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:

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

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

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

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

$$f(a) + f'(a)h$$

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

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

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

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

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

$$f(x) = f(x_1, ..., x_n)$$

Such a function is differentiable at a point $a = (a_1, ..., a_n)$ if "there's a thing that plays the role of the derivative." We could think about a deriative 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

$$\lambda : \mathbb{R}^n \to \mathbb{R}$$

such that

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

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

$$\lambda(h) = \begin{pmatrix} \lambda_1 & ... & \lambda_n \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$

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 it's 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 agebraic 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

$$|h| = \sqrt{h_1^2 + ... + h_n^2}$$

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

$$|x - y|$$

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

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

$$||x|| = \max\{|x_1|, ..., |x_n|\}$$

$$|||x||| = \sum_{i=1}^n |x_i|$$

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

$$(0, r), (0, -r), (r, 0), (-r, 0)$$

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

$$||x|| \leq |x| \leq |||x|||$$

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

Now how do we show that

$$|||x||| \leq a|x|$$

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

$$|||x||| \leq \sqrt{n}|x|$$

or, analytically,

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

where

$$u_i(x) = \text{sgn}(x_i) = \left\{ \begin{array}{rl} 1 & \text{if} \quad x_i \geq 0 \\ -1 & \text{if} \quad x_i < 0 \end{array} \right.$$

Now, how do we show that

$$|x| \leq a||x||$$

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

1.3.1 Linear Transformations

TODO: this

1.4 Topology of $\mathbb{R}^n$

TODO: this

1.4.1 Functions and Continuity

TODO: this

1.4.2 Compactness

TODO: this

2 Differentiation

2.1 Definitions

TODO: this

2.2 Basic Results

TODO: this

2.3 The Inverse Function Theorem

TODO: this

2.4 The Implicit Function Theorem

TODO: this

2.5 Tangent Space

We begin with some examples:

1. Consider an ellipsoid

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

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

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

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

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 <u>orthogonal</u> at (a, b) if

$$\frac{\partial f}{\partial x}\frac{\partial g}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial g}{\partial y} = 0$$

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

$$\frac{\partial f}{\partial x}\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x} = 0$$

   at (a, b).

3. Consider the family of parabolas

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

   Here, p is the parameter. These parabolas open along the x axis, either to the right or two 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

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

   intersect where

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

   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

$$f_x g_x + f_y g_y = 0$$

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

$$f_x g_x = (-2p_1)(-2p_2) = 4p_1 p_2$$

And what's the product of the y derivatives?

$$f_y g_y = (2y)(2y) = 4y^2$$

So we want to see that

$$f_x g_x + f_y g_y = 4p_1 p_2 + 4y^2$$

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:

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

So we have

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

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

2.6 Extremum Problems

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

$$f : U \to \mathbb{R}$$

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

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

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,

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

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 we want to maximize

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

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

$$|\mathbf{e}| |\operatorname{grad} f(a)| \cos\theta = |\operatorname{grad} f(a)| \cos\theta$$

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

    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

$$f(x, y) = x^2 - y^2$$

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 vetices $P_i = (x_i, y_i)$. Find a fourth point $P = (x, y)$ such that the sum of distances to each P_i is as small as possible. We have

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

where

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

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:

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

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

i.e. if

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

TODO: rest

2.7 Multivariate Taylor Series

TODO: this

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 1. *If $f : A \to \mathbb{R}$ is a bounded function on a closed rectangle $A \subset \mathbb{R}^n$, f is <u>integrable</u> on A if, where $\mathcal{P}$ is the set of partitions of A,*

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

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

$$\int_A f$$

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 1. *If $f : A \to \mathbb{R}$ 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 apartition P of A such that*

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

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

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

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

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

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

$\square$

Just like in first year. And also, the kind of first calculations of itnegrals 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

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

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

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

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

2. Let $f : [0,1] \times [0,1] \to \mathbb{R}$. Then
$$f(x,y) = \begin{cases} 0 & \text{if } x \in \mathbb{Q} \\ 1 & \text{otherwise} \end{cases}$$

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
$$f,g : A \to \mathbb{R}$$

Then

1. $f + g$ is integrable and
$$\int_A (f+g) = \int_A f + \int_A g$$

2.
$$f \le g \implies \int_A f \le \int_A g$$

3. $f \cdot g$ is integrable

4. $|f|$ is integrable and
$$\int_A |f| \ge \left| s22 \int_A f \right| \ge \int_A f$$

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

$$A = [a, b] \times [c, d] \subseteq \mathbb{R}^2 \tag{1}$$

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

$$z = f(x, y) \geq 0 \tag{2}$$

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,

$$h(x) = \int_c^d g_x(y)dy = \int_c^d f(x, y)dy \tag{3}$$

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

$$\int_A f = \int_a^b h = \int_a^b \left(\int_c^d f(x, y)dy \right) dx \tag{4}$$

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 $f : A \to \mathbb{R}$ 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 supermum of lower sums and the infimum of lower sums, which is what we'll do.

We'll define the <u>lower</u> and <u>upper integrals</u> of f on A, L $\int_A f$ and 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 1. *Suppose $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^n$ are closed rectangles and $f : A \times B \to \mathbb{R}$ is integrable. For all $x \in A$, define*

$$g_x : B \to \mathbb{R}, g_x(y) = f(x, y) \tag{5}$$

Set

$$\mathcal{L}(x) = \mathrm{L} \int_B g_x = \mathrm{L} \int_B f(x, y)dy, \mathcal{U}(x) = \mathrm{U} \int_B g_x = \mathrm{U} \int_B f(x, y)dy \tag{6}$$

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

$$\int_{A \times B} f = \int_A \mathcal{L} = \int_A \mathcal{U} = \int_A \left(\mathrm{L} \int_B f(x, y)dy \right) dx = \int_A \left(\mathrm{U} \int_B f(x, y)dy \right) dx \tag{7}$$

Before we prove this, some remarks:

1. We also have

$$\int_{A \times B} f = \int_B \left(\mathrm{L} \int_A (f(x, y)dx) \right) dy = \int_B \left(\mathrm{U} \int_A (f(x, y)dx) \right) dy \tag{8}$$

8

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

$$\int_{A \times B} f = \int_A \left(\int_B f(x, y) dy \right) dx \tag{9}$$

3. If

$$A = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n] \tag{10}$$

we can apply Fubini's theorem repeatedly to obtain

$$\int_A f = \int_{a_n}^{b_n} \left(... \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x_1, ..., x_n) dx_1 \right) dx_2 ... \right) dx_n \tag{11}$$

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

$$\int_C f = \int_A f \chi_C \tag{12}$$

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

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

$$C = [-1, 1] \times [-1, 1] \setminus \{x \in \mathbb{R}^2, x_1^2 + x_2^2 < 1\} \tag{13}$$

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

$$\int_C f = \int_{[-1,1]^2} f \chi_C \tag{14}$$

We can write

$$\chi_C = \begin{cases} 1 & \text{if } -1 \le y \le -\sqrt{1 - x^2} \text{ or } \sqrt{1 - x^2} \le y \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

We can write

$$\int_{-1}^1 f(x, y) \chi_C(x, y) dy = \int_{-1}^{-\sqrt{1-x^2}} f(x, y) dy + \int_{\sqrt{1-x^2}}^1 f(x, y) dy \tag{16}$$

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

$$\int_C f = \int_{-1}^1 \left(\int_{-1}^{-\sqrt{1-x^2}} f(x, y) dy + \int_{\sqrt{1-x^2}}^1 f(x, y) dy \right) dx \tag{17}$$

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

$$\int_C f = \int_0^a \left(\int_y^a f(x, y) dx \right) dy \tag{18}$$

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

$$\int_C f = \int_0^a \left(\int_0^x f(x, y) dy \right) dx \tag{19}$$

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

$$\int_0^a \left(\int_0^x f(y) dy \right) dx \tag{20}$$

9

doesn't simplify very much, but

$$\int_0^a \left(\int_y^a f(y)dx \right) dy = \int_0^a (a-y)f(y)dy \tag{21}$$

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 whuch 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 $\varphi(x) \le \psi(x)$ 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

$$S = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : x \in C, y \in [\varphi(x), \psi(x)]\} \tag{22}$$

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

$$\int_S f = \int_C \left(\int_{\varphi(x)}^{\psi(x)} f(x,y)dy \right) dx \tag{23}$$

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

1. Consider

$$\int_0^2 \int_{y/2}^1 ye^{-x^3} dxdy \tag{24}$$

   There's no hope in proceeding naively, since e^{-x^3} doesn't even have an elementary primitive, i.e. you cannot write down it's 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

$$\int_0^1 \left(\int_0^{2x} ye^{-x^3} dy \right) dx = \int_0^1 \frac{y^2}{2} e^{-x^3} \Big|_0^{2x} dx = 2 \int_0^1 x^2 e^{-x^3} dx = -\frac{2}{3} e^{-x^3} \Big|_0^1 = \frac{2}{3}(1 - e^{-1}) \tag{25}$$

   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

$$\int_2^4 \int_{4/x}^{\frac{20-4x}{8-x}} (y-4)dydx \tag{26}$$

   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:

$$y = \frac{20-4x}{8-x} = 4 + \frac{12}{x-8} \implies x - 8 = \frac{12}{y-4} \iff x = 8 + \frac{12}{y-4} \tag{27}$$

Hence we can write the above as

$$\int_1^2 \left(\int_{4/y}^{8+12/(y-4)} (y-4)dx \right) dy = \int_1^2 (y-4)x \Big|_{4/y}^{8+12/(y-4)} dy = \int_1^2 (y-4) \left(8 + \frac{12}{y-4} - \frac{4}{y} \right) dy \quad (28)$$

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

$$S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \le z^2 \wedge x^2 + y^2 + z^2 \le 1\} \quad (29)$$

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.

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

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

$$\int \int_{x^2+y^2 \le \frac{1}{2}} \left(\int_{\sqrt{x^2+y^2}}^{\sqrt{1-(x^2+y^2)}} zdz \right) dxdy$$

$$= \int \int_{x^2+y^2 \le \frac{1}{2}} \frac{1}{2}(1 - (x^2+y^2) - (x^2+y^2))dxdy \quad (30)$$

$$= \frac{1}{2} \int \int_{x^2+y^2 \le \frac{1}{2}} dxdy - \int \int_{x^2+y^2-\frac{1}{2}} (x^2+y^2)dxdy = \frac{1}{2}\pi\frac{1}{2} = \frac{\pi}{4}$$

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 caclulus that we can write

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)du \quad (31)$$

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

$$x = r\cos\theta, y = r\cos\theta, \theta \in [0, 2\pi], r \in [0, \infty) \quad (32)$$

This gives that

$$\frac{\partial(x, y)}{\partial(r, \theta)} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \implies \det \frac{\partial(x, y)}{\partial(r, \theta)} = r \implies dxdy = drd\theta \quad (33)$$

"generalizing" equation 31. On the other hand, we have that

$$x^2 + y^2 = r^2 \quad (34)$$

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

$$\int_0^{2\pi} \int_0^{\frac{1}{\sqrt{2}}} r^2 \cdot rdrd\theta = \int_0^{2\pi} d\theta \int_0^{\frac{1}{\sqrt{2}}} r^3 dr = 2\pi\frac{r^4}{4}\Big|_0^{\frac{1}{\sqrt{2}}} = \frac{\pi}{8} \quad (35)$$

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 $\mathcal{C}^\infty$ function $f : U \to \mathbb{R}$ 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,

$$f(x) = \begin{cases} e^{1/x^2} & \text{if} \quad x \neq 0 \\ 0 & \text{if} \quad x = 0 \end{cases} \tag{36}$$

We proceed as follows, building up by steps:

1. There is a $\mathcal{C}^\infty$ function $g : \mathbb{R} \to \mathbb{R}$ 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 36, as follows:

$$g(x) = \begin{cases} e^{-\frac{1}{(x-1)^2}} e^{-\frac{1}{(x+1)^2}} & \text{if} \quad x \in (-1, 1) \\ 0 & \text{if} \quad x \notin (-1, 1) \end{cases} \tag{37}$$

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 $\mathcal{C}^\infty$ function $g_{a,\delta} : \mathbb{R}^n \to \mathbb{R}$ such that

$$g_{a,\delta} > 0 \text{ on } |x - a| < \delta, g_{a,\delta} = 0 \text{ outside} \tag{38}$$

So how could achieve this? We could use the function defined in 37 to write

$$g_{a,\delta} = g\left(\frac{|x-a|^2}{\delta^2}\right) \tag{39}$$

3. There is a $\mathcal{C}^\infty$ function $h : U \to \mathbb{R}$ 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 $U_1, ..., U_k$ with centers $a_1, ..., a_k$ and radii $\delta_1, ..., \delta_k$ whose closures lie in U. We can do this since C is compact. Once we do this, we can define quite simply

$$h(x) = \sum_{i=1}^{k} g_{a_k,\delta_k}(x) \tag{40}$$

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 $\mathcal{C}^\infty$ 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 $\mathcal{C}^\infty$ function like that in equation 37, 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

$$\psi_\epsilon(x) = \frac{\int_0^x g}{\int_0^\epsilon g} \tag{41}$$

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 40, and we're going to modify it using equation 41 as follows: letting $\epsilon = \min_C h > 0$, which exists since C is compact, let

$$f(x) = \psi_\epsilon \circ h \tag{42}$$

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 2. *Given $A \subset \mathbb{R}^n$ and an open cover $\mathcal{O}$ of A, there is a countable collection Φ of $\mathcal{C}^\infty$ 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 2. *A collection φ satisfying conditions 1, 2, 3 above is called a $\underline{\mathcal{C}^\infty \ \text{partition for } A}$. If condition 4 holds, this partition is called $\underline{\text{subordinate to } \mathcal{O}}$*

This is all $\mathcal{C}^\infty$ 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 continous. 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 $\mathcal{C}^\infty$, and hence using the function given by

$$f(x) = \begin{cases} |x^{r+1}| & \text{on} & [0, \infty) \\ -|x^{r+1}| & \text{on} & (-\infty, 0] \end{cases} \tag{43}$$

Before we try to construct such a $(\mathcal{C}^\infty)$ 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 doen 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 manfiold 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 teh interior of one of these rectangles. Now, if we want to integrate our fucntion 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

$$\int f = \sum \int \varphi f \tag{44}$$

So the key is to take a function, and make it's 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

$$\int_a^b f' = f(a) - f(b) \tag{45}$$

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 manfolds 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 whay we're going to do in the rest of the course. But first, we have to prove Theorem 2, 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 $U_1, ..., U_q$. In this case we're going to construct a *finite* (not countable) partition of unity subordinate to $\{U_1, ..., U_q\}$.

   For each $i \in \{1, ..., q\}$, 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 $D_{x_1}, ..., D_{x_p}$. For each $i \in \{1, ..., q\}$, let

   $$D_i = \bigcup_{D_{x_j} \subset U_i} D_{x_j} \tag{46}$$

   Let ψ_i be a $\mathcal{C}^\infty$ function $0 \le \psi_i \le 1$ 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,

   $$\forall x \in U \supseteq A, \sum_{i=1}^q \psi_i(x) > 0 \tag{47}$$

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

   $$\Phi = \{\varphi_1, ..., \varphi_q\}, \varphi_i = \frac{\psi_i}{\sum_{i=1}^q \psi_i} \tag{48}$$

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 $B_i = A_i \setminus \operatorname{Int} A_{i-1}$. Let $\mathcal{O}_i$ be covers of B_i by the open sets

   $$U \cap (\operatorname{Int} A_{i+1} \setminus A_{i-2}) \tag{49}$$

   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

   $$\sigma = \sum_i \sum_{\psi_i \in \Phi_i} \psi_i \tag{50}$$

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

   $$\varphi = \frac{\psi}{\sigma} \tag{51}$$

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

14

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

$$A_i = \{x : |x| \leq i, d(x, \mathbb{R}^n \setminus A) > i^{-1}\} \tag{52}$$

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 $\mathcal{O}$

$\square$

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

$$\int_A f \tag{53}$$

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 $f : A \to \mathbb{R}$, 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

$$\int_U f = \int_C f \tag{54}$$

So it makes sense to say

$$\int_A f = \int_C f \tag{55}$$

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 55, we can define, $\forall \varphi \in \Phi$,

$$\int_A \varphi|f| \tag{56}$$

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

Definition 3. f *is* <u>*integrable*</u> *on* A *if*

$$\sum_{\varphi \in \Phi} \int_A \varphi|f| \tag{57}$$

converges. Then we define

$$\int_A f = \sum_{\varphi \in \Phi} \varphi f \tag{58}$$

Note that equation 58 must converge, and in fact converges absolutely, if equation 57 converges, because

$$\sum_{\varphi \in \Phi} \left| \int_A \varphi f \right| \leq \sum_{\varphi \in \Phi} \int_A |\varphi||f| = \sum_{\varphi \in \Phi} \int_A \varphi f \tag{59}$$

since φ is nonnegative.

Theorem 3. *Suppose* $A \subset \mathbb{R}$ *is open,* $f : A \to \mathbb{R}$ *is locally bounded and the set of discontinuities of* f *has measure 0.*

15

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 57 converges for Φ, then it converges for Ψ and the integral defined using Φ and the integral defined using Ψ are equal, i.e.*

$$\sum_{\psi \in \Psi} \int_A \psi f = \sum_{\varphi \in \Phi} \int_A \varphi f \tag{60}$$

   Note that equation 60 would not necessarily be true if equation 57 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. Specificially, if both A and f are bounded, whether or not the boundary of A has measure zero,*

$$\sum_{\varphi \in \Phi} \int_A \varphi |f| \tag{61}$$

   (equation 57) converges

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

$$\int_A f = \sum_{\varphi \in \Phi} \int_A \varphi f \tag{62}$$

   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

$$\int_A f = \lim_{i \to \infty} \int_{C_i} f \tag{63}$$

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

$$\sum_{\psi \in \Psi} \psi = 1 \implies \sum_{\varphi \in \Phi} \int_A \varphi f = \sum_{\varphi \in \Phi} \int_A \left(\sum_{\psi \in \Psi} \psi \right) \varphi f = \sum_{\varphi \in \Phi} \sum_{\psi \in \Psi} \int_A \varphi \psi f \tag{64}$$

   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 $|f| \le M$ on A. For any finite subset F of Φ,

$$\sum_{\varphi \in F} \int_A \varphi |f| \le M \int_A \sum_{\varphi \in F} \varphi \le M V(B) \tag{65}$$

   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.   □

 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 parititon 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

$$\int_A f = \sum_{\varphi \in \Phi} \int_A \varphi f \tag{66}$$

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

$$V(A \setminus C) = \int_{A \setminus C} 1 = \int_A \chi_{A \setminus C} < \epsilon \tag{67}$$

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

$$C = A \setminus \text{Int} \bigcup \mathcal{B} \tag{68}$$

(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 is 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

$$\left| \int_A f - \sum_{\varphi \in F} \int_A \varphi f \right| \tag{69}$$

As the absolute value of the integral is less than or equal to the integral of the absolute value, we have that equation 69 is less than or equal to

$$\int_A \left| f - \sum_{\varphi \in F} \varphi f \right| \tag{70}$$

Since f is bounded, we can choose M such that $|f| \leq M$. So equation 70 is less than or equal to

$$M \int_A \left(1 - \sum_{\varphi \in F} \varphi \right) = M \int_A \sum_{\varphi \in \Phi \setminus F} \varphi \tag{71}$$

Since

$$\forall \varphi \in \Phi \setminus F, \forall c \in C, \varphi(c) = 0 \tag{72}$$

we have that equation 71 is less than or equal to

$$MV(A \setminus C) \leq M\epsilon \tag{73}$$

$\square$

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 4. *Let A be open, $f : A \to \mathbb{R}$ be locally bounded and let the set of discontinuties of f have measure zero. Write*

$$A = \bigcup_{n \in \mathbb{N}} C_n \tag{74}$$

17

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

$$\left\{ \int_{C_i} |f| : n \in \mathbb{N} \right\} \tag{75}$$

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

$$\left\{ \int_{C_i} f : n \in \mathbb{N} \right\} \tag{76}$$

converges absolutely, and we can write

$$\int_A f = \lim_{n \to \infty} \int_{C_n} f \tag{77}$$

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

$$\psi_1, ..., \psi_p \tag{78}$$

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

$$\varphi_i = \frac{\psi_i}{\psi_1 + ... \psi_p} \tag{79}$$

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

$$\{ f \cdot \varphi_i : i \in 1, ..., p \} \tag{80}$$

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

$$\int_{g(a)}^{g(b)} f = \int_a^b (f \circ g)g' \tag{81}$$

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 81 can be rewritten as

$$\int_{g([a,b])} f = \int_{[a,b]} (f \circ g)|g'| \tag{82}$$

since otherwise the equaton 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 5 (Change of Variables). *Let $A \subset \mathbb{R}^n$ be open, $g : A \to \mathbb{R}^n$ be one to one, continuously differentiable and let, for all $x \in A$, $g'(x) \neq 0$. Then*

$$f : g(A) \to \mathbb{R} \tag{83}$$

is integrable if and only if

$$f \circ g |\det g'| \tag{84}$$

is integrable on A. In this case,

$$\int_{g(A)} f = \int_A f \circ g |\det g'| \tag{85}$$

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

$$x = r \cos\theta, y = r \sin\theta \tag{86}$$

We have

$$D = \frac{\partial(x,y)}{\partial(r,\theta)} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \implies \det D = r \tag{87}$$

So we can write

$$\iint_A f(x,y)dxdy = \iint_A f(r\cos\theta, r\sin\theta)rdrd\theta \tag{88}$$

We now examine a corollary of Theorem 5.

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 $\mathbb{R}^n$ wich satisfies the conditions of theorem 5 on A, then*

$$f : g(C) \to \mathbb{R} \tag{89}$$

is integrable if and only if

$$f \circ g |\det g'| \tag{90}$$

is iintegrable on C, and in this case

$$\int_{g(C)} f = \int_C f \circ g |\det g'| \tag{91}$$

Lemma 2. *Assume $A \subset \mathbb{R}^n$ is open and $g : A \to \mathbb{R}^n$ 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 $C_1 \subset C_2 \subset ...$

To do so, remember that a countable intersection of measure 0 sets is measure 0. Using $\mathcal{C}_1$, which is more than uniformly continuous, we have that

$$\forall x \in C, \forall y \in U, |g(x) - g(y)| \leq c|x - y| \tag{92}$$

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

$$\int_{g(A)} f = \int_{g(C)} f \tag{93}$$

$$\int_A (f \circ g)|\det g'| = \int_C (f \circ g)|\det g'| \tag{94}$$

giving the desired equality by Theorem 5 $\square$

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

$$x = r \cos \phi \sin \theta, y = r \sin \phi \sin \theta, z = r \cos]theta \tag{95}$$

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

$$D = \frac{\partial(x, y, z)}{\partial(r, \theta, \phi)} = \begin{pmatrix} \cos \phi \sin \theta & r \cos \phi \cos \theta & -r \sin \phi \sin \theta \\ \sin \phi \sin \theta & r \sin \phi \cos \theta & r \cos \phi \sin \theta \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix} \implies \det D = r^2 \sin \theta \tag{96}$$

Now, for some problems:

1. Find the volume of an ellipsoid

$$E = \left\{ (x, y, z) \in \mathbb{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1 \right\} \tag{97}$$

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

$$g : \mathbb{R} \to \mathbb{R}^3, (x, y, z) \mapsto (x/a, y/bz/c) \tag{98}$$

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

$$\iiint_{B_1(\mathbf{0})} 1 dudvdw = \iiint_E \frac{1}{abc} dxdydz = \frac{1}{abc} \iiint_E 1 dxdydz = \frac{4}{3}\pi \implies V(E) = \frac{4}{3}\pi abc \tag{99}$$

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:

$$\int_0^{2\pi} \int_0^{\pi} \int_0^1 r^2 \sin \theta dr d\theta d\phi = \int_0^1 r^2 dr \int_0^{\pi} \sin \theta d\theta \int_0^{2\pi} d\phi = \frac{1}{3} \cdot 2 \cdot 2\pi = \frac{4\pi}{3} \tag{100}$$

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 $T : \mathbb{R}^n \to \mathbb{R}^n$ 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

$$V(B) = \int_B 1 = \int_A 1 \cdot |\det T| = |\det T| \int_A 1 = V(A)|\det T| \tag{101}$$

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

$$\int_{g(A)} f = \int_A (f \circ g)|\det g'| \tag{102}$$

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 $f \cdot g |\det g'|$ 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

$$F \circ g^{-1} |\det(g^{-1})'| = f \circ \cancel{g \circ g^{-1}} \cancel{|\det g' \circ g^{-1}|} \frac{1}{\cancel{|\det g' \circ g^{-1}|}} = f \tag{103}$$

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 $f : g(A) \to \mathbb{R}$ 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 6 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

$$\int_{g(U)} f = \int_U (f \circ g) |\det g'| \tag{104}$$

Then the theorem is true. What we're saying here is it's good enough to, instead of proving the thoerem as stated, just prove it for thse 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

$$\mathcal{O}' = \{g(U) : U \in \mathcal{O}\} \tag{105}$$

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

$$\Phi' = \{\varphi \circ g : \varphi \in \Phi\} \tag{106}$$

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

$$\int_{g(A)} \varphi \circ g = \int_{g(U)} \varphi \circ f = \int_U (\varphi \cdot f) \circ g |\det g'| = \int_A (\varphi \cdot f) \circ g |\det g'| \tag{107}$$

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

$$\int_{g(A)} f = \sum_{\varphi \in \Phi} \int_{g(A)} \varphi f = \sum_{\varphi \in \Phi} \int_A (\varphi \cdot f) \circ g |\det g'| = \int_A f \circ g |\det g'| \tag{108}$$

Note that the same equality as 108 with $|f|$ in place of f shows that the RHS converges, which is our definition of intregrability.

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,

$$\int_V f = \int_{g^{-1}(V)} (f \circ g)|\det g'| \tag{109}$$

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

$$\int_V f = \int_{g^{-1}(V)} (f \circ g)|\det g'| \tag{110}$$

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

$$\mathcal{L}(f, P) = \sum_{S \in P} m_S(f) V(S) = \sum_{S \in P} \int_{\text{Int } S} f_S \tag{111}$$

By this assumption for constant functions, we can rewrite equation 111 as

$$\sum_S \int_{g^{-1}(\text{Int } S)} f_S \circ g |\det g'| \leq \sum_S \int_{g^{-1}(\text{Int } S)} f \circ g |\det g'| \leq \int_{g^{-1}(V)} f \circ g |\det g'| \tag{112}$$

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

$$\int_V f = \sup_P \mathcal{L}(f, P) \leq \int_{g^{-1}(V)} f \circ g |\det g'| \tag{113}$$

The same argument using upper sums gives $\geq$.

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

$$\int_{(h \circ g)(A)} f = \int_{g(A)} (f \circ h)|\det h'| = \int_A ((f \circ h) \circ g)|\det h' \circ g||\det g'| = \int_A f \circ (h \circ g)|\det(h \circ g)'| \tag{114}$$

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

$$\int_{g(U)} 1 = \int_U |\det g'| \iff \text{vol}(g(U)) = |\det g| \text{vol}(U) \tag{115}$$

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 $\mathrm{vol}(g(U)) = \lambda \mathrm{vol}(U) = |\det g| \mathrm{vol}(U)$ 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 integral 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

$$g = T \circ (T^{-1} \circ g) \tag{116}$$

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

$$(T^{-1} \circ g)' = I \tag{117}$$

To complete the proof, we use induction on n:

- $n = 1$: by integration by substitution

- Assume it's true for $n - 1$. It's enought 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 $h : A \to \mathbb{R}^n$ by

$$h(x) = (g_1(x), ..., g_n(x), x_n) \tag{118}$$

  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

$$k : h(U') \to \mathbb{R}^n, k(y) = (y_1, ..., y_{n-1}, g_n(h^{-1}(y_n))) \tag{119}$$

  We have $g = k \circ h$,
$$g'(a) = k'(h(a))h'(a) = I \implies k'(h(a)) = I \tag{120}$$

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

$$\det k'(g) = 0 \tag{121}$$

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

  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

$$g = (x \mapsto (g_1(x), ..., g_{n-1}(x), x_n)) \circ (y \mapsto (y_1, ..., y_{n-1}, g_n(h^{-1}(y)))) \tag{122}$$

  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 nth coordinate of it's input. Using this, we can consider the following:

$$\int_{h(W)} 1 \tag{123}$$

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

$$h_{x_n} : D \to \mathbb{R}^{n-1}, h_{x_n}(x_1, ..., x_{n-1})(g_1(x_1, ..., x_n), ..., g_{n-1}(x_1, ..., x_n)) \tag{124}$$

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

$$(\det h'_{x_n})(x_1, ..., x_{n-1}) \tag{125}$$

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,

$$h(D \times \{x_n\}) = h_{x_n}(D) \times \{x_n\} \tag{126}$$

since, like we said before, the nth coordinate is unaffected. So we can proceed using Fubini's theorem,:

$$\int_{a_n}^{b_n} \left(\int_{h_{x_n}(D)} 1 \right) dx_n \tag{127}$$

By induction, this is equal to

$$\int_{a_n}^{b_n} \int_D 1 |\det h'_{x_n}| = \int_{a_n}^{b_n} \int_D |\det h'| = \int_{D \times [a_n, b_n]} 1 \cdot |\det h'| = \int_W 1 \cdot |\det h'| \tag{128}$$

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 4. *A k-dimensional manifold manifold is a set $M \subseteq \mathbb{R}^n$ such that for all $a \in M$, there is a $\mathcal{C}^r$ mapping $f : U \to \mathbb{R}^{n-k}$ defined on an open neighborhood U of a such that $M \cap U = f^{-1}(0)$. f has rank $n-k$ at every $x \in M \cap U$.*

Definition 5. *A k-dimensional manifold 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 $\mathcal{C}^r$ function $z = g(y)$, $z = (z_1, ..., z_{n-k}), y = (y_1, ..., y_k)$ where*

$$(y, z) \neq (x_{\sigma(1)}, ..., x_{\sigma(n)})$$

where σ is a permutation of $1, ..., n$.

Definition 6. *A k-dimensional manifold 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 $\mathcal{C}^r$ <u>diffeomorphism</u> (a $\mathcal{C}^r$ mapping with a $\mathcal{C}^r$ inverse) $h : U \to V$ such that*

$$h(M \cap U) = V \cap (\mathbb{R}^k \times \{\mathbf{0} \in \mathbb{R}^{n-k}\}) = \{y = (y_1, ..., y_n) \in V : y_{k+1} = ... = y_n\}$$

Claim 1. *The definitions given for a manifold are equivalent*

Proof. • 4 $\implies$ 5: By the implicit function theorem

- 5 $\implies$ 4: Let $h(y,z) = z - g(y)$.

- 5 $\implies$ 6: Let $h(y,z) = (y, z - g(y))$

- 6 $\implies$ 4: Let $f(x) = (h_{k+1}(x), ..., h_n(x))$

$\square$

To illustrate the first definition in action, consider the example of the unit sphere $S^{n-1} \subset \mathbb{R}^n$, given by

$$S^{n-1} = \{x = (x_1, ..., x_n) : f(x) - 1 = 0)\} \text{ where } f(x) = x_1^2 + ... + x_n^2$$

We have

$$\forall x \in S^{n-1}, \text{grad } f(x) = 2(x_1, ..., x_n) = 2x \neq 0$$

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 $f : \mathbb{R}^n \to \mathbb{R}^p$, 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 $\mathbb{R}^n_a$, the copy of $\mathbb{R}^n$ at a, to $\mathbb{R}^p_{f(a)}$, 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 4 is $\text{Ker } Df(a) = Df(a)^{-1}(0)$, a vector subspace of $\mathbb{R}^n_a$

2. To think of the tangent space in terms of Definition 5, 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

$$z - c = Dg(b)(y - b)$$

But we want to think of it as a vector space with a as the origin, a vector subspace of $\mathbb{R}^n_a$. So we should write it as

$$Z = Dg(b)(Y)$$

where $Z \in \mathbb{R}^p_{f(a)}$, $Y \in \mathbb{R}^n_a$. This is because $f(y, g(y)) = 0$, so

$$Df(a) \begin{pmatrix} I \\ Dg(b) \end{pmatrix} = 0 \implies \text{Ker } Df(a) \supseteq \text{Im} \begin{pmatrix} I \\ Dg(b) \end{pmatrix}$$

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

$$Z = Dg(b)(Y)$$

3. In terms of the Definition 6, we can write the tangent space as

$$Dh(a)^{-1}(\mathbb{R}^k \times \{\mathbf{0} \in \mathbb{R}^{n-k}\})$$

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 7 (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 $\mathcal{C}^r$ mapping $\varphi : W \to \mathbb{R}^n$ 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 "$\varphi^{-1} : \varphi(W) = M \cap U \to W$" 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 7 implies Definition 6:

Proof. Say $a = \varphi(b)$ for some $b \in W$. We can assume

$$\frac{\partial(\varphi_1, ..., \varphi_k)}{\partial(y_1, ..., y_k)}$$

has rank k on W. Define $\psi : W \times \mathbb{R}^{n-k} \to \mathbb{R}^n$ by

$$(y, z) \mapsto \varphi(y) + (0, z)$$

Then we get the block matrix

$$\psi'(y, z) = \begin{pmatrix} \frac{\partial(\varphi_1, ..., \varphi_k)}{\partial(y_1, ..., y_k)} & 0 \\ * & I \end{pmatrix}$$

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 $\psi : V_1' \to U_1'$ has a $\mathcal{C}^r$ inverse $\psi^{-1} : U_1' \to V_1'$.

We have that

$$\psi^{-1}(\varphi(y)) = (y, 0) \in V_1' = \varphi(W) \cap \widetilde{U}$$

where U is open in $\mathbb{R}^n$. Take $U_1 = U_1' \cap \widetilde{U}$ and $V_1 = \psi^{-1}(U_1)$. We have

$$M \cap U_1 = \{\varphi(y) : (y, 0) \in V_1\}$$

So

$$h = \psi^{-1}|_{U_1}$$

satisfies the conditions implied by (3) since

$$h(M \cap U_1) = \psi^{-1}(M \cap U_1) = \{(y, 0) : (y, 0) \in V_1\} = V_1 \cap (\mathbb{R}^k \times \{\mathbf{0}\})$$

$\square$

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 6 and 7 given for a manifold in the previous lecture

Definition 8. $TMa = Dh(a)^{-1}(\mathbb{R}^k \times \{\mathbf{0}\})$

Definition 9. $TMa = D\varphi(b)(\mathbb{R}^k_b)$

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 9 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 interesecting open subsets M_1, M_2 of M to spaces W_1, W_2. Then

$$\varphi_1 = \varphi_2 \circ (\varphi_2^{-1} \circ \varphi_1)$$

on $M_1 \cap M_2$, giving

$$D\varphi_1(b_1) = D\varphi_2(b_2) \circ D(\varphi_2^{-1} \circ \varphi_1)(b_1)$$

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 10.

$$TM_a = \{tangent\ vectors\ \gamma'(0)\ of\ \mathcal{C}^\infty\ curves\ \gamma : (-\delta, \delta) \to M\ s.t.\ \gamma(0) = a\}$$

So what does this even mean? Recall the following definition:

Definition 11. *A $\mathcal{C}^1$ curve in $\mathbb{R}^n$ is a $\mathcal{C}^1$ mapping, for some open interval (c, d),*

$$\gamma : (c, d) \to \mathbb{R}^n, \gamma(t) = (\gamma_1(t), ..., \gamma_n(t))$$

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 12. *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 inteval (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,

$$TMa = \operatorname{Ker} Df(a)$$

In our case, we have

$$f(\gamma(t)) = 0 \implies Df(\gamma(t_0)) \cdot \gamma'(t_0) = 0 \implies \gamma'(t_0) \in \operatorname{Ker} Df(a) = TMa$$

So, considering this, why is definition 11 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 9, 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,

$$\gamma'(0) = D\varphi(b) \circ \delta'(0)$$

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 11 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 firs tyear calculus, in that first you dodifferentiation 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 implict 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 13. *A half space $\mathbb{H}^k \subset \mathbb{R}^k$ is given by*

$$\{x = (x_1, ..., x_k) : x_k \geq 0\}$$

Definition 14. *A subset $M \subseteq \mathbb{R}^n$ is a k-dimensional $\mathcal{C}^r$ 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 $\mathcal{C}^r$ diffeomorphism $h : U \to V$ such that either*

(1) *The usual condition for a manifold,*

$$h(M \cap U) = V \cap (\mathbb{R}^k \times \{\mathbf{0} \in \mathbb{R}^{n-k}\})$$

(2) *The manifold looks like a half space around a, i.e.*

$$h(M \cap U) = V \cap (\mathbb{H}^k \times \{\mathbf{0} \in \mathbb{R}^{n-k}\}) = \{y = (y_1, ..., y_n) : y_1, ..., y_k \geq 0 \wedge y_{k+1} = ... = y_n = 0\}$$

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 $h_1 : U_1 \to V_1$ satisfies (1) and $h_2 : U_2 \to V_2$ satisfies (2) for some a. How can we get a contradiction? If we take

$$h_2 \circ h_1^{-1} : V_1 \to V_2$$

we get a $\mathcal{C}^r$ 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 15. *The boundary of M, written ∂M, is the set of points satisfying condition (2) in definition 14*

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 compue 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:

$$\int_a^b f = g(b) - g(a) \tag{129}$$

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

$$\gamma : [a, b] \to \mathbb{R}^n \tag{130}$$

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 $\mathcal{C}^1$. What this means is, except for some finite number of points (since we're working on a compact interval), it's $\mathcal{C}^1$. 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,

$$\gamma(t) : [0, 3\pi] \to \mathbb{R}^2 = t \mapsto (\cos t, \sin t) \tag{131}$$

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

$$\ell(C) = \int_a^b |\gamma'(t)| dt \tag{132}$$

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 $P = \{t_0, t_1, ..., t_k\}$ be a partition of $[a, b]$, and let $\ell_P(C)$ be the length of the associated polygonal approximation:

$$\ell_p(C) = \sum_{j=1}^{k} |\gamma(t_j) - \gamma(t_{j-1})| \tag{133}$$

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 16. *The curve C is $\underline{rectifiable}$ if*

$$\sup_P \ell_P(C) \tag{134}$$

exists, in which case we define 134 it to be the $\underline{length}$ 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 134. 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