MAT257: Real Analysis II

QiLin Xue

Fall 2021

Contents

1 Differentiation

1.1 Inverse Function Theorem

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

$$(f^{-1})'(y) = [f'(f^{-1}(y))]^{-1} \tag{1.1}$$

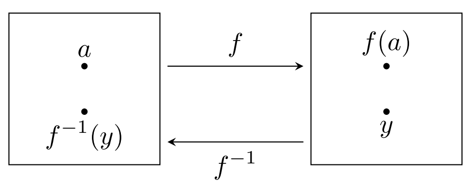
Motivation: In 1D calculus, if a function has a derivative $f'(x) > 0$ at some point $x = a$, then around this point the function is monotone and has an inverse, per the intermediate value theorem. We want something similar for multiple dimensions, however there is no equivalent intermediate value theorem.

We will motivate our proof with the following steps

1. Prove the last step, as it is the easiest.
2. WLOG, make the simplifying assumption that $f'(a) = I$.
3. Define "all-scale fidelity" to describe how vectors in an open neighbourhood around a are *nearly* preserved. Then show that f has all-scale fidelity on some neighbourhood U around a.
4. Given $y \in W$. We can construct a sequence (x_i) such that $\{f(x_i)\}$ is a Cauchy sequence which converges to y.
5. Show that there exists an $x \in V$ such that $f(x) = y$, by invoking continuity. Said differently, $f|_V : V \to W$ is onto.
6. Show that $f|_V : V \to W$ is injective (1-1).
7. We have shown that f^{-1} exists. We now need to show that f^{-1} is continuous.
8. Show that f^{-1} is differentiable at the point $f(a) = b$.
9. Show that f^{-1} is differentiable at a point near b.
10. Show that $f^{-1}(y)$ is continuously differentiable near b.

Let us perform these steps:

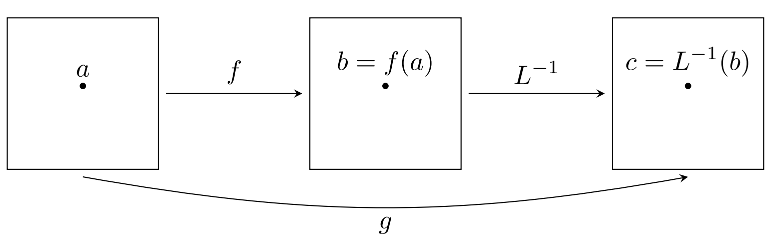
1. Consider the following setup



Recall that $f \circ f^{-1} = I$, so differentiating and applying the chain rule we can write

$$f'(f^{-1}(y)) \cdot (f^{-1})'(y) = I \implies (f^{-1})'(y) = [f'(f^{-1}(y))]^{-1} \tag{1.2}$$

2. We are allowed to write $f'(a) = I$ since every invertible matrix is just the identity with a change of basis. More concretely, consider another composition represented below, where $L = f'(a)$:



By the chain rule, we have $g'(a) = L^{-1} \circ f'(a)$ since L^{-1} is a linear transformation. Note that $f'(a) = L$, so $g'(a) = I$ is the identity.

If the IFT was true for functions whose differential is I, then it's true for g, so there exists g^{-1}. Also, $f^{-1} = g^{-1} \circ L^{-1}$. Therefore, if g^{-1} is continuously differentiable, then f^{-1} would also be continuously differentiable. Thus, it is sufficient to only look at the case where the differential is the identity.

3. Consider a small neighbourhood around a. Intuitively, we should expect vectors and the image of the vectors (which need not start at a) to look roughly the same.

   Specifically, f has **all-scale-fidelity** on some neighbourhood U of a with *fidelity factor* $\frac{1}{257}$. This means for all $x_1, x_2 \in U$, we have
   $$|(x_2 - x_1) - (f(x_2) - f(x_1))| \leq \frac{1}{257}|x_2 - x_1| \tag{1.3}$$

   Proof. From a previous theorem, if we have a function $g : \mathbb{R}^n \to \mathbb{R}^n$ whose derivative $|g'| < M$ is bounded in some open set, then we can write
   $$|g(x_1) - g(x_2)| < n^2 M|x_1 - x_2|. \tag{1.4}$$

   Now consider $g(x) = f(x) - x$. The derivative is $g' = f' - I$ so $g'(a) = 0$. Therefore, there exists an open rectangle U of a where we can say $|g'(x)| \leq \frac{1}{257n^2}$.

   By the previous theorem, for any $x_1, x_2 \in U$, we have
   $$|g(x_1) - g(x_2)| \leq \frac{1}{257}|x_1 - x_2| \tag{1.5}$$

   However, the LHS is just
   $$\begin{aligned} |g(x_1) - g(x_2)| &= |f(x_1) - x_1 - (f(x_2) - x_2)| \\ &= |(x_2 - x_1) - (f(x_2) - f(x_1))| \end{aligned}$$

   □

4. Given $y \in W$, we wish to find an $x \in V$ such that $f(x) = y$. We will travel this direction in W, but we may "miss" by a bit. We can then repeat this process, with each step we travel in the direction of y.

   Put this formally, let $W = B_{r/2}(b)$. Given $y \in W$, we claim that there exists an $x \in B_r(a)$ such that $f(x) = y$. Indeed,
   $$x_1 = a + (y - b) \tag{1.6}$$
   $$x_2 = x_1 + (y - f(x_1)) \tag{1.7}$$
   $$x_3 = x_2 + (y - f(x_2)) \tag{1.8}$$
   $$x_n = x_{n-1} + (y - f(x_{n-1})) \tag{1.9}$$

   But the difference between any two consecutive terms is just the LHS of the all-scale fidelity
   $$|x_n - x_{n-1}| = |(x_{n-1} - x_{n-2}) - (f(x_{n-1} - f(x_{n-2}))| \tag{1.10}$$
   $$\leq \frac{1}{257}|x_{n-1} - x_{n-2}| \tag{1.11}$$
   $$\leq \frac{1}{257^{n-1}}|x_1 - x_0| \tag{1.12}$$
   $$\leq \frac{1}{257^{n-1}}|y - b| \tag{1.13}$$
   $$\leq \frac{1}{257^{n-1}}\frac{r}{2} \tag{1.14}$$

   We now need to show that each x_i is within the ball of radius r around a (since this is only when all-scale fidelity is defined). It can be shown via induction that $|x_n - a| \leq r$.

   Finally, we show that (x_n) is a Cauchy-Sequence. We can immediately show this by noting that
   $$|x_n - x_m| \leq \frac{1}{257^m} r \tag{1.15}$$

   so (x_n) is cauchy.

5. While we have shown that $f(x_n)$ converges to y, we have not yet shown that this is possible, i.e. what if there is a discontinuity? We can invoke the continuity of f to show that there does exist such an x_n. We have

$$|f(x_n) - y| = |x_{n+1} - x_n| \leq \frac{1}{257^n}\frac{r}{2} \to 0 \tag{1.16}$$

so from continuity, there exists an x such that

$$|f(x) - y| = \lim_{n\to\infty} |f(x_n) - y| = 0. \tag{1.17}$$

Therefore, $f(x) = y$. We can now define $V = F^{-1}(W)$ and now

$$F|_V : V \to W \tag{1.18}$$

is onto and 1-1.

6. We have constructed x in one such way. How do we know that if we use a different procedure, we find a different x?

Assume that $f(x_1) = f(x_2)$ where $x_{1,2} \in B_r(a)$. Then by ASF,

$$|(x_1 - x_2) - (f(x_1) - f(x_2))| \leq \frac{1}{257}|x_1 - x_2| \tag{1.19}$$

$$|x_1 - x_2| \leq \frac{1}{257}|x_1 - x_2| \tag{1.20}$$

which is true if and only if $x_1 = x_2$.

7. It might seem that continuity for f^{-1} is cheap. After all, the difference between two vectors in both the input and image space is roughly the same. However, the mistake is written in terms of $|x_1 - x_2|$, so this reasoning becomes circular. We need to reformulate the ASF principle such that the mistake is in terms of $|y_1 - y_2|$.

To simplify things, let $\alpha = x_1 - x_2$ and $\beta = f(x_1) - f(x_2)$. Then ASF says that

$$|\alpha - \beta| \leq \frac{1}{257}|\alpha| \tag{1.21}$$

But $\alpha = \beta + \alpha - \beta$. By the triangle inequality, this becomes

$$|\alpha - \beta| \leq \frac{1}{257}(|\beta| + |\alpha - \beta|) \tag{1.22}$$

$$\frac{256}{257}|\alpha - \beta| \leq \frac{1}{257}|\beta \tag{1.23}$$

$$|\alpha - \beta| \leq \frac{1}{256}\beta \tag{1.24}$$

8. Let us first show that f^{-1} is differentiable at the point b. We can write

$$f^{-1}(b + h) = f^{-1}(b) + I \cdot h + e(h) \tag{1.25}$$

We want to show that $e(h)$ is tiny. We want to rearrange this in a form such that we can apply all scale fidelity. Let $b + h = y_2$ and let $x_2 = f^{-1}(y_2)$. Let $b = y_1$ and $a = x_1$, so the above just becomes

$$x_2 = x_1 + y_2 - y_1 + e(h) \tag{1.26}$$

However the error once rearranged becomes

$$|e(h)| = |(x_2 - x_1) - (y_2 - y_1)| \leq \frac{1}{256}|y_2 - y_1| \tag{1.27}$$

Since $y_2 - y_1 = h$, we end up with

$$\frac{|e(h)|}{|h|} \leq \frac{1}{256} \tag{1.28}$$

Now we have a problem: we want to show that this approaches zero. This condition is not good enough. However, this constant was chosen arbitrarily, so we can make $\dfrac{|e(h)|}{|h|}$ as small as possible.

9. Similarly, the choice of b was also arbitrary. If the conditions for the IFT hold at a, then they have to hold in an open set near a (due to continuity). Therefore, we can rewrite the entire proof by considering points near b and not just b.

10. To show f^{-1} is continuously differentiable, we can use the chain rule:

$$(f^{-1})'(y) = [f'(f^{-1}(y))]^{-1} \tag{1.29}$$

We can conclude that $f^{-1}(y)$ is continuous in y and $f'(x)$ is continuous in x. Therefore, $M \mapsto M^{-1}$ is a continuous operation on matrices. Specifically, it is a function that maps $\mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$. This is not everywhere defined. But where defined, the inverse is continuous by Cramer's Law. There is an explicit formula for the inverse, so this map is continuous.

1.2 Implicit Function Theorem

TBA

2 Integration

2.1 Rigorous Definition

2.1.1 Partitions

Let $R = \prod_{i=1}^{n} ([a_i, b_i])$ be a rectangle in $\mathbb{R}^n$ and let $f : R \to \mathbb{R}$ be bounded. We lay out some definitions:

- A **partition** P of R is a sequence $(P_i)_{i=1}^n$, where P_i is a partition of $[a_i, b_i]$ (i.e. $P = a_i = t_{i_0} \leq t_{i_1} \leq \cdots \leq t_{i_{N_i}} = b_i$)
- When a **sub-rectangle** S is relative to partition P, we use the notation $S \in P$. If we choose integers $1 \leq j_i \leq N_i$ for $i \in \{1, \dots, n\}$ and $j = (j_1, \dots, j_n)$, then it can be written as $S_j = \prod_{i=1}^{n}([t_{i,j_{i-1}}, t_{i,j_i}])$.
- The volume of a rectangle is

$$\text{vol}(R) = |R| = \prod_{i=1}^{n}(b_i - a_i) \tag{2.1}$$

2.1.2 Lower and Upper Sums

Let $f : R \to \mathbb{R}$ be founded. Then

$$L(f, P) = \text{lower sum of F rel P} = \sum_{S \in P}(\text{vol}(S) m_s(f)) \tag{2.2}$$

$$U(f, P) = \text{upper sum of F rel P} = \sum_{S \in P}(\text{vol}(S) M_s(f)) \tag{2.3}$$

where $m_s \leq M_s \implies L(f, P) \leq U(f, P)$. A function is **integrable** if $L(f, P)$ and $U(f, P)$ converge to each other.

Lemma 1: If P is a partition of R, then

$$\sum_{S \in P}(V(S)) = V(R) \tag{2.4}$$

2.1.3 Refinements and Integrability Theorems

Definition: We say that P' **refines** P if every $S' \in P'$ is a subset of some $S \in P$. Equivalently, if $S \in P \implies S = U_{S' \in P', s' \in S}(s')$.

Lemma 2: If a partition P' refines a partition P, then

$$L(f, P) \leq L(f, P') \tag{2.5}$$

$$U(f, P) \geq U(f, P') \tag{2.6}$$

A direct corollary is that if P_1, P_2 are partitions of R then $L(f, P_1) \leq U(f, P_2)$.

Definition: **Upper integral** of f is

$$\inf\{U(f, P)\} =: U(f) = \int_U f \tag{2.7}$$

and the **lower integral** of f is

$$\sup\{L(f, P)\} =: L(f) = \int_L f \tag{2.8}$$

An alternative but equivalent way to say f is **integrable** is when $L(f) = U(f)$.

Theorem: The function f is integrable if and only if for all $\epsilon > 0$, there exists P such that $U(f, P) - L(f, P) < \epsilon$.

Theorem: Continuous functions are integrable.

2.2 Measure Theory

Definition: A is of **measure-0** means that for $\forall \epsilon > 0$, there exists open (alternatively closed) rectangles $(R_i)_{i=1}^{\infty}$ such that

1. $A \subset \bigcup R_i$
2. $\sum \text{vol}(R_i) < \epsilon$.

Note that when we say measure 0, we refer to **Lebesgue measure-0.**

For example, finite & countable sets, along with $\mathbb{R} \subset \mathbb{R}^2$ are of measure 0.

Definition: A set X is called **countably infinite** if there is a surjective function F such that $F : \mathbb{N} \twoheadrightarrow X$

A few facts that follows:

1. Finite sets are countable. (Typically, we exclude this from the definition.)
2. Subsets of countable sets are countable. (Proof: Elements in a countable set can be enumerated. Simply select a new enumeration.)
3. A finite/countable union of countable sets is countable. If A_i is countable for all i, then $\bigcup A_i$ is countable.

   Proof. Consider

$$\begin{array}{lllll} A_1 : & a_{11} & a_{12} & a_{13} & \cdots \\ A_2 : & a_{21} & a_{22} & a_{23} & \cdots \\ A_3 : & a_{31} & a_{32} & a_{33} & \cdots \end{array}$$

   To enumerate the union, we look at the diagonals, i.e.

$$\bigcup A = \{a_{11}, a_{12}, a_{21}, a_{13}, a_{22}, a_{31}, \dots\} \tag{2.9}$$

   □

   Another fact that immediately follows is that the set of integers is countable, since $\mathbb{Z} = (-\mathbb{N}) \cup \{0\} \cup \mathbb{N}$

4. If A, B are countable, then $A \times B$ is countable.

   Proof. We can prove this similar to how a countable union of sets is countable. Alternatively, we can write

$$A \times B = \bigcup_{b \in B} A \times \{b\} \tag{2.10}$$

   And each set $A \times \{b\}$ is coutnable. □

   A fact that follows is that the set of rationals is countable, since $\mathbb{Q} \subset \mathbb{Z} \times \mathbb{Z}$.

It may seem a bit suspicious since up to this point, everything is countable. However, there are sets that are uncountable!

Theorem: $\mathbb{R}$ is not countable. And hence, irrational numbers are not countable, i.e. there are "more" reals than naturals, more irrationals than rationals.

Proof. Assume that $\mathbb{R}$ is countable, that is (a_i) is an enumeration of the real numbers. Let x be a real number whose k^{th} decimal digit is different from the k^{th} decimal digit of a_k. Note that x cannot be any of the a_ks, hence $\{a_k\} \neq \mathbb{R}$. □

We can now state and prove a few statements about measure-0.

1. If A is measure-0 and $B \subset A$, then B is measure-0.

2. A countable union of measure-0 sets is measure-0.

   Proof. Suppose $\forall i$, A_i is of measure 0, so given $\epsilon > 0$, we can cover A_i with countably many rectangles whose $\sum \text{vols} < \frac{\epsilon}{2^i}$. Take all the rectangles above together, as a countable collection of countable sets, this collection of rectangles is countable, and

   $$\sum \text{vols} < \sum_{i=1}^{\infty} \frac{\epsilon}{2^i} = \epsilon. \tag{2.11}$$

   Finally this collection covers

   $$\bigcup A_i = A \tag{2.12}$$

   so A is measure-0. □

   Note that we have stated before that $\mathbb{R} \subset \mathbb{R}^2$

Warning: Countable sets are measure-0, but the converse is not true! For example, $\mathbb{R} \subset \mathbb{R}^2$ has measure-0, but is not countable.

Another Example: Let $\mathcal{C}$ be the **cantor set** $\subset [0, 1]$ which is uncountable, yet of measure-0 in $\mathbb{R}$. Let

$$\mathcal{C} = \{0.C_1C_2C_3C_4 \cdots : C_i \in \{0, 2\}\} \tag{2.13}$$

C is uncountable for the same reason as $\mathbb{R}$. Define C_k to be the union of 2^k intervals of length $\frac{1}{3^k}$. Therefore,

$$C \subset C_k \tag{2.14}$$

where C_k itself is a union of intervals of total length $2^k \cdot \frac{1}{3^k} = \left(\frac{2}{3}\right)^k$, which approaches 0.

Therefore, C is measure-0. As an aside, each C_k is compact, so $\mathcal{C} = \cap C_K$ is therefore also compact.

Theorem: $[a, b] \subset \mathbb{R}$ is *not* measure 0. In fact, $R \subset \mathbb{R}^n$ is not measure 0.

Definition: $A \subset \mathbb{R}^n$ is said to be of **content-0** if $\forall \epsilon > 0$ it is contained in a finite union of rectangles whose sum of volumes is smaller than ϵ.

Note that $\mathbb{Z} \in \mathbb{R}$ is not of content-0. Note that content-0 refers to **Jordan Measure 0.**

A set is **Jordan Measurable** (alternatively **rectifiable**) as follows:

Definition: S is Jordan measurable if and only if S is bounded and the boundary has measure 0.

Definition: S is Jordan measurable if and only if the identity function is integrable over S.

We can relate integrability and measure theory.

Theorem: A function f is integrable if and only if it is continuous except on a set of measure 0.

3 Fubini's Theorem

Currently, we have the tools to define the integral, but we don't have the tools to compute the integral yet. We will start off with a loose example.

Example 1: Integrate xy on $[0,1]_x \times [0,1]_y$. Fubini's theorem loosely tells us that we can fix x, and then fix y. Namely,

$$\int\limits_{[0,1]\times[0,1]} xy \,\mathrm{d}x\,\mathrm{d}y = \int_0^1 x \int_0^1 y \,\mathrm{d}y\,\mathrm{d}x = \frac{1}{4} \tag{3.1}$$

We will now formally state it.

Theorem: **(Tempting but Incorrect)** Let $A \subset \mathbb{R}^n_x$ and $B \subset \mathbb{R}^m_y$ be rectangles, set $R = A \times B \subset \mathbb{R}^{n+m}$. Let

$$F : R \to \mathbb{R} \tag{3.2}$$

be an integrable function. Let

$$g(x) = \int\limits_B f(x,y)\,\mathrm{d}y\,. \tag{3.3}$$

Then

$$\int\limits_R F = \int\limits_A g\,\mathrm{d}x \tag{3.4}$$

Note that this is incorrect for general functions f, but is true if f is continuous.

Warning: Note that we cannot define $g(x) = \int\limits_B f(x,y)\,\mathrm{d}y$ and $\int\limits_A f = \int\limits_A g\,\mathrm{d}x$ since while f is integrable over $A \times B$, it is not necessarily integrable over B. For example, suppose we have a function defined as

$$f(x,y) = \begin{cases} 0 & x < 0.5 \\ 1 & x > 0.5 \\ 1_{\mathbb{Q}} \end{cases} \tag{3.5}$$

where $1_{\mathbb{Q}}$ is the Dirichlet function, defined to be 1 if rational and 0 otherwise. f will be integrable in the region $[0,1] \times [0,1]$ but is not integrable if we restrict it to the line $x = 0.5$. This is because the set of discontinuities is of measure 0 in $\mathbb{R}^2$ but is of measure 1 in $\mathbb{R}$.

While it may be tempting to write the theorem as

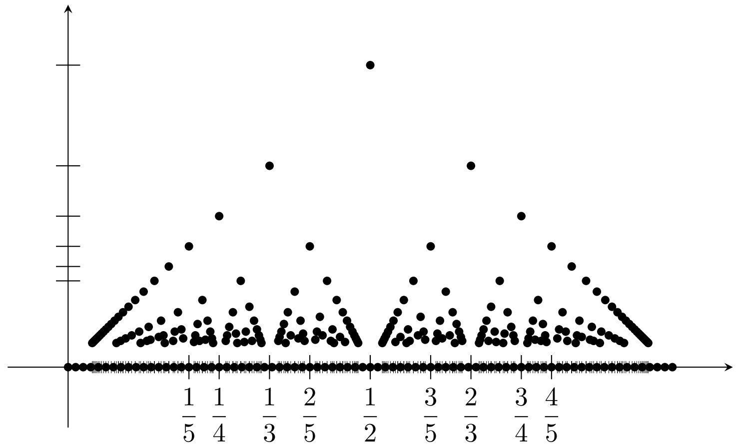
$$g(x) = \begin{cases} \int\limits_B f(x,y)\,\mathrm{d}y & \text{f(x,-) is integrable} \\ 17 & \text{otherwise.} \end{cases} \tag{3.6}$$

and define

$$\int\limits_R f = \int\limits_A g \tag{3.7}$$

which might solve the problem of discontinuities. However, this is still wrong. Here is a counter example.

Consider $h(x) = \begin{cases} \frac{1}{q} & x = \frac{p}{q} \\ 0 & \text{otherwise} \end{cases}$ defined on $[0,1]$, known as Thomae's function, which looks like the below



Note that $h(x)$ is discontinuous on $\mathbb{Q}$ but continuous on $\mathbb{Q}^C$. Since $\mathbb{Q}$ is of measure 0, h is integrable. The integral is

$$\int_0^1 h(x) = 0 \tag{3.8}$$

and we can prove this by cropping the function about some arbitrary $y = c$. Since there are only a finite amount of points above this line, we can "chop" them off. Now we extend this to two variables. Consider

$$f(x,y) = \begin{cases} \frac{1}{q} & x, y \in \mathbb{Q}, x = \frac{p}{q} \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

defined on the set $[0,1] \times [0,1]$. If we try to compute g using the second incorrect attempt, we get

$$g(x) = \begin{cases} 0 & x \notin \mathbb{Q} \\ 17 & x \notin \mathbb{Q} \end{cases} \tag{3.10}$$

but this is a "bed of nails" function with respect to x, and is not integrable, and thus we cannot expect an equality.

Note that the choice for $g(x)$ might sound stupid. After all, we can choose 0 instead of 17, to remove the problematic values. However, we can just shift $f(x,y)$ to create another counterexample.

Let us now state the correct theorem,

Theorem: (Correct Fubini's Theorem) Let $A \subset \mathbb{R}^n_x$ and $B \subset \mathbb{R}^n_y$ be rectangles, set $R = A \times B \subset \mathbb{R}^{n+m}$. Let $f : R \to \mathbb{R}$ be an integrable function and let

$$\underline{g}(x) = \int_L f(x,y)\,\mathrm{d}y = L(f(x,-)) = \sup \text{ lower sums for } f(x,-) \tag{3.11}$$

$$\bar{g}(x) = \int_U f(x,y)\,\mathrm{d}y = U(f(x,-)) = \inf \text{ upper sums} \tag{3.12}$$

Then $\underline{g}$ and $\bar{g}$ are integrable and

$$\int_R f\,\mathrm{d}x\,\mathrm{d}y = \int_A \underline{g}\,\mathrm{d}x = \int_A \bar{g}\,\mathrm{d}x \tag{3.13}$$

Let us go back to our previous counterexample. Now,

$$\underline{g}(x) = \begin{cases} 0 & x \notin \mathbb{Q} \\ 0 & x \in \mathbb{Q} \end{cases} \tag{3.14}$$

so $\underline{g}(x) = 0$ and $\displaystyle\int \underline{g} = \int 0 = 0$. On the other hand,

$$\bar{g}(x) = \begin{cases} 0 & x \notin Q \\ \frac{1}{q} & x \in Q \end{cases} \tag{3.15}$$

which is just $h(x)$ from earlier, which we have computed the integral already to be 0. Now that we have worked through examples, but we have not yet proved the theorem yet.

Before we do so, note that if f is continuous, all that is a non-issue

$$g(x) = \int_B f(x,y)\,\mathrm{d}y = \bar{g}(x) = \underline{g}(x) \tag{3.16}$$

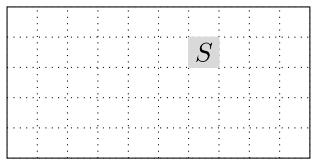
for all x, so the naive Fubini's Theorem holds. Likewise, if $f(x,-)$ is integrable except for finitely many x's, then the second attempt we made holds.

Proof. We sketch out the proof as follows:

1. Bound $L(f,P) \le L(\underline{g}, P_A)$ and $U(f,P) \le U(\bar{g}, P_A)$.
2. Show that $\bar{g}$ and $\underline{g}$ are integrable.

We will now carry out these steps:

1. Consider a partition P of $R \subset \mathbb{R}^{n+m}$, which is illustrated below. We can restrict our attention to the first n and the last m coordinates. We can always write it as $P_A \times P_B$, where P_A is a partition of A and P_B is a partition of B.



If $S \in P$, we can write $S = S' \times S''$ where $S' \in P_A$ and $S'' \in P_B$.

Given this partition $P = P_A \times P_B$ of R, we have

$$\begin{aligned} L(f,P) &= \sum_{S \in P} V(S) \cdot \inf_{(x,y)\in S} f(x,y) && (3.17)\\ &= \sum_{\substack{S' \in P_A \\ S'' \in P_B}} V(S')V(S'') \cdot \inf_{x \in S'} \inf_{y \in S''} f(x,y) && (3.18)\\ &= \sum_{S' \in P_A} V(S') \sum_{S'' \in P_B} V(S'') \inf_{x \in S'} \inf_{y \in S''} f(x,y) && (3.19)\end{aligned}$$

We are aiming to pull the infimums outside. We can do this with the following lemma.

Lemma 3: Let $h_k : X \to \mathbb{R}_1$. Then

$$\sum_k \inf h_k(x) \le \inf \sum_k h_k(x) \tag{3.20}$$

Proof. Note that $\inf_x h_k(x) \le h_k(y)$ for all y given any k. This means that we can sum both sides

$$\sum_k \inf h_k(x) \le \sum_k h_k(y) \tag{3.21}$$

The left hand side is a constant but the right hand side is a function of y. Since this inequality is true for all y, we can just pick the y to minimize the right hand side:

$$\sum_k \inf_x h_k(x) \le \inf_y \sum_y h_k(y) \tag{3.22}$$

which is just the lemma with different variable names. $\square$

Using this lemma, we can bound $L(f,P)$ by

$$L(f,P) \leq \sum_{S' \in P_A} V(S') \inf_{x \in S'} \underbrace{\sum_{S'' \in P_B} V(S'') \inf_{y \in S''} f(x,y)}_{L(f(x,-),P_B)} \tag{3.23}$$

$$\leq \sum_{S' \in P_A} V(S') \inf_{x \in S'} \underline{g}(x) \tag{3.24}$$

$$= L(\underline{g}, P_A) \tag{3.25}$$

Similarly, we can do the same thing with supremums to get

$$L(f,P) \leq L(\underline{g}, P_A) \qquad U(\bar{g}, P_A) \leq U(f,P) \tag{3.26}$$

2. Note that both $L(\bar{g}, P_A)$ and $U(\underline{g}, P_A)$ are bounded by both $L(\underline{g}, P_A)$ (on the lower end) and $U(\bar{g}, P_A)$ (on the upper end). Let us restrict our attention to

$$L(\underline{g}, P_A) \leq U(\underline{g}, P_A). \tag{3.27}$$

This means that $\underline{g}$ is integrable. Similarly, $\bar{g}$ is integrable.

Now assume $\epsilon > 0$ and P was chosen such that $U(f,P) - L(f,P) < \epsilon$ (which is possible since f is integrable), then

$$U(\underline{g}, P_A) - L(\underline{g}, P_A) \leq \epsilon \tag{3.28}$$

and

$$U(\bar{g}, P_A) - L(\bar{g}, P_A) \leq \epsilon \tag{3.29}$$

so $\underline{g}$ and $\bar{g}$ are integrable on A.

3. From the inequalities earlier, we can write

$$L(f,P) \leq \int_A \bar{g}, \int_A \underline{g} \leq U(f,P) \tag{3.30}$$

which is true for every P. This means that we can take the infimum and supremum over all partitions, to get

$$\int f \leq \int \bar{g}, \int \underline{g} \leq \int f \tag{3.31}$$

so $\int f = \int \bar{g} = \int \underline{g}$ and we are done.

$\square$

It turns out that we can ignore the upper and lower bound conditions if we have that $f_x(y)$ is integrable for all x, where $f_x(y)$ is f restricted to a given x.

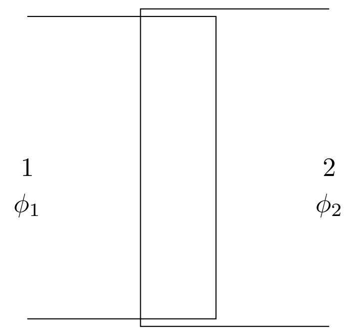
4 Partitions of Unity

4.1 Motivation

We will answer the question of how we can integrate over a non-compact set. Similar issues appear in other areas: We know how things work on a small scale. How do we make it work on a large scale?

For example, we could say that to integrate over an unbounded set, we can partition the set by an infinite amount of intersecting rectangles, integrate over the rectangles, and then sum the integrals. However, this is not so easy to do.

Consider the example of two players trying to divide the work over two regions.



where ϕ_i represents the amount of work each player does. In each non-intersecting section, we want $\phi_i = 1$ and in the shared section, we want to break up the work in some continuous fashion (i.e. close to player 1, player 1 does more work.)

Let us look at the general case where we want to divide up the work $W = \bigcup U_i$ across n players. We can define ϕ_i on three conditions:

1. The function ϕ_i is defined
$$\phi_i : W \to [0,1] \tag{4.1}$$
and we can let it be C^∞. It turns out this condition can be weaker, but this stronger condition works out fine at the end. We also have the condition for all x,
$$\sum \phi_i(x) = 1 \tag{4.2}$$

2. We also want the support of the function to be a subset of U_i, that is $\operatorname{supp} \phi_i \subset U_i$, where the support of a function is defined as follows:
$$\operatorname{supp} \phi_i = \operatorname{cl}\{x : \phi_i(x) \neq 0\} \tag{4.3}$$

3. We also want *local finiteness*, where ever $x \in W$ has a neighbourhood $V \ni x$ such that
$$|\{i : V \cap \operatorname{supp} \phi_i\}| < \infty. \tag{4.4}$$

This is essentially what partitions of unity is about. We want to partition unity via functions $\phi_i(x)$. Our goal at the end is to show that if we have a function f defined on a set $W \subset \bigcup U_i$.

$$\int_W f := \sum_i \int_{U_i} \phi_i f \tag{4.5}$$

4.2 Theorem

We propose the following theorem, which states that all of the above can be achieved:

Theorem: Given $A \subset \mathbb{R}^n$ and given an open cover of A defined as $\mathcal{U} = \{\mathcal{U}_\alpha\}$ where U_α is open and $\bigcup_\alpha U_\alpha \supset A$, we can find a countable collection
$$\Phi = \{\phi_i : W \to [0,1]\} \tag{4.6}$$
of C^∞ functions defined on some open set $W \supset A$ (where W is defined to be slightly bigger than A) such that three conditions hold:

1. **Local Finiteness:** Every $x \in W$ has an open neighbourhood $V \ni x$ such that
$$|\{i : V \cap \operatorname{supp} \phi_i \neq \emptyset\}| < \infty. \tag{4.7}$$
2. **Sum is Unity:** For all $x \in A$, we have
$$\sum_{i=1}^{\infty} \phi_i(x) = 1 \tag{4.8}$$
Note that by the first condition, there are only a finite number of ϕ_i that is nonzero for each i, i.e. it is a "nearly finite sum."
3. **Subordinate:** We have Φ is subordinate to $\mathcal{U}$, i.e. for all i, there exists an (not necessarily unique) α such that
$$\operatorname{supp} \phi_i \subset U_\alpha \tag{4.9}$$
Note that this third condition is slightly different from the one in the motivation. This is because we can have an infinite cover. All this is insisting is that each "worker" has a set in which it does work.

4.3 Proof

We start with some preliminary lemmas.

4.3.1 Preliminary 1: Finding a Smooth Bump

We want to eventually find a function ϕ such that if we have a set $C \in U$, we want to find a function that is 1 inside C and 0 outside C.

Lemma 4: Given a compact C contained in an open U, where $C \subset U$, there exists a C^∞ function $\psi : U \to [0,1]$ such that
1. $\psi|_C = 1$
2. $\operatorname{supp} \psi \subset U$.

That is, if ψ existed, then it could look like the following mountain (where the flat part is C, slanted part is U, and the rest is outside of U)

We should find proving this problematic since even in the simplest case, we don't know any functions that do this! However, we can still prove this.

Proof. We take the following steps:

1. We restrict our attention to one dimension (i.e. a seashore). There exists a function $\sigma : \mathbb{R} \to \mathbb{R}$ such that

$$\begin{cases} \sigma(x) = 0 & x \leq 0 \\ \sigma(x) > 0 & x > 0 \end{cases} \tag{4.10}$$

   where $\sigma \in C^\infty$.

2. There exists a smooth one-dimensional function β_ϵ that represents a bump such that

$$\begin{cases} \beta_\epsilon(x) = 0 & |x| > \epsilon \\ \beta_\epsilon(0) > 0 \end{cases} \tag{4.11}$$

3. We now look at the general case of $\mathbb{R}^n$. We wish to show that there exists an n-dimensional bump. Given $a \in \mathbb{R}^n$ and $\epsilon > 0$, there exists $\beta_{a,\epsilon}$ such that $\beta \in C^\infty$ and

$$\begin{cases} \beta(a) > 0 & |x - a| < \epsilon \\ \beta(x) = 0 & |x - a| \geq \epsilon \end{cases} \tag{4.12}$$

4. Let's look at $\mathbb{R}$ again. There exists a smooth *step* function $\sigma \in C^\infty$ and $\theta : \mathbb{R} \to [0, 1]$ such that

$$\begin{cases} \theta(x) = 0 & x \leq 0 \\ \theta(x) = 1 & x \geq 1 \end{cases} \tag{4.13}$$

5. We can construct a *flaptop* mountain such that for some $C \subset U$, we have $f(x) = 1$ and outside U we have $f(x) = 0$.

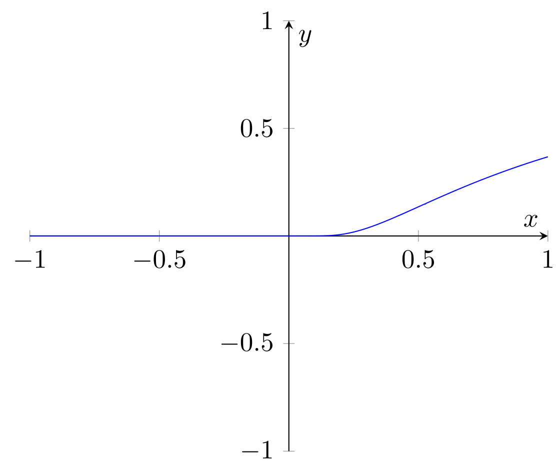
Let us take these steps and prove them:

1. In particular, we claim that

$$\sigma(x) = \begin{cases} e^{-1/x} & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{4.14}$$

   is such a function.

Seashore Function



   We can show σ is smooth via the following: for $x > 0$, we have

$$\sigma' = \frac{1}{x^2} e^{-1/x} \tag{4.15}$$

$$\sigma'' = \left(-\frac{2}{x^3} + \frac{1}{x^4} \right) e^{-1/x} \tag{4.16}$$

$$\sigma^{(n)} = r(x) e^{-1/x} \tag{4.17}$$

   where $r(x)$ is some rational function, which we can prove via induction. We can also show that $\lim\limits_{x \to 0} \sigma^{(n)} = 0$ since exponentials beat polynomials (it is not hard to show this formally). This is the key point to showing that the n^{th} derivative approaches 0 at $x = 0$, and that $\sigma(x)$ is C^∞.

2. We can construct such a $\beta(x)$ by multiplying two seashore functions together. Namely,
$$\beta_\epsilon(x) = \sigma(x+\epsilon)\sigma(\epsilon - x). \tag{4.18}$$
To show this has the desired properties, we just need to apply the properties of σ.
3. We can define
$$\beta_{a,\epsilon}(x) := \beta_{\epsilon^2}(|x-a|^2) \tag{4.19}$$
Since both β_{ϵ^2} and $|x-a|^2$ is smooth, this function is also smooth. Note that we couldn't have $|x-a|$ in the parameter since we want the function to be smooth, and we only know that the composition of two smooth function is smooth.
4. Let us define
$$\theta_0(x) = \int_0^x \beta_{1/2,1/2}(t)\,\mathrm{d}t \tag{4.20}$$
and let
$$\theta(x) = \frac{1}{\theta_0(1)}\theta_0(x) \tag{4.21}$$
5. We can create a finite open cover for C. To show this, for each $x \in C$, we can find $\epsilon_x > 0$ such that $B_{\epsilon_x}(x) \subset U$ which is possible since U is open. Then,
$$\{B_{\epsilon_x}(x)\} \tag{4.22}$$
is an open cover of C, hence it has a finite subcover x_i with $i = 1, \ldots, m$ and
$$\bigcup_{i=1}^m B_{\epsilon_{x_i}}(x_i) \supset C. \tag{4.23}$$
Our first guess for such a flaptop mountain function may be
$$f_0(x) = \sum_{i=1}^m \beta_{x_i,\epsilon_{x_i}}(x) \tag{4.24}$$
But we still need to crop this! Note that f_0 is a continuous function on a compact set bounded below by some $b > 0$ on C. The function we want is
$$f(x) = \theta\left(\frac{1}{b}f_0(x)\right) \tag{4.25}$$

□

4.3.2 Preliminary 2: Separating a Compact Set Contained in Open Set

Intuitively, a compact set contains its boundary while an open set is fuzzy around its boundary. Therefore, we should expect a gap between a compact set C and an open set U.

Lemma 5: Given a compact C and an open U, where $C \subset U \subset \mathbb{R}^n$, there exists a compact D such that
$$C \subset \operatorname{int} D \subset D \subset U \tag{4.26}$$

This lemma is a bit stronger than our intuition. It says that a compact set C can be contained in an open set $\operatorname{int} D$, which is contained in a compact set D and is contained in an open set U.

Proof. The lemma is intuitive: We can construct a finite open cover of C contained in U, since C is compact. We can then "shrink" each ball by a tiny bit and take the closure, so we can also find a finite closed cover of C. Taking the union, we have constructed such a D.

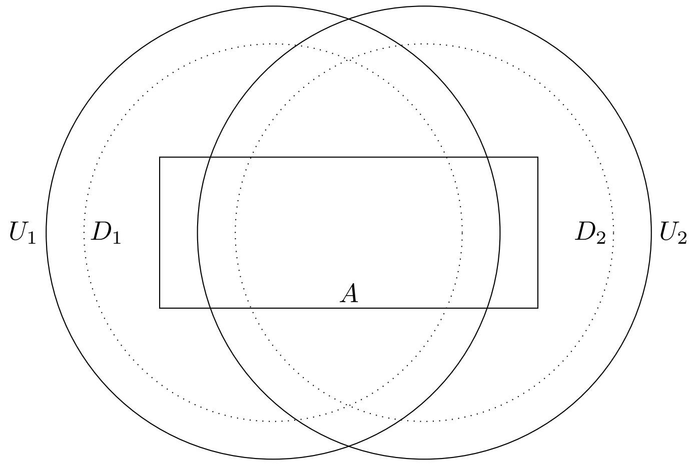
Let's make this more rigorous. For each $x \in C$, we can find an open ball B_x such that $\overline{B_x} \subset U$. Clearly, $\{B_x\}_{x\in C}$ covers C. By compactness, it has a finite subcover B_{x_i} where $i = 1, \ldots, p$. Take
$$D = \bigcup \overline{B_{x_i}}, \tag{4.27}$$
which is a finite union of closed and bounded sets, so D is closed and bounded, hence it is compact. It is also easy to check that
$$\operatorname{int} D = \bigcup B_{x_i} \supset C. \tag{4.28}$$
□

4.3.3 Putting it All Together

We can now put it all together and prove the Partition of Unity theorem, but we will separate it into cases.

Case I: A is compact:

We can visualize this case:



Since A is compact, we can construct a finite open cover of A, i.e. $U_1 \subset U_2$, then by lemma 2, construct $D_1 \subset D_2$ to also be a cover. Intuitively, we want each worker to do some work inside D_1 but no work outside U_1.

Proof. We perform the following steps:

1. Define the exclusive zone of U_1 to be $E_1 = A \setminus \bigcup_{j=2}^{p} U_j$ and show that we can construct it.
2. Show that we can "Shrink" each U_i to a compact set D_i such that $\{\operatorname{int} D_i\}$ still covers A. Namely, we can find compact D_i such that $D_i \subset U_i$ and $\bigcup \operatorname{int} D_i \supset A$.
3. Given the claim, we can find $\Psi_i : \mathbb{R}^n \to [0,1]$ which are C^∞ such that $\Psi_i|_{D_i} \equiv 1$ and $\operatorname{supp} \Psi_i \subset U_i$. The Ψ_i does not have to be C^∞ at this point.
4. Make the Ψ_i a C^∞ function by using preliminary 1.

Let us now prove these statements.

1. Let E_1 denote the exclusive zone of U_1, and is equal to $E_1 = A \setminus \bigcup_{j=2}^{p} U_j$. We know E_1 is compact and $E_1 \subset U_1$.

   By preliminary 2, we can find a compact set D_1 such that

   $$E_1 \subset \operatorname{int} D_1 \subset D_1 \subset U_1. \tag{4.29}$$

2. Let the **exclusivity zone** for U_1 be E_1, defined as

   $$E_1 = A \setminus \bigcup_{i=2}^{p} (U_i) \subseteq U_1. \tag{4.30}$$

   By preliminary 2, we can find a compact D_1 such that $E_1 \subseteq \text{int}(D_1) \subseteq D_1 \subseteq U_1$. Therefore, we have that

   $$\text{int}(D_1) \cup \bigcup_{i=2}^{p} U_i \supseteq A. \tag{4.31}$$

   Suppose $D_1, \ldots, D_{q-1}$ where $2 \leq q \leq p$ have been constructed such that $D_i \subseteq U_i$ and

   $$\bigcup_{i=1}^{q-1} (\text{int} D_i) \cup \bigcup_{i=q}^{p} (U_i) \supset A. \tag{4.32}$$

In words, this says that we can "shrink" some U_i to intD_i, and when joined with the other U_i, it still forms a cover for A. Let us set

$$E_q = A \setminus \left(\bigcup_{i=1}^{q-1} \mathsf{int}(D_i) \bigcup_{i=q+1}^{p} U_i \right). \tag{4.33}$$

Then E_q is compact, $E_q \subseteq U_q$, so we can use preliminary 2 to find D_q such that $E_q \subset \mathsf{int}(D_q) \subseteq D_q \subseteq U_q$. This is true for $q = 2$, and we can repeat until $q = p$.

3. We can define

$$\psi_i(x) = \begin{cases} \dfrac{\Psi_i(x)}{\sum_{j=1}^{p} \Psi_j(x)} & x \in \bigcup \operatorname{int} D_i \supset A \\ 0 & x \notin \bigcup \operatorname{int} D_i \end{cases} \tag{4.34}$$

Note that the denominator cannot be zero when $x \in \bigcup \operatorname{int} D_i \supset A$, since it will contain all the $\Psi_i|_{D_i}$. However, we might have continuity issues on the boundary of $\bigcup \operatorname{int} D_i$, since it may be zero right outside and nonzero directly inside.

4. We can multiply by a function $f(x)$ such that

$$\psi_i(x) = \begin{cases} f(x)\dfrac{\Psi_i(x)}{\sum_{j=1}^{p} \Psi_j(x)} & x \in \bigcup \operatorname{int} D_i \supset A \\ 0 & x \notin \bigcup \operatorname{int} D_i \end{cases} \tag{4.35}$$

where $f(x)$ is smooth and satisfies $f|_A \equiv 1$ and $\operatorname{supp} f \subset \bigcup \operatorname{int} D_i$. We can construct an even smaller cover $\mathcal{V}$ than $\bigcup D_i$ such that outside, $f(x) = 0$ and inside $\mathcal{V}$, we have $f(x) = 1$. This ensures that ψ_i is smooth.

□

Case II: A is in the form of $A = \bigcup_{k=0}^{\infty} A_k$ where A_k is compact and $A_k \subseteq$ int(A_{k+1}): Let us define

$$L_k = A_{k+1} - \mathsf{int}(A_k) \tag{4.36}$$

for $k \geq 0$. Let

$$\mathcal{U}_k = \{U \cap \mathsf{int}(A_{k+2}) \cap A_{k-1}^C | U \in \mathcal{U}\}. \tag{4.37}$$

Note that since the interior is open, U_k is open. We can use step 1 to find a partition of unity $\psi_k = \{\psi_{k_i}\}$ for L_k and subordinate to U_k. Let

$$\Psi = \bigcup_{k=0}^{\infty} \psi_k = \{\psi_i\}, \tag{4.38}$$

which is a countable collection of C^∞ functions. We can set

$$\phi_i(x) = \frac{\psi_i(x)}{\sum_{j=1}^{\infty} \psi_j(x)}. \tag{4.39}$$

Clearly, $\{\phi_i\}$ is the desired partition of unity for $A = \bigcup(A_k)$.

Case III: A is an arbitrary open set: Let us write

$$A = \bigcup_{k \geq 1} (A_k) \tag{4.40}$$

where

$$A_k = \left\{ x \in A : |x| \leq k \text{ and } \mathsf{dist}(x, A^c) \leq \frac{1}{k} \right\}. \tag{4.41}$$

Our claim is then A_k is contained in int(A_{k+1}). Each A_k is compact, and $\bigcup_{k \geq 1} (A_k) = A$. Therefore, by case II, we are done.

5 Integration on Unbounded Sets

We have only defined integration of bounded functions on bounded sets. We can extend this to unbounded functions on unbounded sets. We wish to define a new integration denoted by NT (for new technology) and relate it to the legacy definition of integration.

Our goal is to show that

$$\int_A^{NT} F = \sum_i \int \phi_i f. \tag{5.1}$$

where NT stands for new technology, which will allow us to integrate over unbounded sets.

We start off with a few preliminaries that worked for our legacy definition of integration.

1. Assuming f and g are integrable, we have

$$\int_R (f+g) = \int_R f + \int_R g \tag{5.2}$$

2. If c is a scalar, we have

$$\int_R cf = c \int_R f \tag{5.3}$$

Warning: However, be warned that things become super wack for infinite sums. The **Riemann Rearrangement Theorem** tells us that we can re-arrange these sums to get any sum you want. The reason is as follows: Suppose we have the sum

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} + \cdots, \tag{5.4}$$

we can divide it into two groups:

$$1, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \ldots \qquad -\frac{1}{2}, -\frac{1}{4}, -\frac{1}{6}, -\frac{1}{8}, \ldots \tag{5.5}$$

Note that both the left and right group diverges. I can set this sum equal to 257 by permuting the terms such that we first construct a number slightly above 257 using the left terms, then subtract some right terms to get slightly below 257, and since we have an infinite amount of numbers on both sides, we can have it bounce back and forth and eventually converge to 257. This sequence of steps is always possible if

$$\sum |a_n| = \infty \tag{5.6}$$

and

$$\sum a_n \text{converges} \tag{5.7}$$

Recall from 157 that $\sum a_i$ is absolutely convergent if $\sum |a_i| < \infty$. If this is the case, then the above nonsense does not work. The commutative law for an infinite sum then holds.

5.1 Setup

We want to use our legacy definition of integration on functions acting on unbounded sets. To save work, we want to be able to relate the two.

Let A be an open set that is not necessarily bounded. Let f be a function $f : A \to \mathbb{R}$ with the following properties:

1. f is **locally bounded**

   Definition: A set A is locally bounded if for all $x \in A$, there exists an open $V \ni x$ such that f is bounded on V.

2. f is continuous except on a measure-0 set.

Let $U = \{\mathcal{U}\}$ be an open cover of A by bounded sets on which f is bounded. Then let $\Phi = \{\phi_i\}$ be a partition of unity for A subordinate to U.

We wish to show that

$$\int_A^{NT} f = \sum_i \int \phi_i f. \tag{5.8}$$

Note that we didn't specify the region of integration of the right hand side, but this doesn't matter since Φ is subordinate to $\mathcal{U}$, so we can define the region of integration to a rectangle bigger than $\operatorname{supp}\phi_i$. There are three problems with this that we can fix:

- How do we know that $\int \phi_i f$ is integrable?
- How do we know that the infinite sum converges and doesn't do anything weird?
- How do we know that a different choice of $\mathcal{U}$ and a partition of unity doesn't affect this?

We fix this by saying that f is (U, Φ)-integrable if

$$\sum_i \int \phi_i |f| < \infty \tag{5.9}$$

In that case, we define

$$\int_A^{(U,\Phi)} f := \sum_i \int \phi_i f. \tag{5.10}$$

We want to later show that this doesn't depend on U and Φ. Note that this series is absolutely convergent since

$$\sum \left| \int \phi_i f \right| \leq \sum \int |\phi_i f| = \sum \int_\infty \phi_i |f| < \infty \tag{5.11}$$

so the weird things with rearrangements do not occur; the sum becomes well behaved.

5.2 Choice of PO1 and Open Cover are Irrelevant

We wish to show that

Theorem: If A and f are as before, then

1. If U, U' and Φ, Φ' are open covers and partitions of unity as before, then: f is (U, Φ)-integrable if and only if f is (U', Φ')-integrable.
2. If f is integrable (NT), then

$$\int_A^{(U',\Phi')} f = \int_A^{(U,\Phi)} f. \tag{5.12}$$

In that case,

$$\int_A^{NT} f := \int_A^{(U',\Phi')} f = \int_A^{(U,\Phi)} f. \tag{5.13}$$

We will also show that this new definition of the integral is equivalent to the old definition, when we have bounded functions in bounded sets.

Proof. We will write a chain of equalities:

$$\int_A^{(U,\Phi)} g \underset{(1)}{=} \sum_i \int \phi_i g \tag{5.14}$$

$$\underset{(2)}{=} \sum_i \int \left(\sum_j \phi'_j \right) \phi_i g \tag{5.15}$$

$$\underset{(3)}{=} \sum_i \sum_j \int \phi'_j \phi_i g \tag{5.16}$$

$$\underset{(4)}{=} \sum_j \sum_i \int \phi_i \phi'_j g \tag{5.17}$$

$$\underset{(3)}{=} \sum_j \int \left(\sum_i \phi_i \right) \phi'_j g \tag{5.18}$$

$$\underset{(2)}{=} \sum_j \int \phi'_j g \tag{5.19}$$

$$\underset{(1)}{=} \int_A^{(U',\Phi')} g \tag{5.20}$$

Let us go through the above first with $g = |f|$. For this pass, we just need to show that the sum at the top converges if and only if the sum at the bottom converges.

1. Irrelevant
2. Sum is 1 in a partition of unity
3. Integration is linear, but we have to be a bit more careful when dealing with infinite sums. We need to show that

$$\int \left(\sum \phi'_j \right) h = \sum \int \phi'_j h, \tag{5.21}$$

   where $\operatorname{supp} h \subset \operatorname{supp} \phi_i \subset U \in \mathcal{U}$, so $\operatorname{supp} h$ is bounded and closed, hence it is compact. Therefore, byu compactness of $\operatorname{supp} h$ and local finiteness of Φ'_i, only finitely many i satisfy

$$\operatorname{supp} \phi'_i \cap \operatorname{supp} h \neq \emptyset. \tag{5.22}$$

   Hence, this holds by finite linearity.

4. Note that $\phi_i \phi'_j = \phi'_j \phi_i$ due to commutativity. Now, we need to show that we can switch the sums. We have $|g|, \phi_1 \geq 0, \phi'_j \geq 0$ so all terms of the sums are non-negative. By the following exercise, we're done.

   Exercise: A sum of non-negative terms is convergent if and only if every rearrangement thereof is convergent.

We wish to repeat the above steps for $g = f$. We now work under the assumption that the function is integrable, so the sums are absolutely convergent. The first step is by definition, the next two steps are the same as above, and the last step is true by absolute convergence. □

5.3 NT Agrees with Old for Bounded Sets and Functions

We wish to show that NT and old integration is irrelevant when A and f are both bounded. However, the old integration might not even exist since we can write

$$\int_A^{\text{old}} f = \int_{\text{big R}}^{\text{old}} \chi_A f \tag{5.23}$$

where χ is the characteristic function

$$\chi_A = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \tag{5.24}$$

and we need to make sure χ_A is continuous except on a set of measure-0.

Theorem: The following are true, keeping the same definitions as above:
1. If A and f are bounded, then f is NT-integrable on A.
2. If in addition A is Jordan Measurable, then the new integration is equivalent to the old integration.

Proof. We will prove both parts:

1. Assume $|f| < M$ and $A \subset R$ where R is some rectangle. We need to consider
$$\sum \int \phi_i |f| \tag{5.25}$$
If this is finite, then f is NT-integrable. Let us write the following chain of equalities:
$$\sum \int_R \phi_i |f| = \int_R \sum \phi_i |f| = \int_R |f| \leq \int_R M = M\text{vol}(R). \tag{5.26}$$
The only problem with this is interchanging the integral and summation symbol. However, this is a non-issue since we can restrict our attention to some finite subset of the ϕ_i's. For example, we can write the chain as
$$\sum_{i=1}^{N} \int_R \phi_i |f| = \int_R \sum_{i}^{N} \phi_i |f| \leq \int_R M = M\text{vol}(R). \tag{5.27}$$
An infinite sum is convergent if and only if all the partial sums converge, so we are done.
2. To prove the second part, we need the following lemma.

Lemma 6: For all $\epsilon > 0$, we can find a compact Jordan-measurable set $C \subset A$ such that $\text{vol}(A - C) < \epsilon$.

The proof is given in A8Q1. Assuming lemma, for only finitely many i's, $\text{supp}(\phi_i) \cap C = \emptyset$. We can find an N bigger than all of those i's, such that $\sum_{i=1}^{N} (\phi_i) = 1$ on C. Let us write down the following sequence of equality and inequalities:
$$\left| \int_A^{\text{old}} f - \sum_{i=1}^{N} \int^{\text{old}} (\phi_i f) \right| = \left| \int_A \left(f - \sum_{i=1}^{N} \phi_i f \right) \right| \tag{5.28}$$
$$\leq \int \left(|f| \cdot \left(1 - \sum_{i=1}^{N} \phi_i \right) \right) \tag{5.29}$$
$$\leq M \int \left(1 - \sum_{i=1}^{N} \phi_i \right) \tag{5.30}$$
$$\leq M \int_{A-C} (1) \tag{5.31}$$
$$\leq M \cdot \text{vol}(A - C) \tag{5.32}$$
$$\leq M \cdot \epsilon, \tag{5.33}$$
and we are done.

□

We still have a few more things to show. Namely, NT integration is linear and applying Fubini to NT integration.

5.4 Other Properties of NT Integration

The proof that NT integration is linear. Assuming f and g are integrable, we need to show that $f + g$ is integrable. To do this, pick the same open cover and PO1. The integral in each open ball is the same as the old way of integrating. The sum of two absolutely converging series is absolutely converging, so we are done. Then,
$$\int_A^{NT} (f + g) = \sum_{i=1}^{\infty} \int \phi_i (f + g) = \sum_{i=1}^{\infty} \left(\int \phi_i \cdot f + \int \phi_i \cdot g \right) = \int f + \int g. \tag{5.34}$$

Finally, we will show that Fubini applies to NT-integration as well. If $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^n$ are open, $f : A \times B \to \mathbb{R}$ is locally bounded and continuous, except for finitely many points. Then,

$$\int_{A\times B} (f(x,y))\,\mathrm{d}x\,\mathrm{d}y = \int_A \mathrm{d}x \int_B \mathrm{d}y\, f(x,y), \tag{5.35}$$

assuming f is integrable on $A \times B$. If:

- $\mathcal{U}$ is an open cover of A
- $\mathcal{V}$ is an open cover of B.
- $\Phi = \{\phi_i\}$
- $\Psi = \{\psi_i\}$

then let us take $\mathcal{W} = \{U \times V : U \in \mathcal{U}, V \in \mathcal{V}\}$ is an open cover of $A \times B$ and $\Lambda = \{\lambda_{i,j} = \phi_i(x) \cdot \phi_j(x)\}$ is a partition of unity for $A \times B$ subordinate to W. Note that

$$\sum_{i=1}^{\infty}(\lambda_{i,j}(x,y)) = \sum_{i,j}(\phi_i(x) \cdot \phi_j(y)) = 1 \cdot 1 = 1. \tag{5.36}$$

Once we have this, NT-fubini reduces to old-fubini.

6 Change of Variables

Theorem: Let $A \subset \mathbb{R}^n$ is open, and let $g : A \to \mathbb{R}^n$ is continuously differentiable, 1-1, and $\forall a \in A$, $g'(a)$ is invertible. If $f : g(A) \to \mathbb{R}$ is integrable, then

$$\int_{g(A)} f = \int_A (f \circ g)|\det g'| \tag{6.1}$$

This is a powerful theorem. We can use it to compute the **Gaussian integral**, which people often refer to it as the most important integral in mathematics:

$$\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi} \tag{6.2}$$

Proof. Let us name $I_1 = \int_{\mathbb{R}} e^{-x^2/2}$. Instead of computing this, let us compute

$$I_2 = \int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y\,. \tag{6.3}$$

Then, since NT-integration is linear:

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}} \mathrm{d}x \int_{\mathbb{R}} e^{-(x^2+y^2)/2} \tag{6.4}$$

$$= \int_{\mathbb{R}} \mathrm{d}x\, e^{-(x^2)/2} \int_{\mathbb{R}} e^{-(y^2)/2} \tag{6.5}$$

$$= I_1^2 \tag{6.6}$$

Let us now compute I_2 by switching to polar coordinates. To do so, consider the map

$$g(r,\theta) = \begin{pmatrix} r\cos\theta \\ r\sin\theta. \end{pmatrix} \tag{6.7}$$

Then,

$$\det g' = \det \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta. \end{pmatrix} = r. \tag{6.8}$$

Therefore, applying the COV formula, we have

$$\int_{g(A)} f = \int_A (f \circ g)|\det g'| \tag{6.9}$$

$$= \int \mathrm{d}r\,\mathrm{d}\theta\, e^{-r^2/2} r \tag{6.10}$$

$$= 2\pi. \tag{6.11}$$

Therefore, $I_1 = \sqrt{2\pi}$. □

However in the above proof, we ignored two things. First, we applied Fubini's on integrals with infinite bounds, and secondly we ignored the distinction between open and closed sets. That is, integrating over the following two sets is equivalent:

$$[0,\infty) \times [0, 2\pi] \leftrightarrow (0,\infty) \times (0, 2\pi). \tag{6.12}$$

Let us address these issues:

1. We can rewrite

$$\int_{\mathbb{R}\times\mathbb{R}} e^{-\frac{x^2+y^2}{2}}\,\mathrm{d}x\,\mathrm{d}y = \left(\int_{\mathbb{R}} e^{-x^2/2}\,\mathrm{d}x\right)\left(\int_{\mathbb{R}} e^{-y^2/2}\,\mathrm{d}y\right) \tag{6.13}$$

as

$$a_N = \int\limits_{[-N,N]^2} e^{-\frac{x^2+y^2}{2}} \, dx \, dy = \left(\int\limits_{-N,N} e^{-x^2/2} \, dx \right) \left(\int\limits_{-N,N} e^{-y^2/2} \, dy \right), \tag{6.14}$$

which is true via the old Fubini's Theorem. What remains to show is that as $N \to \infty$, all three integrals converge to each other. Intuitively, this makes sense since the function "descends rapidly," but this is not formal.

To be rigorous, define

$$\int f \equiv \int\limits_{\mathbb{R}\times\mathbb{R}} e^{-(x^2+y^2)/2} \geq 0 \tag{6.15}$$

which is equal to, according to POI:

$$\int f = \sum_{i=1}^{\infty} \phi_i f = \lim_{p\to\infty} \sum_{i}^{p} \int \phi_i f. \tag{6.16}$$

Since the ϕ_i have finite support, so we can interchange the integral and summation:

$$\int f = \int \left(\sum_{i=1}^{p} \phi_i \right) f \equiv b_p. \tag{6.17}$$

We now want to show that a_N and b_P have the same limit. Note that both a_N and b_P are both increasing sequences. If we can show that for all N, there exists an N such that $b_p \geq a_N$ and for all p, there exists N such that $a_N \geq b_P$, then

$$\lim_{p\to\infty} b_p = \lim_{N\to\infty} a_N. \tag{6.18}$$

To show that these two properties are true, we can do the following:

- To show that there exists a P, we can pick a p such that $\sum_{i=1}^{p} \phi_i(x) = 1$ on $[-N, N]^2$, which implies that

$$\sum \phi_i \geq \chi_{[-N,N]^2}. \tag{6.19}$$

  Therefore:

$$b_p = \int \left(\sum_{i=1}^{p} \phi_i \right) f \geq \int \chi_{[-N,N]^2} f = a_N. \tag{6.20}$$

- The proof is exactly the same. For large enough N, the square $[-N, N]^2$ contains the support,

$$[-N, N]^2 \supset \operatorname{supp} \sum_{i=1}^{p} \phi_i \tag{6.21}$$

2. The general idea is that you can always ignore closed sets of measure-0. This is because compact sets of measure-0 are content-0 and thus can be ignored.

> **Warning**: This is not the case with ignoring general sets. If we change the value of a function on a set of measure-0, then we can introduce a lot of issues (i.e. continuity). However, if the set of measure-0 is also closed, we do not have any issues.

6.1 Layer Preserving Map

> **Definition**: A function $g : \mathbb{R} \to \mathbb{R}^n$ is layer preserving if $g(x_1, \dots, x_n) = (\cdots, x_n)$. That is, the last coordinate stays the same.

As an analogy, consider Prof. Dror's stacked books. We can integrate by going one layer at a time and then integrating over the book. We hope this will reduce the COV formula for g to the COV formula for g restricted to a given layer. This suggests that we might be able to prove the COV formula for layer preserving maps using induction.

6.2 Proof

Proof. We will do the following:

1. If the theorem holds for two functions, then it holds for the composition.
2. If the COV formula holds in dimensions $n-1$, that it applies in n dimensions (for layer preserving maps).
3. Show that g can be written as a composition of linear preserving maps.
4. If the COV formula holds for small sets, then it holds for large sets.
5. COV holds for the identity.
6. The restricted COV formula implies the general COV formula holds.

And to prove the above:

1. Consider the following composition of functions:
$$A \underset{h}{\to} h(A) \underset{g}{\to} (g \circ h)(A) \underset{f}{\to} \mathbb{R}. \tag{6.22}$$
Then by the COV formula for g, we can write:
$$\int\limits_{(g\circ h)(A)} f = \int\limits_{h(A)} (f \circ g)|\det g'|. \tag{6.23}$$
Let us define $\bar{f} \equiv (f \circ g)|\det g'|$, and apply the COV formula for h. We get:
$$\int\limits_{h(A)} \bar{f} = \int\limits_{A} (\bar{f} \circ h)|\det h'|. \tag{6.24}$$
Substituting in $\bar{f}$ and using the chain law, we have what is desired.

2. Assume that COV holds for $n-1$ dimensions. Let $g : U \to \mathbb{R}^n$ where U is an open bounded set and let g be a layer-preserving map such that $g(U)$ is bounded. We then want to show the following restricted formula for cov(g) holds: namely if $f : g(U) \to \mathbb{R}$ is continuous and $\operatorname{supp} f \subset g(U)$, then
$$\int\limits_{g(U)} f = \int\limits_{U} (f \circ g)|\det g'|. \tag{6.25}$$
We don't actually need to specify the domain of integration, since the support is contained in $g(U)$. For any $z \in \mathbb{R}$, define $g_Z : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ by
$$g_z(x) = \begin{pmatrix} g_1(x,z) \\ \vdots \\ g_{n-1}(x,z) \end{pmatrix} \tag{6.26}$$
and define $f_Z : \mathbb{R}^{n-1} \to \mathbb{R}$ by $f_z(x) = f(x,z)$. The integral is thus
$$\int f = \int\limits_{\mathbb{R}} \mathrm{d}z \int\limits_{\mathbb{R}^{n-1}} \mathrm{d}x\, f(x,z) = \int\limits_{\mathbb{R}} \mathrm{d}z \int\limits_{\mathbb{R}^{n-1}} f_z(x) \tag{6.27}$$
Using our inductive hypothesis (COV holds for $n-1$ dimensions), we can rewrite the integral as
$$\int\limits_{\mathbb{R}} \mathrm{d}z \int\limits_{\mathbb{R}^{n-1}} (f_z(x) \circ g_z) \cdot |\det g_z'|. \tag{6.28}$$
Recall the derivative is written as the Jacobian
$$g' = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g}{\partial x_{n-1}} & \frac{\partial g_1}{\partial z} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial g_{n-1}}{\partial x_1} & \cdots & \frac{\partial g_{n-1}}{\partial x_{n-1}} & \frac{\partial g_{n-1}}{\partial z} \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_{n-1}}{\partial x_{n-1}} & 1 \end{pmatrix} \tag{6.29}$$

3. Let $g : \mathbb{R}^n_{x_1} \to \mathbb{R}^n_{y_1}$ be an arbitrary function such that it maps $(x_1, \ldots, x_n)$ to $(y_1, \ldots, y_n)$ via $y_i = g_i(x_1, \ldots, x_n)$. We can break up this map in the following way:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \xrightarrow{\alpha} \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1} \\ y_n \end{pmatrix} \xrightarrow{\beta} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \tag{6.30}$$

Note that α and β are invertible, which can be shown using the inverse function theorem.

□

6.3 Baby Sard's Theorem

We will give a brief overview of Sard's Theorem, and revisit it with greater rigor when we discuss Stoke's Theorem. Suppose we want to map $\{(r, \theta) : r \geq 0, 0 \leq \theta \leq 2\pi\}$ to $\mathbb{R}^2$. Note that this does not meet the conditions for COV since $\{(0, \theta) : 0 \leq \theta \leq 2\pi\}$ gets mapped to $(0, 0)$, so this map is not injective. However, intuitively, it feels like we can still use the COV formula in this case. A corollary of Sard's Theorem makes this official.

Namely, we can drop the condition that g' is $1-1$.

Theorem: (Baby Sard Theorem) Suppose we have open $A \subset \mathbb{R}^n$ and $g : A \to \mathbb{R}^n$ is continuously differentiable. Define:

$$C = \text{“critical set of } g\text{”} = \{x \in A | \det g'(x) = 0\}. \tag{6.31}$$

Then $g(C)$ is of measure-0.

The nomenclature comes from in 1-dimension, C is literally the set of critical values, since $\det g'(x) = g'(x)$ in single-variable calculus. It is important to note that we take the image of C. For example, consider a smooth bump. Clearly, the set of x values at which the slope is zero is not of measure-0, since it is a continuous range in $\mathbb{R}$.

Claim: $A \setminus C$ is closed.

Proof. The function $h(x) \equiv \det g'(x)$ is continuous since g is continuously differentiable and the determinant just involves combining terms in the matrix (which are continuous), so h is continuous. Then:

$$C = h^{-1}\{0\}. \tag{6.32}$$

For continuous functions, the pre-image of a closed set is a closed set □

Corollary: In the COV theorem, we can drop the condition g' is 1-1 (equivalently invertible).

Proof. Suppose we map A to $g(A)$. Define $C \in A$ to be the closed set where $g'(C)$ is not invertible. While C may not be of measure-0, $g(C)$ is of measure-0.

First, we show that $g(A)$ is closed. To do this, we can write $g(A) \setminus g(C) = g(A \setminus C)$. Note that $A \setminus C$ is an open set of which g' is invertible. Therefore, the inverse function theorem applies, so $g(A \setminus C)$ is open.

Then:

$$\int_{g(A)} f = \int_{g(A \setminus C)} f. \tag{6.33}$$

and

$$\int_A (f \circ g)|\det g'| = \int_{A \setminus C} (f \circ g)|\det g'|. \tag{6.34}$$

By the old COV theorem, the relevant conditions are satisfied for $\int_{A \setminus C} (f \circ g)|\det g'| = \int_{g(A \setminus C)} f$ and using the equalities above, we can show the new version of COV. □

We can give a *baby* proof of baby Sard. We will not go into details, since this is only a sketch

1. Restrict attention to A is a rectangle, and there is some subset $C \in A$ such that $g'(C)$ is not invertible.
2. Partition A into several subrectangles. For the subrectangles that touch C, then there will be a point where g' is not onto, i.e. it maps to a smaller dimensional subspace. This subrectangles will not be mapped into a $n-1$ dimensional subspace, but will be mapped to a neighbourhood of one.
3. This shows that we can cover $g(C)$ with rectangles where one dimension is very, very narrow, so this suggests it is of measure-0.

6.4 Adult Sard's Theorem

We can generalize Sard's theorem to a map from n dimensions to m dimensions.

Theorem: Let A be an open set and $g : A \subset \mathbb{R}^n \to \mathbb{R}^m$. Define

$$C = \{x \in A | \text{rank} g' < m\}, \tag{6.35}$$

and g is k times continuously differentiable, where $k = \max(1, n-m+1)$. Then $g(C)$ is measure-0.

This is a much harder theorem to prove!

7 K-Vector

Definition: Let V be a vector space over $\mathbb{R}$ and $k \in \mathbb{N} = \mathbb{Z}_{\geq 0}$. Then: $T : V^k \to \mathbb{R}$ is called **multi-linear** or **k-linear** if

$$T(u_1, \ldots, \alpha u_1' + \beta u_1'', \ldots, u_k) = \alpha T(u_1, \ldots, u_1', \ldots, u_k) + \beta T(u_1, \ldots, u_1'', \ldots, u_k). \tag{7.1}$$

Let us look at some examples. An inner product is a 2-linear map $\langle u_1, u_2 \rangle \in \mathbb{R}$. An example of a 1-linear map is a linear map $\phi : V \to \mathbb{R}$ where $\phi \in V^*$. An example of an n-linear map is the determinant. An example of a 0-linear is $w : V^0 = \{()\} \to \mathbb{R}$. Note that V^0 has a single element.

Definition: Define

$$\mathcal{T}^k(V) = \{\text{k-linear maps on V}\}. \tag{7.2}$$

Note that many other sources call it $\mathcal{T}^k(V^*)$.

This means that $\langle , \rangle \in \mathcal{T}^2 V$, $\det \in \mathcal{T}^n V$, $\mathcal{T}^0 V \cong \mathbb{R}$, and $\mathcal{T}' V = V^*$. The set $\mathcal{T}^k(V)$ is related to the things that we can integrate, so we will later think about the dimensions.

Review: Suppose $v_1, \ldots, v_n$ is a basis of V. Then there exists a unique basis $\phi_1, \ldots, \phi_n$ of $V^* = \mathcal{T}^1 V$ such that

$$\phi_i(v_j) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise.} \end{cases} \tag{7.3}$$

Then $\{\phi_i\}$ is called the **the dual basis of $(v_1, \ldots, v_n)$**.

Let us give some more examples. For example, what is the dual basis of $\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\}$? Recall that $\mathbb{R}^2$ can be thought of as the set of column vectors while $(\mathbb{R}^2)^*$ can be thought of as the set of row vectors. The properties that we want are:

$$\phi_1(v_1) = 1 \qquad \phi_1(v_2) = 0 \tag{7.4}$$
$$\phi_2(v_1) = 0 \qquad \phi_2(v_2) = 1. \tag{7.5}$$

These set of equations are equivalent to

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \begin{pmatrix} v_1 & v_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{7.6}$$

Therefore,

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} v_1 & v_2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} -2 & 3/2 \\ 1 & -1/2 \end{pmatrix}, \tag{7.7}$$

so $\phi_1 = (-2, 3/2)$ and $\phi_2 = (1, -1/2)$.

Claim: $\mathcal{T}^k(V)$ is itself a vector space.

Tensor multiplication is denoted with the map $\otimes : \mathcal{T}^k \times \mathcal{T}^\ell \to \mathcal{T}^{k+\ell}$ defined by

$$(T_1 \otimes T_2)(u_1, \dots, u_{k+\ell}) = T_1(u_1, \dots, u_k) \cdots T_2(u_{k+1}, \cdots, u_{k+\ell}). \tag{7.8}$$

Claim: $\mathcal{T}_1 \otimes \mathcal{T}_2 = \mathcal{T}_1\mathcal{T}_2 \in \mathcal{T}^{k+\ell}$.

7.1 Properties and Notation

Here are some properties of vectors

- Associativity: $T_1(T_2T_3) = (T_1T_2)T_3$
- Distributivity: $(T_1 + T_2)T_3 = T_1T_2 + T_1T_3$ (More specifically, bilinear.)

Note that multiplication isn't necessarily commutative. Consider the counterexample $e_1, e_2 \in \mathbb{R}^2$ and $\phi_1, \phi_2 \in \mathcal{T}'(V)$. Then:

$$(\phi_1\phi_2)(e_1, e_2) = \phi_1(e_1)\phi_2(e_2) = 1 \cdot 1 = 1$$
$$(\phi_2\phi_1)(e_1, e_2) = \phi_2(e_1)\phi_1(e_2) = 0 \cdot 0 = 0$$

Definition: Let $\underline{n} = \{1, \dots, n\}$. Then

$$\underline{n}^k = \{\bar{i} = I = (i_1, \dots, i_k) : i_\alpha \in \underline{n}\}. \tag{7.9}$$

This notation is nice since $|\underline{n}^k| = n^k$. Let $(v_j)_{j=1}^n \in V^n$ and $J \in n^k$, where J is known as a **multi-index**. Let us write

$$V_J = (V_{j_1}, V_{j_2}, \dots, V_{j_k}). \tag{7.10}$$

Then $\phi_i \in V^*$ for $i = 1, \dots, n$, $I \in \underline{n}^k$, and $\phi_I = \phi_{i_1} \cdot \dots \cdot \phi_{i_k}$.

- **Example:** Consider our previous counterexample. We can re-write it as

$$\begin{aligned} \phi_1\phi_2 - \phi_{(1,2)} &\qquad \phi_{(1,2)}(e_{(1,2)}) = 1 \\ \phi_2\phi_1 = \phi_{(2,1)} &\qquad \phi_{(2,1)}(e_{(1,2)}) = 0 \end{aligned}$$

Suppose V is a vector space with basis $v_1, \dots, v_n$ and dual basis $\phi_1, \dots, \phi_n$, and $I, J \in \underline{n}^k$. Then

$$\begin{aligned} \mathcal{T}^k &= \phi_I(v_J) && (7.11) \\ &= (\phi_{i_1} \cdots \phi_{i_k})(V_{j_i} \cdots V_{j_k}) && (7.12) \\ &= \prod_{\alpha=1}^{k} \phi_{i_\alpha}(V_{j_\alpha}) && (7.13) \\ &= \prod_{\alpha=1}^{k} \delta_{i_\alpha j_\alpha} && (7.14) \\ &= \begin{cases} 1 & I = J \\ 0 & I \neq J \end{cases} && (7.15) \end{aligned}$$

7.2 Dimension of k-tensors

We are finally ready for our first theorem!

Theorem: Let V be a vector space with respect to a basis $v_1, \dots, v_n$ and dual basis $\phi_1, \dots, \phi_n$. Then:

$$\{\phi_I : I \in \underline{n}^k\} \tag{7.16}$$

is a basis of $\mathcal{T}^k(V)$, and hence, the dimension is

$$\dim \mathcal{T}^k V = n^k. \tag{7.17}$$

This allows us to better understand the space of k-tensors. We prove this step by step through a series of lemmas.

Lemma 7: If $T_1, T_2 \in \mathcal{T}^k$, then $T_1 = T_2 \iff \forall I, T_1(V_I) = T_2(V_I)$.

Proof. The forwards direction is obvious. To prove the backwards direction, assume that for all I, we have $T_1(V_I) = T_2(V_I)$. Then set $T = T_1 - T_2$. Therefore:

$$T(u_1, \dots, u_k) = T\left(\sum_{i_1=1}^{n} a_{1,i_1} v_1, \sum a_{2,i_2} v_2, \dots, \sum a_{k,i_k} v_k\right),$$

and applying linearity, we end up with

$$\begin{aligned} T(u_1, \dots, u_k) &= \text{(sum and product of coefficients)}(T_1(V_I) - T_2(V_I)) = 0 \\ &\implies T = 0 \implies T_1 = T_2. \end{aligned}$$

□

Lemma 8: The set $\{\phi_I\}$ spans $\mathcal{T}^k V$.

Let us verify that this lemma makes sense first. Given $T \in \mathcal{T}^k$, does it make sense that T can be written as a linear combination of ϕ_I, specifically $T \stackrel{?}{=} \sum_I a_I \phi_I$. The trick is to evaluate both sides on V_J, to get

$$\begin{aligned} T(V_J) &= \sum_I a_I \phi_I(V_J) \\ &= \sum_I a_I \delta_{IJ} = a_J. \end{aligned}$$

Note that we found a formula for a_J, so we're nearly done! Let us turn this into a formal proof.

Proof. Given $T \in \mathcal{T}^k$, set $a_I = T(V_I)$, and then I claim that $T = \sum_I a_I \phi_I$. To check two tensors are equal to each other, we can use lemma 1. All we need to show is

$$T(V_J) = \left(\sum a_I \phi_I\right)(V_J).$$

But, by definition, $T(V_J) = a_J$ and the RHS is a_J as well from the above discussion. □

Lemma 9: The $\{\phi_I\}$ are linearly independent.

Proof. Pick $a_I \in \mathbb{R}$ such that $\sum a_I \phi_I = 0$. Therefore, for every $J \in \underline{n}^k$,

$$\left(\sum_I a_I \phi_I\right)(V_J) = \sum_I a_I \phi_I V_J = \sum_I a_I \delta_{IJ} = 0. \tag{7.18}$$

Recall that we also have $\sum_I a_I \delta_{IJ} = a_J$, so $a_J = 0$ for all J. □

A Philosophical Aside

In math, things tend to either push forward or pull back. When we map two spaces, we are transforming points, subsets, paths, functions, etc. For example:

- Points like to be pushed forward (to be pull backed, they need extra conditions)
- Subsets can be pushed forward and be pulled back.
- Paths like to be pushed.
- Functions like to be pulled back. For example, if $f : X \to Y$ and $h : Y \to \mathbb{R}$, then we can write

$$h \circ f = f^* h \tag{7.19}$$

7.3 Pullback

We will now resume our discussion on k-tensors. Suppose $L : V \to W$ is a linear function. Then there exists $L^* : \mathcal{T}^k(W) \to \mathcal{T}^k(V)$ defined by

$$T \mapsto (L^*T)(u_1, \ldots, u_k) = L^*(Tu_1, \ldots, Tu_k) \tag{7.20}$$

where $T \in \mathcal{T}^k$ and $u_i \in V$.

Lemma 10: If $T \in \mathcal{T}^k W$, then $L^*T \in \mathcal{T}^* V$. Namely, L^*T is multi-linear.

Proof. We can compute:

$$\begin{aligned}
(L^*T)(u_1, \ldots, \alpha u_i' + u_i'', \ldots, u_k) &= T(Lu_1, \ldots, L(\alpha u_i' + u_i''), \ldots, Lu_k) && (7.21)\\
&= T(Lu_1, \ldots, \alpha Lu_i' + Lu_i'', \ldots, Lu_k) && (7.22)\\
&= \alpha T(u_1, \ldots, u_i', \ldots, u_k) + T(u_1, \ldots, u_i'', \ldots, u_k) && (7.23)\\
&= \alpha(L^*T)(u_1, \ldots, u_1', \ldots, u_k) + (L^*T)(u_1, \ldots, u_1'', \ldots, u_k). && (7.24)
\end{aligned}$$

□

Lemma 11: The pullback map $L^* : \mathcal{T}^k W \to \mathcal{T}^k V$ is a linear map between vector spaces.

Proof. We need to show that if $T_1, T_2 \in \mathcal{T}^k W$, then:

$$L^*(\alpha T_1 + \beta T_2) = \alpha(L^*T_1) + \beta(L^*T_2), \tag{7.25}$$

which can be shown similarly by direct computation. □

Lemma 12: L^* is compatible with tensor multiplication. Namely, if $T_1 \in \mathcal{T}^k W$ and $T_2 \in \mathcal{T}^\ell W$. Then $T_1T_2 \in \mathcal{T}^{k+\ell}(W)$ and:

$$L^*(T_1T_2) = (L^*T_1)(L^*T_2) \in \mathcal{T}^{k+\ell} V. \tag{7.26}$$

Proof. Trace the definitions and see that it works. □

7.4 Killing Repetitions

Definition: A tensor $T \in \mathcal{T}^k V$ **kills repetitions** if $T(u_1, \ldots, u_k) = 0$ whenever $u_i = u_j$ is satisfied for some i, j.

We wish to study these repetition killing tensors in the same sense that we studied all k tensors.

Definition: $T \in \mathcal{T}^k$ is **alternating** if

$$T(\ldots, u, \ldots, w, \ldots) = -T(\ldots, w, \ldots, u, \ldots). \tag{7.27}$$

We then define

$$\Lambda^k(V) := \{T \in \mathcal{T}^k V : T \text{ is alternating}\} \tag{7.28}$$

Note that $\Lambda^k(V)$ is a subspace of $\mathcal{T}^k(V)$.

Lemma 13: If $T \in \mathcal{T}^k$, then T kills repetitions if and only if T is alternating.

Proof. Suppose that T is alternating. Consider a repetition. Then:

$$T(\dots,u,\dots,u,\dots) = -T(\dots,u,\dots,u,\dots) \implies T(\dots,u,\dots,u,\dots) = 0. \tag{7.29}$$

Now suppose that T kills repetitions. Then: $T(\dots,u+w,\dots,u+w,\dots) = 0$. If we expand the LHS using multilinearity, we get

$$\begin{aligned} 0 &= T(\dots,u,\dots,u,\dots) + T(\dots,u,\dots,w,\dots) + T(\dots,w,\dots,u,\dots) + T(\dots,w,\dots,w,\dots) \\ 0 &= T(\dots,u,\dots,w,\dots) + T(\dots,w,\dots,u,\dots), \end{aligned}$$

which is the multilinearity property. □

- $\det \in \Lambda^n\mathbb{R}^n$
- Suppose $k \leq n$. Then $\lambda_I \in \Lambda^k\mathbb{R}^n$ where $I \in \underline{n}^k$ and maps a $n \times k$ matrix to the determinant of the *minor* corresponding to rows given by I.

Let's explore λ_I a bit more. Notice that it doesn't make any sense if I contains anything repeating, because then it will be zero. Furthermore, λ_I remains the same (up to a sign change) when the elements of I are permuted. Thus, it is sufficient to only look at I where the numbers are increasing:

$$\{I \in \underline{n}^k : i_1 < i_2 < \dots < i_k\} := \underline{n}_a^k. \tag{7.30}$$

Note that $|\underline{n}_a^k| = \binom{n}{k}$. Therefore, we will rename $\underline{n}_a^k$ to $\underline{\binom{n}{k}}$.

Suppose we have an arbitrary $\omega \in \Lambda^k V$, and we feed it some list of vectors, but not in order, i.e. $\omega(u_{\sigma 1}, u_{\sigma 2}, \dots, u_{\sigma k})$ where $\sigma : \underline{k} \to \underline{k}$.

Lemma 14: If $T \in \Lambda^k$ and $\sigma \in S_k$, then $T \circ \sigma^* = (-1)^\sigma T$ where $\sigma^*(v_1,\dots,v_k) = (v_{\sigma_1},\dots,v_{\sigma_k})$. Therefore,

$$(T \circ \sigma^*)(v_1,\dots,v_k) = T(v_{\sigma_1},\dots,v_{\sigma_k}). \tag{7.31}$$

The proof is straightforward if we write out σ as a product of permutations.

We now wish to construct a basis for $\Lambda^k V$.

Definition: If $I \in \underline{n}_a^k$, then

$$\omega_I = \sum_{\sigma \in S_k} (-1)^\sigma \phi_I \circ \sigma^*. \tag{7.32}$$