify, $d : \mathcal{A}^k(U) \to \mathcal{A}^{k+1}(U)$ is a linear map where

   $$g\ dx_{i_1} \wedge \cdots \wedge dx_{i_k} \mapsto dg \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

   The operation d is somewhat like a derivative for k-forms, and in fact, it is commonly known as the *exterior derivative.*

Proposition 21.1. Suppose $U, V \subseteq \mathbb{R}^n$ are open subsets and $\phi \in \mathcal{A}^k(U)$ and $\psi \in A^\ell(U)$. If $f : U \to V$ is a differentiable function, then

1. $\phi \wedge \psi = (-1)^{k\ell}\psi \wedge \phi$,
2. $f^*(\phi \wedge \psi) = f^*(\phi) \wedge f^*(\psi)$,
3. If ϕ and ψ are differentiable, then

   $$d(\phi \wedge \psi) = d\phi \wedge \psi + (-1)^k \phi \wedge d\psi.$$

4. If ϕ is differentiable, then

   $$f^*(d\phi) = d(f^*\phi).$$

5. If ϕ is a C^2 function, then $d(d\phi) = 0$.

Proof.

1.

2.

3.

4.

5. Note that by linearity of d, it suffices to check for $\phi = g \cdot dx_{i_1} \wedge \cdots \wedge dx_{i_k}$. Observe that

$$\begin{aligned} d\phi &= dg \cdot dx_{i_1} \wedge \cdots \wedge dx_{i_k}, \\ &= \left(\sum_{i=1}^{n} \frac{\partial g}{\partial x_i} \cdot dx_i \right) \cdot dx_{i_1} \wedge \cdots \wedge dx_{i_k}, \end{aligned}$$

which implies

$$d(d\phi) = \sum_{i=1}^{n} d\left(\frac{\partial g}{\partial x_i}\right) \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

Note that

$$d\left(\frac{\partial g}{\partial x_i}\right) = \sum_{j=1}^{n} \frac{\partial^2 g}{\partial x_j \partial x_i} dx_j,$$

so we have

$$d(d\phi) = \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^2 g}{\partial x_j \partial x_i} \cdot dx_j \wedge dx_i \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

Recall that partial derivatives commute, so

$$\frac{\partial^s g}{\partial x_j \partial x_i} dx_j \wedge dx_i \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k} = -\frac{\partial^2 g}{\partial x_i \partial x_j} dx_i \wedge dx_j \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

As such, the terms in the sum for $d(d\phi)$ cancel out and we have $d(d\phi) = 0$. $\square$

We conclude this lecture by setting up the tools for our future discussion of *manifolds with boundary* and a more abstract and "instrinsic" definition of smooth manifolds. Just as a manifold (without boundary) locally resembles $\mathbb{R}^n$, we can imagine that a manifold with boundary looks like the upper half space

$$\mathbb{H}^n := \{(x_1, \ldots, x_n) : x_i \in \mathbb{R}, x_n \geq 0\} \subseteq \mathbb{R}^n,$$

whose boundary is given by

$$\partial\mathbb{H}^n := \{(x_1, \ldots, x_{n-1}, 0)\} \subseteq \mathbb{H}^n.$$

Ever since we defined differentiability of functions in $\mathbb{R}^n$, we have assumed open subsets as our domain. Differentiability becomes very weird when we consider closed domains, and this becomes an issue if we want to talk about differentiation on manifolds with boundary. We will adopt the following definition to circumvent this issue.

Definition 21.2. Let U be an open subset of $\mathbb{H}^n$. We say that a function $f : U \to \mathbb{R}^m$ is *differentiable* on U if there is an open subset V of $\mathbb{R}^n$ such that $U = V \cap \mathbb{H}^n$ and f extends to a differentiable function $\tilde{f} : V \to \mathbb{R}^m$.

Remark. Note that the extension $\tilde{f} : V \to \mathbb{R}^m$ is not necessarily unique, but this fine because the derivative at the boundary will still be well-defined.

22 Abstract Manifolds

In this lecture, we present both an "extrinsic" and "intrinsic" definition for *manifolds (with boundary)*. At the end of the last lecture, we defined the upper half space

$$\mathbb{H}^n := \{(x_1, \dots, x_n) : x_n \geq 0\},$$

which is a closed subspace of $\mathbb{R}^n$. We will also consider the following subspaces of $\mathbb{H}^n$.

$$\mathbb{H}^n_+ := \{(x_1, \dots, x_n) : x_n > 0\} \subseteq \mathbb{H}^n,$$
$$\partial\mathbb{H}^n := \{(x_1, \dots, x_n) : x_n = 0\} \subseteq \mathbb{H}^n.$$

We now define manifolds with boundary in reference to the upper half space of $\mathbb{R}^n$:

Definition 22.1. Let $k \leq n$. We say a subset $M \subseteq \mathbb{R}^n$ is an *embedded k-dimensional smooth manifold with boundary* if for all $p \in M$, there exist

1. an open neighborhood $V := V' \cap M$ of p where V' is open in $\mathbb{R}^n$,
2. an open subset $U \subseteq \mathbb{H}^n$,
3. and a diffeomorphism $\phi : V \to U$ such that $D\phi^{-1}(x)$ has rank k for all $x \in U$.

Remark. For the remainder of the course, we adhere to the convention where the local diffeomorphism ϕ is a map from the manifold to Euclidean space.

Example 22.2. The closed ball $\bar{B}(0,1) \subseteq \mathbb{R}^n$ is an n-dimensional manifold with boundary S^{n-1}.

Definition 22.3. If $M \subseteq \mathbb{R}^n$ is a k-dimensional manifold with boundary, then we define the *tangent space* T_pM of M at p to be the image

$$D\phi^{-1}(x_0)(\mathbb{R}^k) \subseteq \mathbb{R}^n,$$

where $\phi : V \to U$ is a diffeoomorphism such that $\phi^{-1}(x_0) = p$.

Previously in Lecture 9 and in Definition 22.1, we defined manifolds as being embedded in some ambient Euclidean space. This is an extrinsic treatment of manifolds, whereas an intrinsic definition would avoid any references to an ambient Euclidean space. To define manifolds intrinsically, we start by defining *topological manifolds*.

Definition 22.4. A *topological k-manifold with boundary* is a topological space M such that

1. M is Hausdorff,
2. M is *second countable*—there exists a collection $\mathcal{A}$ of open sets in M such that every open set is a union of open sets in $\mathcal{A}$—
3. and M is *locally Euclidean with dimension k*, where for all $p \in M$ there exist
   (a) an open neighborhood $U \subseteq M$ of p,

(b) an open subset $\widehat{U} \subseteq \mathbb{H}^k$,

(c) and a homeomorphism $\phi : U \to \widehat{U}$.

We refer to the pair (U, ϕ) as a *chart*, or *coordinate patch*, for M. We refer to the inverse $\phi^{-1} : \widehat{U} \to U$ as the *local parametrization* of U by $\widehat{U}$.

Remark. Note that $\mathbb{R}^n$ is Hausdorff because it is a metric space and second countable because we can consider the collection of balls with rational centers and rational radii.

We now want to endow our topological manifold with some notion of smoothness:

Definition 22.5. A *smooth k-dimensional manifold with boundary* is a topological k-manifold together with a collection $\mathcal{A}$ of charts such that

1. for all $p \in M$, there exists a chart $(U, \phi) \in \mathcal{A}$ containing p,
2. for any two charts (U, ϕ) and (V, ψ), the *transition function*

$$\psi \circ \phi^{-1} : \underbrace{\phi(U \cap V)}_{\text{open in } \mathbb{H}^k} \to \underbrace{\psi(U \cap V)}_{\text{open in } \mathbb{R}^k}$$

   is smooth,

3. and $\mathcal{A}$ is *maximal* under collections of charts that satisfy the previous conditions.

We refer to the collection of charts $\mathcal{A}$ an *atlas* of M.

Remark. Note that the existence of a maximal atlas is guaranteed by Zorn's Lemma.

Once again, the descrepancy between the intrinsic and extrinsic definitions of manifolds—with or without boundary—is whether we assume M is a subset of Euclidean space. We can see that any smooth k-manifold with boundary embedded in $\mathbb{R}^n$ is an abstract k-manifold, but is the converse true as well? Indeed it is, and it is a theorem that we will not prove in this course.

Theorem 22.6 (Whitney Embedding Theorem)**.** Any (abstract) smooth k-manifold with boundary can be realized as a submanifold embedded in $\mathbb{R}^{2k}$.

Definition 22.7. Suppose that M is a smooth k-manifold with boundary. We say $p \in M$ is an

1. *interior point* if there exists a chart (U, ϕ) containing p such that $\phi(p) \in \mathbb{H}^k_+$,
2. *boundary point* if instead $\phi(p) \in \partial\mathbb{H}^k$.

Proposition 22.8. For a smooth manifold M with boundary, the set of interior points and boundary points are disjoint.

Proof. Suppose (U, ϕ) and (V, ψ) are smooth charts in a given collection $\mathcal{A}$ such that $\phi(p) \in H^k_+$ and $\psi(p) \in \partial H^k$. The transition function $\tau : \phi(U \cap V) \to \psi(U \cap V)$ is a smooth function which maps $\phi(p)$ to $\psi(p)$ and has an invertible derivative at $\phi(p)$. As such, we have a diffeomorphism between an open set in $\mathbb{R}^k$ and an open set in $\mathbb{H}^k$ that is not open in $\mathbb{R}^k$, and we arrive at a contradiction due to the Inverse Function Theorem. □

23 Partitions of Unity

In this lecture, we will list some results about *partitions of unity*. Proofs will be omitted as they are available in Munkres' *Analysis on Manifolds*.

Lemma 23.1. If Q is a rectangle in $\mathbb{R}^n$, then there exists a smooth function $\phi : \mathbb{R}^n \to \mathbb{R}$ such that $\phi(x) > 0$ for $x \in \text{int}(Q)$ and $\phi(x) = 0$ otherwise.

Lemma 23.2. Let $\mathcal{A} := \{U_\lambda\}_{\lambda\in\Lambda}$ be a collection of open sets in $\mathbb{R}^n$ and define $A := \bigcup_{\lambda\in\Lambda} U_\lambda$. Then there exists a countable collection of rectangles $\{Q_i\}_{i=1}^\infty$ contained in A such that

1. $\{\text{int}Q_i\}_{i=1}^\infty$ covers A,
2. each Q_i is a subset of some $U_\lambda \in A$,
3. and each point in A has a neighborhood that intersects only finitely many rectangles in $\{Q_i\}_{i=1}^\infty$.

Definition 23.3. If $\phi : \mathbb{R}^n \to \mathbb{R}$ is a function on $\mathbb{R}^n$, then we define the *support* of ϕ as the closure

$$\text{supp}(\phi) := \overline{\{x \in \mathbb{R}^n : \phi(x) \neq 0\}}.$$

Theorem 23.4 (Partitions of Unity)**.** Let $\mathcal{A} := \{U_\lambda\}_{\lambda\in\Lambda}$ be a collection of open sets and define $A := \bigcup_{\lambda\in\Lambda} U_\lambda$. Then there exists a sequence $\phi_1, \phi_2, \ldots$ of smooth functions $\phi_i : \mathbb{R}^n \to \mathbb{R}$ such that

1. $\phi_i(x) \geq 0$ for all $x \in \mathbb{R}^n$,
2. $\text{supp}(\phi_i)$ is a compact subset of some $U_\lambda \in \mathcal{A}$,
3. each point in A has a neighborhood that intersects only finitely many $\text{supp}(\phi_i)$,
4. and $\sum_{i=1}^\infty \phi_i(x) = 1$ for all $x \in A$.

24 Stokes' Theorem: Integration of Forms

In this lecture, we will lay down the final steps that culminate in *Stokes' Theorem*. We will apply some ideas about partitions of unity from the previous lecture to discuss the integration of differential forms. Recall that there was a descrepancy between how we defined integration in $\mathbb{R}^n$ and the single-variable Fundamental Theorem of Calculus, whereby a result like

$$\int_a^b f'(x)dx = f(b) - f(a)$$

indicates some "orientation" when we write $f(b) - f(a)$ instead of $f(a) - f(b)$.

Definition 24.1. Let $M \subseteq \mathbb{R}^n$ be an m-dimensional submanifold with boundary embedded in $\mathbb{R}^n$. We say M is *orientable* if there exists an open neighborhood $U \subseteq \mathbb{R}^n$ containing M and an m-form $\phi : U \to \bigwedge^m(\mathbb{R}^n)^*$ such that for all local parametrizations $f : W \to M$—where W is an open neighborhood of $\mathbb{H}^m$—the pullback $f^*(\phi)$ is an m-form that is nowhere zero.

The pair $\langle\phi, U\rangle$ *determines an orientation* of M, and we say that two pairs $\langle\phi, U\rangle$ and $\langle\psi, V\rangle$ determine the same orientation if for all local parametrizations $f : W \to M$, there exists a positive function

$g : W \to \mathbb{R}$ such that

$$f^*(\psi) = g \circ f^* \circ \phi.$$

As such, an *orientation* of M is an equivalence class of pairs that determine orientation, and we call a manifold with orientation an *oriented manifold*.

Remark. As a friendly reminder, chart homeomorphisms map from manifold to Euclidean space, whereas local parametrizations are the opposite.

To motivate orientation in a slightly different manner, consider a simpler example where we're trying to describe orientation of a vector space. If V is an n-dimensional vector space, then we can pick n linearly independent vectors $v_1, \ldots, v_n$ and define the wedge product $v_1 \wedge \cdots \wedge v_n$ to be positively oriented. Note that $v_2 \wedge v_1 \wedge v_3 \wedge \cdots \wedge v_n = -v_1 \wedge v_2 \wedge \cdots \wedge v_n$, which shows that the ordered set of vectors $\{v_2, v_1, v_3, \ldots, v_n\}$ has a negative orientation relative to $\{v_1, \ldots, v_n\}$. To clarify, orientation is defined by setting a particular set of vectors as positively oriented and defining other sets relative to this "basis" that we started with.

Example 24.2. The following are some examples of orientable manifolds:

1. The interval $= [0, 1]$ is orientable with orientation dx.
2. Euclidean space $\mathbb{R}^n$ is orientable with orientation $dx_1 \wedge \cdots \wedge dx_n$ (a top form).
3. The n-sphere $S^n \subseteq \mathbb{R}^{n+1}$ is orientable. Our choice of a form which describes orientation is

$$\phi = \sum_{i=1}^{n+1} (-1)^i x_i \; dx_1 \wedge \cdots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \cdots \wedge dx_{n+1}.$$

   For instance, for the local parametrization $f(x_1, \ldots, x_n) = (x_1, \ldots, x_n, \sqrt{1 - (x_1^2 + \cdots + x_n^2)})$, we can check that

$$f^*(\phi) = \frac{(-1)^n}{\sqrt{1 - (x_1^2 + \cdots + x_n^2)}} \; dx_1 \wedge \cdots \wedge dx_n.$$

We are interested in forms because there are independent of coordinates, and we have seen that determinants—which we used to define *orientation preserving* maps in Definition 20.5—are intrinsically built into exterior products. To integrate forms on orientable manifolds, we will patch up a manifold using partitions of unity.

Definition 24.3. Let $\mathcal{A}$ be a collection of open sets in $\mathbb{R}^n$ and let $A := \bigcup_{U \in \mathcal{A}} U$. A *partition of unity subordinate to* $\mathcal{A}$ is a countable collection of smooth functions $\{\phi_i : i \in I\}$ such that

1. $\phi_i(x) \geq 0$ for all $x \in \mathbb{R}^n$,
2. ϕ_i has compact support contained by some $U \in \mathcal{A}$,
3. for all $a \in A$, there exists an open neighborhood V of a such that all but finitely many ϕ_i are identically 0 on V,
4. and $\sum_{i \in I} \phi_i(a) = 1$ for all $a \in A$.

Last class, we showed that partitions of unity exist. Observe that if $K \subseteq \mathbb{R}^n$ is a compact subset and $\{U_i\}_{i \in I}$ is a cover of K and $\{\phi_i : i \in I\}$ is a partition of unity subordinate to $\{U_i\}_{i \in I}$, then all but finitely many ϕ_i vanish identically on K.

We will now shift our discussion towards integration on manifolds. Let $M \subseteq \mathbb{R}^n$ be a compact oriented k-dimenionional submanifold with boundary, $U \subseteq \mathbb{R}^n$ an open set containing M, and ω an m-form on U.

We want to define

$$\int_M \omega.$$

For $p \in M$, we can choose a smooth local parametrization $\phi_p : U_p \to M$ that is *compatible* with orientation—in other words, if $\langle \psi, V \rangle$ represents the orientation of M, then

$$\phi_p^*(\psi) = g \; dx_1 \wedge \cdots \wedge dx_m$$

with $g > 0$. And so, it suffices to consider charts on M that are compatible with the orientation of M.

Because ϕ_p is a local parametrization—hence bijective—$\phi_p(U_p)$ is open with respect to M and there exists an open subset $V_p \subseteq \mathbb{R}^n$ such that

$$\phi_p(U_p) = V_p \cap M.$$

Since M is compact, finitely many of these $\{V_1, \ldots, V_n\}$ cover M. We can define the open cover $\{V_p : p \in M\}$ of M, and the compactness of M implies that a finite subcover $\{V_{p_i}\}_{i=1}^n$ exists. Now that we have an open cover of M, we can define a partition of unity $\{\psi_j\}_{j\in J}$ subordinate to $\{V_{p_i}\}_{i=1}^n$. We have that finitely many ψ_j do not vanish on V, and

$$\omega = \left(\sum_{j \in J} \psi_j \right) \omega.$$

We can reasonably expect

$$\int_M \omega = \sum_{j\in J} \int_M \psi_j \cdot \omega.$$

Note that each $\psi_j \cdot \omega$ is supported on some $V_{i(j)}$—in other words, $\phi_{i(j)}^*(\psi_j \cdot \omega)$ is an m-form with compact support in $U_{i(j)} \subseteq \mathbb{H}^m$. And so, we have

$$\int_M \psi_j \cdot \omega = \int_{U_{i(j)}} \phi_{i(j)}^*(\psi_j \cdot \omega)$$

by a change of variables, which thus implies

$$\int_M \omega := \sum_{j\in J} \int_M \psi_j \cdot \omega.$$

In definiting the integral above, we had to pick a partition of unity. We want to make sure that our definition is well-defined and independent of our choices of partitions of unity. To do so we will use change of variables.

Suppose $\mu_k : U_k' \to M$ for $1 \leq k \leq q$ is another set of local parametrizations. Let $\{\nu_\ell\}_{\ell \in L}$ be a partition of unity subordinate to a collection of open sets $\{W_1, \ldots, W_q\}$ where $W_k \subseteq \mathbb{R}^n$ is an open subset such that $M \cap W_k = \mu_k(U_k')$. The closure of support of ν_ℓ is contained in $W_{j(\ell)}$. To show well-definedness, we want to show that

$$\sum_{\ell \in L} \int_{U'_{j(\ell)}} \mu_{j(\ell)}^*(\mu_\ell \cdot \omega) = \sum_{j \in J} \int_{U_{i(j)}} \psi_{i(j)}^*(\psi_j \cdot \omega).$$

First, note that we can rewrite the left sum as

$$\sum_{j,\ell} \int_{U'_{j(\ell)}} \mu_{j(\ell)}^*(\psi_j \cdot \nu_\ell \cdot \omega)$$

because $\nu_\ell \cdot \omega = \left(\sum_{j \in J} \psi_j\right) \cdot \nu_\ell \cdot \omega$. It suffices to show then that

$$\int_{U'_{j(\ell)}} \mu^*_{j(\ell)}(\psi_j \cdot \nu_\ell \cdot \omega) = \int_{U_{i(j)}} \phi^*_{i(j)}(\psi_j \cdot \nu_\ell \cdot \omega).$$

Observe that the map

$$\mu_{j(\ell)} \circ \phi^{-1}_{i(j)} : \phi^{-1}_{i(j)}\left(V_{i(j)} \cap W_{j(\ell)}\right) \to \mu^{-1}_{j(\ell)}(V_{i(j)} \cap W_{j(\ell)})$$

is an orientation preserving diffeomorphism, so we have

$$\mu^*_{j(k)}(\psi_j \cdot \nu_\ell \cdot \omega) = (\mu_{j(k)} \circ \phi^{-1}_{i(j)})^* \circ \phi^*_{i(j)}(\psi_j \cdot \nu_\ell \cdot \omega).$$

By the change of variables formula, we finally have

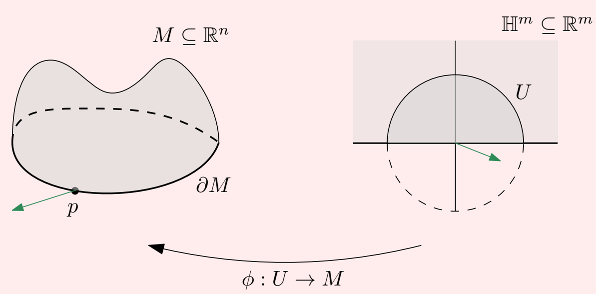
$$\int_{U'_{j(\ell)}} \mu^*_{j(\ell)}(\psi_j \cdot \nu_\ell \cdot \omega) = \int_{U_{i(j)}} \phi^*_{i(j)}(\psi_j \cdot \nu_\ell \cdot \omega).$$

25 Stokes' Theorem: The Conclusion

Last time, we defined a notion of orientation for manifolds and discussed how we integrate m-forms on oriented m-manifolds. In this final lecture, we will use these ideas to prove Stokes' Theorem. But before we continue any further, consider the following definition and lemma.

Definition 25.1. Suppose $M \subset \mathbb{R}^n$ is an oriented m-dimensional manifold with boundary. Then ∂M is an $(m-1)$-dimensional manifold. For $p \in \partial M$, let $U \subseteq \mathbb{H}^m$ be an open neighborhood of 0 and $\phi : U \to M$ a local parametrization such that $\phi(0) = p$.

We say that a *tangent vector* $v \in T_pM = D\phi(0)(\mathbb{R}^m)$ is *outward-pointing* if it is the image of a vector $w \in \mathbb{R}^m$ that does not belong to $\mathbb{H}^m \subseteq \mathbb{R}^m$.



Lemma 25.2. For any compact m-dimensional manifold $M \subseteq \mathbb{R}^n$, there is a smooth map $X : \partial M \to \mathbb{R}^n$ such that $X(p) \in T_pM$ for all $p \in \partial M$ and $X(p)$ is outward-pointing.

Proof. Let $\phi_i : B \cap \mathbb{H}^m \to M$ be a countable collection of local parametrizations of points near ∂M, where B is the unit open ball in $\mathbb{R}^m$. Now define $U_i := \phi_i(\partial H^m \cap B)$, and notice that $\{U_i\}_{i=1}^{\infty}$ is a cover of ∂M. Let $\{\psi_j\}_{j=1}^{\infty}$ be a partition of unity subordinate to $\{U_i\}_{i=1}^{\infty}$, and because ∂M is compact, it suffices to consider a finite subcollection of $\{\psi_j\}_{j=1}^{\infty}$. Now define $X_i : U_i \to \mathbb{R}^n$ by

$$X_i(p) := \underbrace{D(\phi_i)(\phi_i^{-1}(p))(-e_m)}_{\text{always outward-facing}}.$$

Let $\sigma_i := \sum \psi_j$ such that $\operatorname{supp}(\psi_j) \subseteq U_i$, and define $X(p) := \sum_i \varphi_i(p) \cdot X_i(p)$. Notice that $X(p)$ is still outward pointing and is therefore our desired function. $\square$

Consequently, if $M \subseteq \mathbb{R}^n$ is a compact m-manifold with boundary and M is oriented by an m-form ω, then we can define an orientation on ∂M, which is an $(m-1)$-form ν on ∂M. We want to produce a map $\nu : \partial M \to \bigwedge^{k-1}(\mathbb{R}^n)^*$ (should be a neighborhood of ∂M, but we'll take this for now); for $p \in \partial M$, define $\nu(p) := \iota_{X(p)} \circ \omega(p)$ where $\iota_V : \bigwedge^k(V^*) \to \bigwedge^{k-1}(V^*)$ is defined as

$$\iota_V(\phi_i \wedge \cdots \wedge \phi_k) = \sum_{i=1}^{k}(-1)^i \phi_i(v)\ \phi_i \wedge \cdots \wedge \phi_{i-1} \wedge \phi_{i+1} \wedge \cdots \wedge v_k.$$

Example 25.3. Let $M : [0,1]$ with the standard orientation determined by dx. The orientation of $\partial M := \{0,1\}$ will be a 0-form, and at 0, the outward point vector is $-e_1$ and e_1 at 1. Notice that at zero, we have $\iota_{-e_1}\ dx = -1$ and at 1 we have $\iota_{e_1}\ dx = 1$.

Theorem 25.4 (Stokes' Theorem)**.** Let $M \subseteq \mathbb{R}^n$ be a compact, oriented m-dimensional manifold with boundary, and let η be an $(m-1)$-form on M. Then

$$\int_M d\eta = \int_{\partial M} \eta.$$

Here, the orientation on ∂M is induced by that of M.

Proof. For every $p \in M$, there exists a rectangle $Q_p \subseteq \mathbb{H}^k$ and another rectangle $R_p \subseteq \operatorname{int}(Q_p)$ and local parametrization $\phi_p : \operatorname{int}(Q_p) \to M$ that is compatible with orientations on M with $p \in \phi_p(\operatorname{int}(R_p)) := V_p$. We can assume that if $p \in \operatorname{int}(M)$, then the image of ϕ_p is in the interior, and if $p \in \partial M$, then ϕ_p maps $Q_p \cap \partial\mathbb{H}^m$ into ∂M.

Now take a partition of unity $\{\psi_i\}$ subordinate to $\{V_p\}$. Because M is compact, we can take ψ_i to be a finite collection. And so,

$$\eta = \sum \eta_i$$

where $\eta_i = \psi_i \cdot \eta$. It suffices to show the theorem for each η_i. The point is that η_i lies in one of the charts. Let Q be a rectangle $Q_p \in \mathbb{H}^m$ such that the support of η_i is contained in V_p, and let $\phi_i : \int(Q) \to M$ be the corresponding chart. Then

$$\int_M d\eta_i = \int_Q \phi^*(d\eta_i) = \int_Q d(\phi^*\eta_i).$$

Let $\phi^*(\eta_i) := \sum_{j=1}^m f_j(x_1, \ldots, x_m)\ dx_1 \wedge \cdots \wedge dx_{j-1} \wedge dx_{j+1} \wedge \cdots \wedge dx_m$. Then

$$d\phi^*\eta_j = \sum_{j=1}^{n}(-1)^j\ \frac{\partial f_j}{\partial x_j}\ dx_1 \wedge \cdots \wedge dx_m.$$

Let $Q = [a_1, b_1] \times \cdots \times [a_m, b_m]$ and let $Q_j := [a_1, b_1] \times [a_{j-1}, b_{j-1}] \times [a_{j+1}, b_{j+1}] \times \cdots \times [a_m, b_m]$. Then

$$\begin{aligned}\int_Q d(\phi^*\eta_j) &= \sum_{j=1}^{n}(-1)^{j-1}\int_{Q_j}\int_{x_j=a_j}^{x_j=b_j}\frac{\partial f_j}{\partial x_j},\\ &= \sum_{j=1}^{n}(-1)^{j-1}\int_{Q_j}\left(f_j(x_1, \ldots, b_j, \ldots, x_m) - f(x_1, \ldots, a_j, \ldots, x_m)\right).\end{aligned}$$

Note that $f(x_1, \ldots, b_j, \ldots, x) = 0$ because f is on the boundary and the support of $\phi^*\eta_j$ lies in the interior of Q. Notice that $f(x_1, \ldots, b_j, \ldots, x) = 0$ is also true, unless $j = m$ and $a_m = 0$ in that case that $Q \cap \partial\mathbb{H}^m \neq \varnothing$.

Note that the integral is non-zero only if ϕ maps Q_m to the boundary of M. In this case, it gives a local parametrization of the boundary. Moreover, the induced orientation on Q_m is given by

$$\iota_{-e_m}(dx_1 \wedge \cdots \wedge dx_m) = (-1)^m dx_1 \wedge \cdots \wedge dx_{m-1}.$$

Therefore,

$$\begin{aligned}\int_Q d\phi^*(\eta_j) &= (-1)^m \int_{Q_m} f_m(x_1, \ldots, x_{m-1}, 0),\\ &= \int f_m dx_1 \wedge \cdots \wedge dx_{m-1},\\ &= \int_{\partial M} \eta_j.\end{aligned}$$

□

Appendix A: Multilinear Algebra