

MAT257 Notes

Jad Elkhaleq Ghalayini

April 2019

This document is a collection of notes for the course MAT257: Analysis II, as taught by Professor Edward Bierstone in 2018 at the University of Toronto. The notes are a combination of notes I made in class (which can be found in their original form in the **notes** folder in this repository) and scans of handwritten notes which Professor Bierstone has generously given me the permission to use.

1 Introduction

1.1 Differentiability

Let's recall for a moment what it means for a function to be differentiable at a point:

$$\exists f'(a) = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} \quad (1)$$

Of course, we can bring everything to the right and write

$$\lim_{h \rightarrow 0} \frac{f(a+h) - (f(a) + f'(a)h)}{h} = 0 \quad (2)$$

The advantage of stating it this way is that we see that the linear function

$$f(a) + f'(a)h \quad (3)$$

is the *best linear approximation* to $f(a+h)$ at $h=0$. In fact, it's the unique linear function $y = b + mh$ such that

$$\lim_{h \rightarrow 0} \frac{f(a+h) - (b + mh)}{h} = 0 \quad (4)$$

Should this be called linear? I don't know. Maybe it would be better to call it affine.

From the point of view of linear algebra, we would say that a function which takes

$$h \mapsto f'(a) \times h \quad (5)$$

is a linear transformation. So the derivative can be thought of as a number, but it can also be thought of as a linear transformation from the real line to itself. And this is how we're going to define, and work with, derivatives in higher dimensions.

1.2 Functions of several variables

We're not going to be working on the real line, instead, we'll be working on \mathbb{R}^n : the space of n -tuples of real numbers. Instead of functions $f(x)$ of a single real variable x , we will hence be working with functions of n real variables, or if you like, functions of a single vector variable, where $x \in \mathbb{R}^n$, of the form

$$f(x) = f(x_1, \dots, x_n) \quad (6)$$

Such a function is differentiable at a point $a = (a_1, \dots, a_n)$ if "there's a thing that plays the role of the derivative." We could think about a derivative in \mathbb{R} as a real number, but we could also think about it as a

linear transformation, which is what we're going to do. That is, a derivative exists, if there exists a linear transformation

$$\lambda : \mathbb{R}^n \rightarrow \mathbb{R} \quad (7)$$

such that

$$\lim_{h \rightarrow 0} \frac{f(a+h) - (f(a) + \lambda(h))}{|h|} = 0 \quad (8)$$

What's a linear transformation from $\mathbb{R}^n \rightarrow \mathbb{R}$. Well first of all, what's h ? $h = (h_1, \dots, h_n)$ is a vector. So λ can be thought of as a sort of row matrix such that

$$\lambda(h) = (\lambda_1 \quad \dots \quad \lambda_n) \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \quad (9)$$

Sometimes, we'll write $\lambda(h)$ simply as λh . This right away is why linear algebra is important. As you can imagine, if we want to analyze the effect of a given derivative on this function, we're going to need to analyze this matrix, so linear algebra is going to be really crucial. In fact, a lot of the things that we do in the course can be done and were done classically with a really minimal amount of linear algebra. In fact this λ is a matrix, but its entries are what we call the partial derivatives of f , so we can work only with a collection of partial derivatives. However, what we'll see is, from a conceptual perspective, there's a lot to be gained from the linear algebraic approach.

Analysis involves making approximations and estimations, like $\epsilon - \delta$ proofs in calculus. Of course, $\epsilon - \delta$ proofs involve measuring the size of a quantity or distance, and those are going to be really essential concepts in this course. That already shows up in the $|h|$ in the equation above, which is defined

$$|h| = \sqrt{h_1^2 + \dots + h_n^2} \quad (10)$$

the norm of h . It's a measure of the size of h . If we write the norm

$$|x - y| \quad (11)$$

this is a measure of the size of the difference between x and y , or the distance between x and y . But especially in high dimension, even in one dimension but especially in high dimensions, there may be many different ways of measuring distance which are essentially equivalent.

1.3 Equivalent Norms

This is one norm, but in fact there are many different but equivalent norms. For example, some norms include

$$|x| = \left(\sum_{i=1}^n x_i^2 \right)^{1/2} \quad (12)$$

$$||x|| = \max\{|x_1|, \dots, |x_n|\} \quad (13)$$

$$|||x||| = \sum_{i=1}^n |x_i| \quad (14)$$

These are all equivalent, which means that if you pick any two of them, the first one is less than or equal to a constant times the second one. Why is that? What's the relationship between them?

It's a very good idea, in, anything, I guess, to understand things geometrically. And this actually can be seen just on the level of a picture of \mathbb{R}^2 . After all, what if we take our first norm. What does it mean to say that the norm is less than or equal to r ? It means that our point is inside a circle of radius r . On the other hand, what does it mean for the *second* norm to be less than or equal to r ? It means that both norms have to be inside the square of side length r centered at the origin. And finally, what does it mean for the third

norm to be less than or equal to r ? That means we're inside the smaller square, or diamond, formed by the points

$$(0, r), (0, -r), (r, 0), (-r, 0) \quad (15)$$

Of course, there would be n -dimensional versions of this picture. But what does this picture say about the relationship between these norms? It tells us that

$$||x|| \leq |x| \leq |||x||| \quad (16)$$

How do we prove this analytically? We just take the square of everything!

Now how do we show that

$$|||x||| \leq a|x| \quad (17)$$

for some constant a ? We can also think about this geometrically: take our circle of radius r . We need the first diamond which is outside of the circle. And what distance is the corner of the diamond from the origin? In the plane, it's $\sqrt{2}r$. In higher dimensions, it's $\sqrt{n}r$. So what this tells us is that

$$|||x||| \leq \sqrt{n}|x| \quad (18)$$

or, analytically,

$$|||x||| = \langle x, u(x) \rangle \leq |x||u(x)| \quad (19)$$

where

$$u_i(x) = \text{sgn}(x_i) = \begin{cases} 1 & \text{if } x_i \geq 0 \\ -1 & \text{if } x_i < 0 \end{cases} \quad (20)$$

Now, how do we show that

$$|x| \leq a||x|| \quad (21)$$

Well, let's again look at our square of side r . Clearly, the circle circumscribing this square has radius $\sqrt{n}r$, in n dimensions. We should finish this by doing the exercise of proving this analytically

You all know that \mathbb{R}^n is the space of n -tuples of real numbers,

$$x = (x_1, \dots, x_n) \quad (22)$$

\mathbb{R}^n , of course, is a vector space with the operations

$$x + y = (x_1 + y_1, \dots, x_n + y_n) \quad (23)$$

$$ax = (ax_1, \dots, ax_n) \quad (24)$$

These are the vector space operations on \mathbb{R}^n that everyone should be familiar with from linear algebra. We, however, are going to be even more interested in the *metric* properties of such a metric space, and these are going to be dependent on the idea of norm that we talked about last time.

We defined a number of norms, including the norm

$$|x| = \sqrt{x_1^2 + \dots + x_n^2} \quad (25)$$

Of course, this norm is intimately connected with the usual inner product on \mathbb{R}^n ,

$$\langle x, y \rangle = x_1y_1 + \dots + x_ny_n \quad (26)$$

What are the basic relations between this and the vector space structure? These should be things you're familiar with, but let's recall.

Let's consider two vectors $x, y \in \mathbb{R}^n$. We know that

1. $|x| \geq 0$ and $|x| = 0 \iff x = 0$
2. $|\langle x, y \rangle| \leq |x||y|$, with equality iff x, y are linearly dependent.
3. $|x + y| \leq |x| + |y|$, with equality iff one is zero or if they point in the same direction.

Proof. 2. We have equality if x, y are linearly dependent. But what if x, y are linearly independent. Then, by definition

$$\forall \lambda \in \mathbb{R}, y - \lambda x \neq 0 \quad (27)$$

$$\implies 0 < |y - \lambda x|^2 = \langle y, y \rangle - 2\lambda \langle x, y \rangle + \lambda^2 \langle x, x \rangle \quad (28)$$

$$= |y|^2 - 2\lambda \langle x, y \rangle + \lambda^2 |x|^2 \quad (29)$$

This is a quadratic polynomial in λ which is never 0, so hence has a discriminant

$$4 \langle x, y \rangle^2 - 4|x|^2|y|^2 < 0 \implies \langle x, y \rangle^2 < |x|^2|y|^2 \quad (30)$$

$$\implies |\langle x, y \rangle| < |x||y| \quad (31)$$

as desired.

$$3. |x + y|^2 = \langle x + y, x + y \rangle = |x|^2 + |y|^2 + 2\langle x, y \rangle \leq |x|^2 + |y|^2 + 2|x||y| = (|x| + |y|)^2.$$

□

What if we have two vectors x and y , and we want to define the angle between them. How should we do that? We can try

Definition 1.

$$\angle(x, y) = \arccos \frac{\langle x, y \rangle}{|x||y|} \quad (32)$$

Why should this, however, be the definition? It comes from the elementary geometric relationship between the length of $y - x$ and the lengths of x, y when formed into a triangle:

$$|y - x|^2 = |x|^2 + |y|^2 - 2|x||y| \cos \theta \quad (33)$$

which can be rewritten to the familiar

$$\langle x, y \rangle = |x||y| \cos \theta \quad (34)$$

1.4 Linear transformations from $\mathbb{R}^n \rightarrow \mathbb{R}^m$

Usually, we represent linear transformations by a matrix. But the matrix, of course, corresponds to a choice of basis vectors. Let's recall what we usually denote the standard basis of \mathbb{R}^n

1.4.1 Standard basis of \mathbb{R}^n

The standard basis of \mathbb{R}^n is composed of e_1, \dots, e_n , where each e_k is 0 everywhere except in the k^{th} place, where it is 1.

Let's assume $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear transformation. What do we mean by the matrix of T with respect to the standard basis? Let

$$e_1, \dots, e_n \in \mathbb{R}^n \quad (35)$$

$$f_1, \dots, f_m \in \mathbb{R}^m \quad (36)$$

be the standard bases.

Our transformation is a matrix A that we multiply a certain vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ by to get a vector $y = (y_1, \dots, y_m) \in \mathbb{R}^m$. Let's label the elements of A a_{11}, \dots, a_{mn} , getting

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad (37)$$

But note we can write

$$x = \sum_{i=1}^n x_i e_i \quad (38)$$

And note that

$$T(e_i) = \sum_{j=1}^m a_{ji} f_j \quad (39)$$

i.e. the i^{th} column.

So what we're really doing is taking one sum of basis vectors into another.

Now what if we have a separate linear transformation? If $S : \mathbb{R}^m \rightarrow \mathbb{R}^\ell$ is another linear transformation, that means we can compose them: we can apply S to $T(x)$. Then

$$S \circ T : \mathbb{R}^n \rightarrow \mathbb{R}^\ell \quad (40)$$

Let's suppose S has a matrix B . And what's the matrix of $S \circ T$? It is, of course, BA .

Ok, finally, what about *bounding* the size of a matrix applied to a vector? This is going to be a basic bound that we use all the time.

If $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear transformation, then there is a constant $M \geq 0$ such that

$$|Tx| \leq M|x| \quad (41)$$

Why?

$$T(x) = T\left(\sum x_i e_i\right) = \sum x_i T(e_i) \quad (42)$$

Let's estimate the norm:

$$|T(x)| \leq \sum_{i=1}^n |x_i T(e_i)| = \sum_{i=1}^n |x_i| |T(e_i)| \quad (43)$$

So how can we use this to get that M ? We can just take

$$C = \max_i |T(e_i)| \quad (44)$$

then

$$|T(x)| \leq C \sum |x_i| \quad (45)$$

This is not quite what we wanted, because this is not that single barred norm of x . This was a different norm of x . But what was the relationship between the two norms?

$$\sum |x_i| \leq \sqrt{n}|x| \quad (46)$$

So we have

$$|T(x)| \leq \sqrt{n}C|x| \quad (47)$$

\mathbb{R}^n with any of our norms is what's called a *metric space*. All a metric space means is a set X together with a distance function

$$d : X \times X \rightarrow \mathbb{R}, (x, y) \mapsto d(x, y) \quad (48)$$

such that

- $\forall x, y \in X, d(x, y) \geq 0$
- $\forall x, y, z \in X, d(x, z) \leq d(x, y) + d(y, z)$
- $d(x, y) = 0 \iff x = y$

In this course, we're going to talk a little bit about the topology of \mathbb{R}^n , that is, talking about what are called open and closed sets, what the relationship between them is, and how this notion relates to continuity.

In fact, we're only going to be talking about a very special class of topological spaces: metric spaces. In fact, they're even a very special class of metric spaces: \mathbb{R}^n or certain subsets of \mathbb{R}^n with the Euclidean metric.

That's what we're going to be talking about tonight (note: I cannot attend, so no notes are available for this lecture): the topology of \mathbb{R}^n .

TODO: this

1.5 Topology of \mathbb{R}^n

TODO: this

1.5.1 Continuity

Definition 2. Consider a function

$$f : A \rightarrow \mathbb{R}^n \quad (49)$$

f is continuous at A if for every $\epsilon > 0$, $\exists \delta > 0$ such that $|f(x) - f(a)| < \epsilon$ whenever $|x - a| < \delta, x \in A$.

Theorem 1. f is continuous if and only if for every open $U \subset \mathbb{R}^n$, $f^{-1}(U) = A \cap V$ where V is open in \mathbb{R}^n .

Proof. \implies consider U open in \mathbb{R}^n . Let $a \in f^{-1}(U)$. For some $\epsilon > 0$, \exists a ball $B_{f(a)} = B(f(a), \epsilon) \subset U$ because U is open.

Since f is continuous at a , $\exists \delta > 0$ such that $f(x) \in B_{f(a)}$ for every $x \in A \cap B(a, \delta) = B_a$. We can define

$$V = \bigcup_{a \in f^{-1}(U)} B_a \quad (50)$$

\Leftarrow Consider $a \in A$ and let $\epsilon > 0$. Let $U = B(f(a), \epsilon)$. Then $f^{-1}(U) = A \cap V$ for some open $V \subset \mathbb{R}^n$.

$$a \in V \implies \exists \delta > 0, B(a, \delta) \subset V \quad (51)$$

But everything which lies inside V gets mapped inside U which is that first ball that we started with. And this is the condition that we want. □

Corollary. A composite of continuous functions is continuous

Proof. Let

$$f : A \rightarrow \mathbb{R}^n, g : B \rightarrow \mathbb{R}^p \quad (52)$$

be continuous functions where $B \subset \mathbb{R}^n$ and $f(A) \subset B$.

Consider open $U \subset \mathbb{R}^p$. We know

$$\exists V \subset \mathbb{R}^n \text{ open, } g^{-1}(U) = B \cap V \quad (53)$$

So we can write

$$(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U)) = f^{-1}(B \cap V) = f^{-1}(V) \quad (54)$$

We know that

$$\exists W \subset \mathbb{R}^n \text{ open, } f^{-1}(V) = A \cap W = (g \circ f)^{-1}(U) \quad (55)$$

But this is exactly what we wanted to show □

Exercise 1. Let $f = (f_1, \dots, f_n)$ be a function from \mathbb{R} to \mathbb{R}^n . Then f is continuous if and only if $\forall i \in \{1, \dots, n\}$, f_i is continuous.

Proof. \Leftarrow We need to estimate the norm $|f(x) - f(a)|$ with the differences between $f_i(x_i)$ and $f_i(a_i)$. We could use a variety of inequalities for this, including

$$|f(x) - f(a)| \leq \sum_{i=1}^n |f_i(x) - f_i(a)| \quad (56)$$

The full ϵ - δ argument is left as an exercise.

\implies We need to estimate the norm $|f_i(x) - f_i(a)| < |f(x) - f(a)|$. Alternatively, we can do this topologically. The full argument is left as an exercise. □

Exercise 2. A linear transformation $T : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is uniformly continuous

Definition 3. There is a norm $M > 0$ such that $|T(x)| \leq M|x|$. Hence,

$$|T(x) - T(y)| = |T(x - y)| \leq M|x - y| \quad (57)$$

Given ϵ , take $\delta = \frac{\epsilon}{M}$.

Exercise 3. Are the following functions continuous?

1. $f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$ - No

2. $f(x, y) = \frac{x^2 + 3xy + y^2}{x^2 + 4xy + y^2}$ - No

3. $f(x, y) = e^{-\frac{|x-y|}{x^2 - 2xy + y^2}} = e^{-\frac{1}{|x-y|}}$ - Yes

Proof. 1. f is 1 along the line $\{f(x, 0) : x \in \mathbb{R}\}$, but -1 along the line $\{f(0, y) : y \in \mathbb{R}\}$, both of which traverse the origin.

2. No for the same reason, but we can't check on the axes, since each axis is 1. Instead, we can check on the x axis and compare that with any line except the y axis, such as $y = x$, where the value is $\frac{5}{6}$

3. Composite of $z \mapsto e^{\frac{1}{|z|}}$ and $(x, y) \mapsto x - y$

□

Exercise 4. Let $X \subseteq \mathbb{R}^n$. We define the distance function

$$d(x, X) = \inf_{a \in X} |x - a| \quad (58)$$

$\forall X \in \mathcal{P}(\mathbb{R}^n)$, $d(x, X)$ is uniformly continuous on \mathbb{R}^n

1.5.2 Compactness

Definition 4. A subset X of \mathbb{R}^n is compact if every open covering of X has a finite subcover.

We're later going to prove a deep theorem about compactness:

Theorem 2. A subset X of \mathbb{R}^n is compact if and only if X is closed and bounded

We won't do this right now, because it'll involve some work, but we'll use something called the Heine-Borel Theorem: $[0, 1]$ is compact. We're going to prove this exactly the same way as we proved the "Three Hard Theorems" from first year calculus. It's a good exercise to try this.

Today, we'll do half of this theorem, the easy part: compact *implies* closed and bounded.

Lemma 1. If $X \subset \mathbb{R}^n$ is compact then X is closed and bounded.

Proof. • X is closed: to say that X is closed is of course the same thing as saying $\mathbb{R}^n \setminus X$ is open. To show this, we take any point in $\mathbb{R}^n \setminus X$, and show there's some ball centered at that point which is a subset of $\mathbb{R}^n \setminus X$.

So let $x \in \mathbb{R}^n \setminus X$. We want to show that $\exists \delta > 0$, $B(a, \delta) \subset \mathbb{R}^n \setminus X$ for some $\delta > 0$. Let's just look at all possible balls centered at A . Or we might as well just consider, for some $k \in \mathbb{N}$, the closed ball $\overline{B(a, k^{-1})}$. Let's look at the complement of this closed ball,

$$U_k = \mathbb{R}^n \setminus \overline{B(a, k^{-1})} \quad (59)$$

We have that

$$\bigcup_k U_k = \mathbb{R}^n \setminus \{a\} \quad (60)$$

So $\{U_k\}$ is an open cover of X . But X is compact, so this open cover has a finite subcover. This fact tells us that $\exists k, X \subset U_k$. But U_k is the complement of the closed ball of radius $\frac{1}{k}$, meaning the open ball

$$B(a, k^{-1}) \subseteq \mathbb{R}^n \setminus X \quad (61)$$

We can clearly see that this implies $\mathbb{R}^n \setminus X$ is open, implying X is closed.

- X is bounded: consider a cover of X by all open balls of radius 1 in \mathbb{R}^n .

□

Theorem 3. *If $X \subset \mathbb{R}^m$ is compact and $f : X \rightarrow \mathbb{R}^n$ is continuous then $f(X)$ is compact*

Proof. Let \mathcal{O} be an open cover of $f(X)$. For every $U \in \mathcal{O}$, $f^{-1}(U) = X \cap V_U$ where V_U is open in \mathbb{R}^m . So

$$\{V_U : U \in \mathcal{O}\} \quad (62)$$

is an open cover of X . Since X is compact, there is a finite subcover

$$V_{U_1}, \dots, V_{U_k} \quad (63)$$

So U_1, \dots, U_k cover $f(X)$.

□

Theorem 4 (Extreme Value Theorem). *A continuous function $f : X \rightarrow \mathbb{R}$ on a nonempty compact subset X of \mathbb{R}^n attains a maximum and minimum*

Proof. Let $M = \sup\{f(x) : x \in X\}$. $M < \infty$ since $f(X)$ is bounded. What if $M \notin f(X)$? Since $f(X)$ is closed, we can draw a small interval around M which is not in $f(X)$, contradicting the fact that $M = \sup f(X)$.

The proof for minima is analogous.

□

Definition 5. *The ϵ -neighborhood of X is given by*

$$\bigcup_{x \in X} B(x, \epsilon) = \{y \in \mathbb{R}^n : d(y, X) < \epsilon\} \quad (64)$$

Theorem 5 (ϵ -neighborhood theorem). *If $X \subset U \subset \mathbb{R}^n$ where X is compact and U is open, then there is $\epsilon > 0$ such that the ϵ -neighborhood of X lies in U .*

Proof. Define $f : X \rightarrow \mathbb{R}$ by $f(x) = d(x, \mathbb{R}^n \setminus U)$. We showed this is a continuous function, which is strictly positive (since U is open). Hence, by the Extreme Value Theorem, it has a positive minimum. Take ϵ to be this minimum.

□

2 Differentiation

2.1 Definitions

TODO: copy beginning

Examples:

1. Let

$$f(x, y) = \int_a^{x+y} g \quad (65)$$

where $g : \mathbb{R} \rightarrow \mathbb{R}$ is continuous. We compute $Df(c, d)$ as follows: We have that $f = q \circ s$, where

$$q(t) = \int_a^t g, s(x, y) = x + y \quad (66)$$

We have

$$Df(c, d) = g'(s(c, d))s'(c, d) = g(c + d)(c + d) \quad (67)$$

since $q'(t) = g(t)$.

2. Let

$$f(x, y) = \int_a^{x^y} g \quad (68)$$

We have

$$f = g \circ \xi, \xi(x, y) = x^y = e^{y \log x} \quad (69)$$

Hence,

$$D\xi(x, y) = x^y D(y \log x) = x^y ((0, 1) \log x + (1/x, 0)y) = x^y (y/x, \log x) \quad (70)$$

$$\implies Df(c, d) = q'(\xi(c, d))\xi'(c, d) = g(c^d)c^d(d/c, \log c) \quad (71)$$

2.1.1 Higher-order Derivatives

Let $f : U \rightarrow \mathbb{R}$ be a function where $U \subseteq \mathbb{R}^m$ and suppose

$$D_i f = \frac{\partial f}{\partial x_i} : U \rightarrow \mathbb{R} \quad (72)$$

exists for all i . So we could now consider

$$D_j(D_i f) = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) \quad (73)$$

We'll write $D_{ij}f$ to denote the above. Notice i is applied *first*. We may also write $f_{x_i x_j}$ and $\frac{\partial^2 f}{\partial x_j \partial x_i}$. We have that

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \quad (74)$$

if both mixed partials exist and are continuous in a neighborhood of a (proof uses f).

In general, we can consider taking higher order partials as well, as in

$$\frac{\partial^{\alpha_1 + \dots + \alpha_n} f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}} \quad (75)$$

Of course we have to worry about the order, but the order is irrelevant if f is \mathcal{C}^∞ , i.e. that all partial derivatives of all orders exist (and are hence continuous).

2.1.2 Multi-index notation

In multi-index notation, $\alpha = (\alpha_1, \dots, \alpha_m)$ is a vector of non-negative integers. We define the *total order* of α

$$|\alpha| = \alpha_1 + \dots + \alpha_m \quad (76)$$

Furthermore, we define

$$x = (x_1, \dots, x_m) \implies x^\alpha = x_1^{\alpha_1} \dots x_m^{\alpha_m} \quad (77)$$

Once we look at Taylor's theorem, the notation

$$x! = x_1! \dots x_m! \quad (78)$$

will also come in handy.

We can now perform another example:

3. Let

$$f(x, y) = \begin{cases} xy \frac{x^2 - y^2}{x^2 + y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases} \quad (79)$$

Is this function differentiable at the origin? Yes: $f'(0, 0) = (0, 0)$. Let's check: we need to show that

$$\lim_{(x, y) \rightarrow (0, 0)} \frac{f(x, y) - f(0, 0) - (0, 0) \cdot \begin{pmatrix} x \\ y \end{pmatrix}}{\sqrt{x^2 + y^2}} = \lim_{(x, y) \rightarrow (0, 0)} xy \frac{x^2 - y^2}{x^2 + y^2} \frac{1}{\sqrt{x^2 + y^2}} = 0 \quad (80)$$

We have that

$$\left| xy \frac{x^2 - y^2}{x^2 + y^2} \frac{1}{\sqrt{x^2 + y^2}} \right| \leq \frac{|xy|}{\sqrt{x^2 + y^2}} \rightarrow 0 \quad (81)$$

2.2 Partial Derivatives

Theorem 6. If $U \subset \mathbb{R}^m$, $f : U \rightarrow \mathbb{R}^n$ is differentiable at a and $f = (f_1, \dots, f_n)$ then each partial derivative $\frac{\partial f_i}{\partial x_j}(a)$ exists, and

$$f'(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \dots & \frac{\partial f_1}{\partial x_m}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(a) & \dots & \frac{\partial f_n}{\partial x_m}(a) \end{pmatrix} \quad (82)$$

with

$$\frac{\partial f}{\partial x_j}(a) = \frac{\partial}{\partial x} f(a_1, \dots, a_{j-1}, x, a_{j+1}, \dots, a_m) \Big|_{x=a_j} \quad (83)$$

Proof. • Case $n = 1$: Let

$$h(x) = (a_1, \dots, a_{j-1}, x, a_{j+1}, \dots, a_m) : \mathbb{R} \rightarrow \mathbb{R}^m \quad (84)$$

Then

$$\frac{\partial f}{\partial x_j}(a) = D(f \circ h)(a_j) = Df(h(a_j)) \circ Dh(a_j) = f'(a) \cdot \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad (85)$$

where the matrix is $1 \times m$ and 1 in the j^{th} place, i.e. the j^{th} entry in $f'(a)$.

• For n in general,

$$f = (f_1, \dots, f_n), f'(a) = \begin{pmatrix} f'_1(a) \\ \vdots \\ f'_n(a) \end{pmatrix} \quad (86)$$

Hence, we have a row for each f_i , and we can apply the case for $n = 1$ to each row. □

We're going to do some calculations right away with this, since it's very important that this is all straight in your head. One thing we want to look at is the converse of this theorem, which turns out not to be true unless we make additional hypotheses.

But let's first do some computations, and we'll see that this notion of partial derivatives lets us rethink what the chain rule says, which is very important when you want to use the chain rule.

2.2.1 The Chain Rule

Suppose we have

$$F(x) = f(g_1(x), \dots, g_n(x)), x = (x_1, \dots, x_m), g = (g_1, \dots, g_n) \quad (87)$$

with g differentiable at a and f differentiable at $g(a)$. We want to compute $\frac{\partial F}{\partial x_i}(a)$. Let's say that $f = f(y_1, \dots, y_n)$. Then what's the formula?

$$\frac{\partial F}{\partial x_i}(a) = \sum_{j=1}^n \frac{\partial f}{\partial y_j}(g(a)) \frac{\partial g_j}{\partial x_i}(a) \quad (88)$$

Why is this? Well, what does the chain rule say?

$$DF(a) = Df(g(a))Dg(a) = \begin{pmatrix} \frac{\partial f}{\partial y_1}(g(a)) & \dots & \frac{\partial f}{\partial y_n}(g(a)) \end{pmatrix} \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(a) & \dots & \frac{\partial g_1}{\partial x_m}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1}(a) & \dots & \frac{\partial g_n}{\partial x_m}(a) \end{pmatrix} \quad (89)$$

So the result of this is a row matrix, and by taking $\frac{\partial F}{\partial x_i}$ is just indexing into it. You should really know this, you should even memorize it. When you come out of calculus, this should be as natural as addition when you come out of elementary school.

These things, sometimes people write them in a kind of symbolic, simpler way, which is useful. Here we have

$$y_j = g_j(x_1, \dots, x_m) \quad (90)$$

and then we're composing to get

$$z = f(y_1, \dots, y_n) = f(g(x)) \quad (91)$$

So if you want to take the derivative with respect to x_i , you could write that as

$$\frac{\partial z}{\partial x_i} \quad (92)$$

But this is like short form, since z is not a function of x , it's a function of y . So when we write $\frac{\partial z}{\partial x_i}$, we're thinking of it as already composite. And the chain rule thinks of this as

$$\frac{\partial z}{\partial x_i} = \sum_{j=1}^n \frac{\partial z}{\partial y_j} \frac{\partial y_j}{\partial x_i} \quad (93)$$

with evaluation, this becomes

$$\frac{\partial z}{\partial x_i}(a) = \sum_{j=1}^n \frac{\partial z}{\partial y_j} \Big|_{y=g(a)} \frac{\partial y_j}{\partial x_i} \Big|_{x=a} \quad (94)$$

Now how does this notation sometimes show up in a confusing way? Some examples!

1. Suppose $F(t) = f(x(t), y(t), z(t), t)$. Let's call $F(t) = w$. Let's take the derivative with respect to t :

$$\frac{dw}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial z} \frac{dz}{dt} + ? \quad (95)$$

What do we write in the fourth place? We want the derivative of f with respect to the fourth variable, so we just write $\frac{\partial w}{\partial t}$. So we have

$$\frac{dw}{dt} \neq \frac{\partial w}{\partial t} \quad (96)$$

See how this is confusing?

This is *shorthand*, so there can be ambiguity. That ambiguity is somewhat resolved by the different ordinary versus partial derivative notation, but it's still confusing. That's why you've got to understand the full, genuine meaning behind the shorthand notation.

2. Let $w = f(x(s, t), y(s, t), s, t)$. Now it gets even worse...

2.3 Basic Results

TODO: this

2.4 The Inverse Function Theorem

In this course, we have three important theorems, of which this is one. Each is a generalization of something you've seen in first year calculus, but we'll see that these generalizations are very far-reaching and involve new techniques.

Recall: suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is a continuously differentiable function in an open interval containing a , and $f'(a) \neq 0$. Then f' is either greater than or less than zero in an open interval containing a . Therefore f is one-to-one, and so has an inverse defined on an open interval W containing $f(a)$. Moreover, f^{-1} is differentiable, and

$$(f^{-1})'(f(a)) = \frac{1}{f'(a)} \quad (97)$$

We're going to discuss a generalization of this idea to several variables.

Theorem 7 (Inverse Function Theorem). *Let $f : U \rightarrow \mathbb{R}^n$ be a continuously differentiable (\mathcal{C}^1) on an open set $U \subset \mathbb{R}^n$. Let $a \in U$ be such that $\det f'(a) \neq 0$. Then there exist open sets $a \in V, f(a) \in W$ such that $f : V \rightarrow W$ with a continuous inverse $f^{-1} : W \rightarrow V$. Moreover, f^{-1} is differentiable on W and*

$$\forall y \in W, (f^{-1})'(y) = (f'(f^{-1}(y)))^{-1} \quad (98)$$

Remark: if we know already that f^{-1} is differentiable, the formula for it follows from the chain rule:

$$f(f^{-1}(y)) = y \implies f'(f^{-1}(y))(f^{-1})'(y) = I \iff (f^{-1})'(y) = (f'(f^{-1}(y)))^{-1} \quad (99)$$

Corollary. 1. f^{-1} is continuously differentiable (\mathcal{C}^1)

2. If f is \mathcal{C}^r then f^{-1} is also \mathcal{C}^r

Proof. 1. We know that since f' is continuous, and f^{-1} is continuous, $f' \circ f^{-1}$ is continuous. Now what about the inversion? What is the inverse of a matrix? It's given by a formula. So this is really a composite of 3 functions. So what about the formula for the inverse of a matrix? It follows from Cramer's rule that this formula is continuous, since the entries of the inverse matrix B of A are given as rational functions of the entries of A , that is,

$$b_{ij} = \frac{(-1)^{i+j} \det A^{ji}}{\det A} \quad (100)$$

where A^{ji} is the matrix obtained by deleting the j^{th} row and i^{th} column from A . So hence the derivative, being the composition of 3 continuous functions, must be continuous.

2. We proceed by induction on r . Assume that if f is \mathcal{C}^{r-1} , then f^{-1} is \mathcal{C}^{r-1} .

If f is \mathcal{C}^r , then f is \mathcal{C}^{r-1} implying that f^{-1} is \mathcal{C}^{r-1} by the inductive hypothesis. So, using the formula

$$(f^{-1})' = (f'(f^{-1}))^{-1} \quad (101)$$

is \mathcal{C}^{r-1} implying that f^{-1} is \mathcal{C}^r . □

This corollary is just to show that we could have stated the above theorem in a stronger way. Examples:

1. Continuity of f' cannot be removed from the hypotheses: consider $f : \mathbb{R} \rightarrow \mathbb{R}$,

$$f(x) = \begin{cases} x + x^2 \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases} \quad (102)$$

We have that

$$x \neq 0 \implies f'(x) = 1 + 2x \sin \frac{1}{x} - \cos \frac{1}{x} \quad (103)$$

but the limit does not exist as $x \rightarrow 0$.

2. Consider $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$,

$$f(x, y) = (e^x \cos y, e^x \sin y) \quad (104)$$

We have

$$f'(x, y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix} \quad (105)$$

implying that

$$\det f'(x, y) = (e^x)^2 = e^{2x} \neq 0 \forall x \in \mathbb{R} \quad (106)$$

But this function is not 1-1, since it is periodic in y . The inverse function theorem says that we can make some neighborhood around a point where f is one to one, but it *doesn't* say that it's *globally* one to one.

Remarks:

1. f may be invertible even though $f'(a) = 0$

2.5 The Implicit Function Theorem

Assume we have a function $F(x, y) = 0$. Can we solve for y as a function $y = g(x)$ near a point (a, b) such that $F(a, b) = 0$? For example, assume $f(x, y) = x^2 + y^2 - 1$. This relationship, of course, defines the unit circle. Can we solve for y as a function of x at any point on the unit circle? No: we can't do so where $y = 0$, we can only get x as a function of y . If we pick a point $(a, b) \in \mathbb{R}^2$ however, and assume $a \neq \pm 1$, then there are open intervals I, J such that for every $x \in I$, there is a unique $y = g(x) \in J$ such that $f(x, y) = 0$.

Let's restrict our attention to a small enough interval such that $b > 0$. Of course, then, we can explicitly write down what the solution is:

$$g(x) = \sqrt{1 - x^2}$$

Sometimes in this context people say that one of the variables is dependent and the other is independent. Which is which? How should you understand that language? Well, the dependent variable is the one that is determined by the independent one you specify. But that notion, it depends on where you are: at the poles $x = \pm 1$, we can only solve for x , whereas at the poles $y \pm 1$, we can only solve for y .

Furthermore, we may not be able to find a nice formula for g . However, by what's known as "implicit differentiation," we can find a formula for the *derivative of g* .

2.5.1 Implicit Differentiation

Let's assume we have a function

$$F(x, g(x)) = 0$$

Whether or not we have a formula for g , we can use the chain rule to obtain

$$\frac{d}{dx} F(x, g(x)) = \partial_1(x, g(x)) + \partial_2(x, g(x))g'(x) = 0$$

So we find that

$$g'(x) = -\frac{\partial_1(x, g(x))}{\partial_2(x, g(x))}$$

Now this formula depends on g . Is this weird or expected? Well, it makes sense. Consider the example above: we might have two solutions (points on the circle) on top of each other for a given x -coordinate. This formula works for *both* of them.

TODO: fill in here

2.5.2 The Inverse Function Theorem Implies the Implicit Function Theorem

We have to start with the hypotheses of the implicit function theorem: given a \mathcal{C}^r (where $r \geq 1$) function $f : U \rightarrow \mathbb{R}^n$ where $U \in \mathbb{R}^{m+n}$, $f(a, b) = 0$ and $\det M \neq 0$, where

$$M = \left(\frac{\partial f_i}{\partial y_j}(a, b) \right)$$

Define

$$F : U \rightarrow \mathbb{R}^m \times \mathbb{R}^n, (x, y) \mapsto (x, f(x, y))$$

In particular,

$$F(a, b) = (a, 0)$$

This is what we're going to apply the inverse function theorem to. So we've got to show that this function satisfies the hypotheses of the inverse function theorem. So we've got to show that its derivative matrix at the point (a, b) is invertible. So let's compute: the derivative is given by

$$F'(a, b) = \left(\begin{array}{c|c} I & 0 \\ \hline * & M \end{array} \right) \implies \det F'(a, b) = \det I \det M = \det M \neq 0$$

These are the conditions under which we can apply the inverse function theorem. So by the inverse function theorem, there exists an open neighborhood V of (a, b) , and an open neighborhood W of $(0, 0)$ so that

$F : V \rightarrow W$ has a \mathcal{C}^r inverse $F^{-1} : W \rightarrow V$. We can assume $V = A \times B$, where A, B are open neighborhoods of a, b . $F^{-1}(u, v)$ has the form $(u, h(u, v))$. So

$$\begin{aligned} F(F^{-1}(u, v)) &= F(u, h(u, v)) = (u, f(u, h(u, v))) = (u, v) \implies f(u, h(u, v)) = v \\ &\implies f(u, h(u, 0)) = 0 \end{aligned}$$

Let $g(x) = h(x, 0)$, which is \mathcal{C}^r . Then

$$f(x, g(x)) = 0$$

Remark: we can find $g'(x)$ by implicit differentiation. We have

$$\forall i \in \{1, \dots, n\}, f_i(x, g(x)) = 0$$

We can write

$$\frac{\partial f_i}{\partial x_j}(x, g(x)) + \sum_{k=1}^n \frac{\partial f_i}{\partial y_k}(x, g(x)) \frac{\partial g_k}{\partial x_j}(x) = 0$$

We can solve for $\frac{\partial g_k}{\partial x_j}$ because $\left(\frac{\partial f_i}{\partial y_k}(x, y)\right)$ is invertible near (a, b) .

2.5.3 The Implicit Function Theorem Implies the Inverse Function Theorem

This time, we start with the hypotheses of the *inverse* function theorem. So here we have a \mathcal{C}^r function $f : U \rightarrow \mathbb{R}^n$ with $\det f'(a) \neq 0$.

Let $b = f(a)$, and define

$$F(x, y) = y - f(x)$$

This is a \mathcal{C}^r function of (a, b) and $F(a, b) = 0$. We have

$$\frac{\partial F}{\partial x}(a, b) = \det(-f'(a)) \neq 0$$

By the implicit function theorem there exist open neighborhoods A, B of a, b respectively such that for all $y \in A$, there is a unique \mathcal{C}^r $x = g(y)$ in B such that

$$F(g(y), y) = 0 \iff y - f(g(y)) = 0$$

Take $V = f^{-1}(A) \cap B$, $W = A$. Then

$$x \in V \implies f(x) \in A$$

and x is the unique element of B such that $g(f(x)) = 0$, i.e. $x = g(f(x))$.

2.5.4 Proof of the Inverse Function Theorem

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be \mathcal{C}^1 on an open set containing $a \in \mathbb{R}^n$ and let

$$\det f'(a) \neq 0$$

We can assume, as shown in last lecture, that $f'(a) = I$. We show that there are open V, W containing a and $f(a)$ respectively such that $f : V \rightarrow W$ with a continuous inverse $f' : W \rightarrow V$

We will begin by showing that we can't have $f(x) = f(a)$ if $x \neq a$ is sufficiently close to a :

$$f'(a) = I \wedge Ih = h \implies \lim_{h \rightarrow 0} \frac{|f(a+h) - f(a) - h|}{|h|} = 0$$

If $f(a+h) = f(a)$ then the quotient is 1. So the quotient being zero means that there exists a closed ball B centered at a such that

$$x \in B \setminus \{a\} \implies f(x) \neq f(a)$$

But that's not enough to mean that f is one to one in B : it could be, for example, that $f(\partial B)$, the image of the boundary of B , intersects itself, for example. What we want to do is find an even smaller ball where f really is one-to-one. So, since f is \mathcal{C}^1 in a neighborhood of a we can also assume:

1. $x \in B \implies \det f'(x) \neq 0$ since f' is continuous

2. Given ϵ , we can make, for all i, j ,

$$\left| \frac{\partial f_i}{\partial x_j}(x) - \frac{\partial f_i}{\partial x_j}(a) \right| < \epsilon$$

Note that, by our previous assumption that $f'(a) = I$,

$$\frac{\partial f_i}{\partial x_j}(a) = I_{ij}$$

Apply the Mean Value Theorem to $g(x) = f(x) - x$. According to the MVT, we can also assume that

$$|g(x_1) - g(x_2)| \leq cb|x_1 - x_2|$$

where c is a constant and b is a bound on the partial derivatives of g . But what is $\frac{\partial g_i}{\partial x_j}$? It's the same as what's in the absolute value brackets in assumption (2), since

$$\frac{\partial g_i}{\partial x_j} = \delta_{ij} = \frac{\partial f_i}{\partial x_j}$$

So we have that

$$|g(x_1) - g(x_2)| \leq c\epsilon|x_1 - x_2|$$

for some constant c . That was for any given ϵ . Let's choose ϵ such that $c\epsilon = \frac{1}{2}$, giving

$$|g(x_1) - g(x_2)| \leq \frac{1}{2}|x_1 - x_2|$$

If you remember that problem we did on the MVT, we gave a precise bound on ϵ , specifically it suffices to choose $\epsilon = \frac{1}{2n}$. So what does this tell us?

$$|f(x_1) - x_1 - f(x_2) + x_2| \leq \frac{1}{2}|x_1 - x_2|$$

using the Triangle Inequality, we obtain

$$|x_1 - x_2| \leq 2|f(x_1) - f(x_2)|$$

So what does *this* tell us? It tells us that f is one-to-one on B , since if $f(x_1) = f(x_2)$ and $x_1 \neq x_2$ then

$$|x_1 - x_2| > 2|f(x_1) - f(x_2)| = 0$$

We have that $f(\partial B)$ is a compact set not including a . So we have that

$$d = d(f(a), f(\partial B)) > 0$$

Let

$$W = \{y : |y - f(a)| \leq d/2\}$$

Then if $y \in W$, $x \in \partial B$, then what's the relationship between $|y - f(a)|$ and $|y - f(x)|$? Since W is the ball of radius $\frac{d}{2}$, we have that

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

This tells us that W is completely in the image of B . It follows that for all $y \in W$, there is some $x \in \text{Int } B$ such that $f(x) = y$. I could have said unique x , but I didn't bother because it must be unique anyways since we already checked one to one. But let's check this. We're going to do so by solving an extreme value problem. Let's define $h : B \rightarrow \mathbb{R}$ by

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

h is continuous, and therefore it has a minimum on the compact set B . So this is the usual business of a max-min problem: it could be that the minimum occurs on the boundary. So can it? No, since, as we saw above,

$$\forall y \in W, |y - f(a)| < |y - f(x)|$$

So that means it occurs at the interior, and since f and hence $|y - f(x)|^2$ is differentiable, at a critical point, a point where all the partial derivatives are zero, $x \in \text{Int } B$. But what does the following exactly mean:

$$\frac{\partial h}{\partial x_j}(x) = 0$$

Expanding, we get

$$\forall j \in \{1, \dots, n\}, \sum_{i=1}^n 2(y_i - f_i(x)) \frac{\partial f_i}{\partial x_j}(x) = 0$$

So what does this tell us? This is a system of n equations linear in $y_i - f_i(x)$ with coefficients $\frac{\partial f_i}{\partial x_j}(x)$. But since the matrix

$$\left(\frac{\partial f_i}{\partial x_j} \right)$$

is invertible, since $\det f'(x) \neq 0$, we have

$$[\forall i \in \{1, \dots, n\}, (y_i - f_i(x)) = 0] \iff y = f(x)$$

which is what we wanted to show.

So back to what we wanted to do, we wanted to find those open sets V and W . So what should we use as V ?

$$V = f^{-1}(W) \cap \text{Int } B$$

Then $f : V \rightarrow W$ has an inverse $f^{-1} : W \rightarrow V$, since f is one to one. We also wanted to show that the inverse is continuous. So how do we see that? Well, we go back to what we showed above:

$$\forall x_1, x_2 \in V, |x_1 - x_2| \leq 2|f(x_1) - f(x_2)| \implies \forall y_1, y_2 \in W, |f^{-1}(y_1) - f^{-1}(y_2)| \leq 2|y_1 - y_2|$$

implying f^{-1} is continuous. So the only thing remaining is to show that f^{-1} is differentiable. So let's take a point $x_0 \in V$, $y_0 = f(x_0)$. Call $\mu = Df(x_0)$. We'll show f^{-1} is differentiable at y_0 and $(f^{-1})'(y_0) = \mu^{-1}$. So what do we know. We know that

$$f(x) = f(x_0) + \mu(x - x_0) + \varphi(x)$$

where

$$\lim_{x \rightarrow x_0} \frac{|\varphi(x)|}{|x - x_0|} = 0$$

Each $y \in W$ can be written $y = f(x)$, $x \in V$. We have

$$\mu^{-1}(y - y_0) = \mu^{-1}(f(x) - f(x_0)) = x - x_0 + \mu^{-1}\varphi(x)$$

So what this says is that

$$f^{-1}(y) = f^{-1}(y_0) + \mu^{-1}(y - y_0) - \mu^{-1}\varphi(f^{-1}(y))$$

That's like what we need to show that f^{-1} is differentiable. Specifically, we have to show

$$\lim_{y \rightarrow y_0} \frac{\mu^{-1}\varphi(f^{-1}(y))}{|y - y_0|} = 0$$

We have

$$\frac{\varphi(f^{-1}(y))}{|y - y_0|} = \frac{\varphi(f^{-1}(y))}{|f^{-1}(y) - f^{-1}(y_0)|} \cdot \frac{|f^{-1}(y) - f^{-1}(y_0)|}{|y - y_0|} \leq 2 \frac{\varphi(f^{-1}(y_0))}{|f^{-1}(y) - f^{-1}(y_0)|}$$

And that's the proof of the inverse function theorem, and hence the implicit function theorem.

2.6 Tangent Space

We begin with some examples:

1. Consider an ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad (107)$$

The tangent plane at a point (x_0, y_0, z_0) of the ellipsoid is given by

$$\frac{2x_0}{a^2}(x - x_0) + \frac{2y_0}{b^2}(y - y_0) + \frac{2z_0}{c^2}(z - z_0) = 0 \quad (108)$$

$$\frac{x_0}{a^2}x + \frac{y_0}{b^2}y + \frac{z_0}{c^2}z = 1 \quad (109)$$

2. Let $f(x, y) = 0$ and $g(x, y) = 0$ be curves in \mathbb{R}^2 . Assume the curves are smooth and intersect at point (a, b) . The two curves are orthogonal at (a, b) if

$$\frac{\partial f}{\partial x} \frac{\partial g}{\partial x} + \frac{\partial f}{\partial y} \frac{\partial g}{\partial y} = 0 \quad (110)$$

at (a, b) . The two curves are tangent at (a, b) if the gradients are proportional to each other, i.e.

$$\frac{\partial f}{\partial x} \frac{\partial g}{\partial y} - \frac{\partial f}{\partial y} \frac{\partial g}{\partial x} = 0 \quad (111)$$

at (a, b) .

3. Consider the family of parabolas

$$y^2 - 2p(x + p/2) = 0 \quad (112)$$

Here, p is the parameter. These parabolas open along the x axis, either to the right or to the left, with x acting like a function of y .

This kind of family is what's called *confocal*: these parabolas all have the same focus, 0. Of course, they don't all intersect, but all those which open to the right intersect all those which open to the left. Let's look at the intersections.

If $p_1 > 0$ and $p_2 > 0$ then the two parabolas

$$f(x, y) = y^2 - 2p_1(x + p_1/2) = 0, g(x, y) = y^2 - 2p_2(x + p_2/2) = 0 \quad (113)$$

intersect where

$$f(x, y) = y^2 - 2p_1(x + p_1/2) = 0 = g(x, y) = y^2 - 2p_2(x + p_2/2) \quad (114)$$

Let's show that all the intersections among all these possible parabolas are right angles.

Let's try to compute the inner product of the gradients. That is, we want to show

$$f_x g_x + f_y g_y = 0 \quad (115)$$

at an intersection point. So what's the product of the x derivatives? Well it'll just be

$$f_x g_x = (-2p_1)(-2p_2) = 4p_1 p_2 \quad (116)$$

And what's the product of the y derivatives?

$$f_y g_y = (2y)(2y) = 4y^2 \quad (117)$$

So we want to see that

$$f_x g_x + f_y g_y = 4p_1 p_2 + 4y^2 \quad (118)$$

vanishes when both f and g are zero.

If you take p_2 times f , and subtract p_1 times g , you get rid of the x terms:

$$p_2 f - p_1 g = p_2 y^2 - 2p_2 p_1 (x + p_1/2) - p_1 y^2 + 2p_1 p_2 (x + p_2/2) = (p_1 - p_2) y^2 + (p_2 - p_1) p_1 p_2 \quad (119)$$

So we have

$$f_x g_x + f_y g_y = 4p_1 p_2 + 4y^2 = 4 \frac{p_2 f - p_1 g}{p_2 - p_1} = 0 \quad (120)$$

Since $f = g = 0$ and $p_2 - p_1 \neq 0$ at the intersection point.

So these are some examples of computations with tangent spaces.

TODO: this

2.7 Extremum Problems

We're going to be considering functions, for open sets $U \subseteq \mathbb{R}^n$,

$$f : U \rightarrow \mathbb{R} \quad (121)$$

If f has a local maximum or minimum at a point a where

$$\frac{\partial f}{\partial x_i}(a) \quad (122)$$

exists, then it must be zero. This is true because the partial derivative is just the derivative along a one-dimensional subspace: a line. That is,

$$\frac{\partial f}{\partial x_i}(a) = g'_i(0) = 0, g_i(t) = f(a + t\mathbf{e}_i) \quad (123)$$

If f is differentiable at a , what is the direction of maximum increase (we could also ask for decrease) of f at a ? It's the direction of the gradient, which I guess is what's called the gradient. Why is this? If we want to look at the increase of f in a particular direction, we have to look at the value of the directional derivative of f in that particular direction. We want to maximize

$$D_{\mathbf{e}} f(a) = Df(a)(\mathbf{e}) = \langle \text{grad } f(a), \mathbf{e} \rangle \quad (124)$$

for some unit vector \mathbf{e} . In general, of course, what is this inner product? It's

$$|\mathbf{e}| |\text{grad } f(a)| \cos \theta = |\text{grad } f(a)| \cos \theta \quad (125)$$

This, of course, is maximized when $\theta = 0$, i.e. \mathbf{e} points in the direction of $\text{grad } f(a)$.

What about the converse: does a point where the derivative is zero imply an extremum. Here, just like in first year, this is false, however, in this case, the converse is false *even* assuming $g''_i(0) \neq 0$. For example, consider

$$f(x, y) = x^2 - y^2 \quad (126)$$

This graph is "saddle shaped". The second derivative with respect to both x and y are nonzero at the origin, the first partial derivatives are zero, and yet we have neither a local maximum or a minimum: in fact, we have a maximum in the y direction and a minimum in the x direction!

Let's do an example: consider an acute angled triangle with vertices $P_i = (x_i, y_i)$. Find a fourth point $P = (x, y)$ such that the sum of distances to each P_i is as small as possible. We have

$$f(x, y) = r_1 + r_2 + r_3 \quad (127)$$

where

$$r_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (128)$$

One thing that's we can see is that there *is* a minimum, and it occurs on a closed disk whose boundary contains P_i . f is differentiable except at P_i , but the minimum can't occur at P_i , since the triangle is acute, implying we can take the intersection of an edge and the perpendicular from a vertex and get a smaller value for f . This isn't the only place where we'll use the fact that the triangle is acute angled, however.

So we know the maximum must occur at a critical point. At a critical point:

$$\frac{\partial f}{\partial x} = \frac{x - x_1}{r_1} + \frac{x - x_2}{r_2} + \frac{x - x_3}{r_3} = 0 \quad (129)$$

$$\frac{\partial f}{\partial y} = \frac{y - y_1}{r_1} + \frac{y - y_2}{r_2} + \frac{y - y_3}{r_3} = 0 \quad (130)$$

i.e. if

$$u_i = \left(\frac{x - x_i}{r_i}, \frac{y - y_i}{r_i} \right), u_1 + u_2 + u_3 = 0 \quad (131)$$

TODO: rest

2.7.1 Lagrange's Method

Today, we'll do some example problems regarding Lagrange's method:

1. Prove that

$$uv \leq \frac{1}{\alpha}u^\alpha + \frac{1}{\beta}v^\beta$$

for all $u, v \in \mathbb{R}_0^+$, $\alpha, \beta \in \mathbb{R}^+$ such that $\frac{1}{\alpha} + \frac{1}{\beta} = 1$.

Proof. This is trivially true if $u = 0$ or $v = 0$, so we can assume $uv \neq 0$.

The inequality holds for u, v if and only if it holds for $ut^{1/\alpha}$, $vt^{1/\beta}$,

$$(ut^{1/\alpha})(vt^{1/\beta}) = 1 \iff t = \frac{1}{uv}$$

So it's enough to show that

$$1 \leq \frac{1}{\alpha}u^\alpha + \frac{1}{\beta}v^\beta$$

for all $u, v \in \mathbb{R}^+$ such that $uv = 1$. We have that

$$\frac{1}{\alpha}u^\alpha + \frac{1}{\beta}v^\beta \rightarrow \infty$$

on $uv = 1$ when either $u \rightarrow \infty$ or $v \rightarrow \infty$. So there is a minimum at a finite point. The critical points of

$$\frac{1}{\alpha}u^\alpha + \frac{1}{\beta}v^\beta + \lambda(uv - 1)$$

are given by

$$u^{\alpha-1} + \lambda v = 0 \wedge v^{\beta-1} + \lambda u = 0 \implies u^\alpha = v^\beta = -\lambda$$

This gives, combined with condition $uv = 1$, solution

$$u = v = 1$$

So the minimum value of

$$\frac{1}{\alpha}v^\alpha + \frac{1}{\beta}v^\beta$$

is simply

$$\frac{1}{\alpha} + \frac{1}{\beta} = 1$$

which implies the above inequality and hence what we desired to prove. We will use these kinds of techniques to prove a variety of different problems. \square

2. Prove Holder's inequality:

$$\sum_{i=1}^n u_i v_i \leq \left(\sum_{i=1}^n u_i^\alpha \right)^{1/\alpha} \left(\sum_{i=1}^n v_i^\beta \right)^{1/\beta}$$

where $u_i, v_i \in \mathbb{R}_0^+$, $\alpha, \beta \in \mathbb{R}^+$ such that

$$\frac{1}{\alpha} + \frac{1}{\beta} = 1$$

Proof. It's a good observation that this is similar to what we did before (1), and in fact we can deduce it just from what we did before instead of starting from scratch.

We begin by assuming

$$\sum_{i=1}^n u_i^\alpha, \sum_{i=1}^n v_i^\beta \neq 0$$

since if either are zero then the inequality holds trivially.

We can now apply the preceding inequality to

$$u = \frac{u_i}{\left(\sum_{i=1}^n u_i^\alpha \right)^{1/\alpha}}, v = \frac{v_i}{\left(\sum_{i=1}^n v_i^\beta \right)^{1/\beta}}$$

for each u_i, v_i , from which the result follows. \square

3. An example from last year's term test: find the rectangular parallelepiped of greatest volume inscribed in an ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

Some people in MAT257 last year didn't like this question. That was true for most of the questions. But part of the reason why they didn't like this question is that they looked at it in a way which was more difficult than I intended, because I was essentially thinking that you should assume that the rectangular parallelepiped is situated in a standard way within the ellipse. We will do it this way.

In fact, maybe there's an ambiguity in the meaning of the question, because what do you mean by a rectangular parallelepiped inscribed in an ellipse. Like, I thought that meant that all four vertices lied on the ellipse. If you want to do the exercise, it's a good geometric exercise, but it's not a required part of the exercise, to show that if you do have all vertices on the ellipse, it has to be centered at the origin.

So let's begin. Let (x, y, z) be the vertex in the positive orthant (like the positive quadrant but in 3D). So what's the problem we have to solve? First of all, what's the volume in terms of this vertex?

$$(2x)(2y)(2z) = 8xyz$$

So we have to maximize $8xyz$, which is the same as maximizing xyz , on the ellipsoid given, with $x, y, z \in \mathbb{R}^+$.

This means we have to find the critical points. This (the ellipse) is a nice smooth surface of course, technically this is something you need to check, in that its gradient is nonzero at every point. To find the extrema, we hence need to find the critical points of

$$F(x, y, z) = xyz + \lambda \left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 \right)$$

This means we need to find the derivatives in terms of (x, y, z) and set them to zero, which are

$$F_x = yz + 2\lambda \frac{x}{a^2} = F_y = xz + 2\lambda \frac{y}{b^2} = F_z = xy + 2\lambda \frac{z}{c^2} = 0$$

So we have four equations in three unknowns: these equations together with the equation of the ellipsoid.

We multiply the first equation by x , the second by y and the third by z to obtain

$$xyz + 2\lambda \frac{x^2}{a^2} = 0, xyz + 2\lambda \frac{y^2}{b^2} = 0, xyz + 2\lambda \frac{z^2}{c^2} = 0$$

and add them up to obtain

$$3xyz + 2\lambda \left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \right) = 0$$

Plugging in the equation of the ellipse, we get

$$3xyz + 2\lambda = 0$$

To solve this for x, y, z , we begin solving for x by substituting yz for $-2\lambda \frac{x}{a^2}$ (from the first equation) to get

$$-3 \frac{x^2}{a^2} + 1 = 0 \implies x = \frac{a}{\sqrt{3}}$$

Likewise,

$$y = \frac{b}{\sqrt{3}}, z = \frac{c}{\sqrt{3}}$$

We know this is not a minimum, since the volume of the minimum is zero.

2.8 Multivariate Taylor Series

2.8.1 Review of Single Variable Taylor Series

Recall the definition of the n^{th} degree Taylor polynomial of f centered at a

$$P_{n,a}(f) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k$$

We define $R_{n,a}$ to be the “remainder”, the difference between $P_{n,a}$ and f . We can write this remainder in “Lagrange form” as

$$\frac{f^{(n+1)}(t)}{(n+1)!} (x-a)^{n+1}$$

Clearly,

$$\lim_{x \rightarrow a} \frac{R_{n,a}(x)}{(x-a)^n} = 0$$

We can now define the Taylor series of f at a to be

$$T_a f = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k$$

Why these coefficients? Suppose

$$f(x) = \sum_{k=0}^{\infty} a_k (x-a)^k$$

Note

$$\frac{d^n}{dx^n} (x-a)^k = \begin{cases} 0 & \text{if } k < n \\ k! & \text{if } n = k \\ k(k-1)\dots(k-n+1)(x-a)^{k-n} & \text{if } k > n \end{cases}$$

2.8.2 Multivariable Case

For example: What's your favorite polynomial? Zero? I'm sure you meant something more like

$$f(x, y, z) = a_{12,8,2}x^{12}y^8z^2 + a_{0,22,0}y^{22} + a_{0,0,1}z + a_{20,20,20}x^{20}y^{20}z^{20}$$

How to do spot $a_{12,8,2}$? Apply:

$$\left. \frac{\partial^{22} f}{\partial x^2 \partial y^8 \partial z^2} \right|_{(x,y,z)=(0,0,0)} = a_{12,8,2} 12! 8! 2!$$

In general, if f is a sum of terms like

$$a_{\alpha_1, \dots, \alpha_n} x^{\alpha_1} \dots x^{\alpha_n}$$

Then

$$a_{\alpha_1, \dots, \alpha_n} = \left. \frac{\partial^{\alpha_1 + \dots + \alpha_n} f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}} \right|_{(x_1, \dots, x_n) = (0, \dots, 0)} \frac{1}{\alpha_1! \dots \alpha_n!}$$

2.8.3 Multi-index notation

We're mathematicians here, and what do mathematicians do, we generalize. But we generalize in such a way such that the general case looks like the specialized case so we don't have to remember different notation when teaching first-years and when doing research. Multi-index notation is a prime example. Compare:

- Let $f : \mathbb{R} \rightarrow \mathbb{R}$. Then

$$f = \sum_{k \in \mathbb{N}} \frac{1}{k!} f^{(k)}(a) (x - a)^k$$

- Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Then

$$f = \sum_{\alpha \in \mathbb{N}^n} \frac{1}{\alpha!} \frac{\partial^{|\alpha|}}{\partial x^\alpha} f(a) (x - a)^\alpha$$

To make this to work, all we have to do is define, for

$$\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$$

the notation

$$\alpha! = \alpha_1! \dots \alpha_n!$$

$$|\alpha| = \alpha_1 + \dots + \alpha_n$$

$$\partial x^\alpha = \partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}$$

$$(x - a)^\alpha = (x_1 - a_1)^{\alpha_1} \dots (x_n - a_n)^{\alpha_n}$$

Now, for some examples

1.

$$\frac{1}{6! 7!} 12 x^6 y^7 = \frac{12}{(6, 7)!} (x, y)^{6, 7}$$

2.

$$x^5 + x^4 y + x^3 y^2 + \dots + y^5 = \sum_{\beta + \gamma = 5} (x, y)^{(\beta, \gamma)} = \sum_{|\alpha| = 5} (x, y)^\alpha$$

Note that we're assuming here β, γ are non-negative. On good days you can just do what we did above, but if it worries you, write it down.

3.

$$a_{0,0} + a_{1,0}x + a_{0,1}y + a_{1,1}xy + a_{2,0}x^2 + a_{0,2}y^2 = \sum_{|\alpha| \leq 2} a_\alpha (x, y)^\alpha$$

2.8.4 Multivariable Taylor Series

Theorem 8. If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is \mathcal{C}^{k+1} then

$$f(a+h) = \sum_{|\alpha| \leq k} \frac{1}{\alpha!} \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(a) h^\alpha + \sum_{|\alpha| = k+1} \frac{1}{\alpha!} \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(a + \theta h) h^\alpha$$

where $\theta \in (0, 1)$

Proof. The key idea is

$$F(t) = f(a + th) \text{ (Old Taylor + Chain Rule)}$$

More rigorously, we know from single variable calculus that

$$F(t) = F(0) + \frac{F'(0)}{1!}t + \dots + \frac{F^{(k)}(0)}{k!}t^k + \frac{F^{(k+1)}(\theta)}{(k+1)!}t^{k+1}$$

Set $t = 1$ to get $F(1) = f(a+h)$. We know $F(0) = f(a)$. We have

$$\left. \frac{dF}{dt} \right|_{t=0} = \left. \frac{d}{dt} f(a+th) \right|_{t=0} = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a+th) \left. \frac{\partial x_i}{\partial t} \right|_{t=0} = \sum_{i=1}^n \left[\left. \frac{\partial f}{\partial x_i}(a+th) \right|_{t=0} \right] h_i = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) h_i$$

We have

$$\left. \frac{d^2 F}{dt^2} \right|_{t=0} = \sum_{i=1}^n \left[\sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(a) h_j \right] h_i$$

By the magical process of ...,

$$\left. \frac{d^k F}{dt^k} \right|_{t=0} = \sum_{i_1, \dots, i_k=1}^n \frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}}(a) h_{i_1} \dots h_{i_k}$$

Now all we need is some combinatorics, combined with the equality of mixed partials (for \mathcal{C}^{k+1} functions), to complete the proof. Recall the *multinomial coefficient*

$$\binom{|\alpha|}{\alpha_1, \dots, \alpha_n} = \frac{|\alpha|!}{\alpha_1! \dots \alpha_n!}$$

is the number of ways of taking $|\alpha|$ things and making k groups of size $\alpha_1, \dots, \alpha_n$. We can use this to rewrite the above as

$$F^{(S)}(0) = \sum_{|\alpha|=S} \frac{S!}{\alpha_1! \dots \alpha_n!} \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(a) h^\alpha$$

So plugging this into our original expression for the Taylor series of F , where we divide the S^{th} term by $\frac{1}{S!}$, we have

$$F(t) = F(0) + \sum_{i=1}^k \sum_{|\alpha|=i} \frac{1}{\alpha_1! \dots \alpha_n!} \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(a) h^\alpha = f(a) + \sum_{|\alpha| \leq k} \frac{1}{\alpha!} \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(a) h^\alpha$$

as desired. □

TODO: fill in here

3 Integration

3.1 The (Riemann) Integral Over a Rectangle

TODO: beginning

We said what the integral of a closed rectangle on \mathbb{R}^n was last time, in complete analogy with first year calculus:

Definition 6. If $f : A \rightarrow \mathbb{R}$ is a bounded function on a closed rectangle $A \subset \mathbb{R}^n$, f is integrable on A if, where \mathcal{P} is the set of partitions of A ,

$$\sup_{P \in \mathcal{P}} L(f, P) = \inf_{P \in \mathcal{P}} U(f, P) \quad (132)$$

then the common value is called the integral of f over A , written

$$\int_A f \quad (133)$$

Throughout this lecture we'll see that most of the properties you learned in first year calculus will carry over to this more general case. One thing we'll be carrying over today is that instead of expressing things in terms of sup and inf, we can instead express things in terms of ϵ :

Lemma 2. If $f : A \rightarrow \mathbb{R}$ is a bounded function on a closed rectangle A then f is integrable on A if and only if for every $\epsilon > 0$, there is a partition P of A such that

$$U(f, P) - L(f, P) < \epsilon \quad (134)$$

The proof is exactly like first year, and one direction should be immediate:

Proof. • \Leftarrow : if there's a partition P for every ϵ , then there's no "room" between the sup and the inf

• \Rightarrow : if $\sup = \inf$, then there are partitions P, P' such that

$$U(f, P) - \inf_{P \in \mathcal{P}} U(f, P) < \frac{\epsilon}{2} \wedge \sup_{P \in \mathcal{P}} L(f, P) - L(f, P') < \frac{\epsilon}{2} \implies U(f, P'') - L(f, P'') < \epsilon \quad (135)$$

where P'' is a refinement of P and P' .

□

Just like in first year. And also, the kind of first calculations of integrals that you've done in first year work out the same way here. For example,

1. If $f = c$ is a constant, then f is integrable and

$$\int_A f = cv(A) \quad (136)$$

where $v(A)$ denotes the volume of A . This is because, for any subrectangle S of P ,

$$m_S(f) = c = M_S(f) \implies L(f, P) = \sum m_s(f)v(S) = c \sum v(S) = cV(A) \quad (137)$$

with an analogous calculation for $U(f, P)$

2. Let $f : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$. Then

$$f(x, y) = \begin{cases} 0 & \text{if } x \in \mathbb{Q} \\ 1 & \text{otherwise} \end{cases} \quad (138)$$

is not integrable, since for any partition P , $m_S(f) = 0 \wedge M_S(f) = 1 \implies L(f, P) = 0 \wedge U(f, P) = 1$

3.1.1 Basic Properties of the Integral

Suppose we have two integrable functions

$$f, g : A \rightarrow \mathbb{R} \quad (139)$$

Then

1. $f + g$ is integrable and

$$\int_A (f + g) = \int_A f + \int_A g \quad (140)$$

- 2.

$$f \leq g \implies \int_A f \leq \int_A g \quad (141)$$

3. $f \cdot g$ is integrable

4. $|f|$ is integrable and

$$\int_A |f| \geq \left| \int_A f \right| \quad (142)$$

It's enough to show that $|f|$ is integrable

5. Exercise: any increasing function is integrable

3.2 Integrals Over More General Bounded Sets

TODO: this

3.3 Fubini's Theorem

TODO: rewrite

We are now going to talk about Fubini's theorem, which is about how to integrate over a high-dimensional rectangle by repeatedly performing individual integrals. Let's start with an example. Suppose we want to integrate over a rectangle

$$A = [a, b] \times [c, d] \subseteq \mathbb{R}^2 \quad (143)$$

and let's suppose, to make it easy, we have a continuous, non-negative function f , defined on this rectangle, say

$$z = f(x, y) \geq 0 \quad (144)$$

The idea is that if we fix a point on the x -axis, say x , we can consider the "slice" in the y -direction determined by this point. We could find, for example, the area of that slice, and it's reasonable to expect that the integral of f on the rectangle, we could obtain by integrating the area of that slice along the length of the rectangle.

The idea then is to try to find the area of such a slice, which would be in this case, for some x ,

$$h(x) = \int_c^d g_x(y) dy = \int_c^d f(x, y) dy \quad (145)$$

where $g_x(y) = f(x, y)$ fixing x . It's reasonable to expect that

$$\int_A f = \int_a^b h = \int_a^b \left(\int_c^d f(x, y) dy \right) dx \quad (146)$$

So that's the idea of Fubini's theorem: that we should be able to integrate a function over a rectangle by repeated one-dimensional integrals. Of course, we're not interested in integrating only continuous functions, we want to look at more general, integrable functions, but if you think about that, supposing f is integrable, this could run into a problem: one of the functions g_x might not be integrable on $[c, d]$. After all, it's set of discontinuities could be $x_0 \times [c, d]$ for some x_0 , so $\int_c^d g_{x_0}$ makes no sense!

So we'll have to formulate something maybe a little bit more technical, but it's to capture that problem. Suppose we just have $f : A \rightarrow \mathbb{R}$ bounded. The function may or may not be integrable, meaning the supremum of lower sums is equal to the infimum of lower sums. Whether the function is integrable or not, we can still look at the supremum of lower sums and the infimum of lower sums, which is what we'll do.

We'll define the lower and upper integrals of f on A , $\mathcal{L} \int_A f$ and $\mathcal{U} \int_A f$ respectively, to be the supremum of all the lower sums $L(f, \mathcal{P})$ and infimum of all the upper sums $U(f, \mathcal{P})$ respectively. We can now write down our theorem as the previous formula, but taking into account that we don't know that the function g_x mentioned before, we just replace it by the lower or upper integral:

Theorem 9. Suppose $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^n$ are closed rectangles and $f : A \times B \rightarrow \mathbb{R}$ is integrable. For all $x \in A$, define

$$g_x : B \rightarrow \mathbb{R}, g_x(y) = f(x, y) \quad (147)$$

Set

$$\mathcal{L}(x) = \mathcal{L} \int_B g_x = \mathcal{L} \int_B f(x, y) dy, \mathcal{U}(x) = \mathcal{U} \int_B g_x = \mathcal{U} \int_B f(x, y) dy \quad (148)$$

Then $\mathcal{L}(x), \mathcal{U}(x)$ are integrable on A and

$$\int_{A \times B} f = \int_A \mathcal{L} = \int_A \mathcal{U} = \int_A \left(\mathcal{L} \int_B f(x, y) dy \right) dx = \int_A \left(\mathcal{U} \int_B f(x, y) dy \right) dx \quad (149)$$

Before we prove this, some remarks:

1. We also have

$$\int_{A \times B} f = \int_B \left(\mathcal{L} \int_A (f(x, y) dx) \right) dy = \int_B \left(\mathcal{U} \int_A (f(x, y) dx) \right) dy \quad (150)$$

2. If g_x is integrable on B for every $x \in A$, then

$$\int_{A \times B} f = \int_A \left(\int_B f(x, y) dy \right) dx \quad (151)$$

3. If

$$A = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n] \quad (152)$$

we can apply Fubini's theorem repeatedly to obtain

$$\int_A f = \int_{a_n}^{b_n} \left(\dots \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x_1, \dots, x_n) dx_1 \right) dx_2 \dots \right) dx_n \quad (153)$$

if f is "sufficiently nice" (otherwise we'll have to sprinkle in some L's or U's).

Now here's an application to $\int_C f$ when f is integrable on a Jordan-measurable set $C \subset \mathbb{R}^n$. What we're going to use is Fubini's theorem in a rectangle with the formula

$$\int_C f = \int_A f \chi_C \quad (154)$$

where $A \supseteq C$ is a closed rectangle. Let's do some quick examples:

1. Let's say we were integrating on the region

$$C = [-1, 1] \times [-1, 1] \setminus \{x \in \mathbb{R}^2, x_1^2 + x_2^2 < 1\} \quad (155)$$

i.e. the unit square with the unit circle removed from it. We have

$$\int_C f = \int_{[-1, 1]^2} f \chi_C \quad (156)$$

We can write

$$\chi_C = \begin{cases} 1 & \text{if } -1 \leq y \leq -\sqrt{1-x^2} \text{ or } \sqrt{1-x^2} \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (157)$$

We can write

$$\int_{-1}^1 f(x, y) \chi_C(x, y) dy = \int_{-1}^{-\sqrt{1-x^2}} f(x, y) dy + \int_{\sqrt{1-x^2}}^1 f(x, y) dy \quad (158)$$

i.e. the limits of integration correspond to the boundary of C . So we get

$$\int_C f = \int_{-1}^1 \left(\int_{-1}^{-\sqrt{1-x^2}} f(x, y) dy + \int_{\sqrt{1-x^2}}^1 f(x, y) dy \right) dx \quad (159)$$

2. Say we want to integrate over a triangle C , under the line from $(0, 0)$ to (a, a) . We have

$$\int_C f = \int_0^a \left(\int_y^a f(x, y) dx \right) dy \quad (160)$$

We could have done it the other way, integrating first with respect to y and then with respect to x , and get

$$\int_C f = \int_0^a \left(\int_0^x f(x, y) dy \right) dx \quad (161)$$

This shows the form might not be the same in different directions. And it could be important to exploit the difference. For example, what if our function $f(x, y)$ only depends on one of the variables, say y (i.e. f is *independent* of x). Then

$$\int_0^a \left(\int_0^x f(y) dy \right) dx \quad (162)$$

doesn't simplify very much, but

$$\int_0^a \left(\int_y^a f(y) dx \right) dy = \int_0^a (a - y) f(y) dy \quad (163)$$

So this double integral reduces to a single integral with respect to y .

TODO: proof

TODO: polish Before we move on, we will give some more examples of computing integrals, using Fubini's theorem and the change of variables theorems as tools.

Let us begin by recalling some facts about the regions on which we can integrate. What are the kinds of region on which we integrate? One of the kinds of regions would be the region between the graphs of two functions that are defined on a Jordan-measurable set. Specifically, if we have a subset $C \subset \mathbb{R}^n$ that is Jordan-measurable, and two integrable functions $\varphi(x) \leq \psi(x)$ defined on C , we're interested in integrating over the region bounded by the graphs of the two functions. Let's call this region S , that is, define

$$S = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : x \in C, y \in [\varphi(x), \psi(x)]\} \quad (164)$$

We have that S is a Jordan-measurable set. We want to be able to integrate a continuous function on S . If $f(x, y)$ is a bounded continuous on S , then we can compute using Fubini's theorem that

$$\int_S f = \int_C \left(\int_{\varphi(x)}^{\psi(x)} f(x, y) dy \right) dx \quad (165)$$

We went through the proof of this before, but I want to use it to compute some examples.

1. Consider

$$\int_0^2 \int_{y/2}^1 ye^{-x^3} dx dy \quad (166)$$

There's no hope in proceeding naively, since e^{-x^3} doesn't even have an elementary primitive, i.e. you cannot write down its integral with only elementary functions. But let's consider what region we're integrating along: the area under the line between $(0,0)$ and $(2,1)$. So using the above observation, we can rewrite the integral to be

$$\int_0^1 \left(\int_0^{2x} ye^{-x^3} dy \right) dx = \int_0^1 \frac{y^2}{2} e^{-x^3} \Big|_0^{2x} dx = 2 \int_0^1 x^2 e^{-x^3} dx = -\frac{2}{3} e^{-x^3} \Big|_0^1 = \frac{2}{3}(1 - e^{-1}) \quad (167)$$

Here, the change of variable was a useful thing to do, because it enabled us to write down an original integral, which we couldn't do in the original form.

2. Consider

$$\int_2^4 \int_{4/x}^{\frac{20-4x}{8-x}} (y-4) dy dx \quad (168)$$

So here, again, should we just go ahead and do it as it's written? Well then we'll get $\frac{y^2}{2} - 4y$ and we'll substitute those things in and we'll just get a terrible mess. But what's the region we're integrating over? $\frac{4}{x}$ is like a hyperbola, which is 2 when x is 2 and 1 when x is 4. The other function on top, what does it look like? It's also a hyperbola: it's a constant plus something over $8-x$, so it'll open down. And this is the region we're integrating on: the area between a hyperbola open up and another opening down.

If we change the order of integration, we'll be integrating on the outside with respect to y , which goes from 1 to 2, and we'll be integrating on the inside with respect to x with x going from $4/y$ to... let's solve:

$$y = \frac{20-4x}{8-x} = 4 + \frac{12}{x-8} \implies x-8 = \frac{12}{y-4} \iff x = 8 + \frac{12}{y-4} \quad (169)$$

Hence we can write the above as

$$\int_1^2 \left(\int_{4/y}^{8+12/(y-4)} (y-4) dx \right) dy = \int_1^2 (y-4)x \Big|_{4/y}^{8+12/(y-4)} dy = \int_1^2 (y-4) \left(8 + \frac{12}{y-4} - \frac{4}{y} \right) dy \quad (170)$$

which is all stuff that's very easy to integrate

3. Let's see how to integrate a simple function z over a region of 3-space

$$S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq z^2 \wedge x^2 + y^2 + z^2 \leq 1\} \quad (171)$$

So, what is this region? The second equation says we're inside the closed ball of radius 1 centered at the origin. The first equation is the region inside two cones, opening up and down from the origin (kind of like an hour glass). So we want the intersection of these two regions. In terms of the theorem that we wrote down at the beginning, the set C is like the discs inside the circles formed by the intersections of the top and bottom of the cones and circles, and we're integrating on the region between the semicircles above and below the plane, and the cones above and below the plane. So what's this going to be? Well, by symmetry, it's going to be *zero*.

Let's make it a little harder. Let's consider only the top, i.e.

$$S^+ = \{(x, y, z) \in S : z \geq 0\} \quad (172)$$

So, rewriting our integral to be over the disc C , we obtain

$$\begin{aligned} & \int \int_{x^2+y^2 \leq \frac{1}{2}} \left(\int_{\sqrt{x^2+y^2}}^{\sqrt{1-(x^2+y^2)}} z dz \right) dx dy \\ &= \int \int_{x^2+y^2 \leq \frac{1}{2}} \frac{1}{2} (1 - (x^2 + y^2) - (x^2 + y^2)) dx dy \\ &= \frac{1}{2} \int \int_{x^2+y^2 \leq \frac{1}{2}} dx dy - \int \int_{x^2+y^2 \leq \frac{1}{2}} (x^2 + y^2) dx dy = \frac{1}{2} \pi \frac{1}{2} = \frac{\pi}{4} \end{aligned} \quad (173)$$

How do we justify that last, “magical” step? Well, the prettiest way to do so is to change to polar coordinates, but we haven’t justified this quite yet, which is the point of this example. You’ll remember from first year calculus that we can write

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)du \quad (174)$$

This is what we’re going to be doing in several variables. But first: polar coordinates? That means writing

$$x = r \cos \theta, y = r \sin \theta, \theta \in [0, 2\pi], r \in [0, \infty) \quad (175)$$

This gives that

$$\frac{\partial(x, y)}{\partial(r, \theta)} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} \implies \det \frac{\partial(x, y)}{\partial(r, \theta)} = r \implies dx dy = r dr d\theta \quad (176)$$

“generalizing” equation 174. On the other hand, we have that

$$x^2 + y^2 = r^2 \quad (177)$$

Since we have that, for this disc, θ ranges over the whole interval $[0, 2\pi]$ whereas r ranges over $[0, 1/\sqrt{2}]$, we can hence rewrite the above integral as

$$\int_0^{2\pi} \int_0^{\frac{1}{\sqrt{2}}} r^2 \cdot r dr d\theta = \int_0^{2\pi} d\theta \int_0^{\frac{1}{\sqrt{2}}} r^3 dr = 2\pi \frac{r^4}{4} \Big|_0^{\frac{1}{\sqrt{2}}} = \frac{\pi}{8} \quad (178)$$

3.4 Partitions of Unity

3.4.1 Bump Functions

A bump function, in the one dimensional case, is a function which goes up to 1 from 0, stays there for a while, and then returns to zero. But we’re interested in doing this in n -dimensions. How? Given an open $U \subset \mathbb{R}^n$, and compact $C \subset U$, we want to construct a \mathcal{C}^∞ function $f : U \rightarrow \mathbb{R}$ with $0 \leq f \leq 1$ such that $\forall x \in C, f(x) = 1$ and $f(x) = 0$ outside some compact subset of U . So it looks like the 1-dimensional bump function, except over an n -dimensional area C .

How can we construct such a function? We’ll do it as a kind of series of exercises, based on one particular function, which you can imagine ought to be the kind of thing that you have to build this function, that is,

$$f(x) = \begin{cases} e^{1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases} \quad (179)$$

We proceed as follows, building up by steps:

1. There is a \mathcal{C}^∞ function $g : \mathbb{R} \rightarrow \mathbb{R}$ such that $g > 0$ on $(-1, 1)$, 0 elsewhere. It’s easy to write down a formula for such a function using the function given in 179, as follows:

$$g(x) = \begin{cases} e^{-\frac{1}{(x-1)^2}} e^{-\frac{1}{(x+1)^2}} & \text{if } x \in (-1, 1) \\ 0 & \text{if } x \notin (-1, 1) \end{cases} \quad (180)$$

This works because $e^{-\frac{1}{(x-1)^2}}$ becomes completely flat at $x = 1$, whereas $e^{-\frac{1}{(x+1)^2}}$ becomes completely flat at $x = -1$.

2. Given $a \in \mathbb{R}^n, \delta > 0$, there is a C^∞ function $g_{a,\delta} : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$g_{a,\delta} > 0 \text{ on } |x - a| < \delta, g_{a,\delta} = 0 \text{ outside} \quad (181)$$

So how could achieve this? We could use the function defined in 180 to write

$$g_{a,\delta} = g\left(\frac{|x - a|^2}{\delta^2}\right) \quad (182)$$

3. There is a C^∞ function $h : U \rightarrow \mathbb{R}$ such that h is greater than 0 on C and $h = 0$ outside a compact subset of U . So how do we do this?

We could cover C with a finite set of open balls U_1, \dots, U_k with centers a_1, \dots, a_k and radii $\delta_1, \dots, \delta_k$ whose closures lie in U . We can do this since C is compact. Once we do this, we can define quite simply

$$h(x) = \sum_{i=1}^k g_{a_i, \delta_i}(x) \quad (183)$$

4. As a fourth thing, going back to the real line, let's show that if we're given any $\epsilon \in \mathbb{R}^+$, we can find a C^∞ function ψ_ϵ which is zero up to 0, becomes 1 at ϵ , and then remains 1 forever after. How can we do this from what we've already done? We're going to do it starting with a C^∞ function like that in equation 180, but instead of on $(-1, 1)$, on $(0, \epsilon)$. More generally, actually, we can use any function g such that $g > 0$ on $(0, \epsilon)$ and is zero outside. We can then define ψ_ϵ by

$$\psi_\epsilon(x) = \frac{\int_0^x g}{\int_0^\epsilon g} \quad (184)$$

5. Now, putting everything together, how can we define the bump function that we wanted to begin with? We begin with the function h as defined in equation 183, and we're going to modify it using equation 184 as follows: letting $\epsilon = \min_C h > 0$, which exists since C is compact, let

$$f(x) = \psi_\epsilon \circ h \quad (185)$$

Now there are other ways of doing it. One of them is to write down one integral formula from the beginning, but it would be harder to see just where this formula comes from in the first place.

3.4.2 Definition

We can now get to defining partitions of unity:

Theorem 10. *Given $A \subset \mathbb{R}^n$ and an open cover \mathcal{O} of A , there is a countable collection Φ of C^∞ functions φ on \mathbb{R}^n with the following properties:*

1. $0 \leq \varphi \leq 1$
2. *For all $x \in A$, there is an open neighborhood V_x of x in \mathbb{R}^n such that all but finitely many φ vanish on V_x*
3. *For all $x \in A$, $\sum_{\varphi \in \Phi} \varphi(x) = 1$ (sum is finite in a neighborhood of x)*
4. *For all $\varphi \in \Phi$, there is some open $U \in \mathcal{O}$ such that $\varphi = 0$ outside a compact subset of U .*

Definition 7. *A collection φ satisfying conditions 1, 2, 3 above is called a C^∞ partition for A . If condition 4 holds, this partition is called subordinate to \mathcal{O}*

This is all \mathcal{C}^∞ stuff, but we could have made something weaker. We could have asked for a \mathcal{C}^r partition, or even a partition that was merely continuous. So suppose we did that, and only cared about a \mathcal{C}^0 partition. Then, we could also, though the proof will have to come next time, we could use functions that look like trapezoids, using $|x|$ instead of $e^{-\frac{1}{x^2}}$. If we wanted a \mathcal{C}^r partition of unity, then of course, we could do the same thing, but only caring about \mathcal{C}^r instead of \mathcal{C}^∞ , and hence using the function given by

$$f(x) = \begin{cases} |x^{r+1}| & \text{on } [0, \infty) \\ -|x^{r+1}| & \text{on } (-\infty, 0] \end{cases} \quad (186)$$

Before we try to construct such a (\mathcal{C}^∞) partition, let's discuss why this is an interesting idea. We're going to use this for several things in this course:

- To extend the definition of the integral to more general regions
- To prove the change of variables theorem
- To extend the definition of the integral to manifolds
- To prove Stoke's theorem

So why? Suppose we wanted to extend the definition of the integral to manifolds. Take some shape in 3-space, with a boundary, which is a manifold. Right now, we can integrate over a rectangle. Once we prove the change of variables theorem, we'll be able to integrate over a space diffeomorphic to a rectangle, a sort of deformed rectangle. Now, if we have a coordinate chart, a diffeomorphism from \mathbb{R}^n to some part of the manifold, we could use this to extend the definition of the integral from the rectangle to at least some kind of "rectangular" piece of the manifold. The question remains: how can we extend the definition to the *entire* manifold? A reasonable way to try to do this is to try to cover the entire manifold by a grid of rectangles. If we add up the integrals on each of those pieces...

That was the way this was done classically: you'd take a manifold and divide it up by some kind of rectangular grid, and just add up the contributions of every piece. This classical approach, however, ran into a big obstacle, which was: how do you know that you can really cover every manifold with such a grid? How can you really divide up a really complex surface into rectangles? That is an extremely hard problem: it's true you can do it, but it's an extremely hard problem, and was solved way after the extension of integrals to manifolds.

So how can we avoid this issue? The way this was solved was, you do something a million times simpler. When you cover the manifold with rectangles, you don't care that they all fit together in a nice grid. Just, for every point, you take a rectangle containing that point, and we don't care whether they overlap or not. How do we exploit that? We use a partition of unity! Because what we'll do is, we'll pick a partition of unity such that every one of these bump functions sits in the interior of one of these rectangles. Now, if we want to integrate our function f over the manifold, and we have our partition of unity Φ , and we look at φf , it vanishes outside one of these rectangles, so we can just integrate over the rectangle and get the answer for the manifold. We can then write

$$\int f = \sum \int \varphi f \quad (187)$$

So the key is to take a function, and make its support very small, and then work with those small pieces. And we're going to use this in the same way to prove Stoke's theorem. Stoke's theorem is like the FTC. What does the FTC say? It says that

$$\int_a^b f' = f(b) - f(a) \quad (188)$$

So the integral of f over the boundary of $[a, b]$ is the same as the integral of the derivative of f over the whole interval. So Stoke's theorem is that the integral of a function over the boundary of a manifold is the same as the integral of the derivative over the whole manifold.

We can prove this for rectangles. Then, if we want to prove it for manifolds, again we need to try to cover the manifolds with a grid like this. If we can do that, we can prove Stoke's theorem for surfaces that we know can be covered by a grid. But we can do it for general manifolds in a really cheap way, again by multiplying with partitions of unity. So that's what we're going to do in the rest of the course. But first,

we have to prove Theorem 10, because that's what we're going to use to glue together little pieces in a very nice but cheap way.

Proof. The way we're going to do this is by considering a number of cases building up to the general case.

1. Assume A is compact. That means finitely many $U \in \mathcal{O}$ cover A . Call them U_1, \dots, U_q . In this case we're going to construct a *finite* (not countable) partition of unity subordinate to $\{U_1, \dots, U_q\}$.

For each $i \in \{1, \dots, q\}$, the first thing that we're going to do is show that we can find a compact set $D_i \subset U_i$ such that the interiors of the D_i cover A . How?

For all $x \in A$, let D_x be a closed ball with centre x lying inside some U_i . Since A is compact, finitely many of the interiors of the D_x cover A . Number them D_{x_1}, \dots, D_{x_p} . For each $i \in \{1, \dots, q\}$, let

$$D_i = \bigcup_{D_{x_j} \subset U_i} D_{x_j} \quad (189)$$

Let ψ_i be a \mathcal{C}^∞ function $0 \leq \psi_i \leq 1$ such that $\psi_i > 0$ on D_i , and is 0 outside a compact subset of U_i . Notice if we do this, then, since the interior of the D_i 's cover's A ,

$$\forall x \in U \supseteq A, \sum_{i=1}^q \psi_i(x) > 0 \quad (190)$$

where U is an open neighborhood of A . SO we define our partition of unity as

$$\Phi = \{\varphi_1, \dots, \varphi_q\}, \varphi_i = \frac{\psi_i}{\sum_{i=1}^q \psi_i} \quad (191)$$

2. Now assume $A = \bigcup_{i \in \mathbb{N}} A_i$ where each A_i is compact and a subset of the interior of A_{i+1} . Let $B_i = A_i \setminus \text{Int } A_{i-1}$. Let \mathcal{O}_i be covers of B_i by the open sets

$$U \cap (\text{Int } A_{i+1} \setminus A_{i-2}) \quad (192)$$

for each $U \in \mathcal{O}$. By Case 1 there is a finite partition of unity Φ_i for B_i subordinate to \mathcal{O}_i . Consider

$$\sigma = \sum_i \sum_{\psi_i \in \Phi_i} \psi_i \quad (193)$$

This is a finite sum in some neighborhood of every $x \in A$. Take for a partition

$$\varphi = \frac{\psi}{\sigma} \quad (194)$$

for each $\psi \in \Phi_i$, for every i .

3. Assume A is open. Then $A = \bigcup_{i \in \mathbb{N}} A_i$ as in Case 2. How can we get such a representation? Try

$$A_i = \{x : |x| \leq i, d(x, \mathbb{R}^n \setminus A) > i^{-1}\} \quad (195)$$

4. Let A be arbitrary. Apply Case 3 to $\bigcup_{U \in \mathcal{O}} U$. A partition of unity subordinate to \mathcal{O} is also a partition of unity for A subordinate to \mathcal{O}

□

3.5 Extended Definition of the Integral (“Improper Integral”)

Suppose that A is open and f is bounded and continuous. Then under the current definition, the expression

$$\int_A f \quad (196)$$

may not exist, since the boundary may not be Jordan-measurable. So we now want to try and rectify that, by attempting to extend the definition of the integral to open subsets $A \subset \mathbb{R}^n$ and *locally* bounded functions $f : A \rightarrow \mathbb{R}$, i.e. for every point in A , there is an open neighborhood around A such that f is bounded around A . Furthermore, we assume the set of discontinuities is of measure zero.

Just like the improper integral from first year calculus, it won’t always exist. But we’ll want to know the value when it does. We’ll begin with a useful note: if f vanishes outside of a compact subset C of A , then in fact f is integrable on C (since we can put that compact subset in a big rectangle and multiply by the characteristic function). In fact, we can say that f is integrable on any bounded open subset U of A containing C , for the same reason, and

$$\int_U f = \int_C f \quad (197)$$

So it makes sense to say

$$\int_A f = \int_C f \quad (198)$$

This is going to be our starting point. Now to go further, we’re going to talk about a partition of unity. Let \mathcal{O} be an open cover of A such that $U \subset A$ for all $U \in \mathcal{O}$ and let Φ be a partition of unity for A subordinate to \mathcal{O} . Note that this is the same as saying let Φ be a partition of unity for A subordinate to $\{A\}$, since that counts as an open covering of A . As in equation 198, we can define, $\forall \varphi \in \Phi$,

$$\int_A \varphi |f| \quad (199)$$

We can now extend the definition of the integral as follows:

Definition 8. f is integrable on A if

$$\sum_{\varphi \in \Phi} \int_A \varphi |f| \quad (200)$$

converges. Then we define

$$\int_A f = \sum_{\varphi \in \Phi} \int_A \varphi f \quad (201)$$

Note that equation 201 must converge, and in fact converges absolutely, if equation 200 converges, because

$$\sum_{\varphi \in \Phi} \left| \int_A \varphi f \right| \leq \sum_{\varphi \in \Phi} \int_A |\varphi| |f| = \sum_{\varphi \in \Phi} \int_A \varphi |f| \quad (202)$$

since φ is nonnegative.

Theorem 11. Suppose $A \subset \mathbb{R}^n$ is open, $f : A \rightarrow \mathbb{R}$ is locally bounded and the set of discontinuities of f has measure 0.

1. Given partitions Φ, Ψ as above, not necessarily subordinate to the same open covering, though this doesn’t matter since both will be subordinate to $\{A\}$, then if equation 200 converges for Φ , then it converges for Ψ and the integral defined using Φ and the integral defined using Ψ are equal, i.e.

$$\sum_{\psi \in \Psi} \int_A \psi f = \sum_{\varphi \in \Phi} \int_A \varphi f \quad (203)$$

Note that equation 203 would not necessarily be true if equation 200 does not converge.

2. The new definition of the integral always exists if both A and f are bounded, showing it truly generalizes the old definition. Specifically, if both A and f are bounded, whether or not the boundary of A has measure zero,

$$\sum_{\varphi \in \Phi} \int_A \varphi |f| \quad (204)$$

(equation 200) converges

3. If A is Jordan-measurable and f is bounded, then

$$\int_A f = \sum_{\varphi \in \Phi} \int_A \varphi f \quad (205)$$

where the left is as defined before.

There are other ways of defining the integral, some of which look more like the improper integral from first year calculus. Another thing that you could do is the following: since you can always write an open set as a kind of “expanding union” of compact sets $C_i \subseteq C_{i+1}$, you can actually do that in such a way such that each C_i is Jordan-measurable, meaning the integral on each of these C_i ’s exist. We could then define

$$\int_A f = \lim_{i \rightarrow \infty} \int_{C_i} f \quad (206)$$

After we finish proving this theorem, either I’ll prove this or stick it on the next problem set.

Proof. 1. $\varphi \cdot f$ vanishes outside a compact set, and only finitely many ψ are nonzero on this set. So

$$\sum_{\psi \in \Psi} \psi = 1 \implies \sum_{\varphi \in \Phi} \int_A \varphi f = \sum_{\varphi \in \Phi} \int_A \left(\sum_{\psi \in \Psi} \psi \right) \varphi f = \sum_{\varphi \in \Phi} \sum_{\psi \in \Psi} \int_A \varphi \psi f \quad (207)$$

A priori, we don’t know that we can interchange the order in the double sum of the last equal expression, but we could write down exactly the same stuff with the absolute value in it, which tells us we can. Interchanging them, we get the desired result.

2. Let B be a big rectangle containing A , and assume $|f| \leq M$ on A . For any finite subset F of Φ ,

$$\sum_{\varphi \in F} \int_A \varphi |f| \leq M \int_A \sum_{\varphi \in F} \varphi \leq MV(B) \quad (208)$$

implying the desired result. We still have to show that in the situation that $\int_A f$ is defined according to our earlier definition, we get the same result. We’ll do so next time. \square

Let’s recap. What we’ve shown so far is that

1. The existence of the integral and it’s value is independent of the partition of unity chosen to compute it.
2. The integral *always* exists if A and f are both bounded.
3. We *want* to show that if we’re in a situation where the integral exists according to our old definition, so that in particular A and f will be bounded, then our new definition agrees with the old one. That is, if moreover, A is Jordan measurable, then

$$\int_A f = \sum_{\varphi \in \Phi} \int_A \varphi f \quad (209)$$

where the left hand side is computed using the old definition

Proof. For every $\epsilon > 0$, there is a compact Jordan-measurable subset $C \subset A$ such that

$$V(A \setminus C) = \int_{A \setminus C} 1 = \int_A \chi_{A \setminus C} < \epsilon \quad (210)$$

So how do we get this C ? We're given that A is Jordan measurable, so as A is bounded we can cover the boundary of A by the interiors of finitely many closed balls \mathcal{B} (or rectangles) of total volume $< \epsilon$. So we can take

$$C = A \setminus \text{Int} \bigcup \mathcal{B} \quad (211)$$

(we could also take the union of interiors). Why is C Jordan measurable? Since the boundary of C is a subset of the boundary of $\text{Int} \bigcup \mathcal{B}$, which has measure zero, it follows that C has boundary of measure zero and is hence Jordan measurable.

Since C is compact, only finitely many of our partition functions $\varphi \in \Phi$ are nonzero on C . So we'll take a finite sum, look at the difference, and show that that difference goes to zero: consider any finite subset F of Φ including the ones which are nonzero on C . Consider

$$\left| \int_A f - \sum_{\varphi \in F} \int_A \varphi f \right| \quad (212)$$

As the absolute value of the integral is less than or equal to the integral of the absolute value, we have that equation 212 is less than or equal to

$$\int_A \left| f - \sum_{\varphi \in F} \varphi f \right| \quad (213)$$

Since f is bounded, we can choose M such that $|f| \leq M$. So equation 213 is less than or equal to

$$M \int_A \left(1 - \sum_{\varphi \in F} \varphi \right) = M \int_A \sum_{\varphi \in \Phi \setminus F} \varphi \quad (214)$$

Since

$$\forall \varphi \in \Phi \setminus F, \forall c \in C, \varphi(c) = 0 \quad (215)$$

we have that equation 214 is less than or equal to

$$MV(A \setminus C) \leq M\epsilon \quad (216)$$

□

As I mentioned last time, you can do this in another way which perhaps looks more like the improper integral from first year calculus, and that's integrating with bigger and bigger sets. Let me state that, but I won't prove it, and rather put it on the problem set:

Theorem 12. Let A be open, $f : A \rightarrow \mathbb{R}$ be locally bounded and let the set of discontinuities of f have measure zero. Write

$$A = \bigcup_{n \in \mathbb{N}} C_n \quad (217)$$

where each C_n is compact and Jordan Measurable and satisfy $C_n \subset \text{Int } C_{n+1}$. Part of the exercise is to show that we can do this. Then f is integrable on A if and only if

$$\left\{ \int_{C_i} |f| : n \in \mathbb{N} \right\} \quad (218)$$

is bounded. Note that if this is the case, the sequence

$$\left\{ \int_{C_i} f : n \in \mathbb{N} \right\} \quad (219)$$

converges absolutely, and we can write

$$\int_A f = \lim_{n \rightarrow \infty} \int_{C_n} f \quad (220)$$

Proof. Exercise. □

It's good to appreciate that the improper integral which we defined using a partition of unity can be expressed in this way. It's also really good to do the above exercise as an exercise in using partitions of unity.

There's a little gap in the treatment of partitions of unity last time which I want to fill here. In the first case, where A was compact, we defined

$$\psi_1, \dots, \psi_p \quad (221)$$

such that $\psi_1 + \dots + \psi_p > 0$ on an open set $U \supseteq A$. We then let

$$\varphi_i = \frac{\psi_i}{\psi_1 + \dots + \psi_p} \quad (222)$$

This is all well and good, but individual φ_i might be defined on sets that go outside U , and we have to worry about dividing by zero to guarantee that each bump function is \mathcal{C}^∞ everywhere. To deal with this, we can simply multiply everything by another bump function (we can find one which is 1 on A and zero outside a compact set U).

Let $f(x)$ be a \mathcal{C}^∞ bump function such that $f = 1$ on A , $f = 0$ outside a compact subset of U . Then the set of functions

$$\{f \cdot \varphi_i : i \in 1, \dots, p\} \quad (223)$$

is a partition of unity as required.

Before we finish, I want to say something about integration by substitution, or change of variables.

3.6 Change of Variables

Recall the single-variable integration by substitution formula

$$\int_{g(a)}^{g(b)} f = \int_a^b (f \circ g)g' \quad (224)$$

This holds whenever “both sides make sense”, e.g. if f is continuous and g is continuously differentiable. The hypotheses, however, could be weaker. When we talk about integration by substitution, our g is going to be the change of variables. But change of variables means like “invertible”. So not just any g , but in particular 1-to-1. I wanted to point out how, in terms of our formula, we should think about this in the case where g is one-to-one. So suppose it is so. Then equation 224 can be rewritten as

$$\int_{g([a,b])} f = \int_{[a,b]} (f \circ g)|g'| \quad (225)$$

since otherwise the equation would not be correct if g was decreasing. So this is what our change-of-variables formula is going to look like. Let's formalize this into a theorem:

Theorem 13 (Change of Variables). *Let $A \subset \mathbb{R}^n$ be open, $g : A \rightarrow \mathbb{R}^n$ be one to one, continuously differentiable and let, for all $x \in A$, $g'(x) \neq 0$. Then*

$$f : g(A) \rightarrow \mathbb{R} \quad (226)$$

is integrable if and only if

$$f \circ g |\det g'| \quad (227)$$

is integrable on A . In this case,

$$\int_{g(A)} f = \int_A f \circ g |\det g'| \quad (228)$$

3.6.1 Examples

Let's look at some examples, starting with polar coordinates: we use coordinates $r \in \mathbb{R}_0^+, \theta \in [0, 2\pi]$ and write

$$x = r \cos \theta, y = r \sin \theta \quad (229)$$

We have

$$D = \frac{\partial(x, y)}{\partial(r, \theta)} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} \implies \det D = r \quad (230)$$

So we can write

$$\iint_A f(x, y) dx dy = \iint_A f(r \cos \theta, r \sin \theta) r dr d\theta \quad (231)$$

We now examine a corollary of Theorem 13.

Corollary. *Let $A \subset C \subset \mathbb{R}^n$ where A is open, C is compact and Jordan-measurable and $C \setminus A$ has measure zero. If g is a continuously differentiable function from a neighborhood of C to \mathbb{R}^n which satisfies the conditions of theorem 13 on A , then*

$$f : g(C) \rightarrow \mathbb{R} \quad (232)$$

is integrable if and only if

$$f \circ g |\det g'| \quad (233)$$

is integrable on C , and in this case

$$\int_{g(C)} f = \int_C f \circ g |\det g'| \quad (234)$$

Lemma 3. *Assume $A \subset \mathbb{R}^n$ is open and $g : A \rightarrow \mathbb{R}^n$ is continuously differentiable. If $B \subset A$ has measure zero, then $g(B)$ has measure zero.*

Proof. Enough to prove that $g(B \cap C)$ has measure zero for any $C \subset A$ compact, since A has an exhaustion by countably many compact sets $C_1 \subset C_2 \subset \dots$

To do so, remember that a countable intersection of measure 0 sets is measure 0. Using C_1 , which is more than uniformly continuous, we have that

$$\forall x \in C, \forall y \in U, |g(x) - g(y)| \leq c|x - y| \quad (235)$$

where U is some neighborhood of C . So g maps a ball of radius ϵ to a ball of radius $c\epsilon$. \square

We now proceed to prove Corollary 3.6.1

Proof. $g(C) \setminus g(A) \subseteq g(C \setminus A)$, and so it is of measure zero. We hence have that

$$\int_{g(A)} f = \int_{g(C)} f \quad (236)$$

$$\int_A (f \circ g) |\det g'| = \int_C (f \circ g) |\det g'| \quad (237)$$

giving the desired equality by Theorem 13 \square

Let's move on to another example: what are called spherical coordinates. We use coordinates $r \in \mathbb{R}_0^+, \phi \in [0, 2\pi], \theta \in [0, \pi]$ where

$$x = r \cos \phi \sin \theta, y = r \sin \phi \sin \theta, z = r \cos \theta \quad (238)$$

We have... this is going to hurt...

$$D = \frac{\partial(x, y, z)}{\partial(r, \theta, \phi)} = \begin{pmatrix} \cos \phi \sin \theta & r \cos \phi \cos \theta & -r \sin \phi \sin \theta \\ \sin \phi \sin \theta & r \sin \phi \cos \theta & r \cos \phi \sin \theta \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix} \implies \det D = r^2 \sin \theta \quad (239)$$

Now, for some problems:

1. Find the volume of an ellipsoid

$$E = \left\{ (x, y, z) \in \mathbb{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1 \right\} \quad (240)$$

We can make a change of variable to “transform” this into a sphere. Specifically, we can use the transformation

$$g : \mathbb{R}^3 \rightarrow \mathbb{R}^3, (x, y, z) \mapsto (x/a, y/b, z/c) \quad (241)$$

to take the ellipsoid to the unit ball, where $(u, v, w) = g(x, y, z)$. So

$$\iiint_{B_1(0)} 1 du dv dw = \iiint_E \frac{1}{abc} dx dy dz = \frac{1}{abc} \iiint_E 1 dx dy dz = \frac{4}{3}\pi \implies V(E) = \frac{4}{3}\pi abc \quad (242)$$

Now what if we didn’t know the volume of a ball of radius 1? We could figure it out using change of variables, using spherical coordinates:

$$\int_0^{2\pi} \int_0^\pi \int_0^1 r^2 \sin \theta dr d\theta d\phi = \int_0^1 r^2 dr \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi = \frac{1}{3} \cdot 2 \cdot 2\pi = \frac{4\pi}{3} \quad (243)$$

Of course, the volume of the sphere of radius r can be treated as a special case of the ellipsoid where $a = b = c = r$, giving volume $\frac{4}{3}\pi r^3$.

2. What’s the simplest kind of change of variable, save the identity? Of course, a linear change of variable. Consider a linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and suppose $A \subset \mathbb{R}^n$ is Jordan measurable and $B = T(A)$. Show that B is Jordan measurable and $V(B) = V(A)|\det T|$. We split this into two cases:

- (a) Assume T is not invertible, i.e. $\det T = 0$. Then the image I of T is a linear subspace of \mathbb{R}^n of dimension less than N , and $B \subset I$. So all of B has n -dimensional measure zero, and hence $V(B) = 0 = V(A)|\det T|$ as desired.
- (b) Assume T is invertible, i.e. $\det T \neq 0$. Then B is Jordan measurable and now by the change of variable theorem

$$V(B) = \int_B 1 = \int_A 1 \cdot |\det T| = |\det T| \int_A 1 = V(A)|\det T| \quad (244)$$

as desired.

3.6.2 Proof of the Change of Variables Theorem

We’re now going to prove the Change of Variables theorem. Let’s begin by considering just one direction:

Theorem 14 (Change of Variables (One Direction)). *Let $A \subset \mathbb{R}^n$ be open and let $g : A \rightarrow \mathbb{R}^n$ be one to one and continuously differentiable such that $\det g'(x) \neq 0$ on A . Then if $f : g(A) \rightarrow \mathbb{R}$ is integrable, $f \circ g|\det g'|$ is integrable on A and*

$$\int_{g(A)} f = \int_A (f \circ g)|\det g'| \quad (245)$$

Note that the converse, and hence the full theorem, follows, since

- As g is invertible f is locally bounded if and only if $f \circ g|\det g'|$ is locally bounded
- The discontinuities of f have measure zero if and only if the discontinuities of $f \circ g|\det g'|$ have measure zero, since g is differentiable, using the lemma from last time.
- We can use g^{-1} to go from $g(A)$ to A .

We still have to show that if $F = f \circ g| \det g'|$ is integrable on A , then f is integrable on $g(A)$. All we have to do is apply the theorem with F, g^{-1} instead of f, g . Then

$$F \circ g^{-1}| \det(g^{-1})'| = f \circ g \circ g^{-1}| \det g' \circ g^{-1}| \frac{1}{|\det g' \circ g^{-1}|} = f \quad (246)$$

as desired.

We can now begin to prove the theorem in a number of steps. We're going to start by looking at a Jordan-measurable open subset U of A with closure in A , e.g. an open rectangle with closure in A . The first thing that we're going to do is show that it's "good enough" to prove the theorem on such an open subset. How? With a partition of unity! That's going to simplify things a lot, but just one thing I want to note first is that if f is integrable on $g(A)$, then f is integrable on $g(U)$ and $f \circ g| \det g'|$ is integrable on U .

We observe that, if U is a Jordan-measurable subset of A with closure in A , then if $f : g(A) \rightarrow \mathbb{R}$ is integrable then f is integrable on $g(U)$ and $f \circ g| \det g'|$ is integrable on U .

We're now going to attempt to prove Theorem 14 in small steps:

1. Suppose \mathcal{O} is an open cover of A by sets like U above. Suppose that for each U , the theorem holds, i.e. f is integrable on $g(U)$ and

$$\int_{g(U)} f = \int_U (f \circ g)| \det g'| \quad (247)$$

Then the theorem is true. What we're saying here is it's good enough to, instead of proving the theorem as stated, just prove it for these Jordan-measurable U 's. So, why is this true?

If we have an open covering \mathcal{O} of A , taking the images of everything in the cover, the set

$$\mathcal{O}' = \{g(U) : U \in \mathcal{O}\} \quad (248)$$

is an open cover of $g(A)$ of some kind, since g is invertible. Now let Φ be a \mathcal{C}^0 partition of unity for $g(A)$ subordinate to \mathcal{O}' .

That means each partition function $\varphi \in \Phi$ is zero outside of a compact subset of some $g(U)$. But notice that, if this is true, then $\varphi \circ g$ must be zero outside of a compact subset of U . This is because g is invertible. This is the last place in the proof that we're going to use the fact that g is one-to-one.

So this tells us that

$$\Phi' = \{\varphi \circ g : \varphi \in \Phi\} \quad (249)$$

forms a partition of unity for A subordinate to \mathcal{O} . So we have that *by our assumption*

$$\int_{g(A)} \varphi \circ g = \int_{g(U)} \varphi \circ f = \int_U (\varphi \cdot f) \circ g| \det g'| = \int_A (\varphi \cdot f) \circ g| \det g'| \quad (250)$$

by the nature of partitions of unity. So, the thing that we're interested in, the integral of f over $g(A)$, can be written, since Φ is a partition of unity, as

$$\int_{g(A)} f = \sum_{\varphi \in \Phi} \int_{g(A)} \varphi f = \sum_{\varphi \in \Phi} \int_A (\varphi \cdot f) \circ g| \det g'| = \int_A f \circ g| \det g'| \quad (251)$$

Note that the same equality as 251 with $|f|$ in place of f shows that the RHS converges, which is our definition of integrability.

We could make the same conclusion that the theorem is true not by assuming this for an open covering of A but rather looking at the same things for an open covering of $g(A)$. That is, it's equivalent for this claim to suppose that if we have an open covering \mathcal{O} of $g(A)$ by open subsets of this kind (Jordan measurable, closure lying in $g(A)$), then, for all $V \in \mathcal{O}$ and all f integrable on V ,

$$\int_V f = \int_{g^{-1}(V)} (f \circ g)| \det g'| \quad (252)$$

2. It's enough to prove that for every open rectangle V in $g(A)$ with closure in $g(A)$ we have

$$\int_V f = \int_{g^{-1}(V)} (f \circ g) |\det g'| \quad (253)$$

whenever f is integrable on V . If I just said that, that would be no further reduction at all, since this is just what we already said. But that's not what I'm going to say here. What I'm going to say is that it's enough to prove this when f is the constant function 1. So this is a big reduction.

Again, when I say it's "enough" to prove that, the meaning is that if you prove this, then the theorem follows, i.e. it is enough to prove the theorem. Why is this the case, however?

Let P be a partition of V . For every subrectangle $S \in P$, let f_S be the constant function $m_S(f)$. We have

$$\mathcal{L}(f, P) = \sum_{S \in P} m_S(f) V(S) = \sum_{S \in P} \int_{\text{Int } S} f_S \quad (254)$$

By this assumption for constant functions, we can rewrite equation 254 as

$$\sum_S \int_{g^{-1}(\text{Int } S)} f_S \circ g |\det g'| \leq \sum_S \int_{g^{-1}(\text{Int } S)} f \circ g |\det g'| \leq \int_{g^{-1}(V)} f \circ g |\det g'| \quad (255)$$

So what does this tell us about the integral of f over V ? We have that

$$\int_V f = \sup_P \mathcal{L}(f, P) \leq \int_{g^{-1}(V)} f \circ g |\det g'| \quad (256)$$

The same argument using upper sums gives \geq .

3. If the theorem is true for $g : A \rightarrow \mathbb{R}^n$, and $h : g(A) \rightarrow \mathbb{R}^n$, with the same hypotheses (one-to-one, \mathcal{C}^1 , $\det h' \neq 0$) then it is true for $h \circ g$. Why?

$$\int_{(h \circ g)(A)} f = \int_{g(A)} (f \circ h) |\det h'| = \int_A ((f \circ h) \circ g) |\det h' \circ g| |\det g'| = \int_A f \circ (h \circ g) |\det (h \circ g)'| \quad (257)$$

by the Chain Rule.

4. The theorem is true if g is an (invertible) linear transformation. We already showed that, for any g , it's good enough to prove that the theorem is true on an open rectangle U with the constant function $f = 1$, i.e. that if U is an open rectangle in A , then

$$\int_{g(U)} 1 = \int_U |\det g'| \iff \text{vol}(g(U)) = |\det g| \text{vol}(U) \quad (258)$$

Note that we've already proved this in linear algebra, using the volume of parallelepipeds. But let's do it analytically.

We will make use of the fact that any invertible linear transformation (i.e. invertible $n \times n$ matrix) is a composite of elementary transformations (elementary matrices). Last time we showed that if the theorem is true for g, h , then it is true for $g \circ h$. Hence, all we need to do is show that the theorem is true for an arbitrary elementary transformation (matrix). We have 3 kinds of elementary matrices:

- A matrix which has all 1's on the diagonal, except for λ in one place. In this case, $\det g = \lambda$, so we could use Fubini's theorem to scale by λ , giving $\text{vol}(g(U)) = \lambda \text{vol}(U) = |\det g| \text{vol}(U)$ as desired.
- A matrix which has all 1's on the diagonal, except for two adjacent rows which have $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ on the diagonal. We have $\det g = \pm 1$. Since we're just changing two coordinates, the integral remains unchanged.

- The identity matrix, plus 1 in a position i, j with $i \neq j$. In this case, the determinant is just 1. Furthermore, we're just translating the top of our rectangle into a parallelogram. Fubini's theorem tells us that the volume is the same since the limits in terms of y are the same, and for fixed y we are integrating over a shifted interval of the same length, the shifting not having an effect since we are integrating a constant.

We can now return to the proof of the theorem: by 3 and 4, we can assume that, given $a \in A$, $g'(a) = I$, where I is the identity. Why? Let $T = g'(a)$ be an invertible linear transformation. Then we can write

$$g = T \circ (T^{-1} \circ g) \quad (259)$$

By 4, this is true for T . By 3, if this is true for T and $T^{-1} \circ g$ then it's true for g . So it's enough to prove for the case

$$(T^{-1} \circ g)' = I \quad (260)$$

To complete the proof, we use induction on n :

- $n = 1$: by integration by substitution
- Assume it's true for $n - 1$. It's enough to show that, for all $a \in A$, we can find Jordan measurable open $U \subset A$ containing a , with closure in A , so that the theorem is true on U when $f = 1$.

We can assume $g'(A) = I$. We'll express g as the composition of 2 mappings, each of which changes at most $n - 1$ coordinates. There are a number of ways to do this. One is defining $h : A \rightarrow \mathbb{R}^n$ by

$$h(x) = (g_1(x), \dots, g_n(x), x_n) \quad (261)$$

Why is this a change of variable? $h'(a) = I$, so there is an open neighborhood U' of a on which h is one to one and $\det h'(a) \neq 0$. Now define

$$k : h(U') \rightarrow \mathbb{R}^n, k(y) = (y_1, \dots, y_{n-1}, g_n(h^{-1}(y_n))) \quad (262)$$

We have $g = k \circ h$,

$$g'(a) = k'(h(a))h'(a) = I \implies k'(h(a)) = I \quad (263)$$

So there is an open neighborhood V of $h(a)$ in U' on which k is one to one and

$$\det k'(g) = 0 \quad (264)$$

Let $U = h^{-1}(V)$. On U , $g = k \circ h$, and $h : U \rightarrow V$, $k : V \rightarrow \mathbb{R}^n$.

So we have to prove the theorem for h and for k . The proof is going to be the same for both cases, so let's do it for the harder case, h , which changes $n - 1$ coordinates (versus k , which changes only one).

We're going to use Fubini's theorem: we'll integrate with respect to x_n . We've previously found that it's enough to show that, for every $a \in A$, there is a Jordan measurable open neighborhood U of a with closure A so that the result is true on U in the case where we are integrating the function $f = 1$. Furthermore, we have shown that we can assume that $g'(a) = I$. We also expressed g as a composite of two functions, each of which changes fewer than n variables. More precisely, we wrote

$$g = (x \mapsto (g_1(x), \dots, g_{n-1}(x), x_n)) \circ (y \mapsto (y_1, \dots, y_{n-1}, g_n(h^{-1}(y)))) \quad (265)$$

We will proceed by induction, using the single variable case as a base case.

We can work with any Jordan-measurable open neighborhood of a , so let's work with an open rectangle. Consider specifically an open rectangle W containing a with closure in A . Since we're separating one coordinate from the other $n - 1$ coordinates, let's write W as $D \times (a_n, b_n)$ where D is an open rectangle in \mathbb{R}^{n-1} . Note that the transformation h does not change the n^{th} coordinate of its input. Using this, we can consider the following:

$$\int_{h(W)} 1 \quad (266)$$

For a fixed x_n , let's consider the function given by h , depending only on $1, \dots, n-1$, written

$$h_{x_n} : D \rightarrow \mathbb{R}^{n-1}, h_{x_n}(x_1, \dots, x_{n-1})(g_1(x_1, \dots, x_n), \dots, g_{n-1}(x_1, \dots, x_n)) \quad (267)$$

This function is one to one, since we know that h is one to one, and h is just h_{x_n} with x_n appended (for each x_n), so two different values at a given x_n have got to go into different points. On the other hand, what is the determinant

$$(\det h'_{x_n})(x_1, \dots, x_{n-1}) \quad (268)$$

Well, it's the same as the determinant of h' , since the last row of h' is just the standard vector e_n . So let's go back: first of all,

$$h(D \times \{x_n\}) = h_{x_n}(D) \times \{x_n\} \quad (269)$$

since, like we said before, the n^{th} coordinate is unaffected. So we can proceed using Fubini's theorem,:

$$\int_{a_n}^{b_n} \left(\int_{h_{x_n}(D)} 1 \right) dx_n \quad (270)$$

By induction, this is equal to

$$\int_{a_n}^{b_n} \int_D 1 |\det h'_{x_n}| = \int_{a_n}^{b_n} \int_D |\det h'| = \int_{D \times [a_n, b_n]} 1 \cdot |\det h'| = \int_W 1 \cdot |\det h'| \quad (271)$$

By strong induction, the theorem is also true for k , giving the desired result.

Now that we finished this, what to do? We've been talking about the integral on \mathbb{R}^n . In the rest of the course, we're going to be talking about the integral but on a manifold.

4 Manifolds

4.1 What is a manifold?

TODO: rewrite, copy

Consider the following definitions of a manifold:

Definition 9. A k -dimensional manifold is a set $M \subseteq \mathbb{R}^n$ such that for all $a \in M$, there is a C^r mapping $f : U \rightarrow \mathbb{R}^{n-k}$ defined on an open neighborhood U of a such that $M \cap U = f^{-1}(0)$. f has rank $n - k$ at every $y \in M \cap U$.

Definition 10. A k -dimensional manifold is a set $M \subseteq \mathbb{R}^n$ such that every point of M has an open neighborhood U in \mathbb{R}^n such that $M \cap U$ is the graph of a C^r function $z = g(y)$, $z = (z_1, \dots, z_{n-k})$, $y = (y_1, \dots, y_k)$ where

$$(y, z) \neq (x_{\sigma(1)}, \dots, x_{\sigma(n)}) \quad (272)$$

where σ is a permutation of $1, \dots, n$.

Definition 11. A k -dimensional manifold is a set $M \subseteq \mathbb{R}^n$ such that for all $a \in M$, there is an open neighborhood U of a , an open set $V \subset \mathbb{R}^n$ and a C^r diffeomorphism (a C^r mapping with a C^r inverse) $h : U \rightarrow V$ such that

$$h(M \cap U) = V \cap (\mathbb{R}^k \times \{\mathbf{0} \in \mathbb{R}^{n-k}\}) = \{y = (y_1, \dots, y_n) \in V : y_{k+1} = \dots = y_n\} \quad (273)$$

Claim 1. The definitions given for a manifold are equivalent

Proof. • 9 \implies 10: By the implicit function theorem

• 10 \implies 9: Let $h(y, z) = z - g(y)$.

- 10 \implies 11: Let $h(y, z) = (y, z - g(y))$
- 11 \implies 9: Let $f(x) = (h_{k+1}(x), \dots, h_n(x))$

□

To illustrate the first definition in action, consider the example of the unit sphere $S^{n-1} \subset \mathbb{R}^n$, given by

$$S^{n-1} = \{x = (x_1, \dots, x_n) : f(x) - 1 = 0\} \text{ where } f(x) = x_1^2 + \dots + x_n^2 \quad (274)$$

We have

$$\forall x \in S^{n-1}, \text{grad } f(x) = 2(x_1, \dots, x_n) = 2x \neq 0 \quad (275)$$

so f has rank 1 at every point of S^{n-1} . Hence S^{n-1} is a manifold.

4.2 Tangent Spaces

We now discuss the tangent space TM_a or M_a at a point $a \in M$. If we have a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$, then $Df(a)$ is a linear transformation, a function from \mathbb{R}^n to \mathbb{R}^p . But we're going to consider it as a function from \mathbb{R}_a^n , the copy of \mathbb{R}^n at a , to $\mathbb{R}_{f(a)}^p$, the copy of \mathbb{R}^p at $f(a)$. To be clear, these copies of $\mathbb{R}^n, \mathbb{R}^p$ are vector spaces we attach to the points $a, f(a)$, with these points identified with the origin.

1. The definition of the tangent space according to Definition 9 is $\text{Ker } Df(a) = Df(a)^{-1}(0)$, a vector subspace of \mathbb{R}_a^n
2. To think of the tangent space in terms of Definition 10, write $x = (y, z), a = (b, c)$. Then we can think of the tangent space as an affine space through the point (b, c) by writing

$$z - c = Dg(b)(y - b) \quad (276)$$

But we want to think of it as a vector space with a as the origin, a vector subspace of \mathbb{R}_a^n . So we should write it as

$$Z = Dg(b)(Y) \quad (277)$$

where $Z \in \mathbb{R}_{f(a)}^p, Y \in \mathbb{R}_a^n$. This is because $f(y, g(y)) = 0$, so

$$Df(a) \begin{pmatrix} I \\ Dg(b) \end{pmatrix} = 0 \implies \text{Ker } Df(a) \supseteq \text{Im} \begin{pmatrix} I \\ Dg(b) \end{pmatrix} \quad (278)$$

But in fact, they're equal, because their dimensions are the same.

So the kernel of $Df(a)$, and the image of $\begin{pmatrix} I \\ Dg(b) \end{pmatrix}$, are both equal to the set of points such that

$$Z = Dg(b)(Y) \quad (279)$$

3. In terms of the Definition 11, we can write the tangent space as

$$Dh(a)^{-1}(\mathbb{R}^k \times \{\mathbf{0} \in \mathbb{R}^{n-k}\}) \quad (280)$$

4.3 Definition by Coordinate Charts

Let's try defining a (k -dimensional) manifold in another way: we want every point $a \in M$ to have a "neighborhood" around it in M which "looks like" \mathbb{R}^k .

Definition 12 (Coordinate Charts). *A k -dimensional manifold $M \subset \mathbb{R}^n$ is a set such that for all $a \in M$ there is an open neighborhood U of a (in \mathbb{R}^n , **not** in M) and an open set $W \subset \mathbb{R}^k$ and a C^r mapping $\varphi : W \rightarrow \mathbb{R}^n$ such that*

- φ is a bijection

- $\varphi(W) = M \cap U$
- φ has rank k at every point of W , i.e. $D\varphi(x)$ has rank k for every $x \in W$.
- We'd like to say " $\varphi^{-1} : \varphi(W) = M \cap U \rightarrow W$ " is continuous. But we don't know how to say it, because we haven't given M a topology. So we can't talk about a function defined on a piece of M being continuous. So the way we're going to say it is we're going to build the definition of the topology into the phrase that we'll say instead of this.

We require instead that $(\varphi^{-1})^{-1}(\Omega)$, i.e. $\varphi(\Omega)$, is $\varphi(W)$ intersected with an open set of \mathbb{R}^n for all Ω open in W .

Note this first and last condition on φ are equivalent to requiring that φ is a homeomorphism between the neighborhood $M \cap U$ and W .

We'll now show that Definition 12 implies Definition 11:

Proof. Say $a = \varphi(b)$ for some $b \in W$. We can assume

$$\frac{\partial(\varphi_1, \dots, \varphi_k)}{\partial(y_1, \dots, y_k)} \quad (281)$$

has rank k on W . Define $\psi : W \times \mathbb{R}^{n-k} \rightarrow \mathbb{R}^n$ by

$$(y, z) \mapsto \varphi(y) + (0, z) \quad (282)$$

Then we get the block matrix

$$\psi'(y, z) = \begin{pmatrix} \frac{\partial(\varphi_1, \dots, \varphi_k)}{\partial(y_1, \dots, y_k)} & 0 \\ * & I \end{pmatrix} \quad (283)$$

This shows that ψ has rank n for all $y \in W$, since it's determinant is nonzero. But that means that we can apply the inverse function theorem. So by the inverse function theorem, there are open neighborhoods V'_1 of $(b, 0)$ and U'_1 of $\psi(b, 0) = \varphi(b) = a$ such that $\psi : V'_1 \rightarrow U'_1$ has a C^r inverse $\psi^{-1} : U'_1 \rightarrow V'_1$.

We have that

$$\psi^{-1}(\varphi(y)) = (y, 0) \in V'_1 = \varphi(W) \cap \tilde{U} \quad (284)$$

where U is open in \mathbb{R}^n . Take $U_1 = U'_1 \cap \tilde{U}$ and $V_1 = \psi^{-1}(U_1)$. We have

$$M \cap U_1 = \{\varphi(y) : (y, 0) \in V_1\} \quad (285)$$

So

$$h = \psi^{-1}|_{U_1} \quad (286)$$

satisfies the conditions implied by (3) since

$$h(M \cap U_1) = \psi^{-1}(M \cap U_1) = \{(y, 0) : (y, 0) \in V_1\} = V_1 \cap (\mathbb{R}^k \times \{\mathbf{0}\}) \quad (287)$$

□

This is quite a delicate topological argument, and gi**TODO: finish this sentence**

We begin by looking at the following two definitions for the tangent space to a manifold in terms of definitions 11 and 12 given for a manifold in the previous lecture

Definition 13. $TMa = Dh(a)^{-1}(\mathbb{R}^k \times \{\mathbf{0}\})$

Definition 14. $TMa = D\varphi(b)(\mathbb{R}^k_b)$

We're forgetting about the first two definitions, but why are *these* two definitions equivalent? We used h to find a coordinate map φ by taking φ to be h^{-1} restricted to $\mathbb{R}^k \times \{\mathbf{0}\}$. Note that definition 14 is independent from the choice of φ . To understand why, consider two maps $\varphi_1(b_1) = a = \varphi_2(b_2)$ which take some intersecting open subsets M_1, M_2 of M to spaces W_1, W_2 . Then

$$\varphi_1 = \varphi_2 \circ (\varphi_2^{-1} \circ \varphi_1) \quad (288)$$

on $M_1 \cap M_2$, giving

$$D\varphi_1(b_1) = D\varphi_2(b_2) \circ D(\varphi_2^{-1} \circ \varphi_1)(b_1) \quad (289)$$

I want to think of this in a different way now. I want to give another definition for the tangent space that doesn't depend on h , doesn't depend on φ , doesn't depend on any data other than M . We give the following:

Definition 15.

$$TM_a = \{\text{tangent vectors } \gamma'(0) \text{ of } \mathcal{C}^\infty \text{ curves } \gamma : (-\delta, \delta) \rightarrow M \text{ s.t. } \gamma(0) = a\} \quad (290)$$

So what does this even mean? Recall the following definition:

Definition 16. A \mathcal{C}^1 curve in \mathbb{R}^n is a \mathcal{C}^1 mapping, for some open interval (c, d) ,

$$\gamma : (c, d) \rightarrow \mathbb{R}^n, \gamma(t) = (\gamma_1(t), \dots, \gamma_n(t)) \quad (291)$$

If we fix a point on γ , what's the tangent vector to γ at $\gamma(t_0)$? It's $\gamma'(t_0) \in T\mathbb{R}^n_{\gamma(t_0)}$. We now generalize the above definition

Definition 17. Let $M \subset \mathbb{R}^n$ be a \mathcal{C}^r submanifold of dimension k . We say γ is a \mathcal{C}^1 curve in M if $\gamma((a, b)) \subset M$ for some open interval (a, b) . In this case $\gamma'(t_0) \in TM_{\gamma(t_0)}$.

So why is $\gamma'(t_0)$ an element of $TM_{\gamma(t_0)}$? In terms of a previous definition,

$$TM_a = \text{Ker } Df(a) \quad (292)$$

In our case, we have

$$f(\gamma(t)) = 0 \implies Df(\gamma(t_0)) \cdot \gamma'(t_0) = 0 \implies \gamma'(t_0) \in \text{Ker } Df(a) = TM_a \quad (293)$$

So, considering this, why is definition 16 the tangent space? Let's think about that, first of all, in a simple special case. What if $M = \mathbb{R}^k$? Take $a \in \mathbb{R}^k$. We are simply asking that every vector $v \in \mathbb{R}^k_a$ through the point a is the tangent vector to some curve through a , i.e. is $\gamma'(0)$ where $\gamma(t) = a + tv$. So if $M = \mathbb{R}^k$, for sure that's the right definition of the tangent space.

So now why is this true in general? Which definition of the tangent space do you think we should use? Well, the coordinate chart definition, i.e. definition 14, is the one that best fits our purposes. In general, consider a coordinate chart φ for M at a . Say we have a \mathcal{C}^1 curve in W such that $\delta(0) = b$. Then $\gamma(t) = (\varphi \circ \delta)(t)$ is a \mathcal{C}^1 curve in M such that $\gamma(0) = a$. So using the Chain Rule,

$$\gamma'(0) = D\varphi(b) \circ \delta'(0) \quad (294)$$

But any \mathcal{C}^1 curve $\gamma(t)$ in M such that $\gamma(0) = b$ is of the form $\gamma = \varphi \circ \delta$ where δ is a \mathcal{C}^1 curve in W . We get this from the description of φ^{-1} as restricted to M of a \mathcal{C}^r function h .

What's nice about definition 16 is that it depends only on M , not on any external information. So now we know what manifolds and tangent spaces are.

This course is somewhat like first year calculus, in that first you do differentiation and then you do integration, and then you put them together with the fundamental theorem of calculus and examine all the interesting consequences. And we are not at the end of quite an important section of the course about the implicit function theorem. Soon, we'll stop and begin learning about integration, and then later return and put the two together to obtain some interesting results, including Stokes theorem.

But we have one thing left to do. Usually, in mathematics, we don't just define spaces, we also define the functions between them. And we should do that for manifolds: if we have a mapping between two manifolds, what does it mean for this mapping to be continuous? Furthermore, we will introduce the slightly more general concept of a manifold with boundary.

4.4 Functions Between Manifolds

TODO: this

4.5 Manifolds with Boundary

Recall the following definitions:

Definition 18. A half space $\mathbb{H}^k \subset \mathbb{R}^k$ is given by

$$\{x = (x_1, \dots, x_k) : x_k \geq 0\} \quad (295)$$

Definition 19. A subset $M \subseteq \mathbb{R}^n$ is a k -dimensional C^r manifold with boundary if for every $a \in M$, there is an open neighborhood U of a in \mathbb{R}^n , an open subset V of \mathbb{R}^n and a C^r diffeomorphism $h : U \rightarrow V$ such that either

(1) The usual condition for a manifold,

$$h(M \cap U) = V \cap (\mathbb{R}^k \times \{\mathbf{0} \in \mathbb{R}^{n-k}\}) \quad (296)$$

(2) The manifold looks like a half space around a , i.e.

$$h(M \cap U) = V \cap (\mathbb{H}^k \times \{\mathbf{0} \in \mathbb{R}^{n-k}\}) = \{y = (y_1, \dots, y_n) : y_1, \dots, y_k \geq 0 \wedge y_{k+1} = \dots = y_n = 0\} \quad (297)$$

One thing you could ask is whether both conditions (1), (2) could be satisfied by the same point a . The answer is no. To show why, suppose $h_1 : U_1 \rightarrow V_1$ satisfies (1) and $h_2 : U_2 \rightarrow V_2$ satisfies (2) for some a . How can we get a contradiction? If we take

$$h_2 \circ h_1^{-1} : V_1 \rightarrow V_2 \quad (298)$$

we get a C^r diffeomorphism in an open set containing a taking an open subset of \mathbb{R}^k onto an open subset of \mathbb{H}^k which is *not* open in \mathbb{R}^k . What we're contradicting here is the inverse function theorem, which says there would be open neighborhoods $h_1(a)$ and $h_2(a)$ which are mapped together with an inverse. Alternatively, we could use the topological arguments that diffeomorphisms, which are homeomorphisms, are open maps.

Once we've established this result, we can really distinguish these two kinds of points. That means, precisely, that we can make the following definitions

Definition 20. The boundary of M , written ∂M , is the set of points satisfying condition (2) in definition 19

We're going to be interested in manifolds because they generalize the idea of intervals with endpoints in the real line. In fact, an interval with endpoints is a one dimensional manifold with boundary.

4.6 Multilinear Algebra

TODO: this

4.7 Vector Fields and Differential Forms

TODO: this

4.8 The Differential Operator

TODO: this

5 Integration on Manifolds

We haven't talked about manifolds for a while, but I hope you can remember how to imagine one because I'm too lazy to \LaTeX one. So what's the point of integrating on our imaginary manifold picture? We know how to integrate on a line, and how to integrate now in \mathbb{R}^n , and in both cases we're integrating with some notion of n -dimensional volume. But now, on our k -dimensional manifold, we're going to have to try to figure out how to integrate using a k -dimensional measure of volume on it, which is nontrivial.

Another reason is the following: if we go back to one dimensional calculus, and we want to compute the integral of some function f from a to b , the very important and aptly named Fundamental Theorem of Calculus tells us that if f has a primitive g it's integral is very easy to compute:

$$\int_a^b f = g(b) - g(a) \quad (299)$$

We want to try to do the same thing with a manifold: generalize the fundamental theory of calculus to relate the integral of a function on the manifold to the integral of a function on the boundary. But what sense does that make? A function on a manifold, unlike that on a line, depends on many variables. So we don't have a notion of primitive like in single variable calculus, since an n dimensional function has many partial derivatives.

So those are the two things we're going to have to understand to get started:

- What is this k dimensional measure of volume?
- What is the appropriate notion of a primitive?

But before that, we're going to do a bit of motivation, and consider the case of a one-dimensional manifold, or as you know it, a curve:

5.1 Integration of Parametrized Curves

We now consider the case of a parametrized Curve C in \mathbb{R}^n . Let's start from the beginning: what is a parametrized curve? It means C is the image of a mapping

$$\gamma : [a, b] \rightarrow \mathbb{R}^n \quad (300)$$

We could write, for example, $x = \gamma(t)$. We're going to assume that our curves define something almost like a manifold, except we're going to allow that the curves cross each other. So assume that γ is piecewise C^1 . What this means is, except for some finite number of points (since we're working on a compact interval), it's C^1 . Furthermore, we assume that except for isolated points, $\gamma' \neq 0$ and γ is one to one. We don't want to allow, for example,

$$\gamma(t) : [0, 3\pi] \rightarrow \mathbb{R}^2 = t \mapsto (\cos t, \sin t) \quad (301)$$

since we don't want a whole big region (in this case a semicircle) where γ overlaps itself.

So the first item of business is: what's the measure of volume on such a curve? Length, of course. Arc length, specifically. We have that

$$\ell(C) = \int_a^b |\gamma'(t)| dt \quad (302)$$

We're not assuming that people have seen this before, so the question is, why? Let's deal with length in a more concrete, geometric way. How can you do that? In the same way that you define the integral: you try to approximate by something that is easy to compute. So one good way to approximate length is to divide up the integral into a partition and look at the corresponding line segments: we define length as the supremum of the length of polygonal approximations to the curve.

Let's let $P = \{t_0, t_1, \dots, t_k\}$ be a partition of $[a, b]$, and let $\ell_P(C)$ be the length of the associated polygonal approximation:

$$\ell_P(C) = \sum_{j=1}^k |\gamma(t_j) - \gamma(t_{j-1})| \quad (303)$$

We'll say that the curve has length if the supremum of these things exists, and we'll define it as the length. Formally,

Definition 21. *The curve C is rectifiable if*

$$\sup_P \ell_P(C) \tag{304}$$

exists, in which case we define 304 it to be the length of C , written $\ell_P(C)$.

Here's a brief homework exercise: prove that if γ is piecewise \mathcal{C}^1 , then C is rectifiable and $\ell(C)$ is given by equation 304. This is a first year calculus exercise using Riemann sums (I think it's in Spivak), so it shouldn't be too hard.

5.2 Integral of a k -form over a k -cube

TODO: this

5.3 Integration of Differential Forms on Manifolds

TODO: this

5.4 Manifolds with Boundary

TODO: this

5.5 Stoke's Theorem on Manifolds

TODO: this