Spivak Notes

Max Chien

February 2024

Contents

Introduction

This document contains notes taken as personal self-study in Summer 2024 of the book *Calculus on Manifolds*, by Michael Spivak. The notes closely follow the structure of Spivak's text.

Chapter 1

Euclidean Space

1.1 Vector Properties of Euclidean Space

In this course, we study functions over Euclidean space. We will assume knowledge of most of the basic properties of the real numbers, and will only briefly introduce the basic properties of Euclidean space.

Definition 1.1

Euclidean n-space, denoted $\mathbb{R}^n$, is the set of n-tuples

$$(x_1, x_2, \ldots, x_n)$$

such that $x_i \in \mathbb{R}$ for each i.

Euclidean space is intended to align with the "standard" notions of space. That is, $\mathbb{R}^1$ is often referred to as the line, $\mathbb{R}^2$ as the plane, and $\mathbb{R}^3$ as space. Moreover, from linear algebra we can see that $\mathbb{R}^n$ can be considered as an n-dimensional vector space over $\mathbb{R}$, with addition and scalar multiplication defined coordinate-wise, so elements of $\mathbb{R}^n$ will alternately be called points or vectors. In fact, it is the canonical representative of n dimensional vector spaces over $\mathbb{R}$, further justifying its study. We denote by 0 or $\mathbf{0}$ the vector $(0, 0, \ldots, 0)$.

Moreover, $\mathbb{R}^n$ is an example of a *normed* vector space. Specifically, we have

Definition 1.2

Given a vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, define the **norm** of x, denoted $|x|$, by

$$|x| \coloneqq \sqrt{x_1^2 + \ldots + x_n^2}$$

Note that for $n = 1$, the norm aligns with the standard absolute value of real numbers. Briefly, we can verify that the norm as defined here indeed satisfies the definition of a norm on a vector space:

Proposition 1.1

Let $x, y \in \mathbb{R}^n$, and $a \in \mathbb{R}$ be arbitrary. Then we have:

- $|x| \geq 0$, with $|x| = 0$ if and only if $x = 0$.
- $|\sum_{i=1}^n x_i y_i| \leq |x||y|$, with equality if and only if x, y are linearly dependent.
- $|x + y| \leq |x| + |y|$.
- $|ax| = |a||x|$

Beyond being a normed vector space, Euclidean space is also an inner product space. We can define the inner product as follows:

Definition 1.3

Given two vectors $x, y \in \mathbb{R}^n$, define the **inner product** of x and y, denoted $\langle x, y \rangle$, as

$$\langle x, y \rangle \coloneqq \sum_{i=1}^n x_i y_i$$

Similarly, we can verify that this inner product satisfies the definitions of an inner product:

Proposition 1.2

Let $x, x_1, x_2, y, y_1, y_2 \in \mathbb{R}^n$ and $a \in \mathbb{R}$ be arbitrary. Then we have:

- $\langle x, y \rangle = \langle y, x \rangle$ (Symmetric)
- $a \langle x, y \rangle = \langle ax, y \rangle = \langle x, ay \rangle$ (Bilinear)
  $\langle x_1 + x_2, y \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle$
  $\langle x, y_1 + y_2 \rangle = \langle x, y_1 \rangle + \langle x, y_2 \rangle$
- $\langle x, x \rangle \geq 0$, with $\langle x, x \rangle = 0$ if and only if $x = 0$. (Positive definite)

Moreover, given our definitions of the norm and inner product, we can also identify further properties:

Proposition 1.3

Let $x, y \in \mathbb{R}^n$ be arbitrary. Then we have:

- $\langle x, y \rangle \leq |x||y|$ (Cauchy-Schwarz Inequality)
- $|x| = \sqrt{\langle x, x \rangle}$
- $\langle x, y \rangle = \frac{|x+y|^2 - |x-y|^2}{4}$ (Polarization Identity)

Definition 1.4

The **standard basis** of $\mathbb{R}^n$ is given by $\{e_1, \dots, e_n\}$, where $(e_i)_j = \delta_{ij}$, so that e_i has a 1 in the ith coordinate and 0 everywhere else.

Definition 1.5

Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Then denote by $[T]$ the $n \times m$ matrix such that $T(x) = [T]x$ for each $x \in \mathbb{R}^n$. In particular, the ith column of $[T]$ is given by $T(e_i)$.

If $x = (x_1, \dots, x_m)\mathbb{R}^m$ and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, then let us adopt the convention that (x, y) is the concatenation $(x_1, \dots, x_m, y_1, \dots, y_n) \in \mathbb{R}^{m+n}$.

1.2 Topology of Euclidean Space

In many results in single variable analysis, we make use of open and closed intervals, denoted $[a, b]$ and (a, b). The analog of these intervals in $\mathbb{R}^n$ is the notion of a *rectangle* or *k-cell*.

Definition 1.6

Let $A \subseteq \mathbb{R}^m$, $B \subseteq \mathbb{R}^n$. Then define the **Cartesian product** of A and B as $A \times B = \{(a, b) \in \mathbb{R}^{m+n} | a \in A, b \in B\}$. Since this operation is associative, denote by $A_1 \times A_2 \times \ldots \times A_i$ the product of any number of sets.

Definition 1.7

A **closed rectangle**, closed box, or **closed k-cell** is a set of the form $[a_1, b_1] \times \ldots \times [a_n, b_n] \subseteq \mathbb{R}^k$. An **open rectangle**, open box, or **open k-cell** is a set of the form $(a_1, b_1) \times \ldots \times (a_n, b_n) \subseteq \mathbb{R}^k$.

Then similarly to how we use open intervals to define a topology on $\mathbb{R}$, we can use open boxes to define a topology on $\mathbb{R}^n$:

Definition 1.8

A set $U \subseteq \mathbb{R}^n$ is **open** if, for every point $x \in U$, there is some open box $B(x) \subseteq U$ such that $x \in B(x)$. A set $C \subseteq \mathbb{R}^n$ is **closed** if $\mathbb{R}^n \setminus C$ is open.

Remark

Note that because every open box has an open ball inside, and because every open ball has an open box inside, the topology defined by open boxes on $\mathbb{R}^n$ is the same topology defined by open balls on $\mathbb{R}^n$. Thus, for $x \in \mathbb{R}^n$ and $r \in \mathbb{R}$, denote by $B_r(x)$ the **open n-ball** with center x and radius r. That is, $B_r(x) \coloneqq \{y \in \mathbb{R}^n : |x-y| < r\}$. When the dimension is ambiguous, denote this ${}_nB_r(x)$. Then we can alternately use open balls and open boxes as the definition of an open set, depending on which is more convenient.

Definition 1.9

If $A \subseteq \mathbb{R}^n$, then the **interior** of A is the set of points contained in an open rectangle entirely in A.

$$\operatorname{int} A \coloneqq \{x \in \mathbb{R}^n : \text{ there exists an open rectangle } B \text{ s.t. } x \in B \subseteq A\}$$

Define the **exterior** of A to be the set of points contained in an open rectangle entirely in $\mathbb{R}^n \setminus A$.

$$\operatorname{ext} A \coloneqq \{x \in \mathbb{R}^n : \text{ there exists an open rectangle } B \text{ s.t. } x \in B \subseteq \mathbb{R}^n \setminus A$$

Define the **boundary** of A to be the set of points where all open rectangles contain points of both A and $\mathbb{R}^n \setminus A$.

$$\partial A \coloneqq \{x \in \mathbb{R}^n : \forall \text{ open rectangles } B, x \in B \implies B \cap A \neq \varnothing, B \cap \mathbb{R}^n \setminus A \neq \varnothing\}$$

Proposition 1.4

Every set of finitely many points in $\mathbb{R}^n$ is closed.

Proof. Let $C \subseteq \mathbb{R}^n$ be a finite set. Let $x \in \mathbb{R}^n \setminus C$ be arbitrary. Then for each point $y \in C$, $x \neq y$, so $d(x, y) > 0$. Then since there are only finitely many points in C, the quantity $d' = \min\{d(x, y)|y \in C\}$ is defined and greater than 0. So we can define an open ball with radius $d'/2$, which does not contain any points in C. Thus we have an open ball containing x that is a subset of $\mathbb{R}^n \setminus C$. So $\mathbb{R}^n \setminus C$ is open and thus C is closed. $\square$

Definition 1.10

An **open cover** of a set A is a collection $\mathcal{O}$ of open sets such that for any $x \in A$, $x \in U$ for some $U \in \mathcal{O}$. A **subcover** of $\mathcal{O}$ is a subcollection of $\mathcal{O}$ which is also a cover for A.

Definition 1.11

A set K is **compact** if for any open cover $\mathcal{O}$ of K, there exists a finite subcover $\mathcal{U}$ of $\mathcal{O}$.

In particular, we can derive certain theorems to identify compact sets.

Theorem 1.5: Heine-Borel Theorem

The closed interval $[a, b]$ is compact.

Proof. Let $\mathcal{U}$ be some open cover of $[a, b]$. Then consider the set

$$A = \{x \in [a, b] : [a, x] \text{ is covered by a finite number of sets in } \mathcal{U}\}$$

The goal is to prove that $b \in A$. First, consider $\alpha = \sup A$ (since this set is bounded above and nonempty). We have $\alpha \leq b$, so $\alpha \in [a, b]$ and thus $\alpha \in U_1$ for some $U_1 \in \mathcal{U}$. Since U_1 is open and α is the supremum of A, there is some $a \leq x < \alpha$ with $x \in U_1$. Then we have $x \in A$, so some finite number of open sets in $\mathcal{U}$ cover $[a, x]$, and U_1 covers $[x, \alpha]$, so a finite number of sets cover $[a, \alpha]$ and thus $\alpha \in A$.

Now suppose $\alpha < b$. Then there is some $y \in U_1$ such that $\alpha < y < b$. But if $[a, \alpha]$ is covered by a finite number of open sets, then so is $[a, y]$, so $y \in A$, contradicting $\alpha = \sup A$. So we must have $\alpha = b$, completing the proof. □

Note that if $B \in \mathbb{R}^m$ is compact and $x \in \mathbb{R}^n$, then the set $\{x\} \times B$ is clearly compact. gMoreover, given any cover of $\{x\} \times B$, the finite subcovers have a "minimum width":

Theorem 1.6

If $B \subseteq \mathbb{R}^m$ is compact and $x \in \mathbb{R}^n$, then given any open cover $\mathcal{U}$ of $\{x\} \times B$, there is some open set $U \in \mathbb{R}^n$ such that $U \times B$ is covered by a finite number of sets in $\mathcal{U}$.

Proof. Take some finite subcover $\mathcal{U}'$ of $\mathcal{U}$. Then we just need to find a set U such that $U \times B$ is covered by $\mathcal{U}'$.

For each $y \in B$, (x, y) is in some open set $O \in \mathcal{U}'$, so there is an open box $U_x \times V_y$ such that $(x, y) \in U_x \times V_y \subseteq O$. Then consider the collection $(V_y)_{y \in B}$. This set covers B, which is compact, so we can pick a finite number $V_1, \ldots, V_k$. Let $U = \bigcap U_i$. Then for any $(x_1, y_1) \in U \times B$, we have $y_1 \in V_i$ for some $1 \leq i \leq n$, and $x_1 \in U_i$, so $x_1 \in U_i \times V_i \subseteq O'$ for some $O' \in \mathcal{U}'$. Thus $\mathcal{U}'$ covers $U \times B$. □

Corollary

If $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ are compact, then $A \times B \subseteq \mathbb{R}^{n+m}$ is compact.

Proof. Let $\mathcal{O}$ be some open cover of $A \times B$. Then for each $x \in A$, $\mathcal{O}$ covers $\{x\} \times B$, so there is some U_x such that a finite subcover $O_{1x}, \ldots, O_{kx}$ covers $U_x \times B$ and $x \in U_x$. Then the collection $(U_x)_{x \in A}$ covers A, so there is a finite subcover $U_{x_1}, \ldots, U_{x_j}$ that covers A. Then the sets $O_{1x_1}, \ldots, O_{kx_1}, \ldots, O_{1x_j}, \ldots, O_{k'x_j}$ form a finite subcover of $\mathcal{O}$ that covers $A \times B$. So $A \times B$ is compact. □

Corollary

A product $A_1 \times \ldots \times A_k$ is compact if each A_i is. A closed rectangle is compact.

Proof. Induct on k using the previous corollary. □

This gives an important result which allows us to work with compactness much more easily in $\mathbb{R}^n$ (though it is not necessarily true for other topological vector spaces).

Theorem 1.7

A set $K \subseteq \mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof. ($\implies$) Suppose $K \subseteq \mathbb{R}^n$ is compact. The collection of open rectangles $(i-1, i+1) \times (j-1, j+1) \ldots \times (k-1, k+1)$ for $i, j, \ldots, k \in \mathbb{Z}$ covers $\mathbb{R}$, so it covers K. Then a finite number of these boxes covers K, so it is bounded.

($\impliedby$) Suppose $K \subseteq \mathbb{R}^n$ is closed and bounded. Then there exists a closed rectangle B with $K \subseteq B$. From the previous corollary, we know that B is compact. Then take some cover of K, $\mathcal{O} = \{O_1, \ldots\}$. Now let $\mathcal{U}$ consist of all the sets in $\mathcal{O}$, as well as the set $\mathbb{R}^n \setminus K$ (which is open since K is closed). $\mathcal{U}$ covers $\mathbb{R}^n$, so it covers B. Then we can take a finite subcollection $\mathcal{U}'$ of $\mathcal{U}$. Then $\mathcal{U}'$ covers B as well as K, and in order to create a subcollection of $\mathcal{O}$, we simply remove $\mathbb{R}^n \setminus K$ if it is in $\mathcal{U}'$ to get $\mathcal{O}'$. So K is compact. □

1.3 Functions and Continuity

In this section, we study **vector valued functions**, which are functions $f : \mathbb{R}^n \to \mathbb{R}^m$, or more generally, $f : A \to B$ for some $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$. We briefly list a few definitions related to these functions that should be familiar to the reader.

Definition 1.12

If $f : A \to B$, then the **image** of $C \subseteq A$ is $f(C) = \{f(x) : x \in C\}$. The **preimage** of $D \subseteq B$ is $f^{-1}(D) = \{y \in A : f(y) \in D\}$.

Definition 1.13

If $f : A \to \mathbb{R}^m$ and $g : B \to \mathbb{R}^n$ with $B \subseteq \mathbb{R}^m$, then the **composition** is defined as $(g \circ f)(x) = g(f(x))$, with domain $A \cap f^{-1}(B)$.

Definition 1.14

If $f : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one, then the **inverse** of f is the function $f^{-1} : f(\mathbb{R}^n) \to \mathbb{R}^n$ which takes $x \in f(\mathbb{R}^n)$ to the unique $y \in \mathbb{R}^m$ such that $f(y) = x$.

In addition to studying a vector valued function f, we can also study the *component functions* which encode its behavior on each axis individually.

Definition 1.15

If $f : A \to \mathbb{R}^m$, then f defines m **component functions** $f^1, f^2, \ldots, f^m$ such that $f(x) = (f^1(x), \ldots, f^m(x))$. Similarly, for any functions $g_1, \ldots, g_m : A \to \mathbb{R}$, we denote by $(g_1, \ldots, g_m)$ the function $f : A \to \mathbb{R}^m$ which satisfies $f(x) = (g_1(x), \ldots, g_m(x))$.

Note that the above definition implies that we can write $f = (f^1, \ldots, f^m)$.

Definition 1.16

Let $\pi : \mathbb{R}^n \to \mathbb{R}^n$ be the identity function. Then $\pi = (\pi^1, \ldots, \pi^n)$. Then π^i is called the i-th **projection function**, such that $\pi^i(x)$ gives the ith coordinate of x.

With the above out of the way, we now turn our attention to limits of functions, which will prove important as we continue our study of multivariate calculus.

Definition 1.17

We write $\lim_{x\to a} f(x) = b$ (the **functional limit**) if, for any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that whenever $0 < |x - a| < \delta$, we have $|f(x) - b| < \varepsilon$.

Just as the above definition is reproduced from single-variable analysis (with the exception of generalizing the notion of distance in $\mathbb{R}^n$), we have an analogous definition of continuity:

Definition 1.18

A function $f : A \to \mathbb{R}^m$ is **continuous** at a point $a \in A$ if $\lim_{x\to a} f(x) = f(a)$. If f is continuous at each $a \in A$, we simply say that f is continuous.

Alternatively, we can utilize the topological nature of $\mathbb{R}^n$, which we discussed in the last section, to characterize continuity using the topological definition instead.

Proposition 1.8

A function $f : A \to \mathbb{R}^m$ for $A \subseteq \mathbb{R}^n$ is continuous if and only if for every open set $U \subseteq \mathbb{R}^m$, there is an open set $V \subseteq \mathbb{R}^n$ such that $f^{-1}(U) = V \cap A$.

Proof. ($\Longrightarrow$) Suppose f is continuous. Then let $U \subseteq \mathbb{R}^m$. For each point $x \in f^{-1}(U)$, $f(x) \in U$ which is open. Thus, there is an open ball $B_{\varepsilon_x}(f(x)) \subseteq U$, there is an open ball $B_{\varepsilon_x}(f(x)) \subseteq U$, and a corresponding open ball $B_{\delta_x}(x) \subseteq f^{-1}(B_{\varepsilon_x}(f(x)))$. Then the set $V = \bigcup_{x \in f^{-1}(U)} B_{\delta_x}(x)$ is an open set.

Moreover, by construction, for any point $y \in V \cap A$, $y \in B_{\delta_x}(x)$ for some x, implying that $f(y) \in B_{\varepsilon_x}(f(x)) \subseteq U$ (which is defined since $y \in A$). So $V \cap A \subseteq f^{-1}(U)$. For any point $x \in f^{-1}(U)$, $x \in B_{\delta_x}(x)$, so $x \in V$. Moreover, any point in $f^{-1}(U)$ is in the domain of f, so $x \in V \cap A$, and thus $f^{-1}(U) = V \cap A$.

($\Longleftarrow$) Suppose every open set $U \subseteq \mathbb{R}^m$ has an associated open set $V \subseteq \mathbb{R}^n$ such that $f^{-1}(U) = V \cap A$. Then pick a point $a \in A$, and let $\varepsilon > 0$ be arbitrary. Then the open ball $B_\varepsilon(f(a))$ has an associated open set V. Moreover, $a \in B_\varepsilon(f(a)) \implies a \in V \cap A \implies a \in V$, so there exists an open ball $B_\delta(a) \subseteq V$. Then for any $x \in A$ with $|x - a| < \delta$, $x \in B_\delta(a) \subseteq V$, so $x \in f^{-1}(B_\varepsilon(f(a)))$, and thus $f(x) \in B_\varepsilon(f(a))$. So $\lim_{x \to a} f(x) = f(a)$. $\square$

When $A = \mathbb{R}^n$, this condition can be phrased as saying "the preimage of every open set is open." Analogously, a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous if and only if the preimage of every closed set is closed. Note that it is not necessarily true that the *image* of every open set is open. For instance, the function $f(x) = x^2$ maps the open set $\mathbb{R}$ to the set $[0, \infty)$, which is not open. However, this condition does imply that for any open set which is not also closed (the only examples are $\varnothing$ and $\mathbb{R}^n$), the image is not closed. Thus, continuity allows us to infer openness *backward* through the function.

In contrast, compactness is passed *forward* through continuous functions, which is another reason that it is useful for our study.

Theorem 1.9

If $f : A \to \mathbb{R}^m$ is continuous and $A \subseteq \mathbb{R}^n$ is compact, then $f(A)$ is compact.

Proof. Pick an open cover $\mathcal{O}$ of $f(A)$. Then by the proposition, for each open set $O \in \mathcal{O}$ there exists an open set $U \in \mathbb{R}^n$ such that $U \cap A = f^{-1}(O)$. Then the collection $\mathcal{U}$ of all such U covers A, so we pick a finite number $U_1, \ldots, U_n$. Then the finite cover $O_1, \ldots, O_n$ cover $f(A)$. So $f(A)$ is compact. $\square$

One disadvantage of these definitions of continuity is that they are binary in nature: a function is either continuous or discontinuous at a certain point. The following definition allows us to measure how discontinuous a function is at a certain point.

Definition 1.19

Let $f : A \to \mathbb{R}^m$ with $A \subseteq \mathbb{R}^n$ bounded, and let $a \in A$. Define

$M(f, a, \delta) = \sup\{f(x) : x \in A, |x - a| < \delta\}, m(f, a, \delta) = \inf\{f(y) : y \in A, |y - a| < \delta\}$

Then the **oscillation** of f at a, denoted $o(f, a)$, is defined as

$$o(f, a) = \lim_{\delta \to 0} [M(f, a, \delta) - m(f, a, \delta)]$$

which always converges since it decreases as $\delta \to 0$ and is bounded below by 0.

In agreement with the intuition for $o(f, a)$ as measuring the discontinuity of f at a, we have the following theorem:

Theorem 1.10

A function $f : A \to \mathbb{R}^m$ with $A \subseteq \mathbb{R}^n$ bounded is continuous at $a \in A$ if and only if $o(f, a) = 0$.

Proof. ($\implies$) Suppose f is continuous at a. Let $\varepsilon > 0$ be arbitrary. Then there exists $\delta > 0$ such that for any $x \in A$ with $|x - a| < \delta$, we have

$$|f(x) - f(a)| < \varepsilon/2 \implies f(a) - \frac{\varepsilon}{2} < f(x) < f(a) + \frac{\varepsilon}{2}$$

Then $M(f, a, \delta) - m(f, a, \delta) < \varepsilon$. So $o(f, a) < \varepsilon$ for every positive ε, and of course $o(f, a) \geq 0$, so $o(f, a) = 0$.

($\impliedby$) Suppose $o(f, a) = 0$. Then let $\varepsilon > 0$ be arbitrary. Since $\lim_{\delta \to 0}[M(f, a, \delta) - m(f, a, \delta)] = 0$, we can pick δ such that $M(f, a, \delta) - m(f, a, \delta) < \varepsilon$. Then for any $x \in A$ with $|x - a| < \delta$,

$$f(x) \leq M(f, a, \delta) < \varepsilon + m(f, a, \delta) < \varepsilon + f(a)$$

Similarly, $f(x) \geq f(a) - \varepsilon$. So $|f(x) - f(a)| < \varepsilon$. Thus f is continuous at a. $\square$

Proposition 1.11

Let $A \subseteq \mathbb{R}^n$ be closed, and let $f : A \to \mathbb{R}^m$ be bounded. For $\varepsilon > 0$, the set $O_\varepsilon = \{x \in A : o(f, x) \geq \varepsilon\}$ is closed.

Proof. We wish to show that $\mathbb{R}^n \setminus O_\varepsilon$ is open. Pick a point $y \in \mathbb{R}^n \setminus O_\varepsilon$. If $y \notin A$, then $y \in \mathbb{R}^n \setminus A$ open so there exists an open rectangle $B \subseteq \mathbb{R}^n \setminus A \subseteq \mathbb{R}^n \setminus O_\varepsilon$ such that $y \in B$.

If $y \in A$, then $o(f, y) < \varepsilon$. Then there exists $B_\delta(y)$ with $M(f, y, \delta) - m(f, y, \delta) < \varepsilon$. I claim that any point $z \in B_\delta(y)$ has $o(f, z) < \varepsilon$. Pick δ' small enough that $B_{\delta'}(z) \subseteq B_\delta(y)$. Then $M(f, z, \delta') \leq M(f, y, \delta)$ and $m(f, z, \delta') \geq m(f, z, \delta)$, so $M(f, z, \delta') - m(f, z, \delta)' \leq M(f, y, \delta) - m(f, y, \delta) < \varepsilon$. So $o(f, z) < \varepsilon$, and thus $B_\delta(y) \subseteq \mathbb{R}^n \setminus O_\varepsilon$, so $\mathbb{R}^n \setminus O_\varepsilon$ is closed and O_ε is open. $\square$

Chapter 2

Differentiation

2.1 Basic Definitions

We now turn our attention to the first major topic of this book; namely, the generalization of differentiation to functions of the form $f : \mathbb{R}^n \to \mathbb{R}^m$. To do so, first recall that $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ if there exists a number $f'(a)$ such that

$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h} = f'(a)$$

We cannot directly use this formula to define vector valued differentiation. First, the quotient would not even make sense when dividing vectors, and even if absolute value bars are taken, it would often be the case that this limit does not exist. However, we can rearrange this equation as

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - f'(a)h}{h} = 0$$

In other words, our new condition is that there is a linear transformation $\lambda(h) = f'(a)(h)$ such that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - \lambda(h)}{h} = 0$$

Conceptually, this is the statement that f is approximated well near a by $f(a) + \lambda$. This interpretation extends nicely to higher dimensions:

Definition 2.1

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable** at $a \in \mathbb{R}^n$ if there exists a linear transformation $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that

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

In this case, λ is denoted $Df(a)$ and is called the **derivative** of f at a.

To justify uniqueness, we prove the following.

Proposition 2.1

If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^n$ then there exists a unique linear transformation $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that

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

Proof. Existence follows from the definition of differentiability. Suppose that λ, μ are two linear transformations which satisfy the above. Then we have

$$\begin{aligned}
\lim_{h \to 0} \frac{|\lambda(h) - \mu(h)|}{|h|} &= \lim_{h \to 0} \frac{|\lambda(h) + f(a) - f(a+h) - \mu(h) - f(a) + f(a+h)|}{|h|} \\
&\leq \lim_{h \to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} + \lim_{h \to 0} \frac{|f(a+h) - f(a) - \mu(h)|}{|h|} \\
&= 0
\end{aligned}$$

Picking any $x \neq 0 \in \mathbb{R}^n$, and any $t \neq 0$,

$$\frac{|\lambda(x) - \mu(x)|}{|x|} = \frac{t}{t} \frac{|\lambda(x) - \mu(x)|}{|x|} = \frac{|\lambda(tx) - \mu(tx)|}{|tx|}$$

But we just showed that

$$\lim_{t \to 0} \frac{|\lambda(tx) - \mu(tx)|}{|tx|} = 0$$

and $\frac{|\lambda(x) - \mu(x)|}{|x|}$ is constant so it must be 0. Thus

$$\frac{|\lambda(x) - \mu(x)|}{|x|} = 0 \implies \lambda = \mu$$ □

We also are often interested in the matrix form of $Df(a)$, so we give it a special notation.

Definition 2.2

If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable, then the **Jacobian matrix** of f is the $m \times n$ matrix

$$f'(a) \coloneqq [Df(a)]$$

Lastly, we note that although many of the theorems presented in this chapter will assume that f is defined on all of $\mathbb{R}^n$, it is often only necessary that f is defined on an open set containing a, so we lose little generality.

2.2 Basic Theorems

As in single variable analysis, the $\varepsilon - \delta$ definition of continuity is often quite cumbersome to work with in practice. Thus, we present a number of theorems in this section which will allow us to easily prove differentiability and calculate derivatives.

Theorem 2.2: Chain Rule

Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentfiable at a, and suppose $g : \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $f(a)$. Then $g \circ f : \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at a with derivative given by

$$D(g \circ f)(a) = Dg(f(a)) \circ Df(a)$$

which can also be written

$$(g \circ f)'(a) = g'(f(a)) \cdot f'(a)$$

Remark

When $n = m = p = 1$, this reduces to the single variable form of the chain rule.

Proof. Here, it will be more convenient to work with the errors of these functions relative to their derivatives:

$$\begin{cases} \varphi(x) \coloneqq f(x) - f(a) - Df(a)(x-a) \\ \psi(x) \coloneqq g(x) - g(f(a)) - Dg(f(a))(x-a) \\ \rho(x) \coloneqq g(f(x)) - g(f(a)) - Dg(f(a))(Df(a)(x-a)) \end{cases}$$

By the definition of the derivatives, we know that

$$\lim_{x \to a} \frac{|\varphi(x)|}{|x-a|} = 0$$

and

$$\lim_{x \to f(a)} \frac{|\psi(x)|}{|x-f(a)|} = 0$$

We want to show that

$$\lim_{x \to a} \frac{|\rho(x)|}{|x-a|} = 0$$

Expanding and using linearity, we have

$$\begin{aligned} \rho(x) &= g(f(x)) - g(f(a)) - Dg(f(a))(Df(a)(x-a)) \\ &= g(f(x)) - g(f(a)) - Dg(f(a))(f(x) - f(a) - \varphi(x)) \\ &= g(f(x)) - g(f(a)) - Dg(f(a))(f(x) - f(a)) + Dg(f(a))(\varphi(x)) \\ &= \psi(f(x)) + Dg(f(a))(\varphi(x)) \end{aligned}$$

Let $\varepsilon > 0$ be arbitrary. Then there exists $\delta > 0$ such that whenever $|f(x) - f(a)| < \varepsilon$,

$$|\psi(f(x))| < \varepsilon |f(x) - f(a)|$$

Since f is continuous, there exists $\delta' > 0$ such that whenever $|x - a| < \delta'$, $|f(x) - f(a)| < \delta$. Then whenever $|x - a| < \delta'$,

$$\begin{aligned} |\psi(f(x))| &< \varepsilon |f(x) - f(a)| \\ &= \varepsilon |\varphi(x) + Df(a)(x-a)| \\ &\leq \varepsilon |\varphi(x)| + \varepsilon |Df(a)(x-a)| \end{aligned}$$

By Exercise 1-10, there exists M_1 such that

$$|Df(a)(x-a)| \leq M_1|x-a|$$

so we have

$$|\psi(f(x))| \leq \varepsilon(|\varphi(x)| + M_1|x-a|)$$

Thus

$$0 \leq \frac{|\psi(f(x))|}{|x-a|} \leq \varepsilon \frac{|\varphi(x)|}{|x-a|} + \varepsilon M_1$$

so

$$0 \leq \lim_{x \to a} \frac{|\psi(f(x))|}{|x-a|} \leq \varepsilon \lim_{x \to a} \frac{|\varphi(x)|}{|x-a|} + \varepsilon M_1 = \varepsilon M_1$$

for all $\varepsilon > 0$, and thus we have

$$\lim_{x \to a} \frac{|\psi(f(x))|}{|x-a|} = 0$$

For the second term,

$$\lim_{x \to a} \frac{|Dg(f(a))(\varphi(x))|}{|x-a|} = \lim_{x \to a} \frac{|Dg(f(a))(\varphi(x))|}{|\varphi(x)|} \frac{|\varphi(x)|}{|x-a|}$$

Since $Dg(f(a))$ is linear, Exercise 1-10 tells us that there exists $M > 0$ such that for any h

$$\frac{|Dg(f(a))h|}{|h|} < M$$

so the first factor is bounded, and the second goes to zero, so we have

$$\lim_{x \to a} \frac{|Dg(f(a))(\varphi(x))|}{|x-a|} = 0$$

and thus

$$\lim_{x \to a} \frac{|\rho(x)|}{|x-a|} = 0$$

which implies that

$$D(g \circ f)(a) = Dg(f(a)) \circ Df(a) \qquad \square$$

Theorem 2.3

1. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a constant function, then
$$Df(a) = 0$$
2. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then
$$Df(a) = f$$
3. If $f : \mathbb{R}^n \to \mathbb{R}^m$, then f is differentiable at $a \in \mathbb{R}^n$ if and only if each component function f^i is, and in this case
$$Df(a) = (Df^1(a), \ldots, Df^m(a))$$
In matrix form, $f'(a)$ is an $m \times n$ matrix with $(f^i)'(a)$ as its ith row.
4. Let $s : \mathbb{R}^2 \to \mathbb{R}$ be the sum function, defined by $s(x, y) = x + y$. Then
$$Ds(a, b) = s$$
5. Let $p : \mathbb{R}^2 \to \mathbb{R}$ be the product function, defined by $p(x, y) = xy$. Then
$$Dp(a, b)(x, y) = bx + ay$$

1. *Proof.* Suppose f is constant. Let $a \in \mathbb{R}^n$ be arbitrary. Then
$$\lim_{x \to a} \frac{|f(x) - f(a) - 0|}{|x - a|} = \lim_{x \to a} 0 = 0$$
so $Df(a) = 0$. □
2. *Proof.* Suppose f is linear. Let $a \in \mathbb{R}^n$ be arbitrary. Then
$$\lim_{x \to a} \frac{|f(x) - f(a) - f(x - a)|}{|x - a|} = \lim_{x \to a} \frac{|f(x - a) - f(x - a)|}{|x - a|} = 0$$
so $Df(a) = f$. □
3. *Proof.* ($\implies$) Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^m$. Then any component function $f^i : \mathbb{R}^n \to \mathbb{R}$ is simply the composition $\pi^i \circ f$, where π^i is the ith projection function. π^i is linear, so by part 2 of this theorem it is also differentiable, and the chain rule tells us that $f^i = \pi \circ f$ is also differentiable.

   ($\impliedby$) Now suppose each component function is differentiable at $a \in \mathbb{R}^n$, and define
$$\lambda = (Df^1(a), \ldots, Df^m(a))$$
   Then the function $f(a + h) - f(a) - \lambda(h)$ has components
$$(f^1(a + h) - f^1(a) - Df^1(a)(h), \ldots, f^m(a + h) - f^m(a) - Df^m(a)(h))$$

so that

$$|f(a+h) - f(a) - \lambda(h)| \leq \sum_{i=1}^{m} |f^i(a+h) - f^i(a) - Df^i(a)(h)|$$

and thus

$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} \leq \sum_{i=1}^{m} \lim_{h \to 0} \frac{|f^i(a+h) - f^i(a) - Df^i(a)(h)|}{|h|} = 0$$

so that

$$Df(a) = (Df^1(a), \ldots, Df^m(a))$$ □

4. *Proof.* s is linear, so this follows from part 2. □

5. *Proof.* Let $\lambda(x, y) = bx + ay$. Then

$$\begin{aligned}
\lim_{(h,k) \to \mathbf{0}} \frac{|p(a+h, b+k) - p(a,b) - \lambda(h,k)|}{|(h,k)|} &= \lim_{(h,k) \to \mathbf{0}} \frac{|hk|}{|(h,k)|} \\
&\leq \lim_{(h,k) \to \mathbf{0}} \frac{h^2 + k^2}{\sqrt{h^2 + k^2}} \\
&= \lim_{(h,k) \to \mathbf{0}} \sqrt{h^2 + k^2} \\
&= 0
\end{aligned}$$ □

Using the sum and product functions, we can now prove the multivariate equivalent of the sum and product rules from single variable analysis.

Theorem 2.4

If $f, g : \mathbb{R}^n \to \mathbb{R}$ are differentiable at $a \in \mathbb{R}^n$, then

$$D(f+g)(a) = Df(a) + Dg(a)$$

and

$$D(fg)(a) = g(a)Df(a) + f(a)Dg(a)$$

If $g(a) \neq 0$, then

$$D(f/g)(a) = \frac{g(a)Df(a) - f(a)Dg(a)}{[g(a)]^2}$$

Proof. Note that we can express sums and products of ($\mathbb{R}$-valued functions) as compositions of the functions with the functions $s, p : \mathbb{R}^2 \to \mathbb{R}$ from the previous theorem.

Specifically, $f + g = s \circ (f, g)$. Then

$$\begin{aligned}
D(f+g)(a) &= D(s \circ (f,g))(a) \\
&= Ds(f(a), g(a)) \circ D(f,g)(a) \\
&= s \circ (Df(a), Dg(a)) \\
&= Df(a) + Dg(a)
\end{aligned}$$

Similarly, $fg = p \circ (f, g)$, Then

$$\begin{aligned} D(fg)(a) &= D(p \circ (f,g))(a) \\ &= Dp(f(a), g(a)) \circ D(f,g)(a) \\ &= Dp(f(a), g(a)) \circ (Df(a), Dg(a)) \\ &= g(a)Df(a) + f(a)Dg(a) \end{aligned}$$

Finally, let $h : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ be defined by $x \mapsto 1/x$. Since we know $g(a) \neq 0$, then we have $f/g = f * (h \circ g)$. We also know from single variable calculus that $Dh(x) = -\frac{1}{x^2}$. Using the product rule we just derived, we have

$$\begin{aligned} D(f/g)(a) &= D(f * (h \circ g))(a) \\ &= (h \circ g)(a)Df(a) + f(a)D(h \circ g)(a) \\ &= \frac{Df(a)}{g(a)} + f(a)Dh(g(a))Dg(a) \\ &= \frac{g(a)Df(a)}{[g(a)]^2} - \frac{f(a)Dg(a)}{[g(a)]^2} \\ &= \frac{g(a)Df(a) - f(a)Dg(a)}{[g(a)]^2} \end{aligned}$$

$\square$

The above theorems allow us, at least in theory, to differentiate vector-valued functions which have components given by sums, products, and quotients of the input components, as well as of single-variable differentiable functions and compositions thereof. However, using the rules above is not always the most convenient in practice.

Example 2.1

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x, y) = \sin(xy^2) = \sin \circ (\pi^1 \cdot [\pi^2]^2)$$

Then we have

$$\begin{aligned} f'(a, b) &= \sin'(ab^2)(\pi^1 \cdot [\pi^2]^2)'(a, b) \\ &= \cos(ab^2)[b^2(\pi^1)'(a, b) + a([\pi^2]^2)'(a, b)] \\ &= \cos(ab^2)[b^2\pi^1 + a(2\pi^2(a, b))(\pi^2)'(a, b)] \\ &= \cos(ab^2)[b^2\pi^1 + 2ab\pi^2] \\ &= \cos(ab^2) \cdot (b^2, 2ab) \\ &= (b^2 \cos(ab^2), 2ab \cos(ab^2)) \end{aligned}$$

2.3 Partial Derivatives

Although the results of the previous section are helpful in assuring us of differentiability of functions, the application of those theorems is often not very efficient, as can be seen in the example at the end of the previous section. Thus, we instead develop the theory of partial derivatives, which will allows us to differentiate these functions much more quickly.

Definition 2.3

If $f : \mathbb{R}^n \to \mathbb{R}$ and $\overrightarrow{a} \in \mathbb{R}^n$, then the i-th **partial derivative** of f at $\overrightarrow{a}$, if it exists, is the limit

$$D_i f(\overrightarrow{a}) = \lim_{h \to 0} \frac{f(\overrightarrow{a} + he_i) - f(\overrightarrow{a})}{h}$$

In other words, the ith partial derivative is the single variable derivative of the function $g_i(x) = f(a_1, \dots, x, \dots, a_n)$ which is produced by treating all the variables except the ith as constant.

Example 2.2

Let $f(x, y) = \sin(xy^2)$. Then by treating y as constant,

$$D_1 f(x, y) = y^2 \sin(xy^2)$$

and treating x as constant,

$$D_2 f(x, y) = 2xy \sin(xy^2)$$

Example 2.3

Let $f(x, y) = x^y$. Then treating y as constant,

$$D_1 f(x, y) = yx^{y-1}$$

Treating x as constant,

$$D_2 f(x, y) = x^y \ln x$$

Assuming that $D_i f$ exists at all points in $\mathbb{R}^n$, we obtain another function $\mathbb{R}^n \to \mathbb{R}$, and thus we can attempt to take another partial derivative of this function. The notation for repeated partial differentiation is "inside out," that is,

$$D_j(D_i f)(x) = D_{i,j} f(x)$$

However, the order of mixed partial derivatives is irrelevant for many common functions:

Theorem 2.5

If $D_{i,j}f$ and $D_{j,i}f$ are continuous in an open set containing $\overrightarrow{a}$, then

$$D_{i,j}f(\overrightarrow{a}) = D_{j,i}f(\overrightarrow{a})$$

Proof. This proof is Exercise 3-28. □

By repeatedly taking mixed partial derivatives of higher orders, we can continue to apply this theorem. In particular, if each partial derivative of f of each order is continuous, then f is said to be C^∞. In this case, the order of partial differentiation is always irrelevant.

Theorem 2.6

Let $A \subseteq \mathbb{R}^n$. If $f : A \to \mathbb{R}$ attains a maximum (or minimum) at a point $\overrightarrow{a} \in \operatorname{int} A$ and $D_i f(\overrightarrow{a})$ exists, then $D_i f(\overrightarrow{a}) = 0$.

Proof. Let $g_i : \mathbb{R} \to \mathbb{R}$ be defined by

$$g_i(x) = (a_1, \ldots, x, \ldots, a_n)$$

Then g_i is defined in an open interval around a_i, and attains a maximum there, so $g_i'(a_i) = 0$, and thus $D_i f(\overrightarrow{a}) = g_i'(a_i) = 0$. □

As in single variable calculus, the above theorem only gives candidate extremal points. Moreover, we still have to check boundary points separately. However, when in single variable calculus this was only a problem of evaluating a function at 2 points, in multivariable calculus, the boundary may not be discrete at all.

2.4 Derivatives

By computing some partial derivatives of functions and comparing them to their derivatives, the reader may observe a correspondence between the two. Of course, this correspondence, which allows for the easy computation of derivatives, was our original motivation for studying partial derivatives. Thus we are retroactively justified in this study, and this correspondence can be summarized in the following theorem:

Theorem 2.7

If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\overrightarrow{a} \in \mathbb{R}^n$, then $D_j f^i(\overrightarrow{a})$ exists for $1 \leq i \leq m, 1 \leq j \leq n$, and $f'(\overrightarrow{a})$ is the $m \times n$ matrix where $[f'(\overrightarrow{a})]_{ij} = D_j f^i(\overrightarrow{a})$.

Proof. We only need to prove this for the case $m = 1$, since we already know that the ith row of $f'(\overrightarrow{a})$ is given by $(f^i)'(a_i)$.

Fix j, and let $h : \mathbb{R} \to \mathbb{R}^n$ be defined by $h(t) = \overrightarrow{a} + te_j$. Then $D_j f(\overrightarrow{a}) = D(f \circ h)(0)$. By the chain rule,

$$\begin{aligned} D_j f(\overrightarrow{a}) &= (f \circ h)'(0) \\ &= f'(h(0))h'(0) \\ &= f'(\overrightarrow{a}) \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix} \end{aligned}$$

The right side of this equation is the jth entry of $f'(\overrightarrow{a})$, showing that $D_j f(a)$ exists. This extends easily for all m. $\square$

While the converse of this theorem is false, we can add another condition to make it true.

Definition 2.4

If $f : \mathbb{R}^n \to \mathbb{R}^m$, then f is called **continuously differentiable** at a if all $D_j f^i(x)$ exist in an open set containing a and if each function $D_j f^i$ is continuous at a.

Theorem 2.8

If $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable at a, then $Df(a)$ exists.

Proof. Suppose f is continuously differentiable at $\overrightarrow{a}$. Then each $D_j f^i(\overrightarrow{a})$ exists. Define $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ by

$$\lambda(x_1, \ldots, x_n) = \left(\sum_{j=1}^n D_j f^1(\overrightarrow{a}) x_j, \ldots, \sum_{j=1}^n D_j f^m(\overrightarrow{a}) x_j \right)$$

Then we have

$$\lim_{\overrightarrow{h} \to \mathbf{0}} \frac{\left| f(\overrightarrow{a} + \overrightarrow{h}) - f(\overrightarrow{a}) - \lambda(\overrightarrow{h}) \right|}{\left| \overrightarrow{h} \right|} \leq \sum_{i=1}^m \lim_{\overrightarrow{h} \to \mathbf{0}} \frac{\left| f^i(\overrightarrow{a} + \overrightarrow{h}) - f^i(\overrightarrow{a}) - \sum_{j=1}^n D_j f^i(\overrightarrow{a}) h_j \right|}{\left| \overrightarrow{h} \right|}$$

Thus it is sufficient to consider the case $m = 1$. When $\overrightarrow{h} = (h_1, \ldots, h_n)$, define $[\overrightarrow{h}]^k := (h_1, \ldots, h_k, 0, \ldots, 0) \in \mathbb{R}^n$. Then we can telescope:

$$f(\overrightarrow{a} + \overrightarrow{h}) - f(\overrightarrow{a}) = \sum_{k=1}^n f\left(\overrightarrow{a} + [\overrightarrow{h}]^k\right) - f\left(\overrightarrow{a} + [\overrightarrow{h}]^{k-1}\right)$$

So

$$f(\overrightarrow{a}+\overrightarrow{h})-f(\overrightarrow{a})-\sum_{j=1}^{n}D_jf(\overrightarrow{a})h_j=\sum_{j=1}^{n}\left[f\left(\overrightarrow{a}+[\overrightarrow{h}]^j\right)-f\left(\overrightarrow{a}+[\overrightarrow{h}]^{j-1}\right)-D_jf(\overrightarrow{a})h_j\right]$$

Thus we need to prove that

$$\lim_{\overrightarrow{h}\to\mathbf{0}}\frac{\left|f\left(\overrightarrow{a}+[\overrightarrow{h}]^j\right)-f\left(\overrightarrow{a}+[\overrightarrow{h}]^{j-1}\right)-D_jf(\overrightarrow{a})h_j\right|}{\left|\overrightarrow{h}\right|}=0$$

for all j. Fix some j. Then define $g_j:\mathbb{R}^n\to\mathbb{R}^m$ by

$$g_j(x)=f(a_1+h_1,\ldots,a_{j-1}+h_{j-1},a_j+x,a_{j+1},\ldots,a_n)$$

Since f is continuously differentiable, we can pick $\overrightarrow{h}$ small enough that D_jf exists at $\overrightarrow{a}+[\overrightarrow{h}]^{j-1}$. Then $D_jf(\overrightarrow{a}+[\overrightarrow{h}]^{j-1})=g_j'(0)$, so we have

$$\begin{aligned}
\lim_{\overrightarrow{h}\to\mathbf{0}}\frac{\left|f(\overrightarrow{a}+[\overrightarrow{h}]^j)-f(\overrightarrow{a}+[\overrightarrow{h}]^{j-1})-D_jf(\overrightarrow{a})h_j\right|}{\left|\overrightarrow{h}\right|}&=\lim_{\overrightarrow{h}\to\mathbf{0}}\frac{|g_j(h_j)-g_j(0)-D_jf(\overrightarrow{a})h_j|}{\left|\overrightarrow{h}\right|}\\
&=\lim_{h_j\to 0}\frac{|g_j(h_j)-g_j(0)-g_j'(0)h_j+g_j'(0)h_j-D_jf(\overrightarrow{a})h_j|}{|h_j|}\\
&\le\lim_{h_j\to 0}\frac{|g_j(h_j)-g_j(0)-g_j'(0)h_j|}{|h_j|}+\lim_{h_j\to 0}\frac{|g_j'(0)h_j-D_jf(\overrightarrow{a})h_j|}{|h_j|}\\
&=\lim_{h_j\to 0}|g_j'(0)-D_jf(\overrightarrow{a})|\\
&=\lim_{h_j\to 0}|D_jf(\overrightarrow{a}+[\overrightarrow{h}]^{j-1})-D_jf(\overrightarrow{a})|\\
&=0
\end{aligned}$$

where the fourth line follows since g_j is differentiable at 0, and the last equality because D_jf is continuous at a_j. Thus $Df(a)=\lambda$ exists. ☐

The above theorem, in combination with the Chain Rule, allows us to derive a specific version of the Chain Rule that allows us to bypass checking for differentiability when the partial derivatives are known.

Corollary 2.9

Let $g_1,\ldots,g_m:\mathbb{R}^n\to\mathbb{R}$ be continuously differentiable at a, and let $f:\mathbb{R}^m\to\mathbb{R}$ be differentiable at $(g_1(a),\ldots,g_m(a))$. Let $F:\mathbb{R}^n\to\mathbb{R}$ be defined by

$$F(a)=f(g_1(a),\ldots,g_m(a))$$

Then

$$D_iF(a)=\sum_{j=1}^{m}D_jf(g_1(a),\ldots,g_m(a))\cdot D_ig_j(a)$$

Proof. Define $g : \mathbb{R}^n \to \mathbb{R}^m$ by $g = (g_1, \ldots, g_m)$. Then $F = f \circ g$. Since $g_1, \ldots, g_m$ are continuously differentiable, g is continuously differentiable, so it is differentiable. Thus the Chain Rule tells us that

$$F'(a) = (f \circ g)'(a) = f'(g(a))g'(a)$$

Matrix multiplication tells us that

$$[F'(a)]_{1i} = \sum_{j=1}^{m} [f'(g(a))]_{1j} [g'(a)]_{ji}$$

Moreover, Theorem 2.7 tells us that

$$\begin{aligned} [F'(a)]_{1i} &= D_i F(a) \\ [f'(g(a))]_{1j} &= D_j f(g(a)) \\ [g'(a)]_{ji} &= D_i g^j(a) = D_i g_j(a) \end{aligned}$$

Thus we conclude that

$$D_i F(a) = \sum_{j=1}^{m} D_j f(g_1(a), \ldots, g_m(a)) \cdot D_i g_j(a)$$

□

Example 2.4

Let $f(x, y, z) = xyz$, and let $g_1(a, b) = a \sin b$, $g_2(a, b) = b \cos a$, $g_3(a, b) = a^3 b$. Then

$$\begin{aligned} \frac{\partial}{\partial a}(f \circ g)\bigg|_{(a,b)} &= D_1(f \circ g)(a, b) \\ = D_1 f(g(a, b)) D_1 g_1(a, b) &+ D_2 f(g(a, b)) D_1 g_2(a, b) + D_3 f(g(a, b)) D_1 g_3(a, b) \\ &= a^3 b^2 \cos a \sin b - a^4 b^2 \sin b \sin a + 3a^2 b^2 \sin b \cos a \end{aligned}$$

In cases where one or more of the g_i do not explicitly depend on all of the variables, the derivatives with respect to those variables is zero.

Example 2.5

Let $f(x, y, z) = xyz$, and let $g_1(a, b) = ab$, $g_2(a) = a$, $g_3(b) = b$. Replacing D_1 with D_a for clarity, we consider

$$D_a g_3(b) = 0, D_b g_2(a) = 0$$

Thus

$$\begin{aligned} D_a(f \circ g)(a, b) &= D_1 f(g(a, b)) D_a g_1(a, b) + D_2 f(g(a, b)) D_a g_2(a) \\ &= ab^2 + ab^2 \\ &= 2ab^2 \end{aligned}$$

(This can be formally established by writing $\hat{g}_2(a,b) = a$, $\hat{g}_3(a,b) = b$, but this is generally unnecessary.)

2.5 Inverse Functions

In Exercise 2-16, we began our study of inverse functions, showing that in the case that $f : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable with a differentiable inverse $f^{-1} : \mathbb{R}^n \to \mathbb{R}^n$,

$$(f^{-1})(a) = [f'(f^{-1}(a))]^{-1}$$

However, the requirement that f has an inverse, and that both are differentiable is a relatively stringent condition. Thus, it is of interest to us to identify when the above equality may be obtained under weaker conditions. In particular, the requirement that f is invertible is a strong *global* condition. However, it can be weakened by instead requiring that f is invertible locally; that is, the restriction of f to a sufficiently small open set is invertible. Thus, it falls to us to determine the conditions where this occurs.

Consider the case of $f : \mathbb{R} \to \mathbb{R}$. We would like our conditions to be in terms of the differentiability of f, since that is what we have studied so far. One observation that we can make is that if f is strictly increasing or decreasing on a small interval, it is 1-1 on that interval. In other words, if $f'(x) > 0$ in an interval around a, then f is invertible in that interval, and similarly if $f'(x) < 0$. Moreover, if f is continuously differentiable, then $f'(a) > 0$ is sufficient to conclude that $f(x) > 0$ in an interval around a. This leads to our multivariate generaliziation, but it will take some work to arrive there.

Lemma 2.10

Let $A \subseteq \mathbb{R}^n$ be a rectangle and let $f : A \to \mathbb{R}^n$ be continuously differentiable. If there is a number $M > 0$ such that $|D_j f^i(x)| \leq M$ for all $x \in \operatorname{int} A$, then

$$|f(x) - f(y)| \leq n^2 M |x - y|$$

for all $x, y \in A$.

Proof. First, we have

$$|f(x) - f(y)| \leq \sum_{i=1}^{n} |f^i(x) - f^i(y)|$$

Now, let $z = y - x$ and define $h^i z(t) = f^i(x + tz)$, so that $h^i_z(0) = f^i(x)$ and $h^i_z(1) = f^i(y)$. Since f^i is differentiable (this follows from Theorem 2.8), we know that the directional derivative $D_z f^i(x)$ exists, and moreover that $h^{i\prime}(t) = D_z f^i(x + tz)$ (see Exercise 2-35). Thus

$$|f^i(y) - f^i(x)| = |h^i(0) - h^i(1)| = \left| \int_0^1 h^{i\prime}(t) dt \right| = \left| \int_0^1 D_z f^i(x + tz) dt \right|$$

We also showed in Exercise 2-29 that D_* is linear with respect to direction, so we can expand this:

$$\begin{aligned}
\left|\int_0^1 D_z f^i(x+tz)dt\right| &= \left|\int_0^1 \sum_{j=1}^n z_j D_j f^i(x+tz)dt\right| \\
&\leq \sum_{j=1}^n \left|\int_0^1 z_j D_j f^i(x+tz)dt\right| \\
&\leq \sum_{j=1}^n |z_j| \left|\int_0^1 D_j f^i(x+tz)dt\right| \\
&\leq \sum_{j=1}^n |z_j| M \\
&\leq \sum_{j=1}^n |z| M \\
&= nM|y-x|
\end{aligned}$$

Thus we have

$$|f^i(y) - f^i(x)| \leq nM|y-x|$$

Combining this with our first inequality, we have

$$|f(x) - f(y)| \leq \sum_{i=1}^n |f^i(x) - f^i(y)| \leq \sum_{i=1}^n nM|y-x| = n^2 M|y-x| \qquad \square$$

Lemma 2.10 provides the necessary machinery to extend our result about locally invertible functions to the multivariate case:

Theorem 2.11: Inverse Function Theorem

Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable in an open set containing a, and $\det f'(a) \neq 0$. Then there is an open set V containing a and an open set W containing $f(a)$ such that $f : V \to W$ has a continuous inverse $f^{-1} : W \to V$ which is differentiable and for all $y \in W$ satisfies

$$(f^{-1})'(y) = [f'(f^{-1}(y))]^{-1}$$

Briefly speaking, this theorem says that so long as $f'(a)$ is nonsingular, then we can find a restriction to a small open set where f is invertible and the derivative condition is met.

Proof. Let $\lambda = Df(a)$. Since $\det f'(a) \neq 0$, λ is invertible. Now suppose that the theorem is true for $\lambda^{-1} \circ f$. Then letting $\phi = (\lambda^{-1} \circ f)^{-1}$, I claim that $\phi \circ \lambda^{-1} = f^{-1}$. To see this, we check that $\phi \circ \lambda^{-1}$ is both a left and right identity:

$$\begin{aligned}
&(\phi \circ \lambda^{-1}) \circ f = (\lambda^{1-} \circ f)^{-1} \circ (\lambda^{-1} \circ f) = \mathrm{id} \\
&f \circ (\phi \circ \lambda^{-1}) = f \circ f^{-1} \circ \lambda \circ \lambda^{-1} = \mathrm{id}
\end{aligned}$$

Moreover, this composition is continuous and differentiable, so if the theorem holds for $\lambda^{-1} \circ f$, it holds for f. Thus it suffices to prove the case where λ is the identity.

Now we know that
$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} = 0$$
so we can choose a small closed rectangle U containing a such that
$$\frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} < 1$$
Now suppose for contradiction that there exists $x \in U$ with $f(x) = f(a)$. Then we would have
$$\frac{|f(x) - f(a) - \lambda(x-a)|}{|x-a|} = \frac{|x-a|}{|x-a|} = 1$$
which contradicts the inequality we just established for U. So $f(x) \neq f(a)$ for all $x \neq a \in U$.

Now note that $x \mapsto \det f'(x)$ consists of sums and products of continuous functions (each $D_j f^i$ exists and is continuous since f is continuously differentiable), so it is continuous. Thus we can also choose U small enough such that $\det f'(x) \neq 0$ for $x \in U$.

Lastly, since f is continuously differentiable, we can pick U small enough such that for any i, j and $x \in U$ we have
$$|D_j f^i(x) - D_j f^i(a)| < \frac{1}{2n^2}$$
Next, let $g(x) = f(x) - x$. Then since $Df(a) = \text{id}$, for any $x \in \operatorname{int} A$ we have
$$|D_j g^i(x)| = |D_j f^i(x) - D_j \text{id}^i(x)| = |D_j f^i(x) - D_j f^i(a)| < \frac{1}{2n^2}$$
so $|D_j g^i(x)| \leq M = 1/2n^2$ for all i, j and $x \in U$. Thus we may apply Lemma 2.10 to conclude that for any $x, y \in U$,
$$|f(x) - x - (f(y) - y)| = |g(x) - g(y)| \leq n^2 M |x - y| = \frac{|x-y|}{2}$$
Moreover, by the reverse triangle inequality,
$$|x - y| - |f(x) - f(y)| \leq |f(x) - x - (f(y) - y)|$$
so we know that for any $x, y \in U$,
$$|x - y| \leq 2|f(x) - f(y)|$$
Since U is a closed rectangle, $\partial U \subseteq U$, so for any $x \in \partial U$ we know $f(x) \neq f(a)$. Thus $f(a) \notin f(\partial U)$. Moreover, ∂U is compact, so $f(\partial U)$ is compact and there exists $d > 0$ such that $|f(a) - f(x)| \geq d$ for any $x \in \partial U$. Then define
$$W = \left\{ y : |y - f(a)| < \frac{d}{2} \right\}$$

If $y \in W$ and $x \in \partial U$, then

$$|y - f(a)| < |y - f(x)|$$

Then we show that for any $y \in W$, there exists a unique preimage $x \in \operatorname{int} U$ with $f(x) = y$. To prove this, note that defining $g : U \to \mathbb{R}$ by

$$g(x) = |y - f(x)|^2 = \sum_{i=1}^{n} (y_i - f^i(x))^2$$

This function is continuous, so it achieves a minimum on U. But since $|y - f(a)| < |y - f(x)|$ for $x \in \partial U$, we know that $g(a) < g(x)$. So the minimum cannot be in ∂U. Thus there exists $x \in \operatorname{int} U$ such that g is minimized, which allows us to conclude that $D_j g(x) = 0$ for all j. Thus

$$\sum_{i=1}^{n} 2(y_i - f^i(x)) D_j f^i(x) = 0$$

Since this holds for every j, we can rewrite this system of equations as

$$f'(x) \begin{bmatrix} y_1 - f^1(x) \\ \vdots \\ y_n - f^n(x) \end{bmatrix} = 0$$

But $\det f'(x) \neq 0$ so we conclude that $y_i - f^i(x) = 0$ for all i. Thus $y = f(x)$. So we know that a preimage x exists. If another preimage x_2 exists, then we have

$$|x - x_2| \leq 2|f(x) - f(x_2)| = 2|y - y| = 0$$

so $x = x_2$. Thus x is unique as well. Thus, we have shown that f is locally invertible. Letting $V = \operatorname{int} U \cap f^{-1}(W)$, we may write that $f : V \to W$ has an inverse $f^{-1} : W \to V$. Moreover, for any $y_1, y_2 \in W$ with $x_1 = f^{-1}(y_1)$ and $x_2 = f^{-1}(y_2)$, we have

$$|f^{-1}(y_1) - f^{-1}(y_2)| = |x_1 - x_2| \leq 2|f(x_1) - f(x_2)| = 2|y_1 - y_2|$$

So f^{-1} is Lipschitz and is thus continuous.

Now we must show that f^{-1} is differentiable. Let $x \in V$, and write $\mu = Df(x)$. Let $y = f(x) \in W$. Then we show that f^{-1} is differentiable at y with $Df^{-1}(y) = \mu^{-1}$. Let $\varphi(x_1) = f(x_1) - f(x) - \mu(x_1 - x)$, such that

$$f(x_1) = f(x) + \mu(x_1 - x) + \varphi(x_1 - x)$$

Moreover, since f is differentiable at x we have

$$\lim_{x_1 \to x} \frac{|\varphi(x_1 - x)|}{|x_1 - x|} = 0$$

So

$$\mu^{-1}(f(x_1) - f(x)) = x_1 - x + \mu^{-1}(\varphi(x_1 - x))$$

or

$$x_1 = \mu^{-1}(f(x_1) - f(x)) + x - \mu^{-1}(\varphi(x_1 - x))$$

But any $y_1 \in W$ is of the form $f(x_1)$ for $x_1 \in V$, so without loss of generality we may wriet

$$f^{-1}(y_1) = f^{-1}(y) + \mu^{-1}(y_1 - y) - \mu^{-1}(\varphi(f^{-1}(y_1) - f^{-1}(y)))$$

and we only need to show that

$$\lim_{y_1 \to y} \frac{|\mu^{-1}(\varphi(f^{-1}(y_1) - f^{-1}(y)))|}{|y_1 - y|} = 0$$

By Exercise 1-10 the linear transformation μ^{-1} is irrelevant here and we only need to show that

$$\lim_{y_1 \to y} \frac{|\varphi(f^{-1}(y_1) - f^{-1}(y))|}{|y_1 - y|} = 0$$

We can apply a trick here, splitting the fraction:

$$\frac{|\varphi(f^{-1}(y_1) - f^{-1}(y)|)}{|y_1 - y|} = \frac{|\varphi(f^{-1}(y_1) - f^{-1}(y))|}{|f^{-1}(y_1) - f^{-1}(y)|} \frac{|f^{-1}(y_1) - f^{-1}(y)|}{|y_1 - y|}$$

Since f^{-1} is continuous, $f^{-1}(y_1) \to f^{-1}(y)$ as $y_1 \to y$, so

$$\lim_{y_1 \to y} \frac{|\varphi(f^{-1}(y_1) - f^{-1}(y))|}{|f^{-1}(y_1) - f^{-1}(y)|} = \lim_{x_1 \to x} \frac{|\varphi(x_1 - x)|}{|x_1 - x|} = 0$$

and the second factor is bounded by 2, completing the proof. □

2.6 Implicit Functions

Having now proved our major result concerning local invertibility of functions, we will apply it to the study of converting implicit function relations into explicit functions.

Example 2.6

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = x^2 + y^2 - 1$. Let C be the set of points (x, y) with $f(x, y) = 0$ (this defines a **level curve** of f). Then this curve is simply a circle of radius 1 centered at the origin.

To convert this curve into an explicit function, we attempt to answer the following question: given a point $(a, b) \in C$, do there exist intervals A around a and B around b such that for any $x \in A$ there exists exactly one $y \in B$ with $(x, y) \in C$. In the case that there is, we can then define a function $g : A \to B$ which maps each x to that unique y.

If we choose (x, y) such that $x \neq \pm 1$, then we can indeed do so. When $y > 0$, the graph of the function $g(x) = \sqrt{1 - x^2}$ traces out the upper semicircle. When $y < 0$, we instead pick $h(x) = -\sqrt{1 - x^2}$, tracing out the lower circle. In both cases, our choice of g or h is forced. However, when $x = \pm 1$, we cannot pick an interval around x where such a function can be defined.

It is also worth remarking that both g and h are differentiable.

To generalize the above discussion to multiple variables, we consider functions of the form $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. If $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, then we would like to find neighborhoods V around x and W around y such that any $\overline{x} \in V$ corresponds to exactly one $\overline{y} \in W$ with $f(\overline{x}, \overline{y}) = 0$, which allows us to implicitly define a function $g : V \to W$, which maps $\overline{x}$ to $\overline{y}$.

Theorem 2.12: Implicit Function Theorem

Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be continuously differentiable in an open set aroud (a, b), and suppose $f(a, b) = 0$. Let M be an $m \times m$ matrix defined by $M_{ij} = D_{n+j} f^i(a, b)$. If $\det M \neq 0$, then there is an open set $A \subseteq \mathbb{R}^n$ containing a and an open set $B \subseteq \mathbb{R}^m$ containing b, such that for any $x \in A$ there is a unique $y \in B$ such that $f(x, y) = 0$. Moreover, the function g defined by $x \mapsto y$ is differentiable.

Proof. Define $F : \mathbb{R}^n \times \mathbb{R} + summ \to \mathbb{R}^n \times \mathbb{R}^m$ by $F(x, y) = (x, f(x, y))$. Then $F'(a, b)$ is given by a block matrix

$$F'(a, b) = \begin{bmatrix} I & O \\ O & M \end{bmatrix}$$

so $\det F'(a, b) = \det M \neq 0$. Apply the Inverse Function Theorem to produce open sets $V \subseteq \mathbb{R}^n \times \mathbb{R}^m$ containing (a, b) and $W \subseteq \mathbb{R}^n \times \mathbb{R}^m$ containing $F(a, b) = (a, 0)$. We can write $V = A \times B$ (Spivak asserts this but I'm not sure how), and thus the restriction $F : A \times B \to W$ has a differentiable inverse $h : W \to A \times B$. Moreover, since F preserves the first n coordinates, h must also, so that $h(x, y) = (x, k(x, y))$ for some differentiable function k. Then define the projection $\pi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ by $\pi(x, y) = y$, such that $\pi \circ F = f$. Thus

$$f(x, k(x, y)) = f \circ h(x, y) = (\pi \circ F) \circ h(x, y) = \pi \circ (F \circ h)(x, y) = \pi(x, y) = y$$

Then $f(x, k(x, 0)) = 0$. So for any $x \in A$, we can pick $y = k(x, 0) \in B$, and we will have $f(x, y) = 0$. Moreover, if there exists another $y' \in B$ with $f(x, y') = 0$, then we would have

$$F(x, y') = (x, f(x, y')) = (x, 0) = (x, f(x, y)) = F(x, y)$$

But F is invertible so we cannot have $y \neq y'$. Thus our choice of y is unique, and the implicitly defined function k is differentiable. □

Since we know that the implicitly defined g is differentiable, we can calculate its derivative. For any coordinate i, we have $f^i(x, g(x)) = 0$, so

$$D_j f^i(x, g(x)) + \sum_{\alpha=1}^{m} D_{n+\alpha} f^i(x, g(x)) D_j g^\alpha(x) = 0$$

which we can then solve for the various $D_j g^\alpha(x)$ by inverting M (which can be done since $\det M \neq 0$).

We can generalize the Implicit Function Theorem as follows:

Theorem 2.13

Let $f : \mathbb{R}^n \to \mathbb{R}^p$ be continuously differentiable in an open set containing a, where $p \leq n$. If $f(a) = 0$ and the $p \times n$ matrix P with $P_{ij} = D_j f^i(a)$ has rank p, then there is an open set $A \subseteq \mathbb{R}^n$ and a differentiable function $h : A \to \mathbb{R}^n$ with differentiable inverse such that $h(A)$ contains a and

$$f \circ h(x_1, \ldots, x_n) = (x_{n-p+1}, \ldots, x_n)$$

Note: Spivak states that A contains a. This is incorrect.

We can interpret the above theorem by saying that whenever the derivative of f has rank p, then we can find h such that $f \circ h$ acts to embed the last p coordinates of $\overrightarrow{x}$ into $\mathbb{R}^p$.

Proof. Consider f as a function $\mathbb{R}^{n-p} \times \mathbb{R}^p \to \mathbb{R}^p$. Then if P has rank p, it has p linearly independent columns. Let $g : \mathbb{R}^n \to \mathbb{R}^n$ permute the coordinates such that those linearly independent columns are the last p columns. Taking $f \circ g$, the matrix M as defined in the Implicit Function Theorem, which is a $p \times p$ matrix with $M_{ij} = D_{n+j}(f \circ g)^i(a)$, has rank p, and thus has nonzero determinant.

Now, as in the proof of the Implicit Function Theorem, define $F : \mathbb{R}^{n-p} \times \mathbb{R}^p \to \mathbb{R}^{n-p} \times \mathbb{R}^p$ by $F(x, y) = (x, f \circ g(x, y))$. Again, $\det F'(a, b) = \det M \neq 0$, so we apply the Inverse Function Theorem to produce h which is locally an inverse of F. As in the previous proof, we have

$$(f \circ g) \circ h(x, y) = y$$

so taking $g \circ h$ produces the requested function. $\square$

Chapter 3

Integration

3.1 Basic Definitions

The following treatment of the basic definitions of integrals over a closed rectangle $A \subseteq \mathbb{R}^n$ is rapid, as this case is similar to the single variable case of integration over an interval.

Definition 3.1

A **partition** of a closed interval $[a, b]$ is a finite sequence $\{t_0, \ldots, t_k\}$, such that $a = t_0 \leq \ldots \leq t_k = b$, such that $[a, b]$ is divided into k subintervals.

Definition 3.2

Let $A = [a_1, b_1] \times \ldots \times [a_n, b_n] \subseteq \mathbb{R}^n$ be a closed rectangle. A partition of A is a collection of partitions $\mathcal{P} = (\mathcal{P}_1, \ldots, \mathcal{P}_n)$, such that $\mathcal{P}_i$ is a partition of $[a_i, b_i]$. If $\mathcal{P}_i$ divides $[a_i, b_i]$ into N_i subintervals, then $\mathcal{P}$ divides A into $N_1 \times \ldots \times N_n$ **subrectangles** of $\mathcal{P}$. Using a slight abuse of notation, we will write $S \in \mathcal{P}$ to denote that S is a subrectangle of $\mathcal{P}$.

If $A \subseteq \mathbb{R}^n$ is a rectangle, $f : A \to \mathbb{R}$ is bounded, and $\mathcal{P}$ is a partition, then we can define the maximum and minimum values for each subrectangle $S \in \mathcal{P}$:

$$m_S(f) = \inf\{f(x) : x \in S\}$$
$$M_S(f) = \sup\{f(x) : x \in S\}$$

Let $v(S)$ denote the volume of S, defined as the product of the side lengths (regardless of whether S is open or closed). Then the lower and upper sums of f with respect to $\mathcal{P}$ are

$$L(f, \mathcal{P}) = \sum_{S \in \mathcal{P}} m_S(f) v(S)$$
$$U(f, \mathcal{P}) = \sum_{S \in \mathcal{P}} M_S(f) v(S)$$

Since $m_S(f) \leq M_S(f)$ for any s, we then have $L(f, \mathcal{P}) \leq U(f, \mathcal{P})$.

Definition 3.3

A partition $\mathcal{P}'$ is called a **refinement** of a partition $\mathcal{P}$ if each subrectangle of $\mathcal{P}'$ is contained in a subrectangle of $\mathcal{P}$.

Lemma 3.1

Let $\mathcal{P}'$ be a refinement of $\mathcal{P}$. Then

$$L(f, \mathcal{P}) \leq L(f, \mathcal{P}')$$

and

$$U(f, \mathcal{P}) \geq U(f, \mathcal{P}')$$

Proof. Let S be a subrectangle of $\mathcal{P}$. Then it contains subrectangles $S_1, \ldots, S_k \in \mathcal{P}'$ which are disjoint and cover S, so that $\sum_{1 \leq i \leq k} v(S_i) = v(S)$. For each S_i, $m_{S_i}(f) \geq M_S(f)$. Thus

$$\sum_{1 \leq i \leq k} m_{S_i}(f) v(S_i) \geq m_S(f) v(S)$$

Since $\mathcal{P}'$ refines $\mathcal{P}$, each subrectangle of $\mathcal{P}'$ is contained in a subrectangle of $\mathcal{P}$. Thus we have

$$L(f, \mathcal{P}') = \sum_{S' \in \mathcal{P}'} m_{S'}(f) v(S') = \sum_{S \in \mathcal{P}} \sum_{1 \leq i \leq k} m_{S_i}(f) v(S_i) \geq \sum_{S \in \mathcal{P}} m_S(f) v(S) = L(f, \mathcal{P})$$

The proof for the other case is similar. $\square$

In essence, as we refine a given partition, the upper and lower sums will grow closer to one another, and under the appropriate conditions, they will also converge to one another. This provides a candidate value for the integral of f over A; however, it is dependent on our starting partition. Ideally, our integral may be defined independent of a particular choice of partition; to do so we must prove the following:

Corollary 3.2

If $\mathcal{P}$ and $\mathcal{P}'$ are partitions, then $L(f, \mathcal{P}') \leq U(f, \mathcal{P})$.

To prove this, we first introduce an auxiliary construction:

Definition 3.4

Let $\mathcal{P}$ and $\mathcal{P}'$ be partitions of an interval $[a, b]$. Then their **common refinement** $\mathcal{Q}$ is the partition $\mathcal{P} \cup \mathcal{P}'$. If $\mathcal{P} = (\mathcal{P}_1, \ldots, \mathcal{P}_n)$ and $\mathcal{P}' = (\mathcal{P}'_1, \ldots, \mathcal{P}'_n)$ be partitions of a rectangle $A \subseteq \mathbb{R}^n$. Then the common refinement $\mathcal{Q}$ is given by $(\mathcal{P}_1 \cup \mathcal{P}'_1, \ldots, \mathcal{P}_n \cup \mathcal{P}'_n)$.

Proof. Let $\mathcal{Q}$ be the common refinement of $\mathcal{P}$ and $\mathcal{P}'$. Then by Lemma 3.1,

$$L(f,\mathcal{P}') \le L(f,\mathcal{Q}) \le U(f,\mathcal{Q}) \le U(f,\mathcal{P})$$

□

Now let $U = \inf U(f,\mathcal{P})$, where the infimum is taken over all partitions $\mathcal{P}$ of A, and let $L = \sup L(f,\mathcal{P})$. By Corollary 3.2, both U and L exist, and $L \le U$. As mentioned above, if our continued refinements converge to a single value, then this provides a plausible definition of the integral. As Corollary 3.2 shows, this convergence is only possible if $U = L$, and it must converge to that common value. Moreover, the values of U and L are independent of our choice of partition, which allows us to define the integral:

Definition 3.5

Let $f : A \to \mathbb{R}$ be bounded, with $A \subseteq \mathbb{R}^n$ a rectangle. Then f is **integrable** if $U = L$. In this case, we denote the **integral** of f on A by $\int_A f = U = L$, which may alternatively be notated $\int_A f(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n$.

The following theorem gives us an equivalent criterion for integrability.

Theorem 3.3

A bounded function $f : A \to \mathbb{R}$ is integrable if and only if, for every $\varepsilon > 0$ there exists a partition $\mathcal{P}$ of A such that

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon$$

Proof. ($\Longrightarrow$) If f is integrable then U exists, so there exists a partition $\mathcal{P}_1$ with $U(f,\mathcal{P}_1) \le U + \frac{\varepsilon}{2}$. Similarly there exists $\mathcal{P}_2$ with $L(f,\mathcal{P}_2) \ge L - \frac{\varepsilon}{2}$. Let $\mathcal{P}$ be the common refinement of $\mathcal{P}_1, \mathcal{P}_2$. Then

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) \le U(f,\mathcal{P}_1) - L(f,\mathcal{P}_2) \le U + \frac{\varepsilon}{2} - (L - \frac{\varepsilon}{2}) = \varepsilon$$

($\Longleftarrow$) By Corollary 3.2, both U and L exist. Let $\varepsilon > 0$ be arbitrary, and let $\mathcal{P}$ be the partition produced by the condition. Then

$$U - L \le U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon$$

So $U - L < \varepsilon$ for all $\varepsilon > 0$ and thus $U = L$. So f is integrable over A. □

Example 3.1

Let $f : A \to \mathbb{R}$ be constant with $f(x) = c$. Then if $\mathcal{P}$ is a partition and $S \in \mathcal{P}$, $m_S(f) = M_S(f) = c$, so

$$U(f,\mathcal{P}) = \sum_{S\in\mathcal{P}} m_S(f) v(S) = c \sum_{S\in\mathcal{P}} v(S) = cv(A)$$
$$L(f,\mathcal{P}) = cv(A)$$

so $U = L = cv(A)$ and f is integrable with $\int_A f = cv(A)$.

Example 3.2

Let $f : [0,1] \times [0,1] \to \mathbb{R}$ be defined by

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

If $\mathcal{P}$ is a partition and $S \in \mathcal{P}$, by the density of $\mathbb{Q}$ in $\mathbb{R}$ we have $m_S(f) = 0$, and by the density of $\mathbb{I} \in \mathbb{R}$ we have $M_S(f) = 1$. So

$$U(f,\mathcal{P}) = \sum_{S \in \mathcal{P}} M_S(f) v(S) = \sum_{S \in \mathcal{P}} v(S) = v(A)$$
$$L(f,\mathcal{P}) = 0$$

So f is not integrable over any rectangle A with $v(A) > 0$.

3.2 Measure Zero and Content Zero

In this section, we discuss the notions of measure and content zero. These quanity the concept of a set which is small enough to be insignificant in certain contexts. Moreover, in particular with the case of measure zero, this is a special case of a more general technique which serves as the formalization of volume in higher dimensions.

Definition 3.6

A subset $A \subseteq \mathbb{R}^n$ has **measure zero** if for any $\varepsilon > 0$ there exists a cover $\mathcal{O}$ of A by closed rectangles such that $\sum_{O \in \mathcal{O}} v(O) < \varepsilon > 0$.

We may also use open rectangles rather than closed rectangles in the above.

Proposition 3.4

If a set $A \subseteq \mathbb{R}^n$ is countable, then it has measure zero.

Proof. Let $\varepsilon > 0$. Enumerate the points in A as $a_1, a_2, \ldots$. Then for each a_i, pick a closed rectangle O_i containing a_i such that $v(O_i) < \frac{\varepsilon}{2^i}$. Then $\mathcal{O} = \{O_1, O_2, \ldots\}$ covers A, and

$$\sum_{O \in \mathcal{O}} v(O) = \sum_{i=1}^{\infty} v(O_i) \le \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon \sum_{i=1}^{\infty} \frac{1}{2^i} = \varepsilon$$

so A has measure zero. $\square$

Importantly, $\mathbb{Q}$ is countable, and thus has measure zero.

Theorem 3.5

Let $A = \bigcup_{i=1}^{\infty} A_i$ be a countable union of measure zero sets A_i. Then A has measure zero.

Proof. Let $\varepsilon > 0$. For each A_i, pick an open cover $\mathcal{O}_i$ such that

$$\sum_{O \in \mathcal{O}_i} v(O) < \frac{\varepsilon}{2^i}$$

Now let $\mathcal{O} = \bigcup_{i=1}^{\infty} \mathcal{O}_i$. Then $\mathcal{O}$ covers A, and

$$\sum_{O \in \mathcal{O}} v(O) = \sum_{i=1}^{\infty} \sum_{O \in \mathcal{O}_i} v(O_i) \leq \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon$$

so A has measure zero. $\square$

While sets of measure zero are important (and indeed, this notion hints at more important themes in measure theory), there are times when we would prefer to work with a finite cover rather than an open cover. This is analogous to our preference for compact sets. Thus, we have a corresponding notion of measure zero for finite covers:

Definition 3.7

A subset $A \subseteq \mathbb{R}^n$ has **content zero** if for any $\varepsilon > 0$ there exists a *finite* cover $\mathcal{O}$ of A by closed rectangles such that

$$\sum_{O \in \mathcal{O}} v(O) < \varepsilon$$

By definition, a set having content zero is a special case of having measure zero.

Theorem 3.6

A nonsingleton interval $[a, b] \subseteq \mathbb{R}$ does not have content zero. For any finite cover $\{O_1, \ldots, O_n\}$ of $[a, b]$, where each O_i is a closed interval,

$$\sum_{i=1}^{n} v(O_i) \geq b - a$$

Proof. Let $\mathcal{O}$ be a finite cover. We can pick a cover $\mathcal{O}' = \{O_1 \cap [a, b], \ldots, O_n \cap [a, b]\}$, which will be a cover if and only if $\mathcal{O}$ is, and which has smaller total length, so without loss of generality we may consider $\mathcal{O}'$. Let $t_0, \ldots, t_k$ be the endpoints of the O_i', with

$a = \mathcal{O}_0 \leq \ldots \leq \mathcal{O}_k = b$. Then each O'_i contains at least one interval $[t_{i-1}, t_i]$, and each interval is contained in at least one O'_i. Then

$$\sum_{O' \in \mathcal{O}'} v(O') \geq \sum_{j=1}^{k} (t_j - t_{j-1}) = b - a$$

□

The reader should note that the above proof also shows that $[a, b]$ does not have measure zero (as long as $a < b$).

Theorem 3.7

If A is compact and has measure zero, then it has content zero.

Proof. Let $\varepsilon > 0$. There exists an open cover $\mathcal{O}$ of A with

$$\sum_{O \in \mathcal{O}} v(O) < \varepsilon$$

Since A is compact, pick a finite subcover $\mathcal{O}'$. Then

$$\sum_{O' \in \mathcal{O}'} v(O') \leq \sum_{O \in \mathcal{O}} v(O) < \varepsilon$$

so A has content zero. □

Example 3.3

Although we pointed out earlier that $\mathbb{Q}$ has measure zero, it does not have content zero. Let $\mathcal{O} = \{[a_i, b_i]\}$ be a finite cover of $\mathbb{Q} \cap [0, 1]$ by closed intervals. Then by the density of $\mathbb{Q}$, $\mathcal{O}$ must cover $[0, 1]$. But then $\sum_{i=1}^{n} b_i - a_i \geq 1$, so $\mathbb{Q} \cap [0, 1]$ does not have content zero. It follows that $\mathbb{Q}$ does not either.

3.3 Integrable Functions

In this section, we will expand on the theory of which functions may be (Riemann) integrated.

Recall that $o(f, x)$ denotes the oscillation of f at x, defined as

$$\lim_{\delta \to 0} M(x, f, \delta) - m(x, f, \delta)$$

where

$$M(x, f, \delta) = \sup\{f(y) : |x - y| < \delta\}$$
$$m(x, f, \delta) = \inf\{f(y) : |x - y| < \delta\}$$

Lemma 3.8

Let A be a closed rectangle and let $f : A \to \mathbb{R}$ be a bounded function such that $o(f, x) < \varepsilon$ for all $x \in A$. Then there is a partition $\mathcal{P}$ of A with

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) < \varepsilon \cdot v(A)$$

Proof. For each x, because $o(f, x) < \varepsilon$ we may pick a closed rectangle U_x containing x such that $M_{U_x}(f) - m_{U_x}(f) < \varepsilon$. Then the collection of U_x covers A compact, so we can pick a finite subcover $U_1, \ldots, U_k$. Then pick a partition $\mathcal{P}$ such that each subrectangle of $\mathcal{P}$ is entirely contained within one of the U_x. Then for any subrectangle $S \in \mathcal{P}$ we have

$$M_S(f) - m_S(f) \leq M_{U_x}(f) - m_{U_x}(f) < \varepsilon$$

Then

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) = \sum_{S \in \mathcal{P}} v(S)[M_S(f) - m_S(f)] < \varepsilon \sum_{S \in \mathcal{P}} v(S) = \varepsilon v(A) \qquad \square$$

Lemma 3.9

Let $\mathcal{R}$ be a finite collection of closed rectangles $R_1, \ldots, R_k \subseteq \mathbb{R}^n$. Let $A \subseteq \mathbb{R}^n$ be a closed rectangle. Then there exists a partition $\mathcal{P}$ of A such that for each $S \in \mathcal{P}$ and each R_i, exactly one of the following is true: $S \subseteq R_i$ or $S \cap \operatorname{int} R_i = \varnothing$.

Proof. Let $a_{i,j}$ be the left endpoint of R_i in the jth direction and $b_{i,j}$ the right endpoint, such that

$$R_i = [a_{i,1}, b_{i,1}] \times \ldots \times [a_{i,n}, b_{i,n}]$$

Let $\mathcal{P}_j = \{a_{1,j}, b_{1,j}, \ldots, a_{k,j}, b_{k,j}\}$ (not necessarily in order). Suppose that when ordered, $\mathcal{P}_j = \{t_{j,1}, \ldots, t_{j,2k}\}$ (note that the j has switched coordinates). Let $\mathcal{P} = (\mathcal{P}_1, \ldots, \mathcal{P}_n)$. Then for each $S \in \mathcal{P}$,

$$S = [t_{1,i_1-1}, t_{1,i_1}] \times \ldots \times [t_{n,i_n-1}, t_{n,i_n}]$$

for appropriately chosen $i_1, \ldots, i_n$. Any R_j is of the form

$$R_j = [t_{1,i'_1-1}, t_{1,i'_1}] \times \ldots \times [t_{n,i'_n-1}, t_{n,i'_n}]$$

for some other $i'_1, \ldots, i'_n$. Now consider the first coordinate direction. Suppose $t_{1,i'_1} \leq t_{1,i_1-1}$. Then for any $x \in S$ and $y \in \operatorname{int} R_j$, we have

$$y_1 < t_{1,i'_1} \leq t_{1,i_1-1} \leq x_1$$

so $x \neq y$ and thus $S \cap \operatorname{int} R_i = \varnothing$. Similarly, if $t_{1,i_1} \leq t_{1,i'_1-1}$, then we have

$$x_1 \leq t_{1,i_1} \leq t_{1,i'_1-1} < y_1$$

so $x \neq y$ and $S \cap \operatorname{int} R_i = \varnothing$. Thus we either immediately conclude that $S \cap \operatorname{int} R_i = \varnothing$, or we know that

$$t_{1,i'_1} > t_{1,i_1-1}$$
$$t_{1,i_1} > t_{1,i'_1-1}$$

This is equivalent to

$$t_{1,i_1'} \geq t_{1,i_1}$$
$$t_{1,i_1-1} \geq t_{1,i_1'-1}$$

so we either have $S \cap \operatorname{int} R_i = \varnothing$ or

$$t_{1,i_1'-1} \leq t_{1,i_1-1} \leq t_{1,i_1} \leq t_{1,i_1'}$$

We can apply this argument to each coordinate direction $1, \dots, n$, so that it is either the case that $S \cap \operatorname{int} R_i = \varnothing$, or we have

$$\begin{gathered} t_{1,i_1'-1} \leq t_{1,i_1-1} \leq t_{1,i_1} \leq t_{1,i_1'} \\ \vdots \\ t_{n,i_n'-1} \leq t_{n,i_n-1} \leq t_{n,i_n} \leq t_{n,i_n'} \end{gathered}$$

In this case, we have $S \subseteq R_i$. $\square$

In particular, the above statement shows that if $\mathcal{O}$ is a finite collection of rectangles such that their interiors cover some set $B \subseteq A \subseteq \mathbb{R}^n$, with A a closed rectangle, then there exists a partition of A such that each subrectangle is either contained in some $O \in \mathcal{O}$ or does not intersect B. Such a collection may be of interest, for instance, if B has content zero.

Theorem 3.10

Let A be a closed rectangle and let $f : A \to \mathbb{R}$ be a bounded function. Let $B = \{x : f \text{ is not continuous at } x\}$. Then f is integrable if and only if B is a set of measure zero.

Proof. ($\implies$) Suppose that f is integrable. Define $B_\varepsilon : \{x : o(f, x) \geq \varepsilon\}$. I claim that $B_{1/n}$ has measure zero for each n.

To see this, let $\mathcal{P}$ be a partition of A such that

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) < \frac{\varepsilon}{n}$$

Then let $\mathcal{S}$ be the collection of subrectangles $S \in \mathcal{P}$ such that $S \cap B_{1/n} \neq \varnothing$. Then $\mathcal{S}$ covers $B_{1/n}$. Now, for each $S \in \mathcal{S}$ we know that $o(f, x) \geq \frac{1}{n}$ for some $x \in S$, so $M_S(f) - m_S(f) \geq \frac{1}{n}$. So

$$1 \leq n(M_S(f) - m_S(f))$$

Thus

$$\sum_{S \in \mathcal{S}} v(S) \leq \sum_{S \in \mathcal{S}} v(S) n(M_S(f) - m_S(f)) \leq n \sum_{S \in \mathcal{P}} v(S)[M_S(f) - m_S(f)] < \varepsilon$$

So $B_{\frac{1}{n}}$ has measure zero. Thus $B = \bigcup_{n=1}^{\infty} B_{1/n}$ has measure zero.

($\impliedby$) Suppose that B has measure zero. Now let $\varepsilon > 0$. Suppose that $|f(x)| < M$ for all

x. Define $\varepsilon' = \varepsilon/2v(A)$. Define $B_\varepsilon \coloneqq \{x : o(f,x) \geq \varepsilon'\}$. We have previously proved that a set of this form is compact. Then B_ε is compact and has measure zero, so it has content zero. Then there exists a finite cover $\mathcal{O}$ of B_ε by the interior of closed rectangles such that

$$\sum_{O \in \mathcal{O}} v(O) < \frac{\varepsilon}{4M}$$

Apply Lemma 3.9 to produce a partition $\mathcal{P}'$ such that the subrectangles which do not intersect B_ε may be enumerated as $R_1, \ldots, R_k$, and $o(f,x) < \varepsilon' = \varepsilon/2v(A)$ for any x in any of those closed rectangles. Then apply Lemma 3.8 to each R_i to produce a refinement $\mathcal{P}'$ such that for each R_i,

$$\sum_{S \in \mathcal{P}' : S \subseteq R_i} v(S)[M_S(f) - m_S(f)] < \varepsilon' v(R_i) = \frac{\varepsilon}{2v(A)} v(R_i)$$

Now, for each subrectangle $S' \in \mathcal{P}'$, $S' \subseteq S$ for exactly one $S \in \mathcal{P}$. We either have $S \subseteq O$ for some $O \in \mathcal{O}$, or $S = R_i$ for some i. Thus either $S' \subseteq O$ for some $O \in \mathcal{O}$ or $S' \subseteq R_i$ for some i. Denote by $\mathcal{L}$ the collection of S' such that $S' \subseteq O$ for $O \in \mathcal{O}$ and by $\mathcal{R}$ the collection of S' such that $S' \subseteq R_i$ for some i. Then

$$\begin{aligned} U(f,\mathcal{P}') - L(f,\mathcal{P}') &= \sum_{S \in \mathcal{P}'} v(S)[M_S(f) - m_S(f)] \\ &= \sum_{S \in \mathcal{L}} v(S)[M_S(f) - m_S(f)] + \sum_{S \in \mathcal{R}} v(S)[M_S(f) - m_S(f)] \end{aligned}$$

We also have

$$\sum_{S \in \mathcal{L}} v(S)[M_S(f) - m_S(f)] \leq \sum_{O \in \mathcal{O}} v(O)[M_O(f) - m_O(f)]$$

and

$$\sum_{S \in \mathcal{R}} v(S)[M_S(f) - m_S(f)] = \sum_{i=1}^{k} \sum_{S' \in \mathcal{P}' : S' \subseteq R_i} v(S')[M_S(f) - m_S(f)]$$

so that

$$U(f,\mathcal{P}') - L(f,\mathcal{P}') \leq \sum_{O \in \mathcal{O}} v(O)[M_O(f) - m_O(f)] + \sum_{i=1}^{k} \sum_{S' \in \mathcal{P}' : S' \subseteq R_i} v(S')[M_S(f) - m_S(f)]$$

Since f is bounded by M, we must have $M_O(f) - m_O(f) \leq 2M$ for any O. Thus

$$\begin{aligned} \sum_{O \in \mathcal{O}} v(O)[M_O(f) - m_O(f)] + \sum_{i=1}^{k} U(f,\mathcal{P}_i) - L(f,\mathcal{P}_i) &< 2M \sum_{O \in \mathcal{O}} v(O) + \frac{\varepsilon}{2v(A)} \sum_{i=1}^{k} v(R_i) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \frac{v(A)}{v(A)} \\ &= \varepsilon \end{aligned}$$

So f is integrable. $\square$

We have thus presented an extremely useful criterion for determining when a function may be successfully integrated, without requiring the use of partitions to do so.

We will now progress to expanding our theory of integration from integration on rectangles to arbitrary bounded sets, which we define in terms of integrals on rectangles.

Definition 3.8

Let $C \subseteq \mathbb{R}^n$. The **characteristic function** of C is

$$\chi_C(x) = \begin{cases} 1, & x \in C \\ 0, & x \notin C \end{cases}$$

Definition 3.9

Suppose that $C \subseteq \mathbb{R}^n$ is bounded by a closed rectangle A, and $f : A \to \mathbb{R}$ is bounded. Then the integral of f on C is defined as

$$\int_C f = \int_A f\chi_C$$

provided this quantity is defined.

As we can see from the definition, $\int_C f$ is defined whenever $f\chi_C$ is integrable on A. As we prove in Exercise 3-14, the product of integrable functions is integrable, so if χ_C and f are both integrable, then $\int_C f$ is well defined. Since we are mainly concerned with integrating functions which integrable to begin with, the main task for us is to determined when χ_C is integrable.

Theorem 3.11

If $C \subseteq A \subseteq \mathbb{R}^n$, where A is a closed rectangle, then $\chi_C : A \to \mathbb{R}$, is integrable if and only if ∂C has measure zero.

Proof. Note that whenever $x \in \partial C$, in any neighborhood of x there exists $y \in C$, such that $\chi_C(y) = 1$, and $z \notin C$, such that $\chi_C(z) = 0$. Thus χ_C is discontinuous on ∂C. On the other hand, if $x \notin \partial C$, then $x \in \operatorname{int} A$ or $x \in \operatorname{ext} A$. In either case, there exists a neighborhood around x such that χ_C is constant, so χ_C is continuous on $\operatorname{int} A$ and $\operatorname{ext} A$. Thus χ_C is discontinuous precisely on ∂C.

Since χ_C is integrable if and only if it is discontinuous on a set of measure zero, it is integrable if and only if ∂C has measure zero. □

We should note that since ∂C is closed and bounded, it also has content zero.

Definition 3.10

If C is bounded and ∂C has measure zero, then C is called **Jordan-measurable**.

Thus, for any integrable function f, $\int_C f$ is defined if C is Jordan-measurable. It is possible for $\int_C f$ to be defined in other cases (for instance, if f is identically zero), but this is of little interest to us. This also allows us to extend our definition of volume to non-rectangle sets.

Definition 3.11

The **volume** (or **content**) of a Jordan-measurable set C is defined as

$$v(C) = \int_C 1$$

Note that even if C is bounded and closed, it may not be Jordan-measurable, as we showed in Exercise 3-11. Thus, $\int_C f$ may not be defined even in the case of C open and f continuous.

3.4 Fubini's Theorem

As with our study of differentiation, we have so far been able to integrate on a case-by-case basis, and now need to produce a general method that will simplify the computation of integration in a broad class of cases. This section will develop Fubini's Theorem, which allows us to simplify computation of integrals into iterated integrals in single variables.

We will first proceed informally in order to develop intuition for the principle behind this theorem. Consider the case of a "sufficiently nice" function $f : [a, b] \times [c, d] \to \mathbb{R}$. Then we can partition $[a, b]$ by $t_0, \ldots, t_k$. For each t_i, the area under the graph of f above $\{t_i\} \times [c, d]$ is

$$\int_c^d f(t_i, y)\,\mathrm{d}y$$

If f is nice, then we can approximate the volume under the graph of f above $[t_{i-1}, t_i] \times [c, d]$ by

$$\int_{[t_{i-1},t_i]\times[c,d]} f \approx (t_i - t_{i-1}) \int_c^d f(x_i, y)\,\mathrm{d}y$$

for any $x_i \in [t_{i-1}, t_i]$. Thus we can approximate the overall integral by

$$\int_{[a,b]\times[c,d]} f = \sum_{i=1}^{k} \int_{[t_{i-1},t_i]\times[c,d]} f \approx \sum_{i=1}^{k} (t_i - t_{i-1}) \int_c^d f(x_i, y)\,\mathrm{d}y$$

But if we consider the single variable integral $\int_a^b (\int_c^d f(x, y)\,\mathrm{d}y)\,\mathrm{d}x$, then this would be approximated by partitions of $[a, b]$ and sums of the form

$$\sum_{i=1}^{k} (t_i - t_{i-1}) \int_c^d f(x_i, y)\,\mathrm{d}y$$

So it seems that for "sufficiently nice" functions, we should have

$$\int_{[a,b]\times[c,d]} f = \int_a^b \left(\int_c^d f(x,y)\,\mathrm{d}y \right) \mathrm{d}x$$

As it turns out, this indeed is the case, but the classification of which functions are "sufficiently nice" becomes a difficult problem. For instance, if $\int_c^d f(x_i, y)\,\mathrm{d}y$ is not defined, then the above equation doesn't even make sense, although f may still be integrable.

Definition 3.12

Let $f : A \to \mathbb{R}$ be bounded with $A \subseteq \mathbb{R}^n$ a closed rectangle. Then the **lower integral** of f on A is

$$\mathbf{L}\int_A f = \sup_{\mathcal{P}} U(f, \mathcal{P})$$

and the **upper integral** is defined similarly as

$$\mathbf{U}\int_A f = \inf_{\mathcal{P}} L(f, \mathcal{P})$$

regardless of whether f is integrable on A.

Theorem 3.12: Fubini's Theorem

Let $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ be closed rectangles, and let $f : A \times B \to \mathbb{R}$ be integrable. For any $x \in A$ define $g_x : B \to \mathbb{R}$ by $g_x(y) = f(x, y)$. Let

$$\mathcal{L}(x) = \mathbf{L}\int_B g_x = \mathbf{L}\int_B f(x,y)\,\mathrm{d}y$$
$$\mathcal{U}(x) = \mathbf{U}\int_B g_x = \mathbf{U}\int_B f(x,y)\,\mathrm{d}y$$

Then $\mathcal{L}$ and $\mathcal{U}$ are integrable on A and

$$\int_{A\times B} f = \int_A \mathcal{L} = \int_A \left(\mathbf{L}\int_B f(x,y)\,\mathrm{d}y \right) \mathrm{d}x$$
$$\int_{A\times B} f = \int_A \mathcal{U} = \int_A \left(\mathbf{U}\int_B f(x,y)\,\mathrm{d}y \right) \mathrm{d}x$$

We refer to integrals of the form $\int_A \left(\mathbf{L}\int_B f(x,y)\,\mathrm{d}y\right) \mathrm{d}x$ or $\int_A \left(\mathbf{U}\int_B f(x,y)\,\mathrm{d}y\right) \mathrm{d}x$ as **iterated integrals**.

Proof. Pick partitions $\mathcal{P}_A$ of A and $\mathcal{P}_B$ of B. Then $\mathcal{P} = (\mathcal{P}_A, \mathcal{P}_B)$ is a partition of $A \times B$.

Moreover, any subrectangle $S \in \mathcal{P}$ is of the form $S_A \times S_B$ for $S_A \in \mathcal{P}_A$, $S_B \in \mathcal{P}_B$. So

$$\begin{aligned}
L(f, \mathcal{P}) &= \sum_{S \in \mathcal{P}} m_S(f) v(S) \\
&= \sum_{S_A \in \mathcal{P}_A, S_B \in \mathcal{P}_B} m_{S_A \times S_B}(f) v(S_A \times S_B) \\
&= \sum_{S_A \in \mathcal{P}_A} v(S_A) \left(\sum_{S_B \in \mathcal{P}_B} m_{S_A \times S_B}(f) v(S_B) \right)
\end{aligned}$$

For any $x \in S_A$ we have $m_{S_A \times S_B}(f) \le m_{S_B}(g_x)$. So for fixed $x \in S_A$,

$$\sum_{S_B \in \mathcal{P}_B} m_{S_A \times S_B}(f) v(S_B) \le \sum_{S_B \in \mathcal{P}_B} m_{S_B}(g_x) \le \mathbf{L} \int_B g_x = \mathcal{L}(x)$$

and thus

$$L(f, \mathcal{P}) = \sum_{S_A \in \mathcal{P}_A} v(S_A) \left(\sum_{S_B \in \mathcal{P}_B} m_{S_A \times S_B}(f) v(S_B) \right) \le \sum_{S_A \in \mathcal{P}_A} m_{S_A}(\mathcal{L}) v(S_A) = L(\mathcal{L}, \mathcal{P}_A)$$

so that

$$L(f, \mathcal{P}) \le L(\mathcal{L}, \mathcal{P}_A) \le U(\mathcal{L}, \mathcal{P}_A) \le U(\mathcal{U}, \mathcal{P}_A) \le U(f, \mathcal{P})$$

where the third inequality follows because $\mathcal{L} \le \mathcal{U}$ and the fourth by a similar argument to what we just proved. Now, f is integrable, which means that

$$\sup L(f, \mathcal{P}) = \inf U(f, \mathcal{P}) = \int_{A \times B} f$$

So that

$$\sup L(\mathcal{L}, \mathcal{P}_A) = \inf U(\mathcal{L}, \mathcal{P}_A) = \int_{A \times B} f$$

Thus $\mathcal{L}$ is integrable on A and

$$\int_A \mathcal{L} = \int_{A \times B} f$$

and similarly $\mathcal{U}$ is integrable iwth

$$\int_A \mathcal{U} = \int_{A \times B} f$$

□

Corollary

Under the same hypotheses,

$$\int_{A \times B} f = \int_B \left(\mathbf{L} \int_A f(x, y)\, \mathrm{d}x \right) \mathrm{d}y = \int_B \left(\mathbf{U} \int_A f(x, y)\, \mathrm{d}x \right) \mathrm{d}y$$

Proof. Analogous. □

The fact that this proof may be repeated in the other order may seem clear based on simply reading the proof However, the important implication is that, for these sufficiently nice functions, not only may our integral be replaced with an iterated integral, but the order of the iterated integral may be changed.

Remark

If each g_x is integrable, then we may dispense with the functions $\mathcal{L}$ and $\mathcal{U}$ and simply write

$$\int_{A\times B} f = \int_A \left(\int_B f(x,y)\,\mathrm{d}y \right) \mathrm{d}x = \int_B \left(\int_A f(x,y)\,\mathrm{d}x \right) \mathrm{d}y$$

In particular, this is the case if f is continuous.

Alternatively, if all but a finite number of g_x are integrable, then we may still write the same, and arbitrarily define the quantity $\int_B f(x,y)\,\mathrm{d}y$ if g_x is not integrable (since changing the value of $\mathcal{L}$ at a finite number of points will not change its integral).

Example 3.4

Define $f : [0,1] \times [0,1] \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} 1, & x \notin \mathbb{Q} \\ 1, & x \in \mathbb{Q}, y \notin \mathbb{Q} \\ 1 - \frac{1}{q}, & x = \frac{p}{q}, y \in \mathbb{Q} \end{cases}$$

where $x = p/q$ is assumed to be in lowest terms. Then f is integrable with $\int_{[0,1]\times[0,1]} f = 1$. But $\int_0^1 f(x,y)\,\mathrm{d}y = 1$ when $x \in \mathbb{Q}$ and does not exist otherwise. So we cannot arbitrarily set the value of $\int_0^1 f(x,y)\,\mathrm{d}y$ wherever the integral doesn't exist. For instance, defining this as zero gives Dirichlet's function, which is not integrable.

Remark

If $A = [a_1, b_1] \times \ldots \times [a_n, b_n]$ and $f : A \to \mathbb{R}$ is "sufficiently nice," then repeated application of Fubini's theorem gives

$$\int_A f = \int_{a_n}^{b_n} \left(\ldots \left(\int_{a_1}^{b_1} f(x_1, \ldots, x_n)\,\mathrm{d}x_1 \right) \ldots \right) \mathrm{d}x_n$$

An application of Fubini's theorem is to integrate over subsets $C \subseteq A \times B$ by appropriately setting bounds on the iterated integrals.

Example 3.5

Let

$$C = ([-1,1]\times[-1,1]) \setminus \{(x,y) : |(x,y)| < 1\}$$

Then

$$\int_C f = \int_{[-1,1]\times[-1,1]} \chi_C f$$

Assuming that f is integrable, $\chi_C f$ is integrable since C is Jordan-measurable. So we may write

$$\int_{[-1,1]\times[-1,1]} \chi_C f = \int_{-1}^{1} \left(\int_{-1}^{1} f(x,y)\chi_C(x,y)\,\mathrm{d}y \right) \mathrm{d}x$$

We have

$$\chi_C(x,y) = \begin{cases} 1, & |y| > \sqrt{1-x^2} \\ 0, & |y| \leq \sqrt{1-x^2} \end{cases}$$

so

$$\int_{-1}^{1} f(x,y)\chi_C(x,y)\,\mathrm{d}y = \int_{\sqrt{1-x^2}}^{1} f(x,y)\,\mathrm{d}y + \int_{-1}^{-\sqrt{1-x^2}} f(x,y)\,\mathrm{d}y$$

and thus

$$\int_C f = \int_{-1}^{1} \left(\int_{\sqrt{1-x^2}}^{1} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x + \int_{-1}^{1} \left(\int_{-1}^{-\sqrt{1-x^2}} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x$$

In general, the problem of determining bounds for arbitrary $C \subseteq A \times B$ is harder. However, one important result of Fubini's theorem is that these bounds may be set in either the $\mathrm{d}y - \mathrm{d}x$ order or the $\mathrm{d}x - \mathrm{d}y$ order, whichever is easier.

3.5 Partitions of Unity

In this section, we will discussion *partitions of unity*. These are an important tool that will help allow us to combine local results into global results, for instance when developing a theory of integration on manifolds.

Definition 3.13

Let $A \subseteq \mathbb{R}^n$. Then a **partition of unity** for A is a collection Φ of C^∞ functions φ which are defined on an open set containing A, such that

1. For all $x \in A$ and all $\varphi \in \Phi$, $0 \leq \varphi(x) \leq 1$.
2. For all $x \in A$ there exists an open set V containing x such that all but finitely many $\varphi \in \Phi$ are 0 on V.
3. For all $x \in A$ it is the case that $\sum_{\varphi \in \Phi} \varphi(x) = 1$, which is a finite sum by 2).

Definition 3.14

Let φ be a partition of unity for some $A \subseteq \mathbb{R}^n$, and let $\mathcal{O}$ be an open cover of A. Then Φ is **subordinate** to $\mathcal{O}$ if, for each $\varphi \in \Phi$ there exists an open set $O \in \mathcal{O}$ such that $\varphi = 0$ outside of some compact set contained in O.

Note: Spivak only requires that $\varphi = 0$ outside of a closed contained in O, but later he makes assumptions which require this set to be compact.

An important tool in proving the existence of partitions of unity will be the smooth bump functions that we proved the existence of in Exercise 2-26. Exercise 2-26 states that if $O \subseteq \mathbb{R}^n$ is open and $C \subseteq O$ is compact, then there exists a closed set $D \subseteq O$ and a C^∞ function which is positive on C and 0 outside of D.

Theorem 3.13

Let $A \subseteq \mathbb{R}^n$ and let $\mathcal{O}$ be an open cover of A. Then there exists a partition of unity Φ for A which is subordinate to $\mathcal{O}$.

Proof. **Case 1:** A is compact.

Note that any partition of unity subordinate to a subcover of $\mathcal{O}$ is also subordinate to $\mathcal{O}$. Since A is compact, we will simply assume $\mathcal{O} = \{U_1, \ldots, U_k\}$ is finite. Now, we will construct a corresponding set of compact sets $D_i \subseteq U_i$ such that $\{\operatorname{int} D_1, \ldots, \operatorname{int} D_k\}$ is also an open cover for A.

To do so, we apply an inductive argument. Let $D_1, \ldots, D_m$ be compact sets chosen so that $\{\operatorname{int} D_1, \ldots, \operatorname{int} D_m, U_{m+1}, \ldots, U_k\}$ covers A. Then let

$$C_{k+1} = A \setminus \left[\left(\bigcup_{i=1}^{m} \operatorname{int} D_i \right) \cup \left(\bigcup_{j=m+2}^{k} U_j \right) \right]$$

Clearly U_{k+1} covers C_{k+1}, and C_{k+1} is the result of a closed set being finitely intersected with the complement of open sets, and is thus closed. So C_{k+1} is compact. Then by Exercise 1-22, there exists a compact set D_{k+1} that satisfies

$$C_{k+1} \subseteq \operatorname{int} D_{k+1}, D_{k+1} \subseteq U_{k+1}$$

By construction, the collection of C_i will cover A, so the collection of $\operatorname{int} D_i$ do as well, and $D_i \subseteq C_i \subseteq U_i$, so this is our desired set.

Now, by Exercise 2-26, we can construct a C^∞ "bump" function ψ_i : which is nonnegative everywhere, strictly positive on D_i, and 0 outside of a closed set contained in U_i. Now, let

$$U = \bigcup_{i=1}^{k} \operatorname{int} D_i$$

$A \subseteq U$ since the $\operatorname{int} D_i$ cover A. Moreover, for $x \in U$, x is in some D_i, and the rest are nonnegative, so

$$\sum_{i=1}^{k} \psi_i > 0$$

on U. So we may define $\varphi_i : U \to \mathbb{R}$ by

$$\varphi_i(x) = \frac{\psi_i(x)}{\sum_{j=1}^{k} \psi_j(x)}$$

which is also smooth on U. Then the collection $\{\varphi_1, \ldots, \varphi_k\}$ is a partition of unity. However, it must be noted that this collection is not necessarily subordinate to $\mathcal{O}$. Indeed, we know that $\psi_i = 0$ outside of some closed set K contained in U_i. However, it may be the case that K is not completely contained within U. In this case, φ_1 is not even defined on K, let alone outside of it.

Moreover, it is not necessarily that case that φ_1 goes to zero at the boundary of its support. For instance, suppose $k = 1$, so that we have only a single bump function ψ_1. Then ψ_1 goes to zero, but φ_1 is identically 1.

We can remedy this by applying Exercise 2-26 once more to construct a C^∞ function $f : U \to [0, 1]$ which is 1 on A and 0 outside of a closed set K' contained in U. Moreover, we can ensure that K' is bounded since A is, so K' is compact. Then the collection $\Phi = \{f\varphi_1, \ldots, f\varphi_k\}$ is still a partition of unity for A (since $f\varphi_i = \varphi_i$ on A), but this time $f\varphi_i$ is zero outside of the compact set $K \cap K' \subseteq U_i$, so Φ is subordinate to $\mathcal{O}$.

Case 2: $A = \bigcup_{i=1}^{\infty} A_i$, where A_i is compact and $A_i \subseteq \operatorname{int} A_{i+1}$.

Define $B_1 = A_1$ and $B_i = A_i \setminus \operatorname{int} A_{i-1}$ for all $i \geq 2$.

Claim

Suppose $x \in A_i$. Then $x \in B_j$ for some $j \leq i$.

Proof. We prove this by induction. In the base case, $x \in A_1 \implies x \in B_1$ since $A_1 = B_1$.

For $i \geq 2$, if $x \in A_i$ then $x \in B_i$ or $x \in \operatorname{int} A_{i-1}$. But $\operatorname{int} A_{i-1} \subseteq A_{i-1}$, so $x \in A_{i-1}$. By the inductive hypothesis $x \in B_j$ for some $j \leq i - 1 < i$. □

By the claim, we have $A \subseteq \bigcup B_i$, and clearly $\bigcup B_i \subseteq A$, so $\bigcup B_i = A$.

Define the open cover $\mathcal{O}_i$ by

$$\mathcal{O}_i = \begin{cases} \{O \cap \operatorname{int} A_{i+1} : O \in \mathcal{O}\}, & i = 1, 2 \\ \{O \cap (\operatorname{int} A_{i+1} \setminus A_{i-2}) : O \in \mathcal{O}\}, & i \geq 3 \end{cases}$$

We will construct a partition of unity for each B_i subordinate to $\mathcal{O}_i$.

Note that each B_i is compact, and that $\mathcal{O}_i$ covers B_i. So by Case 1 there exists a partition of unity Φ_i for B_i subordinate to $\mathcal{O}_i$, where the functions are defined on some open set U_i containing B_i. Now let $x \in A$. Then $x \in B_i$ for some i. Thus $x \in A_i$. Moreover, for any $j \geq i+2$, $x \in A_i \subseteq A_{j-2}$ so $x \notin O \cap (\operatorname{int} A_{j+1} \setminus A_j)$ for any $O \in \mathcal{O}$, and thus $x \notin O'$ for any $O' \in \mathcal{O}_j$. Since Φ_j is subordinate to $\mathcal{O}_j$, $\varphi(x) = 0$ for any $\varphi \in \Phi_j$ with $j \geq i+2$. As a result, the sum

$$\sigma(x) = \sum_{j=1}^{\infty} \sum_{\varphi \in \Phi_j} \varphi(x) = \sum_{j=1}^{i+2} \sum_{\varphi \in \Phi_j} \varphi(x) \geq 1$$

is a finite sum. Now for any $\varphi \in \Phi_j$ for any j, define $\varphi' : U_i \to \mathbb{R}$ by

$$\varphi'(x) = \frac{\varphi(x)}{\sigma(x)}$$

Moreover, the domain may be extended to $\bigcup U_i$ by simply setting $\varphi' = 0$ outside of U_i.[1] Then the collection $\Phi = \{\varphi' : \varphi \in \Phi_j, j \in \mathbb{N}\}$ satisfies conditions 1 and 3 for being a partition of unity. For condition 2, suppose $x \in A_i$. Then for each $j \leq i+2$, there exists an open set V_j containing x such that all but finitely many $\varphi \in \Phi_j$ are zero on V_j. Let $V = V_1 \cup \ldots \cup V_{i+2}$, which is open. By the argument above, $\varphi(x) = 0$ if $\varphi \in \Phi_j$ for $j > i+2$, so there are only finitely many nonzero φ at x, and thus only finitely many φ' are nonzero at x.

So Φ is a partition of unity. Let $\varphi' \in \Phi$. Then $\varphi \in \Phi_j$ for some j. Φ_j is subordinate to $\mathcal{O}_j$, so there exists $O' = O \cap (\operatorname{int} A_{j+1} \setminus A_{j-2}) \in \mathcal{O}_j$ such that φ is zero outside of a compact set contained in $O' \subseteq O$. Then φ' is also zero outside this set (assured since $U \subseteq \operatorname{int} A_{j+1} \subseteq A$, and φ' is defined on $A \subseteq U$). So Φ' is subordinate to $\mathcal{O}$.

Case 3: A is open.

Let $d(x, \partial A)$ be the distance from x to ∂A as defined in Exercise 1-21 part a). Define

$$A_i : \{x : |x| \leq i, d(x, \partial A) \geq \frac{1}{i}\}$$

For any $x \in A$, $|x| < M$ for some $M \in \mathbb{N}$, and $d(x, \partial A) \geq \frac{1}{N}$ for some other $N \in \mathbb{N}$ since A is open. So $x \in A_i$ for some i and thus $A = \bigcup_{i=1}^{\infty} A_i$. So A is of the type considered in Case 2.

Case 4: A is arbitrary.

Let $B = \bigcup_{O \in \mathcal{O}} O$. Then apply Case 3 to get a partition of unity Φ for B subordinate to $\mathcal{O}$. Then this is also a partition of unity for A. $\square$

[1] For details, see my answer here

Remark 3.1

For any $C \subseteq A$, if Φ is a partition of unity for A, then for $x \in C$, there exists V_x open containing x such that only finitely many φ are nonzero on V_x. Then these V_x are an open cover of C, so by compactness we only need finitely many and thus only finitely many φ are nonzero on C. In particular, if A is compact then we only need finitely many φ (this was already proved in Case 1).

Remark 3.2

Note that our proof shows that we may demand that our partition of unity is countable.

Similarly to compactness, partitions of unity will allow us to make local constructions and combine them into a global result. We will demonstrate this by extending our definition of the integral to general open sets.

Definition 3.15

Let $A \subseteq \mathbb{R}^n$ be open and let $\mathcal{O}$ be an open cover of A. $\mathcal{O}$ is said to be **admissible** if $O \subseteq A$ for each $O \in \mathcal{O}$ (equivalently, if $\bigcup_{O\in\mathcal{O}} O = A$).

Let Φ be a partition of unity for an open set $A \subseteq \mathbb{R}^n$ (not necessarily bounded) subordinate to an admissible open cover $\mathcal{O}$. Suppose also that $f : A \to \mathbb{R}$ is bounded in an open set around each point of A, and that its set of discontinuities has measure zero. Since φ has compact support, let $C_\varphi \subseteq A$ be a closed rectangle such that $\varphi = 0$ outside of C_φ ($C_\varphi \subseteq A$ is guaranteed since Φ is subordinate to $\mathcal{O}$, which is admissible). Since f is bounded in an open neighborhood around each point, we apply compactness to pick a finite number of them and conclude f is boudned on C_φ. $|f|$ is continuous whenever f is, so it is discontinuous on a set of measure zero and thus $|f|$ is integrable on C_φ. φ is also continuous, so $\int_{C_\varphi} \varphi|f|$ exists.

Now, by Remark 3.2, Φ is countable. So we may consider the series

$$\sum_{\varphi\in\Phi} \int_{C_\varphi} \varphi|f|$$

Suppose this series converges. Since $0 \le \varphi \le 1$, $\varphi|f| = |\varphi f|$, and thus by Exercise 3-6,

$$\left|\int_{C_\varphi} \varphi f\right| \le \int_{C_\varphi} |\varphi f| = \int_{C_\varphi} \varphi|f|$$

so the series

$$\sum_{\varphi\in\Phi} \left|\int_{C_\varphi} \varphi f\right|$$

converges absolutely. This means it is independent of our ordering of Φ. Moreover, we will show that this value is also independent of our choices of Φ and $\mathcal{O}$, allowing us to define this value without reference to any specific cover or partition of unity. Noe that this is only the case if the series $\sum_{\varphi\in\Phi}\int_{C_\varphi}\varphi|f|$ converges; the convergence of $\sum_{\varphi\in\Phi}|\int_{C_\varphi}\varphi f|$ is not a sufficient condition.

Definition 3.16

Let $A \subseteq \mathbb{R}^n$ be open. Suppose $f : A \to \mathbb{R}$ is bounded in an open set around each point of A, and its set of discontinuities has measure zero. Let Φ be a partition of unity for A subordinate to an admissible open cover $\mathcal{O}$ of A. For each $\varphi \in \Phi$, let $C_\varphi \subseteq A$ be a closed rectangle such that $\varphi = 0$ outside of C_φ. Then if the series

$$\sum_{\varphi\in\phi}\int_{C_\varphi}\varphi|f|$$

converges, then we say that f is **extended integrable** relative to Φ. Moreover, we define the extended integral of f on A relative to Φ to be

$$\operatorname*{ext}_{\Phi}\int_A f = \sum_{\varphi\in\phi}\int_{C_\varphi}\varphi f$$

Theorem 3.14

Let $A \subseteq \mathbb{R}^n$ be open. Suppose $f : A \to \mathbb{R}$ is bounded in an open set around each point of A, and its set of discontinuities has measure zero. Let Φ be a partition of unity for A subordinate to an admissible open cover $\mathcal{O}$ of A. Let Ψ be another partition of unity for A subordinate to another admissible open cover $\mathcal{O}'$ of A. If f is extended integrable relative to Φ, then it is extended integrable relative to Ψ, and

$$\operatorname*{ext}_{\Phi}\int_A f = \sum_{\varphi\in\Phi}\int_{C_\varphi}\varphi f = \sum_{\psi\in\Psi}\int_{C_\psi}\psi f = \operatorname*{ext}_{\Psi}\int_A f$$

Proof. For each $\varphi \in \Phi$, C_φ is compact, so by Remark 3.1, only finitely many $\psi \in \Psi$ are nonzero on C_φ. Moreover, the finite sum $\sum_{\psi\in\Psi}\psi = 1$ on $C_\varphi \subseteq A$ (the subset follows since $\mathcal{O}$ is admissible), so we have

$$\sum_{\varphi\in\Phi}\int_{C_\varphi}\varphi|f| = \sum_{\varphi\in\Phi}\left(\int_{C_\varphi}\varphi|f|\left(\sum_{\psi\in\Psi}\psi\right)\right) = \sum_{\varphi\in\Phi}\int_{C_\varphi}\sum_{\psi\in\Psi}\varphi\psi|f| = \sum_{\varphi\in\Phi}\sum_{\psi\in\Psi}\int_{C_\varphi}\varphi\psi|f|$$

Now, since the left sides series converges by assumption, the right side series does as well. Since $|\varphi\psi|f|| = \varphi\psi|f|$,

$$\sum_{\varphi\in\Phi}\sum_{\psi\in\Psi}\int_{C_\varphi}\varphi\psi|f|$$

converges absolutely and thus we may switch the order of the sums:

$$\sum_{\varphi\in\Phi}\sum_{\psi\in\Psi}\int_{C_\varphi}\varphi\psi|f| = \sum_{\psi\in\Psi}\sum_{\varphi\in\Phi}\int_{C_\varphi}\varphi\psi|f|$$

Now, since ψ and φ are both zero outside of a compact set, if we let R be a rectangle containing both C_φ and C_ψ, we have

$$\int_{C_\varphi}\varphi\psi|f| = \int_R \varphi\psi|f| = \int_{C_\psi}\varphi\psi|f|$$

So

$$\sum_{\psi\in\Psi}\sum_{\varphi\in\Phi}\int_{C_\varphi}\varphi\psi|f| = \sum_{\psi\in\Psi}\sum_{\varphi\in\Phi}\int_{C_\psi}\varphi\psi|f|$$

By the argument we made at the beginning, the sum $\sum_{\varphi\in\Phi}$ is finite and equal to 1 on C_ψ, so we have

$$\sum_{\psi\in\Psi}\sum_{\varphi\in\Phi}\int_{C_\psi}\varphi\psi|f| = \sum_{\psi\in\Psi}\int_{C_\psi}\psi|f|$$

So we have shown that

$$\sum_{\varphi\in\Phi}\int_{C_\varphi}\varphi|f| = \sum_{\psi\in\Psi}\int_{C_\psi}\psi|f|$$

so the right side converges, and thus f is extended integrable relative to Ψ. Repeating this argument with f substituted for $|f|$ shows that

$$\operatorname*{ext}_{\Phi}\int_A f = \sum_{\varphi\in\Phi}\int_{C_\varphi}\varphi f = \sum_{\psi\in\Psi}\int_{C_\psi}\psi f = \operatorname*{ext}_{\Psi}\int_A f \qquad \square$$

By Theorem 3.14, our choice of partition is irrelevant when considering extended integrability and the value of the integral, so long as f is extended integrable with respect to some partition. Thus we may define this without reference to a particular partition.

Definition 3.17

Let $A \subseteq \mathbb{R}^n$ be open. Suppose $f : A \to \mathbb{R}$ is bounded in an open set around each point of A, and its set of discontinuities has measure zero. Then f is **extended integrable** if it is extended integrable relative to some partition of unity Φ, and the extended integral of f on A is

$$\operatorname{ext}\int_A f = \operatorname*{ext}_{\Phi}\int_A f$$

Theorem 3.15

If $A \subseteq \mathbb{R}^n$ is open and bounded, $f : A \to \mathbb{R}$ is bounded, and its set of discontinuities is a set of measure zero, then f is extended integrable.

Proof. Let $\Phi = \{\varphi_1, \varphi_2, \ldots\}$ be a countable (by Remark 3.2) partition of unity subordinate to some admissible cover $\mathcal{O}$. Suppose $|f| \leq M$ on A. Then let

$$S_k = \sum_{i=1}^{k} \int_{C_{\varphi_i}} \varphi_i |f|$$

be the kth partial sum of the corresponding infinite series. Since $\varphi_i |f| \geq 0$,

$$\int_{C_{\varphi_i}} \varphi_i |f| \geq 0$$

for each i, and thus (S_k) is increasing. Let B be some rectangle containing A. Since Φ is subordinate to an admissible cover, $C_\varphi \subseteq A \subseteq B$, and thus

$$\int_B \varphi = \int_{C_\varphi} \varphi$$

and thus

$$S_k = \sum_{i=1}^{k} \int_{C_{\varphi_i}} \varphi_i |f| \leq \sum_{i=1}^{k} M \int_{C_{\varphi_i}} \varphi = M \int_B \sum_{i=1}^{k} \varphi \leq M \int_B 1 = Mv(B)$$

which is constant. So (S_k) is increasing and bounded above, so it is convergent and thus f is extended integrable. □

Theorem 3.16

Let $A \subseteq \mathbb{R}^n$ be open and Jordan-measurable. Let $f : A \to \mathbb{R}$ be bounded, and suppose its set of discontinuities has measure zero. Then

$$\int_A f = \operatorname{ext} \int_A f$$

Proof. Note that since A is Jordan-measurable, it is bounded and thus f is extended integrable by Theorem 3.15.

Let $\varepsilon > 0$, and let Φ be an arbitrary partition of unity subordinate to an admissible open cover $\mathcal{O}$. Let M be such that $|f| \leq M$. Then by Exercise 3-22, there exists a compact Jordan-measurable set $C \subseteq A$ such that

$$\int_{A \setminus C} 1 < \frac{\varepsilon}{M}$$

By Remark 3.1, the subpartition Φ' of those $\varphi \in \Phi$ which are nonzero on C is finite. Then we have

$$\left| \int_A f - \operatorname{ext} \int_A f \right| = \left| \int_A f - \sum_{\varphi \in \Phi'} \int_{C_\varphi} \varphi f \right|$$

Since $\mathcal{O}$ is admissible, $C_\varphi \subseteq A$ for each $\varphi \in \Phi'$ and thus

$$\int_A \varphi f = \int_{C_\varphi} \varphi f$$

so that

$$\left| \int_A f - \sum_{\varphi \in \Phi'} \int_{C_\varphi} \varphi f \right| = \left| \int_A f - \int_A \sum_{\varphi \in \Phi'} \varphi f \right| \leq \int_A |f| \left(1 - \sum_{\varphi \in \Phi'} \varphi \right)$$

Now, we have

$$\sum_{\varphi \in \Phi} \varphi = 1$$

on A, so we may write

$$\int_A |f| \left(1 - \sum_{\varphi \in \Phi'} \varphi \right) = \int_A |f| \left(\sum_{\varphi \in \Phi} \varphi - \sum_{\varphi' \in \Phi'} \varphi' \right) \leq M \int_A \left(\sum_{\varphi \in \Phi} \varphi - \sum_{\varphi' \in \Phi'} \varphi' \right)$$

Let Ψ be the collection of $\varphi \in \Phi$ such that $\varphi \notin \Phi'$. In other words, Ψ is the collection of φ which are zero on C. Then

$$M \int_A \left(\sum_{\varphi \in \Phi} \varphi - \sum_{\varphi' \in \Phi'} \varphi' \right) = M \int_A \sum_{\psi \in \Psi} \psi$$

Since the $\psi \in \Psi$ are zero on C, they are only nonzero on $A \setminus C$. Thus

$$M \int_A \sum_{\psi \in \Psi} \psi \leq M \int_{A \setminus C} \sum_{\psi \in \Psi} \psi \leq M \int_{A \setminus C} 1 < \varepsilon$$

So

$$\int_A f = \operatorname{ext} \int_A f$$

□

3.6 Change of Variables

Consider the "u-substitution" strategy employed in single variable calculus. If $u : [a, b] \to \mathbb{R}$ is a continuously differentiable function and $f : \mathbb{R} \to \mathbb{R}$ is continuous, then let F be such that $F' = f$. By the chain rule, $(F \circ u)' = (f \circ u)u'$. Thus

$$\int_{u(a)}^{u(b)} f = \int_{u(a)}^{u(b)} F' = F(u(b)) - F(u(a)) = \int_a^b (F \circ u)' = \int_a^b (f \circ u)u'$$

For instance, this strategy could be used computationally as follows:

$$\int_0^3 2x \sin(x^2)\, \mathrm{d}x = \int_0^9 \sin u \,\mathrm{d}u = \cos 0 - \cos 9 = 1 - \cos 9$$

Claim

Let $u : [a,b] \to \mathbb{R}$ be continuously differentiable and injective. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous. Then

$$\int_{u(a,b)} f = \int_{(a,b)} (f \circ u)|u'|$$

Proof. Since u is continuous and injective, it is strictly monotone. Suppose it is strictly increasing. Then $|u'| = u'$ and $u(b) > u(a)$, so $u(a,b) = (u(a), u(b))$ and the claim follows directly from the equality above.

If u is decreasing, then $|u'| = -u'$, and $u(b) < u(a)$, so that $u(a,b) = (u(b), u(a))$. Thus

$$\int_{u(a,b)} f = \int_{u(b)}^{u(a)} f = -\int_{u(a)}^{u(b)} f = \int_a^b -(f \circ u)u' = \int_a^b (f \circ u)|u'|$$

□

This method is invaluable for computational calculus, which motivates the development of an equivalent technique in multiple dimensions. We will do so by first proving it for linear transformations.

Lemma 3.17

Let $A \subseteq \mathbb{R}^n$ be open and let $u : A \to \mathbb{R}^n$ be injective and continuously differentiable with $\det u'(x) \neq 0$ on A. Suppose there exists an admissible cover $\mathcal{O}$ for A such that for all $U \in \mathcal{O}$ and $f : U \to \mathbb{R}$ integrable it is the case that

$$\text{ext} \int_{u(U)} f = \text{ext} \int_U (f \circ u)|\det u'|$$

Then

$$\text{ext} \int_{u(A)} f = \text{ext} \int_A (f \circ u)|\det u'|$$

Proof. The collection $\mathcal{U} = \{u(O)\}_{O \in \mathcal{O}}$ is an open cover for $u(A)$, so we may pick a partition of unity Φ for $u(A)$ subordinate to $\mathcal{U}$. Suppose that $U_\varphi \in \mathcal{U}$ contains C_φ for each φ. Then we have

$$\begin{aligned}
\text{ext} \int_{u(A)} f &= \sum_{\varphi \in \Phi} \int_{U_\varphi} \varphi f \\
&= \sum_{\varphi \in \Phi} \int_{u^{-1}(U_\varphi)} (\varphi \circ u)(f \circ u)|\det u'| \\
&= \sum_{\varphi \in \Phi} \int_A (\varphi \circ u)(f \circ u)|\det u'|
\end{aligned}$$

Let Ψ be the partition of unity for A given by $\{\varphi \circ u\}_{\varphi \in \Phi}$. Then Ψ is thus subordinate to $\mathcal{O}$. Then

$$\sum_{\varphi \in \Phi} \int_A (\varphi \circ u)(f \circ u)|\det u| = \sum_{\psi \in \Psi} \int_A \psi(f \circ u)|\det u'| = \text{ext} \int_A (f \circ u)|\det u| \qquad \square$$

Lemma 3.18

Let $A \subseteq \mathbb{R}^n$ be open and let $u : A \to \mathbb{R}^n$ be linear, with $\det u(x) \neq 0$ on A. Then if $f : u(A) \to \mathbb{R}$ is integrable, we have

$$\text{ext} \int_{u(A)} f = \text{ext} \int_A (f \circ u)|\det u|$$

Proof. First note that by Exercise 3-35, if f is the constant function 1 and U is an open rectangle, then

$$\int_{u(U)} 1 = v(u(U)) = |\det u| v(U) = |\det u| \int_U 1 = \int_U |\det u|$$

We can make an analogous argument for u^{-1}, so for any open rectangle U we have

$$\int_{u^{-1}(U)} 1 = \int_U |\det u^{-1}| \implies \int_U 1 = \int_{u^{-1}(U)} |\det u|$$

Now suppose f is arbitrary. Let $V \subseteq u(A)$ be a rectangle, and let $\mathcal{P}$ be a partition of V.

$$\begin{aligned} L(f, \mathcal{P}) &= \sum_{S \in \mathcal{P}} v(S) m_S(f) \\ &= \sum_{S \in \mathcal{P}} m_S(f) \int_S 1 \\ &= \sum_{S \in \mathcal{P}} m_S(f) \int_{u^{-1}(\text{int}\, S)} |\det u| \\ &= \sum_{S \in \mathcal{P}} \int_{u^{-1}(\text{int}\, S)} m_S(f) |\det u| \end{aligned}$$

For each $S \in \mathcal{P}$, define $f|_S : S \to \mathbb{R}$ to be the constant function $f|_S(x) = m_S(f)$. Then we have

$$\begin{aligned} \sum_{S \in \mathcal{P}} \int_{u^{-1}(\text{int}\, S)} m_S(f)|\det u| &= \sum_{S \in \mathcal{P}} \int_{u^{-1}(\text{int}\, V)} (f|_S \circ u)|\det u| \\ &\leq \sum_{S \in \mathcal{P}} \int_{u^{-1}(\text{int}\, S)} (f \circ u)|\det u| \\ &\leq \int_{u^{-1}(V)} (f \circ u)|\det u| \end{aligned}$$

So $\int_{u^{-1}(V)}(f \circ u)|\det u|$ is an upper bound for all $L(f, \mathcal{P})$, but $\int_V f$ is the least such upper bound, so we have

$$\int_V f \leq \int_{u^{-1}(V)} (f \circ u)|\det u|$$

An analogous argument shows the reverse inequality, so we conclude that

$$\int_V f = \int_{u^{-1}(V)} (f \circ u)|\det u|$$

for every $V \subseteq u(A)$ and any f.

Since A is open, and u is a continuous injection, $u(A)$ is open. So for each $\alpha \in u(A)$, we may pick $V_\alpha \subseteq u(A)$ containing $u(A)$. Then the collection $\mathcal{V} = \{V_\alpha\}_{\alpha \in u(A)}$ is an admissible open cover for $u(A)$, and the hypothesis of Lemma 3.17 applies. So we conclude that

$$\operatorname{ext} \int_{u(A)} f = \operatorname{ext} \int_A (f \circ u)|\det u| \qquad \square$$

We now progress to the general case. To do so, we will need to replace u with u' (which are equal in the linear case).

Theorem 3.19

Let $A \subseteq \mathbb{R}^n$ be open and let $u : A \to \mathbb{R}^n$ be one-to-one and continuously differentiable. Moreover, suppose that $\det u'(x) \neq 0$ on A. Then for any integrable $f : u(A) \to \mathbb{R}$, we have

$$\operatorname{ext} \int_{u(A)} f = \operatorname{ext} \int_A (f \circ u)|\det u'|$$

We first prove one simplifying lemma.

Lemma 3.20

Suppose that the conclusion of Theorem 3.19 holds for two change-of-variable functions $g : A \to \mathbb{R}^n$ and $h : B \to \mathbb{R}^n$. Moreover, assume that $g(A) \subseteq B$. Then the theorem holds for $h \circ g$.

Proof. We have

$$\begin{aligned}
\operatorname{ext} \int_{(h \circ g)(A)} f &= \operatorname{ext} \int_{g(A)} (f \circ h)|\det h'| \\
&= \operatorname{ext} \int_A (f \circ h \circ g)[|\det h'| \circ g]|\det g'| \\
&= \operatorname{ext} \int_A (f \circ (h \circ g))|\det(h \circ g)'|
\end{aligned} \qquad \square$$

Returning to the main proof,

Proof of Theorem 3.19. We induct on n. For the base case $n = 1$, we can form an admissible open cover of A by open intervals, and the result follows from the discussion beginning this section combined with Lemma 3.17.

Suppose the theorem is proved for $n - 1$. Then for n, we will attempt to find an open set $U_\alpha \subseteq A$ containing α for each $\alpha \in A$ such that

$$\int_{u(U_\alpha)} f = \int_{U_\alpha} (f \circ u)|\det u'|$$

Then the theorem follows from Lemma 3.17. Thus, fix some $\alpha \in A$. Then

$$(Du(\alpha)^{-1} \circ u)'(\alpha) = \frac{u(\alpha)^{-1}}{'} u'(\alpha) = I$$

Note that Lemma 3.18 implies that the theorem is true for $Du(\alpha)^{-1}$. If the theorem is true for $(Du(\alpha)^{-1} \circ u)$, then it follows from Lemma 3.20 that it is true for u. So we may assume that $Du(\alpha) = \mathrm{id}$.

Define the function $h : A \to \mathbb{R}^n$ by

$$h(x) = (u_1(x), \ldots, u_{n-1}(x), x_n)$$

Then $h'(\alpha) = I$. h is continuously differentiable, so there exists an open set $U' \subseteq A$ containing α where h is injective and invertible. Then define $k : h(U') \to \mathbb{R}$ by

$$k(x) = (x_1, \ldots, x_{n-1}, u_n(h^{-1}(x)))$$

so that $u = k \circ h$. Both of these functions only change at most $n - 1$ variables, so we will be able to apply the inductive hypothesis. Afterward, we would now like to apply Lemma 3.20; however, we cannot be assured that k is injective with k' invertible.

To remedy this, note that

$$(g^n \circ h^{-1})'(\alpha) = (g^n)'(h^{-1}(h(\alpha))) \underbrace{[h'(h^{-1}(h(\alpha)))]^{-1}}_{\text{Inverse Function Thm}} = (g^n)'(\alpha)[h'(\alpha)]^{-1} = (g^n)'(\alpha)$$

So $D_n(g^n \circ h^{-1})(h(\alpha)) = D_n g^n(\alpha) = 1$, and thus $k'(h(\alpha)) = I$. So we can find an open set $V \subseteq h(U)$ containing $h(\alpha)$ where k is injective and k' is invertible. We can then restrict h to $U = k^{-1}(V)$, and then h, k satisfy the hypotheses of Lemma 3.20.

We now prove that the theorem applies to h. The proof for k is easier. Pick an open rectangle $W \subseteq U$, and suppose $W = D \times [a_n, b_n]$, with $D \subseteq \mathbb{R}^{n-1}$. Because h does not change the n-th coordinate, Fubini's Theorem gives

$$\int_{h(W)} 1 = \int_{[a_n, b_n]} \left(\int_{h(D \times \{x_n\})} 1 \, \mathrm{d}x_1 \ldots \mathrm{d}x_{n-1} \right) \mathrm{d}x_n$$

For each x_n, define $h_{x_n} : D \to \mathbb{R}^{n-1}$ by

$$h_{x_n}(x_1, \ldots, x_{n-1}) = h(u_1(x_1, \ldots, x_n), \ldots, u_{n-1}(x_1, \ldots, x_n))$$

so that

$$\det h'_{x_n}(x_1, \ldots, x_{n-1}) = \det h'(x_1, \ldots, x_n) \neq 0$$

Moreover, h_{x_n} is injective, so the inductive hypothesis applies. We also have

$$\int_{h(D\times\{x_n\})} 1 \, \mathrm{d}x_1 \ldots \mathrm{d}x_{n-1} = \int_{h_{x_n}(D)} 1$$

Then using the inductive hypothesis, we have

$$\begin{aligned}
\int_{h(W)} 1 &= \int_{[a_n,b_n]} \left(\int_{h(D\times\{x_n\})} 1 \, \mathrm{d}x_1 \ldots \mathrm{d}x_{n-1} \right) \mathrm{d}x_n \\
&= \int_{[a_n,b_n]} \left(\int_{h_{x_n}(D)} 1 \right) \mathrm{d}x_n \\
&= \int_{[a_n,b_n]} \left(\int_D |\det h'_{x_n}| \right) \mathrm{d}_{x_n} \\
&= \int_{[a_n,b_n]} \left(\int_D |\det h'(x_1, \ldots, x_n)| \right) \mathrm{d}_{x_n} \\
&= \int_W |\det h'|
\end{aligned}$$

From the proof for Lemma 3.18, it is sufficient to prove the theorem for the constant function 1. So we conclude that the theorem holds for h. A similar argument holds for k. By Lemma 3.20, it holds for u. □

We will now prove a simple version of an important theorem.

Theorem 3.21: Sard's Theorem

Suppose $g : A \to \mathbb{R}^n$ is continuously differentiable, with $A \subseteq \mathbb{R}^n$ open. Let $B = \{x \in A : \det g'(x) = 0\}$ be the set of critical values of g. Then $g(B)$ has measure zero.

Proof. Suppose that $U \subseteq A$ is a closed n-cube with side length ℓ. Since U is compact, each $D_j g^i$ is uniformly continuous on U. Thus there exists N large enough such that when U is divided into N^n subcubes, then for any x, y which are both in the same subcube and any i, j we have

$$|D_j g^i(y) - D_j g^i(x)| < \frac{\varepsilon}{n^2}$$

Fix some subcube S and some $x \in S$. Define $f(z) = Dg(x)(z) - g(z)$, so that its partial derivatives are bounded:

$$|D_j f^i(z)| = |D_j g^i(x) - D_j g^i(z)| < \frac{\varepsilon}{n^2}$$

Then by Lemma 2.10, for any $y \in S$ we have

$$|Dg(x)(y - x) - g(y) + g(x)| = |f(y) - f(x)| < \varepsilon |y - x| \leq \varepsilon \sqrt{n} \frac{\ell}{N}$$

We can repeat this for each x, so this holds whenever x, y are in the same subcube. If $S \cap B \neq \varnothing$, then fix $x \in S \cap B$. Then we have $\det g'(x) = 0$, so $Dg(x)(S)$ is a subset of an $n-1$ dimensional subspace V of $\mathbb{R}^n$. Then every point $\{g(y) - g(x) : y \in S\}$ is contained within $\varepsilon\sqrt{n}(\ell/N)$ of V, meaning that $g(S)$ is contained within $\varepsilon\sqrt{n}\ell/N$ of $V + g(x)$. Moreover, each $D_j g^i$ is uniformly continuous on S, so they are boudned by some M. Then by Lemma 2.10, we have

$$|g(x) - g(y)| < n^2 M|x - y| \leq n^2 M\sqrt{n}\frac{\ell}{N}$$

Thus $g(S)$ lies within a cylinder with height $2\varepsilon\sqrt{n}\ell/N$ and base given by a $n-1$-sphere with radius $n^2 M\sqrt{n}\ell/N$, which has volume bounded by $C(\ell/N)^n\varepsilon$ for an appropriate constant C. Then the total volume of these cylinders (which covers $g(U \cap B)$ for each S is $C\ell^n\varepsilon$. So $g(U \cap B)$ has measure zero. Now, we can produce a cover of A (countable by Exercise 3-13) and repeat this process, so $g(B)$ is the countable union of measure zero sets, and thus measure zero. $\square$

Sard's Theorem, among many other applications, allows us prove Theorem 3.19 without the assumption $\det u'(x) \neq 0$. This is the content of Exercise 3-39.

Chapter 4

Integration on Chains

4.1 Algebraic Preliminaries

In this chapter, we will begin to develop our theory of integration over objects with richer structure than pure subsets of $\mathbb{R}^n$. This will allow us to define integrals over parameterized objects, such as line integrals and surface integrals, and we will prove a version of Stokes' Theorem for this setting. We will also set th egroundwork for the development of a similar theory for manifolds in Chapter 5.

Definition 4.1

Let V be a real vector space, and let $V^k = V \times \ldots \times V$ k times. A **multilinear** function $T : V^k \to \mathbb{R}$ is a function such that, for each $1 \leq i \leq k$ and each $v = (v_1, \ldots, v_k) \in V^k$ the function $T_v^i : V \to \mathbb{R}$ defined by

$$T_v^i(y) = T(v_1, \ldots, \underbrace{y}_{i}, \ldots, v_k)$$

is linear. Such a function is also called a k**-tensor** on V.

Definition 4.2

The set of all k-tensors on a fixed vector space V is denoted $\mathfrak{J}^k(V)$. $\mathfrak{J}^k(V)$ is a real vector space if the operations are defined as

$$\begin{aligned}(S+T)(v_1, \ldots, v_k) &= S(v_1, \ldots, v_k) + T(v_1, \ldots, v_k) \\ (aS)(v_1, \ldots, v_k) &= a(S(v_1, \ldots, v_k))\end{aligned}$$

Definition 4.3

Suppose $S \in \mathfrak{J}^k(V)$ and $T \in \mathfrak{J}^k(V)$. Then the **tensor product** of S and T is a $k+l$-tensor $S \otimes T$ defined by

$$(S \otimes T)(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+l}) = S(v_1, \ldots, v_k) \cdot T(v_{k+1}, \ldots, v_{k+l})$$

Note that the tensor product is clearly not commutative. Because tensors are maps into $\mathbb{R}$, we may use properties of $\mathbb{R}$ to derive similar properties for tensors.

Proposition 4.1

The following are properties of the tensor product:

1. $(S_1 + S_2) \otimes T = S_1 \otimes T + S_2 \otimes T$
2. $S \otimes (T_1 + T_2) = S \otimes T_1 + S \otimes T_2$
3. $(aS) \otimes T = S \otimes (aT) = a(S \otimes T)$
4. $S \otimes (T \otimes U) = (S \otimes T) \otimes U$

Proof. 1. Let $S_1, S_2 \in \mathfrak{J}^k(V)$ and $T \in \mathfrak{J}^l(V)$. Let $v = (v_1, \ldots, v_{k+l}) \in V^{k+l}$. Let $v^k = (v_1, \ldots, v_k)$ and $v^l = (v_{k+1}, \ldots, v_{k+l})$. Then

$$\begin{aligned}
((S_1 + S_2) \otimes T)(v) &= (S_1 + S_2)(v^k) \cdot T(v^l) \\
&= (S_1(v^k) + S_2(v^k)) \cdot T(v^l) \\
&= S_1(v^k) \cdot T(v^l) + S_2(v^k) \cdot T(v^l) \\
&= (S_1 \otimes T)(v) + (S_2 \otimes T)(v) \\
&= (S_1 \otimes T + S_2 \otimes T)(v)
\end{aligned}$$

2. Let $S \in \mathfrak{J}^k(V)$ and $T_1, T_2 \in \mathfrak{J}^l(V)$. Using the same notation,

$$\begin{aligned}
(S \otimes (T_1 + T_2))(v) &= S(v^k) \cdot (T_1 + T_2)(v^l) \\
&= S(v^k) \cdot (T_1(v^l) + T_2(v^l)) \\
&= S(v^k) \cdot T_1(v^l) + S(v^k) \cdot T_2(v^l) \\
&= (S \otimes T_1)(v) + (S \otimes T_2)(v) \\
&= (S \otimes T_1 + S \otimes T_2)(v)
\end{aligned}$$

3. Let $a \in \mathbb{R}$, $S \in \mathfrak{J}^k(V)$, and $T \in \mathfrak{J}^l(V)$. Then

$$\begin{aligned}
((aS) \otimes T)(v) &= (aS)(v^k) \cdot T(v^l) \\
&= a(S(v^k) \cdot T(v^l)) \\
&= a(S \otimes T(v)) \\
&= S(v^k) \cdot a(T(v^l)) \\
&= (S \otimes (aT))(v)
\end{aligned}$$

4. Let $S \in \mathfrak{J}^k(V)$, $T \in \mathfrak{J}^l(V)$, and $U \in \mathfrak{J}^m(V)$. Let $v = (v_1, \ldots, v_{k+l+m})$, and let $v^k = (v_1, \ldots, v_k)$, $v^l = (v_{k+1}, \ldots, v_{k+l})$, and $v^m = (v_{k+l+1}, \ldots, v_{k+l+m})$. Then

$$\begin{aligned}
(S \otimes (T \otimes U))(v) &= S(v^k) \cdot (T \otimes U)(v^l, v^m) \\
&= S(v^k) \cdot (T(v^l) \cdot U(v^m)) \\
&= (S(v^k) \cdot T(v^l)) \cdot U(v^m) \\
&= (S \otimes T)(v^k, v^l) \cdot U(v^m) \\
&= ((S \otimes T) \otimes U)(v)
\end{aligned}$$

□

Since the tensor product is associative, we will drop the parentheses in general. Note that we already know how to describe $\mathfrak{J}^1(V)$: since it is the set of all linear maps from $V \to \mathbb{R}$, it is precisely the dual space V^*. We can use this to help us understand higher order $\mathfrak{J}^k(V)$.

Theorem 4.2

Let $v_1, \ldots, v_n$ be a basis for V. Let $\varphi_1, \ldots, \varphi_n$ be the natural dual basis given by $\varphi_i(v_j) = \delta_{ij}$. Then the set of k-tensors of the form

$$\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_k}$$

where $1 \leq i_j \leq n$ for each index, is a basis of $\mathfrak{J}^k(V)$.

Proof. We first show that this collection spans $\mathfrak{J}^k(V)$.

Let $T \in \mathfrak{J}^k(V)$. Suppose that $w_1, \ldots, w_k \in V$, and $w_i = \sum_{j=1}^n a_{i,j} v_j$. Then

$$\begin{aligned}
T(w_1, \ldots, w_k) &= T(\sum_{j_1=1}^n a_{1,j_1} v_{j_1}, w_2, \ldots, w_k) \\
&= \sum_{j_1=1}^n a_{1,j_1} T(v_{j_1}, w_2, \ldots, w_k) \\
&\vdots \\
&= \sum_{j_1=1}^n \ldots \sum_{j_k=1}^n a_{1,j_1} \cdot \ldots \cdot a_{k,j_k} T(v_{j_1}, \ldots, v_{j_k}) \qquad (*)
\end{aligned}$$

Now, we have

$$\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_k}(v_{j_1}, \ldots, v_{j_k}) = \delta_{i_1,j_1} \cdot \ldots \cdot \delta_{i_k,j_k} = \begin{cases} 1, & i_1 = j_1, \ldots, i_k = j_k \\ 0 \end{cases} \qquad (**)$$

Since $\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_k} \in \mathfrak{J}^k(V)$, we can use $(*)$ and $(**)$ to conclude that

$$\begin{aligned}
\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_k}(w_1, \ldots, w_k) &= \sum_{j_1=1}^{n} \ldots \sum_{j_k=1}^{n} a_{1,j_1} \cdot \ldots \cdot a_{k,j_k} (\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_k})(v_{j_1}, \ldots, v_{j_k}) \\
&= \sum_{j_1=1}^{n} \ldots \sum_{j_{k-1}=1}^{n} a_{1,j_1} \cdot \ldots \cdot a_{k-1,j_{k-1}} a_{k,i_k} (\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_k})(v_{j_1}, \ldots, v_{i_k}) \\
&= \vdots \\
&= a_{1,i_1} \cdot \ldots \cdot a_{k,i_k}
\end{aligned}$$

Substituting into $(*)$, we have

$$T(w_1, \ldots, w_k) = \sum_{j_1=1}^{n} \ldots \sum_{j_k=1}^{n} T(v_{j_1}, \ldots, v_{j_k})(\varphi_{j_1} \otimes \ldots \otimes \varphi_{j_k})(w_1, \ldots, w_k)$$

So

$$T = \sum_{j_1=1}^{n} \ldots \sum_{j_k=1}^{n} T(v_{j_1}, \ldots, v_{j_k})(\varphi_{j_1} \otimes \ldots \otimes \varphi_{j_k})$$

so T is a linear combination of the $\varphi_{j_1} \otimes \ldots \otimes \varphi_{j_k}$.

To show that the $\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_k}$ are linearly independent, suppose that

$$\sum_{i_1=1}^{n} \ldots \sum_{i_k=1}^{n} a_{i_1,\ldots,i_k} (\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_k}) = 0$$

Then plugging in some combination of basis vectors $(v_{j_1}, \ldots, v_{j_k})$, by $(**)$ we have

$$0 = \sum_{i_1=1}^{n} \ldots \sum_{i_k=1}^{n} a_{i_1,\ldots,i_k} (\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_k})(v_{j_1}, \ldots, v_{j_k}) = a_{j_1,\ldots,j_k}$$

Repeating this with each combination of basis vectors shows that the linear combination is trivial. So the $\varphi_{i_1} \otimes \ldots \otimes \varphi_{i_k}$ are linearly independent and thus a basis. $\square$

Recall that if $T : V \to W$ is a linear transformation, then its adjoint $T^* : W^* \to V^*$ is the linear operator defined such that for any $\Phi \in W^*$ it is the case that $T^*(\Phi) = \Phi \circ T$. Then we can extend this notion to arbitrary $\mathfrak{J}^k(V)$.

Definition 4.4

Let $f : V \to W$ be linear. Then define the (k-tensor) **pullback** of f to be the linear transformation $f^* : \mathfrak{J}^k(W) \to \mathfrak{J}^k(V)$ by

$$(f^*(T))(v_1, \ldots, v_k) = T(f(v_1), \ldots, f(v_k))$$

where $T \in \mathfrak{J}^k(W)$ and $v_1, \ldots, v_k \in V$.

Proposition 4.3

If $S \in \mathfrak{J}^k(V)$ and $T \in \mathfrak{J}^l(V)$, and $f : V \to W$, then

$$f^*(S \otimes T) = f^*S \otimes f^*T$$

Proof. Let $v^k = (v_1, \ldots, v_k) \in V^k$ and $v^l = (v_{k+1}, \ldots, v_{k+l}) \in V^l$. Then

$$\begin{aligned} f^*(S \otimes T)(v^k, v^l) &= (S \otimes T)(f(v_1), \ldots, f(v_k), f(v_{k+1}), \ldots, f(v_{k+l})) \\ &= S(f(v_1), \ldots, f(v_k)) \cdot T(f(v_{k+1}), \ldots, f(v_{k+l})) \\ &= f^*S(v^k) \cdot f^*T(v^l) \\ &= (f^*S \otimes f^*T)(v^k, v^l) \end{aligned}$$

□

An example of a k-tensor which is not a linear functional is the dot product on $\mathbb{R}^n$, which is a 2-tensor. We can use this language to make an equivalent definition for arbitrary real inner products.

Definition 4.5

An **inner product** on a real vector space V is a 2-tensor $T \in \mathfrak{J}^2(V)$ which satisfies the following:

- $T(v, w) = T(w, v)$ (symmetric)
- $T(v, v) > 0$ if $v \neq 0$ (positive definite)

We can similarly reproduce some theorems from linear algebra.

Definition 4.6

A basis $v_1, \ldots, v_n$ for a real vector space V is **orthonormal** with respect to an inner product $T \in \mathfrak{J}^2(V)$ if $T(v_i, v_j) = \delta_{ij}$.

Theorem 4.4

For any inner product T on V, there is an orthonormal basis with respect to T.

Proof. Pick a basis and apply Gram-Schmidt. □

Corollary 4.5

If T is an inner product on V, then there exists an isomorphism $f : \mathbb{R}^n \to V$ such that $T(f(x), f(y)) = x \cdot y$, or equivalently so that f^*T is the dot product on $\mathbb{R}^n$.

Proof. Let $v_1, \ldots, v_k$ be an orthonormal basis for T. Define f by $f(e_i) = v_i$. Then if $x = (a_1, \ldots, a_n)$ and $y = (b_1, \ldots, b_n)$, we have

$$\begin{aligned} T(f(x), f(y)) &= T(\sum_{i=1}^{n} a_i v_i, \sum_{j=1}^{n} b_j v_j) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j T(v_i, v_j) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j \delta_{ij} \\ &= \sum_{i=1}^{n} a_i b_i \\ &= x \cdot y \end{aligned}$$

□

Suppose we consider a square $n \times n$ matrix as a vector whose entries are column vectors. That is, we will associate $M_{n\times n}(\mathbb{R})$ with $(\mathbb{R}^n)^n$. Then $\det : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ may be considered as a k-tensor for $\mathbb{R}^n$. Recall that one definition of the determinant defines it as the unique alternating multilinear map with $\det I = 1$. Let us attempt to generalize this notion.

Definition 4.7

A k-tensor $T \in \mathfrak{J}^k(V)$ is **alternating** if, for every pair $i < j$, we have

$$T(v_1, \ldots, v_k) = -T(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_{j-1}, v_i, v_{j+1}, \ldots, v_k)$$

In other words, switching the role of two entries also switches the sign of T.

Definition 4.8

The set of all alternating k-tensors on V is denoted $A^k(V)$.[a]

[a] Spivak uses the notation $\Lambda^k(V)$, but writes in the Addenda that $\Omega^k(V)$ should be used instead. This definition, if V is finite dimensional, is naturally isomorphic to $\bigwedge(V^*)$. See here for why neither of these are quite accurate.

One can quickly verify that $A^k(V)$ is a subspace of $\mathfrak{J}^k(V)$. The close relationship of alternating tensors with signed quantities will help us to define oriented objects. Due to this, it is of interest to us to investigate how to consistently represent elements of $A^k(V)$.

Recall that the **sign** of a permutation σ, denoted $\operatorname{sgn}\sigma$, is $+1$ if σ is even (that is, it is composed of an even number of transpositions), and -1 if it is odd.

Definition 4.9

Let $T \in \mathfrak{J}^k(V)$. Then $\operatorname{Alt}(T) \in \mathfrak{J}^k(V)$ is defined by

$$\operatorname{Alt}(T)(v_1, \ldots, v_k) = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn} \sigma \cdot T(v_{\sigma(1)}, \ldots, v_{\sigma(k)})$$

We can see Alt as a kind of projection from $\mathfrak{J}^k(V)$ into $A^k(V)$:

Theorem 4.6

Let V be a real vector space.

1. If $T \in \mathfrak{J}^k(V)$, then $\operatorname{Alt}(T) \in A^k(V)$.
2. If $\omega \in A^k(V)$, then $\operatorname{Alt}(\omega) = \omega$.
3. If $T \in \mathfrak{J}^k(V)$, then $\operatorname{Alt}(\operatorname{Alt}(T)) = \operatorname{Alt}(T)$.

Proof. 1. Fix i, j, and let (i, j) be the transposition of i and j. For each $\sigma \in S_k$, write $\sigma' = \sigma \cdot (i, j)$. We have $S_k(i, j) = S_k$. So

$$\begin{aligned}
\operatorname{Alt}(T)(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k) &= \operatorname{Alt}(T)(v_{(i,j)(1)}, \ldots, v_{(i,j)(k)}) \\
&= \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn} \sigma \cdot T(v_{\sigma((i,j)(1))}, \ldots, v_{\sigma((i,j)(k))}) \\
&= \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn} \sigma \cdot T(v_{\sigma'(1)}, \ldots, v_{\sigma'(k)}) \\
&= \frac{1}{k!} \sum_{\sigma' \in S_k \cdot (i,j)} -\operatorname{sgn} \sigma' \cdot T(v_{\sigma'(1)}, \ldots, v_{\sigma'(k)}) \\
&= -\frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn} \sigma \cdot T(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) \\
&= -\operatorname{Alt}(T)(v_1, \ldots, v_k)
\end{aligned}$$

2. Let ω be alternating. For a transposition $\sigma = (i, j)$, we have

$$\omega(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) = -\omega(v_1, \ldots, v_k) = \operatorname{sgn} \sigma \cdot \omega(v_1, \ldots, v_k) \qquad (*)$$

For arbitrary permutations σ, σ can be decomposed into a product of transpositions $\sigma_1, \ldots, \sigma_m$. Since $\operatorname{sgn}(\sigma_1 \circ \ldots \circ \sigma_m) = \operatorname{sgn} \sigma_1 \cdot \ldots \cdot \operatorname{sgn} \sigma_m$, we simply apply $(*)$ m times to see that $(*)$ holds when σ is arbitrary. Now,

$$\begin{aligned}\operatorname{Alt}(\omega)(v_1,\ldots,v_k) &= \frac{1}{k!}\sum_{\sigma\in S_k}\omega(v_{\sigma(1)},\ldots,v_{\sigma(k)})\\ &= \frac{1}{k!}\sum_{\sigma\in S_k}\operatorname{sgn}\sigma\cdot\operatorname{sgn}\sigma\cdot\omega(v_1,\ldots,v_k)\\ &= \frac{1}{k!}\sum_{\sigma\in S_k}\omega(v_1,\ldots,v_k)\\ &= \omega(v_1,\ldots,v_k)\end{aligned}$$

3. Follows from 1) and 2). □

One way of describing alternating tensors would be to produce a basis of $A^k(V)$. Note that we cannot necessarily apply Theorem 4.2. This is because if $\omega \in A^k(V)$ and $\eta \in A^k(V)$, it is not necessarily the case that $\omega \otimes \eta$ is alternating (consider a transposition which swaps entries in the ω and η domains). Thus, we will need to define an analogous product which takes alternating tensors to alternating tensors.

Definition 4.10

Let $\omega \in A^k(V)$ and $\eta \in A^l(V)$. Then the **wedge product** of ω and η, denoted $\omega \wedge \eta \in A^k(V)$, is defined by

$$\omega \wedge \eta = \frac{(k+l)!}{k!l!}\operatorname{Alt}(\omega \otimes \eta)$$

We can prove properties of $\wedge$ using similar methods as we did for $\otimes$.

Proposition 4.7

Let $\omega, \omega_1, \omega_2 \in A^k(V)$, $\eta, \eta_1, \eta_2 \in A^l(V)$, and $a \in \mathbb{R}$. Then

1. $(\omega_1 + \omega_2) \wedge \eta = \omega_1 \wedge \eta + \omega_2 \wedge \eta$
2. $\omega \wedge (\eta_1 + \eta_2) = \omega \wedge \eta_1 + \omega \wedge \eta_2$
3. $(a\omega) \wedge \eta = \omega \wedge (a\eta) = a(\omega \wedge \eta)$

Proof. 1. Let $v^k = (v_1, \ldots, v_k) \in V^k$ and $v^l = (v_{k+1}, \ldots, v_{k+l}) \in V^l$. Write $\sigma(v^k, v^l) =$

$(v_{\sigma(1)}, \ldots, v_{\sigma(k+l)})$. Then

$$\begin{aligned}
[(\omega_1 + \omega_2) \wedge \eta](v^k, v^l) &= \frac{(k+l)!}{k!l!} \operatorname{Alt}((\omega_1 + \omega_2) \otimes \eta)(v^k, v^l) \\
&= \frac{(k+l)!}{k!l!} \operatorname{Alt}(\omega_1 \otimes \eta + \omega_2 \otimes \eta)(v^k, v^l) \\
&= \frac{(k+l)!}{k!l!} \cdot \frac{1}{(k+l)!} \sum_{\sigma \in S_k} \operatorname{sgn} \sigma \cdot (\omega_1 \otimes \eta + \omega_2 \otimes \eta)(\sigma(v^k, v^l)) \\
&= \frac{(k+l)!}{k!l!} \cdot \frac{1}{(k+l)!} \sum_{\sigma \in S_k} \operatorname{sgn} \sigma \cdot (\omega_1 \otimes \eta(\sigma(v^k, v^l)) + \omega_2 \otimes \eta(\sigma(v^k, v^l))) \\
&= \frac{(k+l)!}{k!l!} (\operatorname{Alt}(\omega_1 \otimes \eta) + \operatorname{Alt}(\omega_2 \otimes \eta)) \\
&= \omega_1 \wedge \eta + \omega_2 \wedge \eta
\end{aligned}$$

2. Analogous.

3. We prove the first and third expressions are equal. The other equality is proved analogously. Then

$$\begin{aligned}
((a\omega) \wedge \eta)(v^k, v^l) &= \frac{(k+l)!}{k!l!} \operatorname{Alt}((a\omega) \otimes \eta)(v^k, v^l) \\
&= \frac{(k+l)!}{k!l!} \operatorname{Alt}(a(\omega \otimes \eta))(v^k, v^l) \\
&= \frac{(k+l)!}{k!l!} \cdot \frac{1}{(k+l)!} \sum_{\sigma \in S_k} \operatorname{sgn} \sigma \cdot a(\omega \otimes \eta)(\sigma(v^k, v^l)) \\
&= a\frac{(k+l)!}{k!l!} \cdot \frac{1}{(k+l)!} \sum_{\sigma \in S_k} \operatorname{sgn} \sigma \cdot (\omega \otimes \eta)(\sigma(v^k, v^l)) \\
&= a\frac{(k+l)!}{k!l!} \operatorname{Alt}(\omega \otimes \eta)(v^k, v^l) \\
&= a(\omega \wedge \eta)(v^k, v^l)
\end{aligned}$$

□

We can also take advantage of the alternating nature of these tensors to prove additional properties.

Proposition 4.8

Let $\omega \in A^k(V)$ and $\eta \in A^l(V)$. Let $f : V \to V$ be linear. Then

1. $\omega \wedge \eta = (-1)^{kl}(\eta \wedge \omega)$
2. $f^*(\omega \wedge \eta) = f^*(\omega) \wedge f^*(\eta)$.

Proof. 1. Let $v = (v_1, \ldots, v_{k+l}) \in V^{k+l}$. Let $\sigma^* \in S_{k+l}$ be a permutation which sends $\{1, 2, \ldots, k+l\}$ to $\{k+l, 1, 2, \ldots, k+l-1\}$. Note that this can be achieved using

$k+l-1$ permutations (as $(k+l, k+l-1) \cdot (k+l, k+l-2) \cdot \dots \cdot (k+l, 1)$). Then $(\sigma^*)^l$ is the permutation which takes $\{1, 2, \dots, k+l\}$ to $\{k+1, \dots, k+l, 1, \dots, k\}$, and

$$\operatorname{sgn}(\sigma^*)^l = (\operatorname{sgn}(\sigma^*))^l = ((-1)^{k+l-1})^l = (-1)^{kl+l^2-l} = (-1)^{kl}$$

Then we have

$$\begin{aligned}
(\omega \wedge \eta)(v) &= \frac{(k+l)!}{k!l!} \operatorname{Alt}(\omega \otimes \eta)(v) \\
&= \frac{(k+l)!}{k!l!} \frac{1}{(k+l)!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn} \sigma \cdot (\omega \otimes \eta)(v_{\sigma(1)}, \dots, v_{\sigma(k)}, v_{\sigma(k+1)}, \dots, v_{\sigma(k+l)}) \\
&= (-1)^{kl} \frac{(k+l)!}{k!l!} \frac{1}{(k+l)!} \sum_{\sigma \cdot (\sigma^*)^l} \operatorname{sgn}(\sigma) \cdot (\omega \otimes \eta)(v_{\sigma(k+1)}, \dots, v_{\sigma(k+l)}, v_{\sigma(1)}, \dots, v_{\sigma(k)}) \\
&= (-1)^{kl} \frac{(k+l)!}{k!l!} \frac{1}{(k+l)!} \sum_{\sigma} \operatorname{sgn}(\sigma) \cdot (\eta \otimes \omega)(v_{\sigma(1)}, \dots, v_{\sigma(k)}, v_{\sigma(k+1)}, \dots, v_{\sigma(k+l)}) \\
&= (-1)^{kl} \frac{(k+l)!}{k!l!} \operatorname{Alt}(\eta \otimes \omega)(v) \\
&= (-1)^{kl} (\eta \wedge \omega)(v)
\end{aligned}$$

2. This follows from Proposition 4.3:

$$\begin{aligned}
f^*(\omega \wedge \eta) &= \frac{(k+l)!}{k!l!} \operatorname{Alt}(f^*(\omega \otimes \eta)) \\
&= \frac{(k+l)!}{k!l!} \operatorname{Alt}(f^*\omega \wedge f^*\eta) \\
&= f^*\omega \wedge f^*\eta
\end{aligned}$$

□

We can also prove associativity of the wedge product:

Theorem 4.9

1. Let $S \in \mathfrak{J}^k(V)$, $T \in \mathfrak{J}^l(V)$, and suppose $\operatorname{Alt}(S) = 0$. Then

$$\operatorname{Alt}(S \otimes T) = \operatorname{Alt}(T \otimes S) = 0$$

2. $\operatorname{Alt}(\operatorname{Alt}(\omega \otimes \eta) \otimes \theta) = \operatorname{Alt}(\omega \otimes \eta \otimes \theta) = \operatorname{Alt}(\omega \otimes \operatorname{Alt}(\eta \otimes \theta))$
3. If $\omega \in A^k(V)$, $\eta \in A^l(V)$, and $\theta \in A^m(V)$, then

$$(\omega \wedge \eta) \wedge \theta = \omega \wedge (\eta \wedge \theta) = \frac{(k+l+m)!}{k!l!m!} \operatorname{Alt}(\omega \otimes \eta \otimes \theta)$$

Proof. 1. Let $G < S_{k+l}$ be the set of all permutations which fix the k-th through $k+l$-th elements. This is a subgroup of S_{k+l}, so we may consider the set of right cosets $G\sigma'$

for $\sigma' \in S_{k+l}$. Let $v^k = (v_1, \ldots, v_k) \in V^k$, $v^l = (v_{k+1}, \ldots, v_{k+l})$, and let $\sigma(v) = (v_{\sigma(1)}, \ldots, v_{\sigma(k+l)})$. Then

$$\begin{aligned}
\operatorname{Alt}(S \otimes T)(v^k, v^l) &= \frac{1}{(k+l)!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn} \sigma \cdot (S \otimes T)(\sigma(v^k, v^l)) \\
&= \frac{1}{(k+l)!} \sum_{G\sigma'} \sum_{\sigma \in G} \operatorname{sgn}(\sigma\sigma') \cdot S(\sigma\sigma'(v^k)) \cdot T(\sigma\sigma'(v^l)) \\
&= \frac{1}{(k+l)!} \sum_{G\sigma'} \operatorname{sgn}(\sigma') T(\sigma'(v^l)) \sum_{\sigma \in G} \operatorname{sgn}(\sigma) \cdot S(\sigma\sigma'(v^k))
\end{aligned}$$

Write $\sigma'(v^k) = w^k$. Noting that $G \cong S_k$, we have

$$\sum_{\sigma \in G} \operatorname{sgn}(\sigma) \cdot S(\sigma\sigma'(v^k)) = \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) \cdot S(\sigma(w^k)) = k! \operatorname{Alt}(S)(w^k) = 0$$

So $\operatorname{Alt}(S \otimes T) = 0$ and similarly $\operatorname{Alt}(T \otimes S) = 0$.

2. Noting that Alt is linear, we know that

$$\operatorname{Alt}(\operatorname{Alt}(\omega \otimes \eta) - \omega \otimes \eta) = \operatorname{Alt}(\operatorname{Alt}(\omega \otimes \eta)) - \operatorname{Alt}(\omega \otimes \eta) = \operatorname{Alt}(\omega \otimes \eta) - \operatorname{Alt}(\omega \otimes \eta) = 0$$

Applying part 1),

$$0 = \operatorname{Alt}((\operatorname{Alt}(\omega \otimes \eta) - \omega \otimes \eta) \otimes \theta) = \operatorname{Alt}(\operatorname{Alt}(\omega \otimes \eta) \otimes \theta) - \operatorname{Alt}(\omega \otimes \eta \otimes \theta)$$

so

$$(\operatorname{Alt}(\omega \otimes \eta) \otimes \theta) = \operatorname{Alt}(\omega \otimes \theta \otimes \eta)$$

and the other equality is similar.

3. We have

$$\begin{aligned}
(\omega \wedge \eta) \wedge \theta &= \frac{(k+l+m)!}{(k+l)!m!} \operatorname{Alt}((\omega \wedge \eta) \otimes \theta) \\
&= \frac{(k+l+m)!}{(k+l)!m!} \operatorname{Alt}(\frac{(k+l)!}{k!l!} \operatorname{Alt}(\omega \otimes \eta) \otimes \theta) \\
&= \frac{(k+l+m)!}{k!l!m!} \operatorname{Alt}(\operatorname{Alt}(\omega \otimes \eta) \otimes \theta) \\
&= \frac{(k+l+m)!}{k!l!m!} \operatorname{Alt}(\omega \otimes \eta \otimes \theta)
\end{aligned}$$

□

As a result, we will also drop the parentheses when discussing wedge products.

Theorem 4.10

Let V be a real vector space with basis $v_1, \ldots, v_n$, and let $\varphi_1, \ldots, \varphi_n$ be the induced dual basis. Then the collection of all k-tensors of the form

$$\varphi_{i_1} \wedge \ldots \wedge \varphi_{i_k}$$

where $1 \leq i_1 < \ldots < i_k \leq n$, is a basis for $A^k(V)$.

Proof. Let $\omega \in A^k(V)$. Then $\omega \in \mathfrak{J}^k(V)$. By Theorem 4.2, the collection of $\varphi_{j_1} \otimes \ldots \otimes \varphi_{j_k}$ is a basis for $\mathfrak{J}^k(V)$ and we have

$$\omega = \sum a_{j_1,\ldots,j_k} \varphi_{j_1} \otimes \ldots \otimes \varphi_{j_k}$$

Since ω is alternating, we have

$$\omega = \mathrm{Alt}(\omega) = \sum a_{j_1,\ldots,j_k} \mathrm{Alt}(\varphi_{j_1} \otimes \ldots \otimes \varphi_{j_k}) = \sum a_{j_1,\ldots,j_k} \frac{(nk)!}{(k!)^n} \varphi_{j_1} \wedge \ldots \wedge \varphi_{j_k}$$

Let $j'_1, \ldots, j'_k$ be a reordering of $j_1, \ldots, j_k$ such that $j'_1 \leq \ldots \leq j'_k$. This may be accomplished by a series of transpositions, each of which changes only the sign of the wedge product by Proposition 4.8. Moreover, if any two of the $\varphi_{j'_1}$ are equal, then the entire wedge product is zero. So we may assume $j'_1 < \ldots < j'_k$, and we have

$$\sum a_{j_1,\ldots,j_k} \frac{(nk)!}{(k!)^n} (-1)^{M_{j_1,\ldots,j_k}} \varphi_{j'_1} \wedge \ldots \wedge \varphi_{j'_k}$$

So the $\varphi_{j'_1} \wedge \ldots \wedge \varphi_{j'_k}$ span $A^k(V)$.

To show linear independence, suppose we have some linear combination

$$\omega = \sum a_{j_1,\ldots,j_k} \varphi_{j_1} \wedge \ldots \wedge \varphi_{j_k}$$

$\square$

Definitions