MATH 25B

Professor: Bena Tshishiku

Michele Tienni
Lowell House
Cambridge, MA 02138
micheletienni@college.harvard.edu

Please note that these notes are not official and they are not to be considered as a substitute for your notes. Specifically, you should not cite these notes in any of the work submitted for the class (problem sets, exams). The author does not guarantee the accuracy of the content of this document. If you find a typo (which will probably happen) feel free to email me.

Contents

Date: April 25, 2018.

1. January 22

Math 25b is a course about real analysis. At a glance, the main contents of the course will be:

- the theory of differentiation;
- the theory of integration;
- Stokes' theorem (which relates the two).

There will be an emphasis on theory.

1.1. Functions.

As usual, $\mathbb{R}^k$ will denote the set of k-tuples of real numbers, e.g.

$$\mathbb{R}^k = \{(x_1, \dots, x_k) : x_i \in \mathbb{R}\}.$$

As a rule, we will generally be interested in functions of the form

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

and we will define

$$Fun(\mathbb{R}^n, \mathbb{R}^m) := \{f : \mathbb{R}^n \to \mathbb{R}^m\}$$

as last semester. In linear algebra we focused on a subset of this set, namely that of *linear functions* $L(\mathbb{R}^n, \mathbb{R}^m)$. Another important case is the case $n = m = 1$, and an even more restrictive example is that of functions defined in terms of polynomials:

$$Fun(\mathbb{R}, \mathbb{R}) \supset Poly(\mathbb{R}) = \{f(x) = a_n x^n + \cdots + a_1 x + a_0\}.$$

We will see that the cases of linear maps and polynomial functions are easy to study. Therefore, our goal is to understand to what extent we can understand any function $f \in Fun(\mathbb{R}^n, \mathbb{R}^m)$ in terms of linear functions (i.e. derivatives) or polynomials (e.g. Taylor series).

Remark. In general this problem is hopeless because a function can be badly behaved.

Example 1.1. The function

$$f(x) = \sin\left(\frac{1}{x}\right)$$

is not well behaved. In fact, recall that the definition of $\sin(x)$ is the vertical coordinate of the point where a line at an angle x meets the unit circle. Therefore, $\sin(x)$ is equal to 0 if and only if $x = k\pi$ where k is an integer. This means that $\sin(1/x)$ is equal to 0 when $x = 1/k\pi$. Similarly, $\sin(1/x) = 1$ if and only if $x = 2/(4k+1)\pi$ and $\sin(x) = -1$ if and only if $x = 2/(4k+3)\pi$. If we graph this function we see that the above means that the graph oscillates more and more as x approaches 0. This function is "badly behaved" at $x = 0$, in a sense that will be made precise at the end of the lecture. Regardless of the definition, the intuitive motivation is that the function oscillates increasingly as x approaches 0.

Example 1.2. Another example of a badly behaved function is

$$g(x) = \begin{cases} x & x \text{ is rational} \\ 0 & x \text{ is irrational.} \end{cases}$$

This function is hard to draw, but you can think of it as the graph of the function $f(x) = x$ with holes in it at irrational coordinates.

Eventually, we will restrict to "nice" fuctions (namely continuous, differentiable, C^r).

1.2. **Limits.** Limits refer to the local behavior of a function. Fix $f : \mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$. An informative definition of a limit is that f *approaches the limit* L *near* a is we can make $f(x)$ as close as we like to L by requiring x to be close, but not equal, to a. By "close" we mean with respect to the distance on $\mathbb{R}$, i.e. we can make $|f(x) - L|$ small whenever $|x - a|$ is small.

Example 1.3. Let $f(x) = 2x \cdot \sin(1/x) + 1$. From a graph we can see that $f(x)$ "approaches" 1 as x goes to 0. To show this, we can use the above working definition of limit. For example, say we want $|f(x) - 1|, 1/10$. We know that

$$\begin{aligned} |f(x) - 1| =& |2x \sin(1/x)| \\ =& 2|x||\sin(1/x)| \\ \leq& 2|x| \end{aligned}$$

so that it suffices to require $x < 1/20$. Similarly, if we want $|f(x) - L| < 1/10^k$ we can achieve this by requiring $|x| < 1/2 \cdot 10^k$.

Example 1.4. Let's return to the example $f(x) = \sin(1/x)$. Does f approach 0 near $a = 0$? Namely, can we force $|f(x)| < 1/10$ by forcing $|x|$ to be small? The answer is no, because the points $2/\pi, 2/5\pi, 2/9\pi, \ldots$ get arbitrarily close to 0 but $f(x) = 1$ remains bounded away from $L = 0$.

We will now give the formal definition of limit.

Definition 1.5. Given $f : \mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$ we say that f *approaches* L *near* a if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - L| < \varepsilon$ whenever $|x - a| < \delta$. In this case we write

$$\lim_{x \to a} f(x) = L.$$

Example 1.6. We want to show that

$$\lim_{x \to 0} 2x \sin\left(\frac{1}{x}\right) + 1 = 1.$$

Proof. Fix $\varepsilon > 0$. Take $\delta = \varepsilon/2$. Then $0 < |x - 0| < \delta$ implies $|f(x) - 1| = |2x \sin(1/x)| \leq 2|x| < 2 \cdot \varepsilon/2 = \varepsilon$. □

Example 1.7. We want to show that

$$\lim_{x \to 0} \sin\left(\frac{1}{x}\right) \neq 0.$$

We note that $\lim_{x \to a} f(x) \neq 0$ means that there exists $\varepsilon > 0$ such that for all $\delta > 0$ there is x with $0 < |x - a| < \delta$ and $|f(x) - L| > \varepsilon$.

Proof. Fix $\varepsilon = 1/10$. Given $\delta > 0$ we take k large enough so that $x = 2/(4k+1)\pi < \delta$. Then $|f(x) - 0| = 1 > \varepsilon$ and $0 < |x - 0| < \delta$ by construction. □

Remark. For $a \neq 0$, $\lim_{x \to a} \sin(1/x)$ *does* exist and is equal to $\sin(1/a)$.

Exercise. Show

(1) $\lim_{x \to a} c = c$
(2) $\lim_{x \to a} x = a$.

2. January 24

2.1. Theorems about limits.
Last time we saw that

Definition 2.1. Given $f : \mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$ we say that f *approaches* L *near* a if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - L| < \varepsilon$ whenever $|x - a| < \delta$. In this case we write

$$\lim_{x \to a} f(x) = L.$$

Remark. We can make this definition more general, and in fact we will do that.

Recall. Last semester we saw how to define a distance function on $\mathbb{R}^k$. For $z, w \in \mathbb{R}^k$ we defined

$$\begin{aligned} |z - w| &= \sqrt{(z_1 - w_1)^2 + \cdots + (z_k - w_k)^2} \\ &= \langle z - w, z - w \rangle^{1/2} \end{aligned}$$

in terms of the standard inner product on $\mathbb{R}^k$ (last line). This for example tells us that $|z - w| \leq |z - u| + |u - w|$ for all $u, z, w \in \mathbb{R}^k$.

Remark. This definition holds more generally for functions $f : X \to Y$ between *metric spaces* (as we will see at some point on the homework).

The plan for today is to answer the following questions:

- are limits unique?
- how do limits behave under sums, products, and quotients?

Theorem 2.2 (Limit uniqueness)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$. *If* $\lim_{x \to a} f(x) = L$ *and* $\lim_{x \to a} = K$ *then* $L = K$.

Proof. We will prove this by contradiction. Suppose $K \neq L$, and without loss of generality $K > L$. Take $\varepsilon = \frac{K-L}{2}$. Then there exists $\delta > 0$ so that if $0 < |x - a| < \delta$ implies $|f(x) - L| < \varepsilon$ and $|f(x) - K| < \varepsilon$. These two inequalities imply that

$$\begin{aligned} |f(x)| &< L + \frac{K - L}{2} = \frac{K + L}{2} \\ |f(x)| &> K - \frac{K - L}{2} = \frac{K + L}{2} \end{aligned}$$

which is a contradiction. Therefore $K = L$. We can generalize the proof for the case $f : \mathbb{R}^n \to \mathbb{R}^m$ by setting $\varepsilon = \frac{|K-L|}{2}$ and using the triangle inequality. $\square$

Theorem 2.3 (Algebra of limits)**.** *Let* $f, g : \mathbb{R}^n \to \mathbb{R}$. *Suppose* $\lim_{x \to a} f(x) = K$ *and* $\lim_{x \to a} g(x) = L$. *Then*

(i) $\lim_{x \to a} f(x) + g(x) = K + L$;
(ii) $\lim_{x \to a} f(x)g(x) = KL$;
(iii) if $K \neq 0$, $\lim_{x \to a} 1/f(x) = 1/K$.

Remark. This theorem allows us to compute many limits without resorting to the $\varepsilon - \delta$ definition. For example, we can compute

$$\lim_{x \to a} \frac{x^{14} - 3x^{100}}{x^{10} + 1} = \frac{a^{14} - 3a^{100}}{a^{10} + 1}$$

just by using the above theorem togehter with the facts (exercises) $\lim_{x\to a} c = c$ and $\lim_{x\to a} x = a$.

Remark. We *do* need to use the $\varepsilon - \delta$ definition to prove Theorem 2.3.

Remark. When proving limit statements using the $\varepsilon - \delta$ definition there are usually 3 steps to follow:

(1) algebra
(2) estimates
(3) write the proof.

The first two points are mostly scratchwork. Let's see this at work in the proof of Theorem 2.3.

Proof. We start by proving (i). The algebra is as follows: we know that we can make $|f(x) - K|$ and $|g(x) - L|$ small as we wish. Therefore, by the triange inequality,

$$\begin{aligned} |(f(x) + g(x)) - (K + L)| &= |f(x) - K + g(x) - L| \\ &\leq |f(x) - K| + |g(x) - L|. \end{aligned}$$

Since we want to make the left hand side smaller than ε, we win if we make each summand of the last line smaller than $\varepsilon/2$. We now get to the second step, namely estimates. By assumption we can choose $\delta_f, \delta_g > 0$ such that $|f(x) - K| < \varepsilon/2$ if $0 < |x - a| < \delta_f$ and similarly for g. We are now ready to write down a proof.

Given $\varepsilon > 0$, choose δ_f, δ_g as shown before. Set $\delta = \min\left\{\delta_f, \delta_g\right\}$. Then $0 < |x - a| < \delta$ implies that

$$\begin{aligned} |f(x) + g(x) - (K + L)| &\leq |f(x) - K| + |g(x) - L| \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

We will now prove part (ii). The algebra part is as follows:

$$\begin{aligned} |f(x)g(x) - KL| &= |f(x)g(x) - Kg(x) + Kg(x) - KL| \\ &\leq |f(x)g(x) - Kg(x)| + |Kg(x) - KL| \\ &= |f(x) - K||g(x)| + |K||g(x) - L|. \end{aligned}$$

We now go on to estimates.

Choose δ_g so that $0 < |x - a| < \delta_g$ implies $|g(x) - L| < \varepsilon/2|K|$. By a problem in the homework, $\lim_{x\to a} g(x) = L$ implies that we can find δ_0 and $M > 0$ so that $0 < |x - a| < \delta_0$ implies that $|g(x)| \leq M$. We finish by choosing δ_f such that $0 < |x_a| < \delta_f$ implies $|f(x) - K| < \varepsilon/2M$. Combining the above follows in a straightforward way, so in the interest of time we will omit the proof. □

2.2. **Continuity.**

Definition 2.4. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is **continuous at** $a \in \mathbb{R}^n$ if $\lim_{x\to a} f(x) = f(a)$. If f is continuous at every $a \in \mathbb{R}^n$, we say that f is **continuous**.

Remark. There are many ways in which a function can fail to be continuous.

- $f(a)$ may not be defined (e.g. $f(a) = x\sin(1/x)$ at $a = 0$)
- $\lim_{x\to a} f(x)$ may not exist (e.g. $f(x) = \sin(1/x)$ at $a = 0$)

- $\lim_{x\to a} f(x) = f(a)$, as in the case
$$f(x) = \begin{cases} x\sin(1/x) & x \neq 0 \\ 1 & x = 0 \end{cases}$$
at $a = 0$.

Example 2.5. By Theorem 2.3 we have that polynomials and rational funcitons (ratios of polynomials) are continuous where defined. Similarly, sums products and quotients of continuous functiosn are continuous.

Example 2.6. Consider the function $f : [0,1] \to \mathbb{R}$ defined as
$$f(x) = \begin{cases} \frac{1}{q} & x = \frac{p}{q} \text{ lowest terms} \\ 0 & x \text{ irrational.} \end{cases}$$
This functions is not continuous at any rational points. However, we claim that f is continuous at every irrational point.

Proof. Fix a irrational. There are finitely many $q \in \mathbb{N}$ such that $\frac{1}{q} < \varepsilon$. Therefore we can choose δ so as to avoid these finitely many points. Therefore $\lim_{x\to a} f(x) = 0$. $\square$

3.1. **Continuity theorems.** Recall from last time that

Definition 3.1. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is **continuous at** $a \in \mathbb{R}^n$ if $\lim_{x\to a} f(x) = f(a)$. If f is continuous at every $a \in \mathbb{R}^n$, we say that f is **continuous**.

We also saw that continuity at a point is not very restrictive, as we could see from the last example in last lecture.

Today we will focus on functions that are not only continuous at a poitn but are instead continuous on a closed interval. These functions have very interesting properties. As usual, we define the interval $[a,b]$ to be the set $\{x \in \mathbb{R}\colon a \le x \le b\}$. With this in mind, we will see that if $f : [a,b] \to \mathbb{R}$ is ontinuous at every $x \in [a,b]$ then f has nice properties.

Remark. For $f : [a,b] \to \mathbb{R}$, we say that f is continuous at a (resp. b) if $\lim_{x\to a^+} f(x) = f(a)$ (resp. $\lim_{x\to b^-} f(x) = f(b)$). For the notation, see homework 1, problem 4. Having specified this, we will probe the following:

Theorem 3.2 (Intermediate Value Theorem (IVT))**.** *If $f : [a,b] \to \mathbb{R}$ is continuous and $f(a) < d < f(b)$ then there exists $c \in [a,b]$ so that $f(c) = d$.*

Theorem 3.3 (Boundedness theorem)**.** *If $f : [a,b] \to \mathbb{R}$ is continuous then f is **bounded from above**, i.e. there exists M such that $f(x) \le M$ for all $x \in [a,b]$. Similarly f is **bounded from below** (analogous definition).*

Theorem 3.4 (Maximum value theorem)**.** *If $f : [a,b] \to \mathbb{R}$ is continuous then there exists $z \in [a,b]$ so that $f(x) \le f(z)$ for all $x \in [a,b]$. In a similar way, there exists $w \in [a,b]$ so that $f(x) \ge f(w)$ for all $x \in [a,b]$.*

Remark.
(1) These theorems are false if f is not continuous at *every* point. For example, consider the function

$$\begin{aligned} f : [-1,1] &\to \mathbb{R} \\ x &\mapsto \begin{cases} \frac{1}{x} & x \neq 0 \\ 1 & x = 0 \end{cases}. \end{aligned}$$

We see that f is not continuous at $= a0$ (HW 1 # 12), f violates the intermediate value theorem (because $f(-1) = -1 < 0 < 1 = f(1)$ and f is never equal to 0), and it violates the boundedness theorem (and hence the maximum value theorem).

(2) These theorems are false if we replace $[a,b]$ with (a,b) (defined as the set $\{x \in \mathbb{R} : a < x < b\}$). An example of this is the function $f : (0,1) \to \mathbb{R}$ given by $f(x) = x^2$, which violates the maximum value theorem.

(3) These theorems are very useful, and we can immediately see some corollaries:

Corollary 3.4.1. *Every positive real number has a positive square root.*

Corollary 3.4.2. *Fix $p = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. Then*
(i) if n is odd then p has a real root;
(ii) if n is even then p has a minimum value, i.e. there exists $c \in \mathbb{R}$ such that $p(x) \ge p(c)$ for all $x \in \mathbb{R}$.

We proved both of these theorems last semester using the fundamental theorem of algebra. Later we will prove the fundamental theorem of algebra using Stokes' theorem.

Proof of 3.4.1. Fix $d > 0$. We want to show that there exists $c > 0$ such that $c^2 = d$. Define $f(x) = x^2$. Choose b such that $b^2 > d$. Then $f : [0, b] \to \mathbb{R}$ is continuous and $f(0) = 0 < d < b^2 = f(b)$. Therefore there exists c such that $c^2 = f(c) = d$. $\square$

Proof of 3.4.2. For (i), we claim that there exist $a < 0$ and $b > 0$ so that $p(a) < 0 < p(b)$. Given this claim, we can restrict p to the interval $[a, b]$ so that the IVT applied to the function $p : [a, b] \to \mathbb{R}$ tells us that there exists $c \in [a, b]$ with $p(c) = 0$. Why is the claim true? The idea is that if x is very large in absolute value, then the leading term (x^n) is going to dominate, and hence determine the sign of p. To prove the claim write

$$p(x) = x^n \left(1 + \frac{a_{n-1}}{x} + \cdots + \frac{a_1}{x^{n-1}} + \frac{a_0}{x^n}\right).$$

We can find $|x| \gg 0$ such that the term inside the parentheses is as close to 1 as we wish (as per definition of limit), and therefore the first term determines the sign of p. $\square$

Proof of Theorem 3.2. The general case follows from

Claim. If $f : [a, b] \to \mathbb{R}$ is continuous and $f(a) < 0 < f(b)$ then there exists $c \in [a, b]$ such that $f(c) = 0$. What follows in an incomplete proof; it is up to you to find where it lacks in rigor:

Incomplete proof: the idea is to try and identify the "last" point where f is negative. Consider $S = \{s \in [a, b] : f(s) < 0\}$ and $T = \{t \in [a, b] : t > s \forall s \in S\}$. Let T_0 to be the smallest element of T. Let's show $f(t_0) = 0$ by contradiction. If t_0 is positive, then $f > 0$ on a small neighborhood of t_0 (HW 1 #2). In particular, there exists some $t_1 < t_0$ such that $t_1 \in T$, thus contradicting minimality of t_0. If $f(t_0) < 0$, then $f < 0$ near t_0 so that there exists $s_1 > t$ such that $f(s_1) < 0$, and therefore $s_1 \in S$ which contradicts $t_0 \in T$.

Remark. Somewhere we used something special about $\mathbb{R}$. For example, consider the function $f(x) = x^2 - 2$ on the set

$$[0, 2]_{\mathbb{Q}} = \{x \in \mathbb{Q} : 0 \leq x \leq 2\}.$$

If the argument given above were actually correct, it would apply to this f as well, which would mean that 2 has a rational square root. But this is not the case. In fact, the property of $\mathbb{R}$ that we used is the **least upper bound property**.

$\square$

4. January 29 – Notes by Kim

I was sick, so no notes from me! Thanks a lot to Kim for sharing notes of Monday lectures. What follows will pretty much be a copy of her notes (mainly for me to go over what was done in class). Original notes can be found on the course website.

4.1. **Least Upper Bound property (the secret sauce of $\mathbb{R}$).**

Last time we introduced three theorems about continuous funcitons $f : [a,b] \to \mathbb{R}$, namely

- the Intermediate Value Theorem (IVT);
- the Boundedness Theorem;
- the Maximum Value Theorem.

This week we will prove these theorems and understand them in a more general context (i.e. $\mathbb{R}^n$). To do so we will need to learn new things about $\mathbb{R}$.

So far, we know that $\mathbb{R}$ is an ordered field. We introduce the following deifinitions:

Definition 4.1. For a subset $A \subset \mathbb{R}$, we say that A is **bounded above** if there exists $z \in \mathbb{R}$ such that $a \leq z$ for all $a \in A$. We call such z an **upper bound**.

Definition 4.2. For a subset $A \in \mathbb{R}$, we say that $z \in \mathbb{R}$ is the **least upper bound** of A if z is an upper bound of A and $z \leq z'$ for any other upper bound z' of A. In this case we write $z = \sup A$.

Definition 4.3. We similarly define notions of **bounded below**, **lower bound** and **greatest lower bound** of a subset. The latter is denoted by $\inf A$.

Remark. We say that $f : [a,b] \to \mathbb{R}$ is bounded above (resp. below) if the set $\{f(x) : x \in [a,b]\}$ is bounded above (resp. below).

Example 4.4. We give different examples of subsets of $\mathbb{R}$ and their respective least upper bounds and greatest lower bounds.

Set A	$\sup A$	$\inf A$
$\{1,2,3\}$	3	1
$\mathbb{N}$	1	N/A
$\emptyset$	N/A	N/A
$\{1/n : n \in \mathbb{N}\}$	1	0
$\left\{x \in \mathbb{Q} : 0 < x < \sqrt{2}\right\}$	$\sqrt{2}$	0

Theorem 4.5 (Least upper bound property)**.** *If $A \subset \mathbb{R}$ is nonempty and is bounded above, then A has a least upper bound, i.e.* $\sup A$ *exists.*

Theorem 4.6 (Characterization of $\mathbb{R}$)**.** *There exists a* unique *ordered field with the least upper bound property.*

Remark.

- $\mathbb{Q}$ is an ordered field but it does not have the least upper bound property (LUB). In fact, the set $A = \left\{x \in \mathbb{Q} : x^2 < 2\right\}$ is bounded above, but has no least upper bound in $\mathbb{Q}$.
- The set

$$\mathrm{Rat}(\mathbb{R}) = \left\{\frac{p(x)}{q(x)} : p, q \in \mathrm{Poly}(\mathbb{R}), q \neq 0\right\}$$

is an ordered field. In fact, for $f, g \in \mathrm{Rat}(\mathbb{R})$ we can define $f \leq g$ if there exists $\delta > 0$ such that $f(x) \leq g(x)$ for $0 < x < \delta$. By theorem 4.6, $\mathrm{Rat}(\mathbb{R})$ does not have the LUB property.

As an example of the usefulness of Theorem 4.5, let's consider the following lemma:

Lemma 4.6.1. *$\mathbb{N} \subset \mathbb{R}$ is not bounded above.*

Remark. Why is the above lemma not obvious? Because every ordered field F contains a copy of $\mathbb{N}$ where $n \in \mathbb{N}$ is given by $1_F + \cdots + 1_F$. In $\mathrm{Rat}(\mathbb{R})$, every n corresponds to a constant polynomial. Then we see that $\mathbb{N} \subset \mathrm{Rat}(\mathbb{R})$ is actually bounded above, for example by $1/x$, and the set $\{1/n : n \in \mathbb{N}\}$ does not have 0 as its greater lower bound (as opposed to the same subset sitting in $\mathbb{R}$), seeing as x is a lower bound and $0 \leq x$.

Proof of Lemma 4.6.1. We prove this by contradiction. Suppose $\mathbb{N} \subset \mathbb{R}$ is bounded above. By Theorem 4.5, it has a least upper bound. Let $z = \sup \mathbb{N}$. Then for all $n \in \mathbb{N}$ we have that both $n \leq z$ and $n + 1 \leq z$. This implies that $n \leq z - 1$ for all $n \in \mathbb{N}$, which contradicts the fact that z is the *least* upper bound. □

4.2. **Proof of the intermediate value theorem.**

Theorem 4.7 (Intermediate Value Theorem (IVT))**.** *If $f : [a, b,] \to \mathbb{R}$ is continuous and $f(a) < 0 < f(b)$ then there exists $c \in [a, b,]$ so that $f(c) = 0$.*

Proof. Let's finding the "last" point where $f < 0$. Consider the set $A = \{x \in [a, b] : f(x) < 0\}$. We know this is nonempty since $a \in A$. Moreover, b is an upper bound, and therefore A has a LUB. Let $c = \sup A$. We will prove by contradiction that $f(c) = 0$.

- Suppose $f(c) > 0$. Then by continuity there exists $\delta > 0$ such that $f(x) > 0$ for all x such that $0 < |x - c| < \delta$. Then there exists $y < c$ with $x \leq y$for all $x \in A$. This contradicts $c = \sup A$.
- Suppose $f(c) < 0$. Then by continuity there exists $\delta > 0$ such that $f(x) < 0$ for all x such that $0 < |x - c| < \delta$. Then there exists $y > c$ with $f(y) < 0$, which contradicts the fact that c is an upper bound of A.

We conclude that $f(c) = 0$. □

4.3. **Proof of the boundedness theorem.**

Theorem 4.8 (Boundedness theorem)**.** *If $f : [a, b] \to \mathbb{R}$ is continuous then f is bounded above.*

Before proving the above theorem, we state a useful fact: if $z = \sup A$, then for every $\delta > 0$ there exists $x \in A$ such that $z - x < \delta$.

Proof. By contradiction. Suppose that f is not bounded above. We want to find the "last" point z such that f is bounded on $[a, z]$. Consider the set $A = \{x \in [a, b] : f$ is bounded on $[a, x]\}$. This set is not empty since $a \in A$, and is bounded above by b. Let $z = \sup A$ (by LUB property). We will show by contradiction that $z = b$. Suppose $z < b$. Then f is continuous at z, and so it is bounded near z. Then there exists $y > z$ such that f is bounded on $[a, y]$. Then $z \neq \sup A$, which is a contradiction. Therefore $z = b$. □

Remark. In the above proof, we did not prove whether or not $b \in A$.

The argument above is important.

5. January 31 – Notes by Natalia

Couldn't come to class again. Many thanks to Nat for sharing notes from today's class! Like I did for Monday, I'm going to copy them here, mainly for me to get a sense of what happened in class (original notes can be found on the class website).

5.1. **Generalizing the boundedness theorem.** When can we expect for a continuous function to be bounded? For example, $f:[0,1]\to\mathbb{R}$ is always bounded but $f:(0,1)\to\mathbb{R}$ is not always bounded.

Example 5.1. Which ones of the following sets X are such that every continuous function $f:X\to\mathbb{R}$ is bounded?

1. $X = S^2 = \left\{x\in\mathbb{R}^3 : |x| = 1\right\}$ (the unit sphere in $\mathbb{R}^3$)
2. $X = \left\{\text{ linear isometries } A:\mathbb{R}^3\to\mathbb{R}^3\right\} = \{A\in M_3(\mathbb{R}) : A^tA = I\}\subset M_3(\mathbb{R})\cong\mathbb{R}^9$ (an example of a continuous map $f:X\to\mathbb{R}$ is $f(A)=\mathrm{Tr}(A)$)
3. $X = \{(x,\sin(1/x)) : x\in[1,-1]\}\cup\{(0,y) : y\in[-1,1]\}$
4. X =Sierpinski triangle
5. $X = C(\mathbb{R},[0,1]) = \{\text{ continuous functions } f:\mathbb{R}\to[0,1]\}\subset\mathrm{Fun}(\mathbb{R},[0,1]) = [0,1]^{\mathbb{R}}$

We actually don't know the answer to these questions yet. To answer this question we need more tools. The key tool is compactness.

5.2. **Subsets of $\mathbb{R}^n$.**

Definition 5.2. An **open rectangle** in $\mathbb{R}^n$ is a set of the form

$$\prod_{i=1}^n (a_i,b_i).$$

A **closed rectangle** is a set of the form

$$\prod_{i=1}^n [a_i,b_i].$$

Definition 5.3. We say that $U\subset\mathbb{R}^n$ is **open** if for every $u\in U$, there exists an open rectangle Q with $u\in Q\subset U$. We say that $A\subset\mathbb{R}^n$ is **closed** if A^c is open.

Example 5.4. For any $r>0$, the open ball $B_r = \{x\in\mathbb{R}^n : |x|<r\}$ is open.

Exercise. Are the following sets open or closed in $\mathbb{R}^2$?

- $\mathbb{Z} = \{(x,y) : x,y\in\mathbb{Z}\}$ (closed)
- $[0,1)\times[0,1)$ (neither)
- $\mathbb{R}^2$ (both)
- $\emptyset$ (both, vacuously).

Some important properites are:

1. Suppose U_β is open for all $\beta\in B$ (where B is some index set). Then $\bigcup_{\beta\in B}U_\beta$ is open, namely arbitrary unions of open sets are open.
2. Suppose $U_1,\dots,U_r$ are open (with $r\in\mathbb{N}$). Then $\bigcap_{i=1}^r U_i$ is open, i.e. finite intersections of open sets are open.

Infinite intersection of open sets is not necessarily open. For example,

$$\{0\} = \bigcap_{n \in \mathbb{N}} \left(-\frac{1}{n}, \frac{1}{n}\right) \subset \mathbb{R}$$

is closed. Since $\left(\bigcap A_\beta\right)^c = \bigcup A_\beta^c$ and $(\bigcup A_i)^c = \bigcap A_i^c$, arbitrary intersections and finite unions of closed sets are closed.

Lemma 5.4.1 (Subset trichotomy)**.** *Let X be a subset of $\mathbb{R}^n$ and $z \in \mathbb{R}^n$. Then exactly one of the following holds:*

1. *there exists an open rectangle Q with $z \in Q \subset X$. We say that z is in the* ***interior*** *of X.*
2. *there exists an open rectangle Q with $z \in Q \subset X^c$. We say that z is in the* ***exterior*** *of X.*
3. *for all open rectangles such that $z \in Q$, we have $Q \cap X$ and $Q \cap X^c$ are both nonempty. We say that z lies on the* ***boundary*** *of X.*

For example, the closed ball $\overline{B}_r = \{x \in \mathbb{R}^n : |x| \leq r\}$ is closed. Its interior is B_r, its exterior is $\overline{B}_r^c$ and its boundary is $\{|x| = r\}$.

Exercise. Let $X = S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\} \subset \mathbb{R}^n$. Find

1. $\mathrm{int}\left(S^{n-1}\right)$ (answer: $\emptyset$)
2. $\mathrm{ext}(X)$ (answer: X^c)
3. $\mathrm{bd}(X)$ (answer: X)

5.3. **Compactness.**

Definition 5.5. An **open cover** of $X \subset \mathbb{R}^n$ is a collection U_β with $\beta \in B$ of open sets that cover X, i.e. such that $X \subset \bigcup_{\beta \in B} U_\beta$.

Some examples of open covers are:

1. $\{(n, n+2) : n \in \mathbb{Z}\}$ is a cover of $\mathbb{R}$
2. $\{(0, 3/4), (1/4, 1)\}$ is a cover of $(0, 1)$
3. $\{(1/n, 1) : n \in \mathbb{N}\}$ is also a cover of $(0, 1)$.

Definition 5.6. A subset $C \subset \mathbb{R}^n$ is **compact set** if every open cover of C has a finite subcover, i.e. given $C = \bigcup_{\beta \in B} U_\beta$ there exist $\beta_1, \ldots, \beta_r \in B$ such that $C \subset U_{\beta_1} \cup \cdots \cup U_{\beta_r}$.

To show that a subset $X \subset \mathbb{R}^n$ is *not* compact, it suffices to find *one* open cover with no finite subcover. For example, $\mathbb{R} = \bigcup (n, n+2)$ has no finite subcover, which implies that $\mathbb{R}$ is not compact. $(0, 1) = \bigcup (1/n, 1)$ also has no finite subcover, so $(0, 1)$ is not compact.

6.1. Onion ring theorem.

Definition 6.1. Given an onion O, a **center** is a point $x \in O$ which is contained in every ring.

Does every onion have a center? What about onions with infinitely many rings? Let's make this a little more formal. Let's think of an onion as a collection of nested rectangles. Then we have a theorem.

Theorem 6.2 (Onion ring theorem)**.** *Let $Q_k \subset \mathbb{R}^n$ be closed rectangles such that $Q_{k+1} \subset Q_k$. Then*

$$\bigcap_{k\geq 1} Q_k \neq \emptyset.$$

This theorem is also known as the **nested interval theorem**.

The proof reduces to the 1-dimensional case $I_k = [a_k, b_k] \subset \mathbb{R}$, where it states that $\bigcap I_k \neq \emptyset$. This is because the statement of the theorem really is a statement for every coordinate of $\mathbb{R}^n$.

Remark. It is important to use *closed* rectangles. In fact, $\bigcap(0, 1/n) = \emptyset$.

Proof of the 1*d case.* Consider the set

$$A = \{a_k : k \geq 1\},$$

and let $z = \sup A$. Why does z exist? A is nonempty since $a_1 \in A$, and is also bounded above by any value of b_k. By definition, $a_k \leq z$ for all k. If we show that $z \leq b_k$ for all k, then $z \in \bigcap I_k$. Note that each b_k is an upper bound on A. In fact, if we fix some a_r, then $a_r \leq b_k$ for any k, since $a_r \leq a_{r+k} \leq b_{r+k} \leq b_k$ for all $r, k \geq 1$. Then since z is the *least* upper bound, it follows that $z \leq b_k$ for all k. ☐

Recall from last time the following definitions:

Definition 6.3. We say that $U \subset \mathbb{R}^n$ is **open** if for every $u \in U$, there exists an open rectangle Q with $u \in Q \subset U$. We say that $A \subset \mathbb{R}^n$ is **closed** if A^c is open.

Definition 6.4. A subset $C \subset \mathbb{R}^n$ is **compact** if every open cover of C has a finite subcover, i.e. given $C = \bigcup_{\beta \in B} U_\beta$ there exist $\beta_1, \ldots, \beta_r \in B$ such that $C \subset U_{\beta_1} \cup \cdots \cup U_{\beta_r}$.

Last time we saw that neither $\mathbb{R}^n$ nor $(0,1)^n$ are compact. Today we will see examples of compact sets. The first theorem we will prove asserts that closed rectangles are compact. The second theorem (Heine-Borel theorem) states that $X \subset \mathbb{R}^n$ is compact if and only if X is closed and bounded (X is said to be **bounded** if there exists a closed rectangle R such that $X \subset R$). (we won't prove the Heine-Borel theorem today)

A sample corollary of the H-B theorem is the following:

Corollary 6.4.1. *The set*

$$S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$$

is compact.

In fact all of the sets X that we saw at the beginning of last lecture are compact. This leads to an improved boundedness theorem:

Theorem 6.5 (Improved boundedness theorem)**.** *A continuous function on a compact set is bounded, i.e.* $f : \mathbb{R}^n \to \mathbb{R}$ *is continuous and* $X \subset \mathbb{R}^n$ *is compact implies that there exists* M *such that* $f(x) \le M$ *for all* $x \in X$.

6.2. **Proof of Theorem 6.5.** Recall that if a function is continuous at $x \in X$ then f is bounded near x, i.e. for all $x \in X$ there exists $r_x \ge 0$ and M_x such that $f(y) \le M_x$ if $|y-x| < r_x$. The amounts r_x and M_x might vary with $x \in X$. Thus we need to show that the function is indeed bounded on all of X.

Consider

$$U_x = \{z \in \mathbb{R}^n : |z-x| < r_x\}.$$

The U_x's are open, and $X \subset \bigcup_{x \in X} U_x$, i.e. U_x for an open cover of X. We know that X is compact, which by definition means that we can find an open subcover $U_{x_1} \cup \cdots \cup U_{x_r}$ of X. Set $M = \max\left\{M_{x_1}, \dots, M_{x_r}\right\}$. Then $f(x) \le M$ for all $x \in X$. $\square$

Remark. In this proof, compactness helped us go from some local information (f is bounded on each point individually) to a *global* piece of information (f is bounded on the whole set).

6.3. **Compactness of closed rectangles.** For concreteness we will show that $Q = [0,1] \times [0,1]$ is compact, and you'll see that the argument works for any closed rectangle.

Proof. Suppose for a contradiction tat $\left\{U_\beta : \beta \in B\right\}$ is an open cover with no finite subcover. The idea is to divide Q into smaller rectangles (actually, into quadrants) so that our assumption that the cover has no finite subcover means that for a quadrant Q_1 this cover has no finite subcover either. We can repeat this argument inductively to get closed rectangles Q_k with the properties

(i) $Q_{k+1} \subset Q_k$
(ii) Q_k is not covered by a finite subset of $\left\{U_\beta\right\}_{\beta \in B}$
(iii) if $x, y \in Q_k$ then $|x-y| \le 2^{-k+\frac{1}{2}}$.

By the onion ring theorem (Theorem 6.2) there exists $z \in \bigcap Q_k$. Since $\left\{U_\beta\right\}_{\beta \in B}$ is an open cover, then $z \in Q_\alpha$ for some $\alpha \in B$. Since U_α is open, $Q_k \subset U_\alpha$ for k sufficiently large. This is because by definition of open sets there exists an open rectangle R such that $z \in R \subset U_\alpha$, and then we know that R has some finite size so that we can find a Q_k that fits in it. This contradicts property (ii) of the Q_k's. $\square$

6.4. **Further applications of compactness.**

- Improvement on the maximum value theorem:

  Theorem 6.6. *If* $f : \mathbb{R}^n \to \mathbb{R}$ *is continuous and* $X \subset \mathbb{R}^n$ *is compact, then there exists* $z \in X$ *such that* $f(x) \le f(z)$ *for all* $x \in X$. *Equivalently, given* $\mathrm{im}(f) = \{f(x) : x \in X\} \subset \mathbb{R}$, *we have that* $\sup(\mathrm{im}(f)) \in \mathrm{im}(f)$.

- We can define an operator norm. For $A : \mathbb{R}^n \to \mathbb{R}^n$ linear, consider $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = |Ax|$. The boundedness theorem/max value theorem says that there exists $u \in S^{n-1} \subset \mathbb{R}^n$ such that $|Ax| \leq |Au|$ for all $x \in S^{n-1}$. Therefore we can define $\|A\| = \sup\left\{|Ax| : x \in S^{n-1}\right\}$. This is the **operator norm**, and makes $M_n\mathbb{R}$ a metric space. If A is symmetric, u will be a eigenvector.

7. February 5

7.1. **Derivatives.** Let us start with a simple questions: given a function $f : \mathbb{R} \to \mathbb{R}$, which linear function $\ell(x) = mx + b$ best approximates f near $a \in \mathbb{R}$? For exaple, at $a = 0$ we have $f(0) = \ell(0) = b$, so that we know that ℓ must have the form $mx + f(0)$. To find m we consider secant lines, namely, for any h consider the line through $(0, f(0))$ and $(h, f(h))$. If your function is nicely behvaed, you would expect the slope to converge to some value. This slope is equal to

$$\frac{f(h) - f(0)}{h},$$

so that we expect that

$$\lim_{h \to 0} \frac{f(h) - f(0)}{h}$$

exists.

Definition 7.1. We say that a function $f : \mathbb{R} \to \mathbb{R}$ is **differentiable** at $a \in \mathbb{R}$ if

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

exists. We denote this limit $f'(a)$, and we view f' as a function defined where f is differentiable.

Example 7.2. Let $f(x) = x^2$. Then

$$\begin{aligned} f'(x) =& \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ =& \lim_{h \to 0} \frac{2xh + h^2}{h} \\ =& 2x. \end{aligned}$$

Example 7.3. Let $f(x) = \sin(1/x)$. Is f differentiable at 0? The first issue is that $\sin(1/0)$ is not defined, and so f is not defined at 0. Even if we defined $f(0) = 0$, the function would not be continuous.

Lemma 7.3.1 (Differentiable implies continuous)**.** *If f is differentiable at a point a, then f is continuous at a.*

Proof. We want to show that $\lim_{x \to a} f(x) = f(a)$, or equivalently, $\lim_{x \to a} f(x) - f(a) = 0$. To do so we use the limit in the definition of differentiability:

$$\begin{aligned} \lim_{x \to a} f(x) - f(a) =& \lim_{h \to 0} f(a+h) - f(a) \\ =& \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \cdot h \\ =& 0. \end{aligned}$$

$\square$

Example 7.4. Let

$$f(x) = \begin{cases} x\sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

We showed in Example 1.3 that f is continuous at 0. But is it differentiable? We see that

$$\begin{aligned} f'(0) &= \lim_{h\to 0} \frac{f(h) - f(0)}{h} \\ &= \lim_{h\to 0} \sin\left(\frac{1}{h}\right) \end{aligned}$$

which does not exists. Therefore f is not differentiable at 0.

Remark. As we can see from lemma 7.3.1, differentiability is a refinement of what it means for a function to be "nice".

Exercise. Let

$$f(x) = \begin{cases} x^2 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}.$$

Is f differentiable at 0? We see that

$$\frac{f(h) - f(0)}{h} = \begin{cases} h & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$$

and both converge to 0.

For the functions we have seen so far, we needed to apply the definition at every point. However, we can appeal to rules of differentiation.

7.2. **Computing derivatives.** We expect $f(x) = x\sin(1/x)$ to be differentiable at $x \neq 0$. This is because $x\sin x$ and $1/x$ are.

Theorem 7.5 (Algebra of derivatives)**.** *Let f, g be differentiable at a. Then*

(1) Sum rule: $(f+g)')(a) = f'(a) + g'(a)$
(2) Product rule: $(fg)'(a) = f(a)g'(a) + f'(a)g(a)$
(3) Quotient rule: for $f(a) \neq 0$,

$$\left(\frac{1}{f}\right)'(a) = -\frac{f'(a)}{f(a)^2}.$$

Theorem 7.6 (Chain rule)**.** *Let g be differentiable at a, and f be differentiable at $g(a)$. Then*

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

Proof of product rule. We see that

$$\begin{aligned}\lim_{x\to a}\frac{f(x)g(x)-f(a)g(a)}{x-a} &= \lim_{x\to a}\frac{f(x)g(x)-f(a)g(x)+f(a)g(x)-f(a)g(a)}{x-a}\\ &= \lim_{x\to a}\left(\frac{f(x)-f(a)}{x-a}g(x)+f(a)\frac{g(x)-g(a)}{x-a}\right)\\ &= f'(a)g(a)+f(a)g'(a)\end{aligned}$$

seeing as $g(x)$ approaches $g(a)$ from Lemma 7.3.1, and the two quotients approach the derivative by definition. □

For the chain rule, we are going to introduce a flawed proof first.

Incomplete proof of the chain rule. To compute $(f\circ g)'(a)$ we need to compute

$$\lim_{h\to 0}\frac{f(g(a+h))-f(g(a))}{h} = \lim_{h\to 0}\frac{f(g(a+h))-f(g(a))}{g(a+h)-g(a)}\cdot\frac{g(a+h)-g(a)}{h}. \tag{†}$$

What does it mean for f to be differentiable at $g(a)$? By definition,

$$\lim_{k\to 0}\frac{f(g(a)+k)-f(g(a))}{k}$$

exists. Then define $k(h)=g(a+h)-g(a)$, so that $f(g(a+h))=f(g(a)+k(h))$. Then

$$\begin{aligned}\lim_{h\to 0}\frac{f(g(a+h))-f(g(a))}{g(a+h)-g(a)} &= \lim_{h\to 0}\frac{f(g(a)+k(h))-f(g(a))}{k(h)}\\ &= f'(g(a)). \tag{‡}\end{aligned}$$

Where is the flaw in this argument? First of all, equation (†) is suspicious. What if $g(x)$ is constant? In this case, the proof does not hold but the chain rule still does hold (the derivative is 0). What if $g(x)$ is equal to 0 infinitely often around 0? And how do we justify going to line (‡)? □

8. February 7

8.1. Chain rule. Recall. We define the derivative at a as

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

There are three main notations for derivatives:

- the one due to Lagrange, f';
- the one due to Euler, Df (since differentiation is a linear operator);
- the one due to Leibniz, $\frac{df}{dx}(a)$, namely the "value" of $\frac{f(a+dx)-f(a)}{dx}$.

Recall that the chain rule theorem states

Theorem 8.1 (Chain rule)**.** *Let g be differentiable at a, and f be differentiable at $g(a)$. Then*

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

Proof. Last time we wrote

$$\lim_{h\to 0}\frac{f(g(a+h))-f(g(a))}{h}=\lim_{h\to 0}\frac{f(g(a+h))-f(g(a))}{g(a+h)-g(a)}\cdot\frac{g(a+h)-g(a)}{h}. \tag{†}$$

, which is problematic when $g(a+h)=g(a)$. To fix this, we replace (†) with $\varphi(h)$ such that

$$\varphi(h)\frac{g(a+h)-g(a)}{h}=\frac{f(g(a_h))-f(g(a))}{h}\text{ for all }h \tag{$*$}$$

$$\lim_{h\to 0}\varphi(h)=f'(g(a)). \tag{$**$}$$

Once this is done, the proof is easy:

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

We define $\varphi(h)$ as

$$\varphi(h)=\begin{cases}\frac{f(g(a+h))-f(g(a))}{g(a+h)-g(a)} & g(a+h)-g(a)\neq 0\\ f'(g(a)) & \text{else}\end{cases}.$$

Defined this way, it is clear that φ satisfies ($*$). For it to satisfy ($**$) we need to show that $\lim_{h\to 0}\varphi(h)=f'(g(a))$. Define $k(h)=g(a+h)-g(a)$. Then φ can also be defined as

$$\varphi(h)=\begin{cases}\frac{f(g(a)+k(h))-f(g(a))}{k(h)} & k(h)\neq 0\\ f'(g(a)) & k(h)=0\end{cases}.$$

Roughly, we see that if h is small, then $k(h)$ is small, and so $\varphi(h)$ is near $f'(g(a))$ (since f is differentiable). We now prove this rigorously. Fix $\varepsilon>0$. Choose $\zeta>0$ so that $0<|\ell|<\zeta$ implies

$$\left|\frac{f(g(a)+\ell)-f(g(a))}{\ell}-f'(g(a))\right|<\varepsilon.$$

Now choose $\delta > 0$ so that $0 < |h| < \delta$ implies $|k(h)| < \zeta$. Then $0 < |h| < \delta$ implies $|\varphi(h) - f'(g(a))| < \varepsilon$. $\square$

Exercise. Compute f' for the following f:

(1) $f(x) = \sin(\sin(\sin x))$ (answer: $f'(x)) = \cos(\sin(\sin x))\cos(\sin x)\cos x$j)

(2) $f(x) = \sin(6\cos(6\sin 6x))$ (answer: $f'(x) = \cos(6\cos(6\sin 6x))\cdots$)

8.2. **Meaning of f'.** Consider

$$C^1(\mathbb{R},\mathbb{R}) \subset C(\mathbb{R},\mathbb{R}) \subset \mathrm{Fun}(\mathbb{R},\mathbb{R}),$$

where C^1 indicates differentiable f with f' continuous, and C indicates continuous f. These inclusions are actually vector subspace inclusions (by algebra of limits and derivatives). Moreover, the map

$$\begin{aligned} D : C^1(\mathbb{R},\mathbb{R}) &\to C(\mathbb{R},\mathbb{R}) \\ f &\mapsto Df \equiv f' \end{aligned}$$

is linear (since $(f+g)' = f' + g'$ and $(cf)' = cf'$). What are $\ker D$ and $\operatorname{Im} D$? We see that $\{f \text{ constant}\} \subset \ker D$. Is there anything more to it (in other words, is the inclusion proper)? To answer this question we need to do a bit more work.

Definition 8.2. Let $f : (a,b) \to \mathbb{R}$ and $y \in (a,b)$. We say that y is a **local maximum** if there exists $\delta > 0$ so that $f(y) \geq f(x)$ for all $x \in (y-\delta, y+\delta)$.

Theorem 8.3. *Let $y \in (a,b)$ be a local maximum or a local minimum (analogous definition), and let f be differentiable at y. Then $f'(y) = 0$.*

Remark. The converse is false, e.g. consider x^3 at $y = 0$.

Proof. Suppose y is a local maximum. Set $F(h) = \frac{f(y+h)-f(h)}{h}$. Then $f'(y) = \lim_{h\to 0} F(h)$. Choose δ as in the definition of local maximum. Then if $0 < h < \delta$ we have that $F(h) \leq 0$ and therefore $\lim_{h\to 0^+} F(h) \leq 0$; if $-\delta < h < 0$ then $F(h) \geq 0$ and therefore $\lim_{h\to 0^-} F(h) \geq 0$. Thus

$$0 \geq \lim_{h\to 0^+} F(h) = \lim_{h\to 0} F(h) = \lim_{h\to 0^-} F(h) \geq 0$$

and therefore $f'(y) = 0$. $\square$

Theorem 8.4 (Rolle's theorem). *Let $f : [a,b] \to \mathbb{R}$ be continuous and differentiable on (a,b). Then $f(a) = f(b)$ implies that there exists $c \in (a,b)$ such that $f'(c) = 0$*

Remark. This theorem is false if f' does not exist at *every point* in the interval (consider $|x| : [-1,1] \to \mathbb{R}$, which is not differentiable at $x = 0$).

Proof. By the maximum value theorem, f has a maximum c_1 and a mininum c_2 on $[a,b]$. If c_1 or $c_2 \in (a,b)$ then $f'(c_i) = 0$ by Theorem 8.3. If c_1, c_2 are at a, b then f is constant, and $f'(c) = 0$ for all $c \in (a,b)$. $\square$

Theorem 8.5 (Mean Value Theorem). *Let $f : [a,b] \to \mathbb{R}$ continuous on $[a,b]$ and differentiable on (a,b). Then there exists $c \in (a,b)$ such that*

$$\frac{f(b)-f(a)}{b-a} = f'(c).$$

Remark. If $f(x)$ is the position of a particle and x is time, then the left hand side is the average velocity and the right hand side is the instantaneous velocity at c.

Remark. The mean value theorem (MVT) is a generalization of Rolle's theorem, and we'll use Rolle to prove MVT.

Proof. Define

$$g(x) = \left(\frac{f(b) - f(a)}{b - a}\right)(x - a) + f(a).$$

We see that $g(a) = f(a)$ and $g(b) = f(b)$, and so by defining $h(x) = f(x) - g(x)$ we see that $h(a) = h(b) = 0$ and therefore we can use Rolle's theorem to see that there exists c such that

$$0 = h'(c) = f'(c) - g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}$$

□

Corollary 8.5.1. *The kernel of D is precisely the subspace of constant functions.*

9.1. **Polynomial approximation.** Last time we saw that $f'(0)$ if and only if f is constant. We did not really prove this but we are going to see a lemma which has a very similar proof.

Lemma 9.0.1. *If $f'(x) > 0$ for all x then f is increasing.*

Proof. The MVT states that given $a, b \in \mathbb{R}$ there exists $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b-a).$$

By assumption, both $b-a$ and $f'(c)$ are positive, and therefore so is $f(b) - f(a)$. □

The intuition to the last proof is that $f'(a)$ is defined so that the funcion $p(x) = f(a) + f'(a)(x-a)$ is the best linear approximation of f near a. Thus, if the slope is positive at every point, such functions will be increasing, and we should expect the same from f.

Definition 9.1. Two functions $f, g : \mathbb{R} \to \mathbb{R}$ **agree up to order** k **at** a if

$$\lim_{h \to 0} \frac{f(a+h) - g(a+h)}{h^k} = 0.$$

Example 9.2. Let $p(x) = f(a) + f'(a)(x-a)$. This polynomial agrees with f up to order 1. In fact,

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

Example 9.3. Let

$$\begin{aligned} P(x) &= a_n x^n + \cdots + a_1 x + a_0 \\ Q(x) &= b_n x^n + \cdots + b_1 x + b_0. \end{aligned}$$

Then

$$\begin{aligned} \frac{P(h) - Q(h)}{h^k} =& (a_n - b_n) h^{n-k} + \cdots + (a_{k+1} - b_{k+1}) h \\ &+ (a_k - b_k) + (a_{k-1} - b_{k-1}) \frac{1}{h} + \cdots + (a_0 - b_0) \frac{1}{h^k}. \end{aligned}$$

Therefore P, Q agree up to order k at 0 if and only if $a_i = b_i$ for all $0 \leq i \leq k$.

We are interested in the following approximation problem: given $f : \mathbb{R} \to \mathbb{R}$, $k \geq 1$ and $a \in \mathbb{R}$ we want to find a polynomial $P_{f,k,a}(x)$ that agrees with f up to order k at a. We already saw the solution in the case $k = 1$, in which case $P_{f,1,a} = f(a) + f'(a)(x-a)$, and $p(a) = f(a), p'(a) = f'(a)$. We have the following theorem:

Theorem 9.4 (Taylor's theorem). *If $f'(a), \ldots, f^{(k)}(a)$ exist then*

$$P_{f,k,a}(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k$$

solves the approximation problem.

Remark. We have that $P^{(i)}_{f,k,a}(a) = f^{(i)}(a)$ for all $0 \leq i \leq k$. P is called the k-th order **Taylor polynomial** of f at a.

9.2. **Derivative magic wands.** Last time we saw that the MVT is a very powerful tool. There is a similar, stronger result:

Theorem 9.5 (Cauchy MVT)**.** *If f and g satisfy the same hypotheses as in the MVT, then there exists $c \in (a, b)$ such that*

$$[f(b) - f(a)]\, g'(c) = [g(b) - g(a)]\, f'(c).$$

Remark. If $g(b) - g(a) \neq 0$ and $g'(c) \neq 0$ we can write this as

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{b - a} \frac{b - a}{g(b) - g(a)} = \frac{f'(c_1)}{g'(c_2)},$$

where c_1 and c_2 come from the MVT applies to each function.

Proof. Found in Spivak's Calculus; similar to MVT. $\square$

Theorem 9.6 (L'Hospital's rule)**.** *Let f, g be continuous with $f(a) = g(a) = 0$. If $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ exists, then*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

Sample application. We can apply this rule to a problem from the first homework:

$$\lim_{x\to 0} \frac{1 - \sqrt{1 - x^2}}{x^2} = \lim_{x\to 0} \frac{\frac{x}{\sqrt{1-x^2}}}{2x} = 1/2.$$

Proof of l'Hospital's rule. Write

$$L = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

We want to show that given $\varepsilon > 0$ we can find $\delta > 0$ such that $0 < |x - a| < \delta$ implies $\left|\frac{f(x)}{g(x)} - L\right| < \varepsilon$. By Cauchy's MVT, we can choose $\delta > 0$ so that $0 < |x-a| < \delta$ then $\left|\frac{f'(x)}{g'(x)} - L\right| < \varepsilon$. Then for some c such that $0 < |c - a| < |x - a| < \delta$ we have that

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(a)}.$$

Therefore

$$\left|\frac{f(x)}{g(x)} - L\right| = \left|\frac{f'(c)}{g'(c)} - L\right| < \varepsilon.$$

$\square$

9.3. **Taylor's Theorem.** We are now ready to prove Taylor's theorem. We will prove if for the case $n = 2$, and the rest will follow by induction.

Proof of Taylor's theorem. In the case $k = 2$ we have that

$$P(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2.$$

To evaluate the limit

$$\lim_{h\to 0} \frac{f(a+h) - P(a+h)}{h^2}$$

we use l'Hopital's rule so that

$$\begin{aligned}\lim_{h\to 0} \frac{f(a+h) - P(a+h)}{h^2} =& \lim_{h\to 0} \frac{f'(a+h) - f'(a) - f''(a)h}{2h} \\ =& \frac{1}{2}\left(\lim_{h\to 0} \frac{f'(a+h) - f'(a)}{h} - f''(a)\right) \\ =& 0.\end{aligned}$$

Note that you can't use l'Hopital's once again, seeing as we would get to the expression

$$\frac{f''(a+h) - f''(a)}{2}$$

which is not defined seeing as f need only be differentiable up to order 2 at a. $\square$

Remark. By definiton, if $f''(a)$ exists then f' exists *near* a.

9.4. **Application.** Last time we saw that if c is a local max/min of f then this implies that $f'(c) = 0$ and we saw that the reverse implication is not true. There's a result that tells you when the reverse statement it is true, which is called the **second derivative test**.

Theorem 9.7 (Second derivative test)**.** *Let f be differentiable, $f'(c) = 0$. Then*

- *if $f''(c) > 0$ then c is a local miminum;*
- *if $f''(c) < 0$ then c is a local maximum.*

Proof. Consider $P(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2$, the Taylor polynomial to order 2 at a. By Taylor's theorem,

$$0 = \lim_{x\to a} \frac{f(x) - P(x)}{(x-a)^2} = \lim_{x\to a} \frac{f(x) - f(a)}{(x-a^2)} - \frac{f''(a)}{2}.$$

Therfore the two have the same signs, and the result follows. $\square$

10.1. **Directional derivatives.** Our goal for today is to generalize our definition of derivative for functions $f : \mathbb{R}^n \to \mathbb{R}^m$.

Example 10.1. Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x_1, x_2) \mapsto x_1 - x_2 + x_1 x_2$.

There are many ways of plotting this function. One might just draw it in 3d, but also a contour plot is a good way [shows a Mathemaica file with a contour plot]. The contour plot by definition plot plots the intersection between planes parallel to the $x_1 - x_2$ plane and our function, thereby allowing us to visualize points on the plot that have a common value.

To start addressing the problem of a derivative on such a function, we will focus at $x = 0$. For a vector (direction) $v \in \mathbb{R}^2$ we restrict f to $\{tv : t \in \mathbb{R}\}$ to get a function

$$\begin{aligned} g : \mathbb{R} &\to \mathbb{R} \\ t &\mapsto f(tv) = tv_1 - tv_2 + t^2 v_1 v_2. \end{aligned}$$

This is a function like the ones we've seen before, and the derivative along v at 0 is therefore equal to

$$g'(0) = v_1 - v_2.$$

Definition 10.2. For $f : \mathbb{R}^n \to \mathbb{R}$ let $a, v \in \mathbb{R}^n$. The **directional derivative** at a is defined as

$$D_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t}.$$

Example 10.3. For our first example, $D_v f(0) = v_1 - v_2$.

Remark. The directional derivatives in the direction of basis vectors e_i are called **partial derivatives** and are denoted by $D_{e_i} f(a) = D_i f(a)$. In our example, $D_1 f(0) = 1, D_2 f(0) = -1$ and $D_{(1,1)} f(0) = 0$ (the function is constant along the vector $(1,1)$).

Example 10.4. Define

$$\begin{aligned} g : \mathbb{R}^2 &\to \mathbb{R} \\ g(x) &= \begin{cases} \frac{x_1^2 x_2}{x_1^4 + x_2^2} & x \neq 0 \\ 0 & x = 0 \end{cases}. \end{aligned}$$

We want to calculate the directional derivative at $x = 0$. Fix $v = (v_1, v_2)$. Then

$$\begin{aligned} \lim_{t \to 0} \frac{f(tv)}{t} &= \lim_{t \to 0} \frac{1}{t} \left[\frac{t^3 v_1^2 v_2}{t^4 v_1^4 + t^2 v_2^2} \right] \\ &= \lim_{t \to 0} \frac{v_1^2 v_2}{t^2 v_1^4 + v_2^2} \\ &= \begin{cases} \frac{v_1^2}{v_2} & v_2 \neq 0 \\ v_2 & v_2 = 0 \end{cases}. \end{aligned}$$

The two example that we saw are quite different from one another. In particular, for the first example we have that $D_v f(0) = v_1 - v_2$ is linear, whereas in the second case $D_v g(0) = v_1^2/v_2$ is not. In fact, if we zoom the first contour plot at $x = 0$ we will see a linear plot, while the second gives no such thing (which is bad behavior). In fact, the second function g is not continuous at $x = 0$. In fact one can see from a contour plot that a lot of different level lines converge at 0. Along $(t, 0)$ we have that $g = 0$ and along (t, t^2) we have that $g(t, t^2) = 1/2$, and so there is a discontinuity.

10.2. **The derivative.**

Definition 10.5. We say that $f : \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable at** $a \in \mathbb{R}^n$ if there exists a linear map $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ so that

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

We write $Df(a) = \lambda$.

Remark. For $n = m = 1$ we have $Df(a) = f'(a)$ (this is a linear map in the way that a linear map $\mathbb{R} \to \mathbb{R}$ is given by multiplication by a scalar). The linear approximation of f at a is equal to

$$L(x) = f(a) + Df(a)(x - a).$$

Moreover, note that numerator and denominator are vectors.

Easy exercise.

(1) If f is constant, then $Df = 0$ (the zero linear map) for all $a \in \mathbb{R}^n$.
(2) If f is linear, then $Df = f$ at all a.
(3) $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable if and only if the coordinate function $f_i : \mathbb{R}^n \to \mathbb{R}$ is differentiable for all i.
(4) If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at a, then $D_v f(a) = Df(a) \cdot v$ (this is a homework exercise).

Corollary 10.5.1. *If $Df(a)$ exists then so do its partial derivatives.*

Proof of corollary. By definition,

$$\lim_{t \to 0} \frac{|f(a + te_i) - f(a) - Df(a)te_i|}{|te_i|} = \lim_{t \to 0} \left| \frac{f(a + te_i) - f(a)}{t} - Df(a)e_i \right| = 0$$

and so the two are equal. □

(5) If $Df(a)$ exists then f is continuous at a. (the function g from the second example is not differentiable at 0, despite the fact that all of the directional derivative exist)

Remark. The last point tells us that differentiability is a stronger claim than knowing that all of the directional derivatives exist.

A problem we want to address is the following: how do we decide (in practice) if $Df(a)$ exists?

Example 10.6. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $(x,y) \mapsto xy$. We see that $D_1 f(x,y) = y$ and $D_2 f(x,y)$. We might wonder wheteher or not $Df(x,y) = \begin{pmatrix} y & x \end{pmatrix}$. In fact,

$$\begin{aligned}
\lim_{h\to 0} \frac{|f(x+h) - f(x) - \begin{pmatrix} y & x \end{pmatrix} h|}{|h|} &= \lim_{h\to 0} \frac{\left|(x_1+h_1)(x_2+h_2) - x_1x_2 - \begin{pmatrix} y & x \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}\right|}{|h|} \\
&= \lim_{h\to 0} \frac{h_1h_2}{|h|} = 0
\end{aligned}$$

and therefore our guess was right.

11. February 14

11.1. **Continuous partials theorem.** From last time, recall that

Definition 11.1. We say that $f : \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable at** $a \in \mathbb{R}^n$ if there exists a linear map $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ so that

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

We write $Df(a) = \lambda$.

We also saw that if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable then

$$Df(a) = \begin{pmatrix} D_1 f(a) & \cdots & D_n f(a) \end{pmatrix}.$$

Remark. Computing $D_i f$ is just as easy as computing derivatives for funcions $\mathbb{R} \to \mathbb{R}$.

The problem with this is that existence of partial derivatives is not sufficient to guarantee that f is differentiable at a. However, the following theorem comes to the rescue.

Theorem 11.2 (Continuous partials theorem)**.** *Let $A \subset \mathbb{R}^n$ be open, and $f : A \to \mathbb{R}^m$ with components $f = (f_1, \dots, f_m)$. If $D_j f_i(x)$ exist and are continuous (as functions $A \to \mathbb{R}$) for all $x \in A, 1 \le i \le m$, $1 \le j \le n$, then f is differentiable on A, and*

$$Df(x) = \begin{pmatrix} D_1 f_1(x) & \cdots & D_n f_1(x) \\ \vdots & \ddots & \vdots \\ D_1 f_m(x) & \cdots & D_n f_m(x) \end{pmatrix}.$$

Example 11.3. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $f(r,\theta) = (r\cos\theta, r\sin\theta)$, i.e. the polar coordinate transformation. We see that $f_1(r,\theta) = r\cos\theta$ and $f_2(r,\theta) = r\sin\theta$, so

$$D_1 f_1(r,\theta) = \cos\theta \qquad D_2 f_1 = -r\sin\theta$$
$$D_1 f_2(r,\theta) = \sin\theta \qquad D_2 f_2 = r\cos\theta.$$

These functions are continuous on $\mathbb{R}^2$ so by theorem 11.2 f is differentiable, with

$$Df(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}.$$

Lemma 11.3.1. *If $f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ are differentiable then $f = (f_1, \dots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ is also differentiable.*

Proof. We will show that

$$Df(a) = \begin{pmatrix} Df_1(a) \\ \vdots \\ Df_m(a) \end{pmatrix}.$$

□

Let

$$\Delta = \begin{pmatrix} Df_1(a) \\ \vdots \\ Df_m(a) \end{pmatrix}.$$

We see that

$$\begin{aligned}&\frac{f(a+h)-f(a)-\Delta\cdot h}{|h|}\\ =&\frac{1}{|h|}\Big(f_1(a+h)-f_1(a)-Df_1(a)\cdot h \quad \cdots \quad f_m(a+h)-f_m(a)-Df_m(a)\cdot h\Big)\end{aligned}$$

and since f_i is differentiable for $1 \le i \le m$ the above goes to 0 as $h \to 0$.

We can now proceed to prove the continuous partials theorem.

Proof of CPT. By the lemma it suffices to consider $m = 1$. Let $f : A \to \mathbb{R}$. For concreteness we will do $n = 3$. Fix $a \in A$ and let

$$Df(a) = (D_1 f(a), \ldots, D_n f(a))$$

and define

$$F(h) = \frac{f(a+h)-f(a)-Df(a)\cdot h}{|h|}.$$

We want to show that $\lim_{h\to 0}|F(h)| = 0$. Since we can compute directional derivatives along coordinates axes (partial derivatives) we write $h = h_1e_1 + h_2e_2 + h_3e_3$. Let

$$\begin{aligned} p_0 = ap_1 &= a + h_1e_1 \\ p_2 &= p_1 + h_2e_2 \\ p_3 = p_2 + h_3e_3 &= a + h \end{aligned}$$

so that we go from a to $a+h$ along one coorinate axis at a time. Then

$$f(a+h)-f(a) = f(p_3)-f(p_2)+f(p_2)-f(p_1)+f(p_1)-f(p_0) = \sum_{i=1}^{n} f(p_i)-f(p_{i-1}).$$

This means that

$$F(h) = \frac{\sum_i f(p_i)-f(p_{i-1})-D_if(a)h_i}{|h|}.$$

Now consider the function

$$g_i(t) = f(p_{i-1}+te_i).$$

Note that

$$\begin{aligned} g_i(0) &= f(p_{i-1}) & g_i(h_i) &= f(p_1) \\ g_i'(t) &= D_if(p_{i-1}+te_i). \end{aligned}$$

By the mean value theorem there exists $t_i \in (0, h_i)$ such that

$$g_i(h_i) - g_i(0) = g_i'(t_i)h_i$$

and therefore

$$f(p_i) - f(p_{i-1}) = D_i f\,(p_{i-1} + t_i h_i)\,h_i$$

so that

$$|F(h)| = \frac{\left|\sum_i [D_i f(p_{i-1} + t_i h_i) - D_i f(a)] \cdot h_i\right|}{|h|} \qquad \leq \sum_i |D_i f(p_{i-1} + t_i h_i) - D_i f(a)| \cdot \frac{|h_i|}{|h|}.$$

Note that as $h \to 0$ then $h_i \to 0$ and $p_{i-1} + t_i h_i \to a$. Since $D_i f$ is continuous it follows that $D_i f(p_{i-1} + t_i h_i) \to D_i f(a)$. Moreover, $|h_i|/|h| \leq 1$ and therefore bounded, and so $\lim_{h\to 0} |F(h)| = 0$. □

Remark. Most functions we'll meet will have continuous partials; these functions are called **continuously differentiable**. The space of continuously differentiable functions is denoted as C^1.

11.2. **Mixed partials.** If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable we can compute the partials of $D_i f : \mathbb{R}^n \to \mathbb{R}$ (if they exist). The functions $D_j(D_i f)$ are called **second order partials** and if they are continuous we say that $f \in C^2$.

Example 11.4. For $f(x,y) = x^2 + y^2$ and $g(x,y) = xy$ compute $Df(0)$, $Dg(0)$, $D_j D_i f(0)$, and $D_j D_i g(0)$. We see that $Df(x,y) = \begin{pmatrix} 2x & 2y \end{pmatrix}$, $Dg(x,y) = \begin{pmatrix} y & x \end{pmatrix}$ and

$$D_j D_i f = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$D_j D_i g = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The matrix $Hf(a) := D_j D_i f(a)$ is called the Hessian. It is a symmetric matrix, and its eigenvectors (which form an orthonormal basis by the spectral theorem) give us information about the behavior of the function.

12. February 16

12.1. Multivariable chain rule.

Theorem 12.1 (Chain rule)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^m \to \mathbb{R}^p$. If $Df(a)$, $Dg(f(a))$ exist then*

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

Remark. When we write $DDg(f(a)) \cdot Df(a)$, the multiplication is matrix multiplication.

Example 12.2. Consider $f : \mathbb{R} \to \mathbb{R}^n, g : \mathbb{R}^n \to \mathbb{R}$ with $f, g \in C^1$ and $h = g \circ f$. Then

$$Df(a) = \begin{pmatrix} f_1'(a) \\ \vdots \\ f_n'(a) \end{pmatrix}, \qquad Dg(b) = \begin{pmatrix} D_1 g(b) & \cdots & D_n g(b) \end{pmatrix}$$

and therefore

$$Dh(a) = \sum_{i=1}^{n} D_i g(f(a)) f_i'(a), j$$

which could be written in Leibniz notation as

$$\frac{dh}{dx}(a) = \sum_{i=1}^{n} \frac{dg}{dx_i} \frac{df_i}{dx}.$$

Corollary 12.2.1. *If $f, g \in C^1$ then $g \circ f \in C^1$.*

Proof. We need to show that $D_j(g \circ f)_i$ is continuous for all $1 \le i \le p, 1 \le j \le n$. In fact,

$$\begin{aligned} D_j(gf)_i =& D(gf)_{ij} \\ =& (Dg \cdot Df)_{ij} \\ =& \sum_{i=1}^{m} (Dg)_{ir}(Df)_{rj} \\ =& \sum_{i=1}^{m} D_r g_i \cdot Djf_r \end{aligned}$$

and the last line is continuous by continuity of the factors. □

Corollary 12.2.2. *We can derive 1-dimensional derivative rules, for example the product rule (namely, for $\psi, \varphi : \mathbb{R} \to \mathbb{R}$ we have $(\varphi\psi)' = \varphi'\psi + \varphi\psi'$).*

Proof. Consider the maps $f : \mathbb{R} \to \mathbb{R}^2$, $g : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x) = (\varphi(x), \psi(x))$ and $g(x, y) = xy$. Then

$$\begin{aligned} (\varphi\psi)'(x) =& D(gf)(x) \\ =& Dg(f(x))Df(x)Df(x) \\ =& \varphi'\psi + \psi'\varphi. \end{aligned}$$

□

Another example of an application is the case $\varphi, \psi : \mathbb{R} \to \mathbb{R}^n$, $h(x) = \langle \varphi(x), \psi(x) \rangle$, in which case one analogously shows that $h'(x) = \langle \varphi'(x), \psi(x) \rangle + \langle \varphi(x), \psi'(x) \rangle$.

Corollary 12.2.3 (Multivariable MVT)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ differentiable. Given $a, a+h$, there eixsts $c = a + sh$ with $s \in (0,1)$ so that $f(a+h) - f(a) = Df(c) \cdot h$.*

Proof. We reduce the proof to the 1-dimensional case. Consider the map $g(t) = a + th$. Then $H = f \circ g : [0,1] \to \mathbb{R}$. Then there exists $s \in (0,1)$ such that

$$f(a+h) - f(a) = H(1) = H(0) = H'(s) \cdot 1 = Df(g(s)) \cdot Dg(s) = Df(c)h$$

with $c = g(s)$. □

12.2. **Proof of the chain rule.** **Notation:** Let $b = f(a)$, and let

$$\begin{aligned} F(h) &= \frac{f(a+h) - f(a) - Df(a) \cdot h}{|h|} \\ G(k) &= \frac{g(b+k) - g(b) - Dg(b) \cdot k}{|k|}. \end{aligned}$$

We know that $\lim_{h \to 0} |F(h)| = \lim_{k \to 0} |G(k)| = 0$. We want to show

$$\lim_{h \to 0} = \frac{|g(f(a+h)) - g(f(a)) - Dg(b)Df(a) \cdot h|}{|h|} = 0.$$

The first step will be to use algebra to rewrite the above expression in terms of F, G and other things we know. Let $k(h) = f(a+h) - f(a)$. Then

$$\begin{aligned} g(f(a+h)) - g(f(a)) &= g(b + k(h)) - g(b) \\ &= |k(h)||G(k(h))| + Dg(b)k(h) \end{aligned}$$

and

$$\begin{aligned} Df(a) \cdot h &= f(a+h) - f(a) - |h|F(h) \\ &= k(h) - |h| \cdot F(h). \end{aligned}$$

Now we can rewrite

$$\begin{aligned} \frac{|g(f(a+h)) - g(f(a)) - Dg(b)Df(a)h|}{|h|} &= \frac{\big||k(h)|G(k(h)) + Dg(b)k(h) - Dg(b)\left[k(h) - |h| \cdot F(h)\right]\big|}{|h|} \\ &\le \frac{|k(h)|}{|h|}|G(k(h))| + |Dg(h) \cdot F(h)|. \end{aligned}$$

We see that $G(k(h))$ goes to 0 by continuity of f. The term $|Dg(h)F(h)|$ is bounded by $M|F(h)|$ by HW #2.2, and $|F(h)| \to 0$ by differentiability of f. We now need to show $|k(h)|/|h|$ is bounded. In fact,

$$\frac{|k(h)|}{|h|} = \frac{|F(h)|h| + |Df(a) \cdot h||}{|h|} \le |F(h)| + M'\frac{|h|}{|h|}$$

and so this term is bounded, which proves the theorem. □

Corollary 12.2.4. *If $f : \mathbb{R}^n \to \mathbb{R}^n$ has inverse f^{-1} and both are differentiable then*

$$Df^{-1}(x) = Df\left(f^{-1}(x)\right)^{-1}.$$

Proof. We know $f \circ f^{-1} = \mathrm{id}$. Then
$$I = D\,\mathrm{id}(x) = Df(f^{-1}(x)) \cdot Df^{-1}(x).$$
$\square$

13. February 21

13.1. **Inverse function theorem.** When does a differentiable function $f : \mathbb{R}^n \to \mathbb{R}^n$ have an inverse? And when is the inverse differentiable?

Example 13.1. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $(x,y) \mapsto \big(\sin(x+y), x^2 - y^2\big)$. Given $(a,b) \in \mathbb{R}^2$, can we solve $\sin(x+y) = a$ and $x^2 - y^2 = b$ uniquely? Does the solution depend smoothly on (a,b)?

Theorem 13.2 (1-dimensional inverse function theorem)**.** *Let $f : [a,b] \to \mathbb{R}$, with f' continuous, $f'(x) \neq 0$ for all $x \in (a,b)$. Then f is invertible and $f^{-1} : f([a,b]) \to [a,b]$ is C^1.*

Proof. We know from the intermediate value theorem that either $f'(x) > 0$ or $f'(x) < 0$ for all $x \in (a,b)$ (since f' is continuous). Therefore, by the mean value theorem f is either strictly increasing or decreasing, and so is injective. For $y = f(x)$, define $g(y) = x$, so $g(f(x)) = x$. If g is differentiable, by the chain rule

$$g'(y) = \frac{1}{f'(x)}.$$

□

Remark. The issue of how we determine whether or not g is differentiable is going to be addressed later.

Example 13.3. Let $f : [-\pi, \pi] \to \mathbb{R}$ given by $x \mapsto 2x + \sin x$. In this case, $f'(x) = 2 + \cos x \geq 1$. Moreover, $f(-\pi) = -2\pi, f(\pi) = 2\pi$. Therefore f has an inverse $g : [-2\pi, 2\pi] \to [-\pi, \pi]$. There is no good formula for g. For example, to compute $g(1)$ we must solve the equation $2x + \sin x = 1$. We could approximate $g(1)$ by the bisection method: we start by finding a, b such that $f(a) < 1 < f(b)$, and set $x_0 = \frac{a+b}{2}$. Then either $f(x_0) = 1$ (which is unlikely), or if $f(x_0) < 1$ we replace a by $\frac{a+b}{2}$ and if $f(x_0) > 1$ we replace b by $\frac{a+b}{2}$. Then we repeat these steps.

Remark. Even if f is not increasing or decreasing on some interval $[a,b]$, it may have an inverse locally. For example, $f(x) = x^2$ is not increasing (or decreasing) on $\mathbb{R}$, but an inverse exists on $(-\infty, 0)$ and $(0, \infty)$.

Remark. If f is differentiable and invertible, the inverse f^{-1} may not be differentiable (for example $f(x) = x^3$ has an inverse, $g(y) = y^{1/3}$, which is not differentiable at 0).

The slogan for the inverse function theorem is that if $Df(a)$ is invertible then f is invertible near a.

Theorem 13.4 (Inverse function theorem)**.** *Suppose $A \subset \mathbb{R}^n$ is open and $f : A \to \mathbb{R}^n$. If f is C^1 and $Df(a)$ is invertible, then there exists an open set $a \in U \subset A$ such that*

- *f is injective on U*
- *$f(U)$ is open*
- *$f^{-1} : f(U) \to U$ is C^1.*

Example 13.5. Consider $f:\mathbb{R}^2\to\mathbb{R}^2$ given by $(x,y)\mapsto(xy,x^2-y^2)$. Then

$$Df(x,y)=\begin{pmatrix} y & x\\ 2x & 2y\end{pmatrix}.$$

We see that $\det Df(x,y)=-2(x^2+y^2)$ which is nonzero as long as $(x,y)\neq(0,0)$. For example, $f(-1,1)=(1,0)$. The inverse function theorem tells us there is an open set U around $(1,1)$ such that the above conditions are satisfied. Note that the inverse function theorem is silent about the behavior of f at $(0,0)$. It also says noting about global invertibility, and in fact $f(1,1)=f(-1,-1)$.

13.2. **Toward the proof of the IFT.**

Lemma 13.5.1. *Let $f:A\to\mathbb{R}^n$ in C^1, with $Df(a)$ invertible. Then there exists an open rectangle Q with $a\in Q\subset A$ such that f is injective on Q.*

The intuition for this is that if f is linear, i.e. if $f(x)=Tx$ for some $T\in M_n(\mathbb{R})$, then $Df(a)=T$ and therefore $Df(a)$ is invertible if and only f is. Generally, we can approximate $f(x)$ as

$$f(x)\sim f(a)+Df(a)(x-a).$$

Therefore one would expect that f is invertible close to a. We saw that in the 1-dimensional case this follows from the MVT. To prove this lemma we will need a bit of work. Recall that

- the multivariable MVT states that for $g:A\to\mathbb{R}$ there exists c on the segment from x to y such that $g(y)-g(x)=Df(c)(y-x)$. Moreover,
- for $B\in M_{n\times m}\mathbb{R}\cong\mathbb{R}^{nm}$ we can define a sup norm as $\|B\|=\max_{i,j}|b_{ij}|$. As an easy exercise one can check that for $B\in M_{n\times m}\mathbb{R}, C\in M_{m\times p}\mathbb{R}$ we have

$$\|BC\|\le m\|B\|\cdot\|C\|.$$

Proof of Lemma 13.5.1. We will find $\alpha>0$ and Q such that $\|f(x_1)-f(x_0)\|\ge\alpha\|x_1-x_0\|$ for all $x_1,x_0\in Q$ (here the norm is the sup norm). Let $E=Df(a)$ and let $h(x)=f(x)-E(x)$. By definition, $Dh(a)=0$. We want to know the coordinate functions of $h=(h_1,\dots,h_n)$.

Estimate 1. For $x,y\in A$ we have that $h_i(y)-h_i(x)=Dh_i(c_i)(y-x)$ for some $c_i\in A$. Therefore $\|h(y)-h(x)\|\le\beta\|y-x\|$ where $\beta=\max\|Dh_i(c_i)\|$. Therefore

$$\begin{aligned}\|h(y)-h(x)\| &=\|f(y)-f(x)-E(y-x)\|\\ &\ge\Big|\|f(y)-f(x)\|-\|E(y-x)\|\Big|.\end{aligned}$$

E is invertible and therefore $\|y-x\|=\|E^{-1}E(y-x)\|\le n\cdot\|E^{-1}\|\cdot\|E(y-x)\|$. Therefore

$$\|E(y-x)\|\ge\frac{1}{n\|E^{-1}\|}\|y-x\|.$$

Combining all together, we find that

$$\frac{1}{n\|E^{-1}\|}\|y-x\|-\|f(y)-f(x)\|\le\beta\|y-x\|$$

and thus

$$\|f(y) - f(x)\| \geq \left(\frac{1}{n\|E^{-1}\|} - \beta\right)\|y - x\|.$$

Since $Dh(a) = 0$ we can choose Q around a. $\square$

14. February 23

Last time, we proved

Lemma 14.0.1. *Let $f : A \to \mathbb{R}^n$ in C^1, with $Df(a)$ invertible. Then there exists an open rectangle Q with $a \in Q \subset A$ such that f is injective on Q.*

Today we will prove what is missing from proving the inverse function theorem, namely that if $f : U \to \mathbb{R}^n$ is C^1 and injective, with $Df(x)$ invertible for all $x \in U$, then

(1) $f(U)$ is open
(2) $f^{-1} : f(U) \to U$ is continuous
(3) f^{-1} is differentiable.

As a warmup, we will prove the following:

Lemma 14.0.2. *Let $Q \subset \mathbb{R}^n$ be an open rectangle, and $\varphi : Q \to \mathbb{R}$ be differentiable. If $p \in Q$ is a minimum (namely $\varphi(p) \leq \varphi(q)$ for all $q \in Q$) then $D\varphi(p) = 0$.*

Proof. Write $D\varphi(p) = (D_1\varphi(0), \ldots, D_n\varphi(0))$. Then the proof follows as in the 1-dimensional case: the quotient

$$\frac{\varphi(p + te_i) - \varphi(p)}{t}$$

has a numerator which is always positive. The denominator is negative for $t < 0$ and positive otherwise. Since the two limits converge, the limit is 0. $\square$

Remark. This lemma does not apply if we remove the condition that Q is open. In fact, consider

$$\begin{aligned} \varphi : [0,1]^2 &\to \mathbb{R} \\ x &\mapsto |x|^2. \end{aligned}$$

This has a maximum at $(1,1)$ but the derivative is not 0 there.

We now proceed to prove (1), (2), and (3).

Proof of (1). We want to show that $f(U) := V$ is open. Fix $v = f(u) \in V$. We will show that there exists $\delta > 0$ such that $B_\delta(v) \subset V$. Since U is open, there exists an open rectangle $u \in Q \subset U$. We can assume $\overline{Q} \subset U$ (if our first guess doesn't work, shrink Q a little). Consider now w near v. We now consider the map $\varphi : \overline{Q} \to \mathbb{R}$ given by $x \mapsto |f(x) - w|^2 = \sum_{i=1}^n (f_i(x) - w_i)$. $\square$

Therefore, $D\phi(p) = 0$. By the chain rule,

$$0 = D\phi(p) = 2(f(p) - w) \cdot Df(p).$$

Therefore $f(p) - w = 0$ (since $Df(p)$ is invertible). The issue is that p might lie on the boundary of $\overline{Q}$. We know that the boundary of Q is compact, and by a problem on the homework $f(\operatorname{bd} Q)$ is compact as well, hence closed. Again by a homeword problem we know that there exists some d so that $B_d(v) \cap f(\operatorname{bd} Q) = \emptyset$. If we choose $\delta = d/2$ then if $w \in B_\delta(v)$ we get

$$\phi(u) = |f(u) - v|^2 = |v - w|^2 < \delta^2$$

and $\phi\big|_{\mathrm{bd}\,Q} > \delta$. Therefore the minimum value of ϕ can't occur on $\mathrm{bd}\,Q$. Therefore, if $w \in B_\delta(v)$ there exists $p \in U$ such that $f(p) = w$. Therefore $B_\delta(v) \subset f(U) = V$ and V is open.

Proof of (2). We want to show that $f^{-1} : f(U) \to U$ is continuous. This means that for every open $W \subset U$ we have that $\left(f^{-1}\right)^{-1}(W)$ is open. We just showed that for all $W \subset U$ the image $f(W)$ is open, so f^{-1} is continuous. □

Remark. Parts (1) and (2) are a weaker version of

Theorem 14.1 (Brower's invariance of domain). *If $A \subset \mathbb{R}^n$ is open and $f : A \to \mathbb{R}^n$ is continuous and injective, then $f(A)$ is open.*

The difference with what we proved is that we considered C^1 functions with invertible differential. This general case can be proved with algebraic topology.

Proof of (3). We want to show that f^{-1} is differentiable and the inverse is continuous. Fix a and let $b = f(a)$. Let $g = f^{-1}$ and $E = Df(a)$.
Recall. By the chain rule if g is differentiable at b then

$$Dg(b) = Df(a)^{-1} = E^{-1}.$$

Thus we want to show

$$\lim_{k\to 0} \frac{g(b+k) - g(b) - E^{-1}k}{|k|} = 0.$$

We know

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

Now comes some heroic algebra. We set $\Delta(k) = g(b+k) - g(b)$. Then

$$\begin{aligned}\frac{g(b+k) - g(b) - E^{-1}k}{|k|} =& \frac{\Delta(k) - E^{-1}k}{|k|} \\ =& \frac{E^{-1}\left(E\Delta(k) - k\right)}{|k|}.\end{aligned}$$

We have that

$$\begin{aligned} k =& (b+k) - b \\ =& fg(b+k) - fg(b) \\ =& f(g(b) + \Delta(k)) - f(g(k)) \\ =& f(a + \Delta(k)) - f(a)\end{aligned}$$

and so the above becomes

$$\frac{E^{-1}\left(E\Delta(k) - k\right)}{|k|} = E^{-1}\frac{E\Delta(k) + f(a) - f(a+\Delta(k))}{|\Delta(k)|} \cdot \frac{|\Delta(k)|}{|k|}.j$$

We know that the middle factor goes to 0 as $k \to 0$ because f is differentiable and g is continuous (hence $\Delta(k) \to 0$). Therefore, applying a linear map still makes it approach 0. Now,

$$\frac{|\Delta(k)|}{|k|} = \frac{|g(b+k) - g(b)|}{|f(g(b+k)) - f(g(b))|}$$

is bounded because f is differentiable (the proof of this is left as an exercise). There is an issue, namely if $\Delta(k) = 0$. However, we know that this is not possible since g is bijective. This completes the proof. $\square$

15. February 26

15.1. **Least upper bound property revisited.**

Definition 15.1. An **ordered field** F is a field equipped with an **ordering** $<$ such that

(i) for all $\alpha, \beta, \gamma \in F$, if $\beta < \gamma$ then $\alpha + \beta < \alpha + \gamma$;
(i) if $xy > 0$, then $x < 0$ implies $y > 0$.

Example 15.2. The rational numbers $\mathbb{Q}$ with the ueusal order is an ordered field, and so is $\mathbb{Q}(\sqrt{2})$. However, $\mathbb{Q}(i)$ is not.

Theorem 15.3. *There exists an ordered field $\mathbb{R}$ with the least upper bound property. Furthermore $\mathbb{Q} \subset \mathbb{R}$. (This theorem can be found in Rudin's* Principles of Analysis, *pp 17-21)*

The idea of the proof is to contstruct such a field using "cuts", namely sets of rational numbers that satisfy the LUB property.

Proof. Step 1. We introduce Dedekind cuts.

Definition 15.4. A **cut** is a subset $\alpha \subset \mathbb{R}$ satisfying

(I) $\alpha \neq 0$ and $\alpha \neq \mathbb{Q}$;
(I) if $p \in \alpha$, then $q \in \alpha$ for all $q < p$;
(I) if $p \in \alpha$, then there exists $r \in \alpha$ such that $r > p$.

Remark. Property (III) guarantees inclusion $\mathbb{Q} \subset \mathbb{R}$.

Denote the set of all cuts by $\mathbb{R}$.

Step 2. We now define an order on $\mathbb{R}$. For $\alpha, \beta \in \mathbb{R}$, declare $\alpha < \beta$ if α is a proper subset of β. We now need to check this is actually an order. Namely we need to check

(i) if $\alpha < \beta$ and $\beta < \gamma$ then $\alpha < \gamma$;
(i) exacyly one of $\alpha < \beta$, $\alpha = \beta$, $\beta < \alpha$ holds (trichotomy).

You should convince yourself that (i) holds. For (ii), we know that at most one of these is true. Now, assume that the first two do not hold, namely there exists $a \in \alpha$ such that $a \notin \beta$. Property (II) implies that $b < a$ for all $b \in \beta$, which means (again by (II)) that $\beta < \alpha$. This shows that $\mathbb{R}$ together with $<$ is an ordered set.

Step 3. We now prove the LUB property. Take $A \subset \mathbb{R}$, and suppose A has upper bound β. Let

$$\gamma = \bigcup_{\alpha \in A} \alpha.$$

Then $\gamma \in \mathbb{R}$. To verify this we need to check axioms (I) to (III). By construction γ is not empty nor is it the whole of $\mathbb{Q}$ (because it is bounded). Therefore it satisfies (I). For (II), if we take any element $c \in \gamma$, then $c \in \alpha$ for some $\alpha \in A$, and so $q \in A$ for all $q < p$. (III) follows similarly.

Claim. γ is a least upper bound, i.e. anu $\delta < \beta$ is not an upper bound.

Proof of the claim. Take $a \in \gamma$ such that $a \notin \delta$. Then $a \in \alpha$ for some $\alpha \in A$, and now δ is not an upper bound, since otherwise $\alpha < \delta$ and hence $a \in \delta$. ☐

Step 4. We now define addition on $\mathbb{R}$. For $\alpha, \beta \in \mathbb{R}$ we define

$$\alpha + \beta := \{a + b : a \in \alpha, b \in \beta\}.$$

First of all we need to verify that $\alpha + \beta \in \mathbb{R}$. Propery (I) is true since neither α or β are $\emptyset$ or $\mathbb{Q}$ (for the latter one, $\alpha + \beta \neq \mathbb{Q}$ because there are $M, N \in Q$ such that they bound α and β and therefore $M + N$ bounds their sum). Axioms (II) and (III) are left as an exercise. We now need to check the axioms for addition in a field:

(A1) $\alpha + \beta \in \mathbb{R}$
(A1) $\alpha + \beta = \beta + \alpha$
(A1) addition is associative
(A1) define $0^* \in \mathbb{R}$ to be the set of negative rational numbers. Then $0^* + \alpha = \alpha$
(A1) for all $\alpha \in \mathbb{R}$ we want to define an additive inverse by

$$\beta = \{p \in \mathbb{Q} : \text{ there exists } r > 0 \text{ such that } -p - r \notin \alpha\}$$

such that $\alpha + \beta = 0$ (additive inverse)

Step 5. We need to check that addition is compatible with the ordering, namely if $\beta < \gamma$ then $\alpha + \beta < \alpha + \gamma$ for all $\alpha \in \mathbb{R}$.

Step 6. We now define multiplication. Given $\alpha, \beta > 0^*$ we define

$$\alpha\beta := \{p \in \mathbb{Q} : p < ab \text{ for some } a \in \alpha, b \in \beta \text{ with } a, b > 0\}.$$

Again, we need to check the axioms for multiplication (same as addition, after removing 0^*). Moreover, we need to check the distributivity axiom.

Step 6.5. We extend multiplication to the whole of $\mathbb{R}$ by defining

$$\alpha\beta := \begin{cases} -(-\alpha)\beta & \text{if } \alpha < 0^*, \beta > 0^* \\ & \vdots \end{cases}$$

Thus we have shown that $(\mathbb{R}, <)$ is an ordered field with the least upper bound property.

Step 7. We now need to prove that $\mathbb{Q} \subset \mathbb{R}$ as an ordered subfield. To do this, for all $r \in Q$ we define

$$r^* = \{p \in \mathbb{Q} : p < r\}$$

and verify that the induced order is the order on $\mathbb{Q}$. We also need to check that all the operations coincide with the usual notions on $\mathbb{Q}$. $\square$

16. March 2

16.1. **Robots & Topology.** Consider four linked rigid rod, and consider how the movement of the rod is constrained. Specifically, we might think about what the space

$$X = \{\text{ possible positions of the joints }\}$$

looks like. We face problems like these in many occasions, for example when we need to model the configuration of a robot. To get an understanding of how to think about X and what X looks like, we need to answer questions like:

- what is "dimension" (degrees of freedom)?
- How does X depend on lengths?
- Is X connected?

A tool to answer these questions is manifolds.

16.2. **Manifolds in $\mathbb{R}^n$.** An informal definition of a manifold is the following:
Informal definition. A **manifold** is a subset $M \subset \mathbb{R}^n$ that is locally the graph of a function.

Example 16.1. Consider

$$M = \left\{(x,y) : x^2 + y^2 = 1\right\} \subset \mathbb{R}^2.$$

Any point in M lies on one of the following graphs:

$$y = f_{\pm}(x) = \sqrt{1-x^2}$$
$$x = g_{\pm}(x) = \sqrt{1-y^2}.$$

Definition 16.2. A 1-**dimensional manifold** in $\mathbb{R}^2$ is $M \subset \mathbb{R}^2$ such that for all $(a,b) \in M$ there exists an open rectangle $Q = (a-\delta, a+\delta) \times (b-\delta, b+\delta)$ and C^1 functions either

$$h : (a-\delta, a+\delta) \to \mathbb{R}$$
$$\text{or } h : (b-\delta, b+\delta) \to \mathbb{R}$$

so that $M \cap Q = \text{graph}(h) = \{(x, h(y))\}$ or $\{(y, h(y))\}$.

We can have a more precise definition.

Definition 16.3. An $(n-1)$-**dimesnional manifold** in $\mathbb{R}^n$ is a set $M \subset \mathbb{R}^n$ so that for all $c \in M$ there exists an open rectangle

$$Q = \prod (c_i - \delta, c_i + \delta)$$

and an index j and C^1 function

$$h : \prod_{i \neq j} (c_i - \delta, c_i + \delta) \to \mathbb{R}$$

so that $Q \cap M$ is equal to the graph of h.

Definition 16.4. Let $M \subset \mathbb{R}^n$ be an $(n-1)$-dimensional manifold. The **tangent space** at $c \in M$ is defined as

$$T_cM = \left\{\gamma'(0) \middle| C^1 \ni \gamma : \mathbb{R} \to M, \gamma(0) = c\right\}.$$

Remark. This is a subspace of $\mathbb{R}^n$, although we won't prove this now.

Example 16.5. Consider the set

$$M = \left\{(x,y,z) : x^2 - y^2 - z = 0\right\}.$$

Is this a manifold? We can define

$$\begin{aligned} h : \mathbb{R}^2 &\to \mathbb{R} \\ (x,y) &\mapsto x^2 - y^2 \end{aligned}$$

and so M equals the graph of h. This surface looks like a saddle centered at the origin, and we can ask what the tangent space is. Consider

$$\begin{aligned} H : \mathbb{R}^2 &\to \mathbb{R}^3 \\ (x,y) &\mapsto (x,y,h(x,y)). \end{aligned}$$

Then for $c \in \mathbb{R}^2$

$$T_{c,h(c)} = \operatorname{Im} DH(c).$$

In particular,

$$DH(c) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ D_1 h(c) & D_2 h(c) \end{pmatrix}.$$

Example 16.6. Let's go back to the four linked rods. Is X a manifold? We start by fixing the length of the rods as being $\ell_1, \ldots, \ell_4 > 0$. Then we can define

$$X = \left\{(u_1, \ldots, u_4) \in \left(\mathbb{R}^2\right)^4 : |u_i - u_{i+1}| = \ell_i\right\} \subset \mathbb{R}^8.$$

If we fix u_1, u_3, how many positions can u_2 and u_4 take? The answer is 4. In fact, each of the two points can occupy opposite positions on the plane. In general, X is given by a system of equations, and there are 8 variables and 4 equations. For example, we can set $u_i = (x_i, y_i)$ so that our equations have the form

$$(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 = \ell_i^2$$

(with the convention that i is modulo 4). In particular, we can expect that locally we can freely move u_1 and u_3 and this will fix (locally) u_2 and u_4. Therefore we obtain that X is the graph of a function $\mathbb{R}^4 \to \mathbb{R}^4$, so that X is a 4-manifold. However, there are exceptions. For example, we can pull u_1 and u_3 as far apart as we can so that u_2 is in the same line. In this case, u_2 is fixed. This is a problem with our graph, but in this case we just define the graph in terms of u_2 and u_4.

16.3. **Manifold recognition.** When does a system of equations determine a manifold?

Theorem 16.7. *Let $g : \mathbb{R}^n \to \mathbb{R} \in C^1$ and*

$$Z = \{z \in \mathbb{R}^n : g(z) = 0\}.$$

If $Dg(z) \neq 0$ for all $z \in Z$, then Z is an $(n-1)$-manifold in $\mathbb{R}^n$.

Proof. Fix $c \in Z$. We know that

$$\begin{aligned} 0 \neq & Dg(c) \\ = & (D_1 g(c), \ldots, D_n g(n)) \end{aligned}$$

and therefore $D_i g(c) \neq 0$ for some i. Without loss of generality let $i = n$. Choose coordinates $(x, y) \in \mathbb{R}^{n-1} \times \mathbb{R}$. We will show that there exists open $A \times B$ around $c = (a, b)$ and $h : A \to \mathbb{R}$ such that $g(x, h(x)) = 0$ for all $x \in A$. Consider

$$\begin{aligned} G : \mathbb{R}^n \to & \mathbb{R}^n \\ (x, y) \mapsto & (x, g(x, y)). \end{aligned}$$

Then

$$DG(c) = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ D_1 g(c) & D_2 g(c) & \cdots & D_{n-1} g(c) & D_n g(c) \end{pmatrix}.$$

If $\det G(c) \neq 0$ then $DG(c)$ is invertible, and therefore there exists $U = A \times B$ around $c = (a, b)$ and $K : G(u) \to U$ such that $KG = GK = \mathrm{id}$. (we are short on times, but check Munkres for details). $\square$

17.1. **Manifolds & tangent spaces.** Last time we saw that a k-dimensional manifold in $\mathbb{R}^n$ is a subset $M \subset \mathbb{R}^n$ that is locally the graph of $\mathbb{R}^k \to \mathbb{R}^n$. Moreover, for $a \in M$, we defined the tangent space

$$T_a M = \left\{ \gamma'(0) : \gamma : C^1 \ni \mathbb{R} \to M, \gamma(0) = a \right\}.$$

Manifolds are some of the best behaved subsets of Euclidean space, and one of the reason why they are so interesting is that they arise as level sets of functions $g : \mathbb{R}^n \to \mathbb{R}$. Another important application is to Lagrange multipliers.

Theorem 17.1. *Let $f, g : \mathbb{R}^n \to \mathbb{R}$. If c is a maximum of f on $z = \{g = 0\}$, then $Df(c) = \gamma Dg(c)$ for some $\lambda \in \mathbb{R}$. This λ is called **Lagrange multiplier**.*

Proof. We will postpone the proof; it will be based on tangen spaces. □

Lemma 17.1.1 (Tangent space of a graph)**.** *Let*

$$M = \operatorname{graph}(h) = \left\{ (z, h(z)) : z \in \mathbb{R}^k \right\},$$

with $h : \mathbb{R}^k \to \mathbb{R}^{n-k}$ ($M \subset \mathbb{R}^n$ is a k-manifold). Let $c = (a, b) \in M$. Then for

$$\begin{aligned} H : \mathbb{R}^k &\to \mathbb{R}^n \\ z &\mapsto (z, h(z)) \end{aligned}$$

we have

$$T_c M = \operatorname{Im}\left[DH(a) : \mathbb{R}^k \to \mathbb{R}^n \right].$$

Corollary 17.1.1. *The tangent space $T_c M \subset \mathbb{R}^n$ is a linear subspace of dim k.*

Proof. We just write

$$DH(c) = \left(\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \hline & & & \\ & & Dh(c) & \end{array} \right)$$

and see the first k vectors in the image are linearly independent. □

Proof of Lemma 17.1.1. Consider the projection

$$\begin{aligned} \pi : \mathbb{R}^k \times \mathbb{R}^{n-k} &\to \mathbb{R}^k \\ (z, w) &\mapsto z. \end{aligned}$$

Note that $H \circ \pi = operatorname{id}_M$ on M. We need to show that $\operatorname{Im} DH(c) \subset T_c M$. Given $v \in \mathbb{R}^k$ define

$$\begin{aligned} \gamma : \mathbb{R} &\to \mathbb{R}^k \to^H M \\ t &\mapsto a + tv \mapsto H(a + tv). \end{aligned}$$

By the chain rule,

$$\gamma(0) = D\left(H(a+tv)\right)(0) = DH(a)v.$$

We now need to show $T_cM \subset \operatorname{Im} DH(a)$. Fix $\gamma : \mathbb{R} \to M$ with $\gamma(0) = a$. We need v such that $\gamma'(0) = DH(a)v$. Since $\gamma = H \circ \pi \circ \gamma$ on M, we have by the chain rule

$$\begin{aligned}\gamma'(0) =& DH\left(\pi(\gamma(0))\right) \cdot D(\pi\gamma(0)) \\ =& DH(a) \cdot D\left(\pi \cdot \gamma\right)(0)\end{aligned}$$

so if we set $v = D\left(\pi \cdot \gamma\right)(0)$ we are done. □

Example 17.2. Let

$$\begin{aligned}M =& \left\{(x,y,z) : z = x^2 - y^2\right\} \subset \mathbb{R}^3 \\ =& \operatorname{graph}\left(h(x,y) = x^2 - y^2\right).\end{aligned}$$

Then

$$DH(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2x & -2y \end{pmatrix}.$$

Then for $c = (x, y, x^2 - y^2) \in M$ we get

$$T_cM = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 0 \\ 2x \end{pmatrix}, \begin{pmatrix} 0 & 1 & -2y \end{pmatrix}\right\}.$$

Example 17.3. For $M = \{(x,y,z) : x = 0\}$ we have that

$$T_{(0,1,1)}M = \{(x,y,z) : x = 0\}.$$

17.2. **Manifold recognition.**

Theorem 17.4 (Implicit function theorem)**.** *Let $g : \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}^m$ be C^1. Suppose $g(a,b) = 0$ and $\frac{\partial g}{\partial y}(a,b)$ is invertible. Then there eixists a neighborhood U of a and an unique, C^1 function $h : U \to \mathbb{R}^m$ so that $g(x, h(x)) = 0$ for all $x \in U$. Here*

$$Dg(a,b) = \left(\underbrace{\frac{\partial g}{\partial x}(a,b)}_{m\times k} \middle| \underbrace{\frac{\partial g}{\partial y}(a,b)}_{m\times m}\right).$$

Example 17.5. Let $g(x,y) = x^2 + y^2 - 5$. Then $Dg(x,y) = \begin{pmatrix} 2x & 2y \end{pmatrix}$. Also, $g(1,2) = 0$, and $Dg(1,2) = (2,4)$, and $4 \neq 0$. However, even if $g(\sqrt{5}, 0) = 0$ the theorem says nothing about a neighborhood of $(\sqrt{5}, 0)$.

Example 17.6. Let $g(x,y) = x^2 - y^3$. Since $Dg(0) = 0$ the implicit function theorem does not say anything, but at 0 we have $y = x^{2/3}$ which is not C^1.

Example 17.7. Let $g(x,y) = y^2 - x^4$. Again $Dg(0,0) = 0$, so we cannot apply the theorem. In particular, at $(0,0)$ this is not a manifold, because we can solve for y but not uniquely.

Lemma 17.7.1 (Tangent space of a level set)**.** *Let $g : \mathbb{R}^n \to \mathbb{R}$ be C^1, and assume that $Z = \{z \in \mathbb{R}^n\} : g(z) = 0$ is a manifold. Then*

$$T_a Z = \ker Dg(a).$$

Example 17.8. Consider $Z = \left\{(x, y, z) : z - x^2 + y^2 = 0\right\}$. Then

$$Dg(x, y, z) = \begin{pmatrix} -2x & 2y & 1 \end{pmatrix}$$

and at $a = (1, 2, -3)$ we have that

$$\begin{aligned} T_a Z &= \ker Dg(1, 2, -3) \\ &= \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} : \begin{pmatrix} -2 & 4 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \right\} \\ &= \left\{ (x, y, z) \in \mathbb{R}^3 : -2x + 4y + z = 0 \right\} \end{aligned}$$

18.1. **Tangent spaces.**

Lemma 18.0.1 (Tangent space of a level set)**.** *Let $g : \mathbb{R}^n \to \mathbb{R}$ be C^1 and let*

$$a \in Z = \{x \in \mathbb{R}^n : g(x) = 0\}, \quad Dg(a) = 0.$$

Then

(i) $T_a Z = \ker Dg(a)$
(ii) $Dg(a)^T$ (viewed as a column vector in $\mathbb{R}^n$) is orthogonal to $T_a Z$
(iii) $Dg(a)^T$ points in the direction where g increases fastest.

Example 18.1. Let

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

so that $g(x) = x^2 + y^2 - 1$. Then

$$\begin{aligned} Dg(x, y) &= \begin{pmatrix} 2x & 2y \end{pmatrix} \\ T_{(x,y)} Z &= \{(u, v) : xu + yv = 0\}. \end{aligned}$$

One can see this graphically by drawing the line tangent to the unit circle at each point.

Proof of lemma 18.0.1. Proof of (i). We first show that $T_a Z \subset \ker Dg(a)$. By definition,

$$T_a Z = \{\gamma'(0) : \gamma : \mathbb{R} \to Z, \gamma(0) = a\}.$$

Given γ with $\gamma'(0) \in T_a Z$ we have that $g \circ \gamma = 0$ and so by the chain rule

$$0 = (g \circ \gamma)'(0) = Dg(a) \cdot \gamma'(0),$$

namely $\gamma'(0) \in \ker Dg(a)$. The dimensions of both spaces is $n - 1$ and therefore $T_a Z = \ker Dg(a)$.

Proof of (ii). We rewrite

$$0 = Dg(a)\gamma'(0) = Dg(a)^T \cdot \gamma'(0)$$

where the last operation is the dot product.

Proof of (iii). We are looking for a unit vector u so that $D_u g(a)$ is maximized. We see that

$$\begin{aligned} D_u g(a) =& Dg(a)u \\ =& Dg(a)^T \cdot u \\ =& |Dg(a)^T| \cdot |u| \cos\theta \\ =& |Dg(a)^T| \cos\theta. \end{aligned}$$

Therefore this value is largest when $\theta = k\pi$, i.e. parallel to $Dg(a)^T$. □

Addendum. If $g = (g_1, \ldots, g_\ell) : \mathbb{R}^n \to \mathbb{R}^\ell$ then for all

$$a \in Z = \{g = 0\} = \bigcap_{i=1}^{\ell} \{g_i = 0\}$$

we have

$$T_a Z = \bigcap_{i=1}^{\ell} \ker Dg_i(a)$$

and therefore

$$Dg_i(a)^T \in (T_a Z)^{\perp}$$

for all $1 \le i \le \ell$.

18.2. **Lagrange multipliers.** We have the following problem: we are given two squares, S_1 and S_2, with $\text{area}(S_1) + \text{area}(S_2) = 1$. We want to find the smallest A such that any two such squares fit into a rectangle with area A. Let x be the side of S_1 and y be the side of S_2. Without loss of generality we can assume $x \ge y$. Then the squares fit side by side in a rectangle of area $x(x+y)$, and therefore by maximizing this area with the constraint $x^2 + y^2 = 1$ we are going to solve the problem. The general form of this problem is as follows:

General problem (maximum under constraint). Let $M \subset \mathbb{R}^n$ be a k-manifold, and let $\phi : \mathbb{R}^n \to \mathbb{R} \in C^1$. Find the maximum/minimum of $\phi\big|_M$.

Remark. A max/min need not exist, but will always exist if M is compact. Moreover, if $a \in \mathbb{R}^n$ is a local maximum of ϕ, then $D\phi(a) = 0$, but with constraint $a \in M$ this need not be true. For example, if $\phi(x,y) = x$ and $M = S^1$ (the unit circle), then ϕ has a maximum on S_1 at $(1,0)$ but the derivative is never 0.

Theorem 18.2 (Lagrange multiplier)**.** *Let $g_1, \dots, g_\ell : \mathbb{R}^n \to \mathbb{R}$ of class C^1. Assume that*

$$M = \{x \in \mathbb{R}^n : g_1(x) = \cdots = g_\ell(x) = 0\}$$

*is a manifold, and let $\phi : \mathbb{R}^n \to \mathbb{R}$ be C^1. If $a \in M$ is a max/min, then $D\phi(a) = \lambda_1 Dg_1(a) + \cdots + \lambda_\ell Dg_\ell(a)$ for some $\lambda_1, \dots, \lambda_\ell \in \mathbb{R}$. The λ_i's are called **Lagrange multipliers**.*

Remark. The special case $\ell = 1$ was stated in Theorem 17.1.

Proof. Let

$$a \in M = \bigcap_{i=1}^{\ell} \{g_i = 0\} \subset \mathbb{R}^n$$

be a max/min of $\phi : \mathbb{R}^n \to \mathbb{R}$ (restricted to M). We want to show that $D\phi(a)^T$ is orthogonal to $T_a M$. Fix $\gamma'(0) \in T_a M$. Then if a is a max of ϕ it follows that 0 is a max of $\phi \circ \gamma$. Therefore, by using facts about derivatives of functions $\mathbb{R} \to \mathbb{R}$ we have

$$\begin{aligned} 0 &= (\phi \circ \gamma)'(0) \\ &= D\phi(a) \cdot \gamma'(0). \end{aligned}$$

□

19.1. **Manifold recognition.** We are interested in recognizing whether a certain space is a manifold.

Example 19.1. For which $a, b \in \mathbb{R}$ is

$$Z_{a,b} = \left\{(x,y,z) : x^2 + y^3 + z = a, x + y + z = b\right\}$$

a 1-manifold in $\mathbb{R}^3$? Note that $Z_{a,b}$ is the intersection of the graphs of $f(x,y) = a - x^2 - y^3$ and $g(x,z) = b - x - z$ respectively, both of which are 2-manifolds.

Consider

$$\begin{aligned} h : \mathbb{R}^3 &\to \mathbb{R}^2 \\ (x,y,z) &\mapsto \left(x^2 + y^3 + z, x + y + z\right). \end{aligned}$$

Then

$$Dh(x,y,z) = \begin{pmatrix} 2x & 3y^2 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

By the implicit function theorem, Z is a manifold near (x,y,z) if $Dh(x,y,z)$ has an invertible submatrix among

$$\begin{pmatrix} 2x & 3y^2 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 2x & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 3y^2 & 1 \\ 1 & 1 \end{pmatrix}.$$

If none of these is invertible we get by taking determinants that

$$2x - 3y^2 = 0, \quad 2x - 1 = 0, \quad 3y^2 - 1 = 0.$$

This means that

$$x = \frac{1}{2} \qquad y = \pm\frac{1}{\sqrt{3}}.$$

Then it follows that

$$b - a = \frac{1}{4} \pm \frac{2}{3\sqrt{3}}.$$

It follows that if

$$b - a \neq \frac{1}{4} \pm \frac{2}{3\sqrt{3}}$$

then $Z_{a,b}$ is a manifold.

Remark. This procedure works in general by considering

$$f : \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}^m$$

given by the intersection and asking when $(\partial f / \partial y)$ is invertible.

19.2. **Lagrange multipliers.** Recall the following one dimensional version of the Lagrange multiplier theorem:

Theorem 19.2. *Let*

$$M = \{x \in \mathbb{R}^n : g(x) = 0\}$$

be a manifold, and let $\phi : \mathbb{R}^n \to \mathbb{R}$ is C^1. If $a \in M$ is a maximum of $\phi\big|_M$ then $D\phi(a) = \lambda Dg(a)$ for some $\lambda \in \mathbb{R}$ (the Lagrange multiplier).

Example 19.3. Last time we saw the example of the rectangle. In this case

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

and $\phi(x, y) = x^2 + xy$. Then

$$\begin{aligned} D\phi(x,y) &= (2x + y, x) \\ Dg(x,y) &= (2x, 2y). \end{aligned}$$

Therefore the Lagrange multiplier equations in the case that (x, y) is a maximum of $\phi\big|_M$ are

$$\begin{aligned} 2x + y &= 2\lambda x \\ x &= 2\lambda y \\ x^2 + y^2 &= 1 \end{aligned}$$

whose solution is

$$\begin{aligned} \lambda &= \frac{1 \pm \sqrt{2}}{2} \\ y_{\pm} &= \pm \frac{1}{\sqrt{4 + 2\sqrt{2}}} \\ x_{\pm} &= \pm \frac{1 + \sqrt{2}}{\sqrt{4 + 2\sqrt{2}}}. \end{aligned}$$

Therefore we have maxima at

$$\begin{aligned} \phi(x_+, y_+) &= \frac{4 + 3\sqrt{2}}{4 + 2\sqrt{2}} \approx 1.2 \\ \phi(x_-, y_-) &= \frac{4 - 3\sqrt{2}}{4 - 2\sqrt{2}} \approx -0.2 \end{aligned}$$

and we are only concerned with the first one.

Example 19.4. Given $f : \mathbb{R}^n \to \mathbb{R}$ we know that $Df(a)$ points in the direction where f incresases the fastest. Fix $a \in \mathbb{R}^n$. To verify this, we want to maximize

$$\phi(x) = D_x f(a) = Df(a) \cdot x$$

with constraint

$$g(x) = x_1^2 + \cdots + x_n^2 - 1 = 0.$$

Then the lagrange equation yields

$$Df(a) = D\phi(x) = \lambda Dg(x) = 2\lambda x^T$$

which means that x is parallel to $Df(a)^T$.

Example 19.5. There is an application to nuclear meltdowns. Suppose that the nuclear reactor in Seabrook, NH. If you are given the flux (level sets) of the nuclear meltdown. If you haven't taken 25b you might be tempted to consider the straight line from Seabrook to Boston and run away from Seabrook in this direction. However, the best choice is to run perpendicular to level sets.

19.3. **Spectral theorem.** (a.k.a. hooray manifolds, hooray derivatives)

Recall the spectral theorem for self-adjoint operators:

Theorem 19.6. *If $A \in M_n\mathbb{R}$ is symmetric then A has an eigenvector $v \in \mathbb{R}^n$ such that $Av = \lambda v$ for $\lambda \in \mathbb{R}$ (in fact this implies that A has an orthonormal basis of eigenvalues, which you find by induction).*

The idea is to consider the function

$$x \mapsto \langle x, Ax \rangle = |x||Ax| \cos\theta.$$

If we restrict our attention to the case $|x| = 1$, in which case $x \mapsto |Ax|\cos\theta$. So we might hope that $\langle x, Ax\rangle$ is maximized when $\theta = 0$ so that then $Ax = \lambda x$ for some $\lambda \in \mathbb{R}$. This however is not true in general. In fact, we are not using the constraint that A is symmetric. Therefore we are going to keep this in mind in the proof.

Proof. We want to maximize

$$\phi(x) = \langle x, Ax \rangle = x^T Ax$$

with constraint

$$g(x) = x_1^2 + \cdots + x_n^2 - 1 = 0.$$

If x is a maximum then

$$D\phi(x) = \lambda Dg(x) = 2\lambda x^T.$$

We need to compute $D\phi(x)$. Note that given $\psi(x,y) = x^T Ay$ and $\Delta(x) = (x,x)$ we have $\phi = \psi \circ \Delta$. Since ψ is bilinear we have (HW 4, problem 8) that

$$D\psi(x,y)(u,v) = \psi(x,v) + \psi(u,y) = x^T Av + u^T Ay.$$

Moreover Δ is linear and therefore $d\Delta(x) = \Delta$. By the chain rule,

$$\begin{aligned} D\phi(u) &= D\psi(x,x) \cdot \Delta(u) \\ &= D\psi(x,x)(u,u) \\ &= x^T Au + u^T Ax \\ &= 2x^T Au \end{aligned}$$

where we used symmetry in the last line. Therefore

$$2x^T Au = 2\lambda x^T u$$

and thus $Ax = \lambda x$. $\square$

20. March 19

Today we are going to start to discuss integrals.

20.1. **Computing area.**

Suppose we are interested in computing the area of the circle. You probably know that the area of a circle of radius r is equal to πr^2. But why? An informal argument might be the following. Suppose we inscribe an equilateral n-gon inside our circle. Then, with q the perimeter of the n-gon and h the height of the triangular sections of the circle, we have that the area A_n of the n-gon is equal to

$$A_n = n\left(\frac{hq}{2n}\right) = \frac{hq}{2}.$$

Then we can take $h \to r$ and $q = c = 2\pi r$ (where c is the circumference), and therefore

$$A = \frac{rc}{2} = \pi r^2.$$

Note that this is valid because $\pi = c/2r$ is constant.

Theorem 20.1 (Archimedes)**.** $A = \frac{cr}{2}$.

Proof. We will prove $A \leq G$ and $A \geq G$. By contradiction, suppose $A > G$, and let $E = A - G$. Choose $n \gg 0$ such that $A - A_n < E$. Then

$$\frac{qh}{2} = A_n > A - E = G = \frac{cr}{2}$$

which is a contradiction since $h < r$ and $q < c$. The reverse direction is proved analogously by using circumscribed polygons. $\square$

This method is called the "method of exhaustion". Our goal will be to compute areas, volumes, etc. more generally. In particular, we will focus on maps of the form $f : [a,b] \to \mathbb{R}$, or, more generally, of the form $f : Q \to \mathbb{R}$ where Q is a closed rectangle in $\mathbb{R}^n$.

20.2. **Defining the Riemann integral.**

Remark. In the previous example we used a notion of area. But how is area (or volume) defined?

Following Archimedes, we can start by first defining the volume of a closed rectangle as

$$V([a_1,b_1] \times \cdots \times [a_n,b_n]) = (b_1 - a_1)\cdots(b_n - a_n).$$

and then define the volume of more complicated regions by approximating with circumscribed and inscribed polygons.

Consider now a function $f : [a,b] \to \mathbb{R}$, and consider a finite partition $P \subset [a,b]$. Thus P decomposes $[a,b]$ into sub-intervals. On an interval R we define

$$\min_R f = \inf\{f(x) : x \in R\}$$
$$\max_R f = \sup\{f(x) : x \in R\}.$$

We can then define the **upper sum** as

$$U(f,P) = \sum_R \max_R(f) \cdot v(R)$$

and the **lower sum** as

$$L(f,P) = \sum_R \min_R(f) \cdot v(R).$$

Note that $L(f,P) \leq U(f,P)$.
Exercise: If $P' \supset P$, then how are $U(f,P')$ and $U(f,P)$ related? What about lower sums?
The answer to this is that $U(f,P') \leq U(f,P)$ and $L(f,P') \geq L(f,P)$.

Definition 20.2. If $P \subset P'$ we call P' a **refinement** of P.

Definition 20.3. We define the **lower integral** (of $f : [a,b] \to \mathbb{R}$) as

$$\underline{\int}_{[a,b]} f = \sup\{L(f,P) : P \text{ is a partition}\}$$

and similarly we define the **upper integral** as

$$\overline{\int}_{[a,b]} f = \inf\{U(f,P)\}.$$

We say that f is **integrable** if

$$\underline{\int}_{[a,b]} f = \overline{\int}_{[a,b]} f =: \int_{[a,b]} f.$$

Example 20.4. Consider $f : [0,1] \to \mathbb{R}$ defined as

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

Then $U(f,P) = 1$ and $L(f,P) = 0$ for all P and thus f is not integrable.

There is an issue with our definition. For a set X, in order for $\inf X$ or $\sup X$ we need to know that X is bounded below/above. In particular, to define the upper and lower sums we need f to be *bounded*.

Example 20.5. Let $f : [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1/x & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

Then $\max_{0,r}$ is undefined, and so the upper integral is also undefined.

Lemma 20.5.1. *Let $f : [a,b] \to \mathbb{R}$ be bounded. For any two partitions P, P' we have that*

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

In particular $\{L(f,P)\}_P$ is bounded above (by any $U(f,P')$).

Proof. The partition $P'' = P \cap P'$ refines both P and P'. Therefore

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

$\square$

Let's summarize what we have seen so far: for $f : [a,b] \to \mathbb{R}$ bounded we define

$$\underline{\int}_{[a,b]} f = \sup\{L(f,P)\} \leq \inf\{U(f,P)\} = \overline{\int}_{[a,b]} f.$$

Moreover, f is integrable if

$$\underline{\int}_{[a,b]} f = \overline{\int}_{[a,b]} f = \int_{[a,b]} f.$$

Example 20.6. Let $f : [a,b] \to \mathbb{R}$ be constant with value c. Then

$$U(f,P) = L(f,P) = c(b-a)$$

for any P and thus

$$\int_{[a,b]} f = c(b-a).$$

Example 20.7. Let $f : [0,1] \to \mathbb{R}$ be given by $f(x) = x$. Let $P = \{0, 1/n, 2/n, \ldots, 1-1/n, 1\}$. Then

$$\begin{aligned} U(f,P) &= \sum_{k=1}^{n} f\left(\frac{k}{n}\right) \cdot \frac{1}{n} \\ &= \frac{1}{n^2} \sum_{k=1}^{n} k \\ &= \frac{1}{n^2}\left(\frac{n(n+1)}{2}\right) \\ &= \frac{1}{2}\left(1 + \frac{1}{n}\right). \end{aligned}$$

Similarly,

$$L(f,P) = \frac{1}{2}\left(1 - \frac{1}{n}\right).$$

Therefore

$$\int_{[0,1]} f = \frac{1}{2}.$$

Next we will see how bad f can f and still be integrable.

21. March 21

21.1. **Which functions are integrable?** Last time we defined

$$\int_{[a,b]} f.$$

Similarly, we can define

$$\int_Q f$$

for a closed rectangle $Q \subset \mathbb{R}^n$ and $f : Q \to \mathbb{R}$ bounded. Given a partition P of Q consisting of subrectangles we can define upper and lower sums analogously to the 1-dimensional case:

$$U(f,P) = \sum_R \max_R(f) \cdot v(R) \qquad L(f,P) = \sum_R \min_R(f) \cdot v(R),$$

and the upper and lower integrals as

$$\underline{\int}_Q f = \sup\{L(f,P)\} \leq \overline{\int}_Q f = \inf\{U(f,P)\}.$$

We say that f is integrable if

$$\underline{\int}_Q f = \overline{\int}_Q f =: \int_Q f.$$

Lemma 21.0.1. *The integral $\int_Q f$ exists if and only if for all $\varepsilon > 0$ there exists a partition P so that $U(f,P) - L(f,P) < \varepsilon$.*

Proof. "Only if" direction. Given $\varepsilon > 0$ there exist P, P' such that

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

(this follows from the definition of supremum and infimum). Let now $P'' = P \cup P'$ be the common refinement of these two partitions. Then

$$U(f,P'') - L(f,P'') < \varepsilon.$$

"If" direction. Suppose $\int_Q f$ does not exist. Then there exists some $\varepsilon > 0$ such that

$$\overline{\int}_Q f - \underline{\int}_Q = 2\varepsilon.$$

Therefore

$$U(f,P) - L(f,P) > \varepsilon$$

for all P. $\square$

Which bounded functions are integrable? Lat time we saw that the map

$$f(x) = \begin{cases} 1/x & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Is not integrable. We will also see that f is continuous only if it is integrable. The converse is false, as can be seen by considering a function with one step. Is Thomae's function integrable? Recall that this is the function defined by

$$f(x) = \begin{cases} \frac{1}{q} & x = \frac{p}{q} \\ 0 & x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

To asnwer this question, we need to introduce the notion of measure.

21.2. Measure and content.

Definition 21.1. We say that a subset $A \subset \mathbb{R}^n$ has **measure** 0 if there is a covering of A by countably many closed rectangles $Q_1, Q_2, \dots$ such that

$$\sum_i v(Q_i) < \varepsilon.$$

If we replace "countable" by "finite" then we say that A has **content** 0.

21.3. Integrability criterion.

Theorem 21.2. *Let $Q \subset \mathbb{R}^n$ be a closed rectangle and $f : Q \to \mathbb{R}$ be bounded. Define*

$$B := \{x \in Q : f \text{ is not continuous at } x\}.$$

Then $\int_Q$ exists if and only if B has measure 0.

Example 21.3. This theorem tells us that Thomae's function is integrable, because in this case $B = \mathbb{Q} \cap [0,1]$ and is countable.

Corollary 21.3.1. *If $f : Q \to \mathbb{R}$ is integrable and $E = \{x \in Q : f(x) \neq 0\}$ has measure 0 then*

$$\int_Q f = 0.$$

Remark. The above implies that Thomae's function has integral 0.

Application. We can apply this to the area of complicated shapes.

Definition 21.4. For $C \subset \mathbb{R}^n$ we define the **characteristic function**

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

If $C \subset Q$ for some Q define

$$\int_C f := \int_Q \chi_C f.$$

For example, if $f = 1$ this defines the area of C. Another corollary of theorem 21.2 is the following:

Corollary 21.4.1. *If* $\chi_C : Q \to \mathbb{R}$ *is integrable then* $\mathrm{bd}(C)$ *has measure* 0*. If* $\mathrm{bd}(C)$ *has measure* 0*, we say that* C *is* ***rectifiable****.*

Proof of Theorem 21.2. We will show that if B has measure 0 then $\int_Q f$ exists. We want to use the criterion we showed at the beginning of class, namely for all $\varepsilon > 0$ there exists a partition P such that

$$U(f,P) - L(f,P) < \varepsilon.$$

Since B has measure 0 we can choose $R_1, R_2, \dots$ such that they cover B and

$$\sum_i v(R_i) < \varepsilon'$$

(we will choose the precise value of ε') later. Also for $a \notin B$ there exists $R_a \ni a$ such that for all $x \in R_a$ we have

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

by continuity. All these rectangles cover Q and therefore we can select a finite subcover $R_1, \dots, R_k, R'_1, \dots, R'_\ell$ of Q. We use these finitely many rectangles to get a partition P. For each subrectangle R of P, either $R \subset R_i$ or $R \subset R'_j$. Then

$$U(f,P) - L(f,P) = \sum_{R \subset R_i} \left(\max_R - \min_R\right) v(R) + \sum_{R \subset R'_j} \left(\max_R - \min_R\right) v(R).$$

We will finish next time. ∎

22. March 23

22.1. **Integrability criterion.** Last time we stated (and almost prove) the following:

Theorem 22.1. *Let $Q \subset \mathbb{R}^n$ be a closed rectangle, with $f : Q \to \mathbb{R}$ bounded. Then $\int_Q f$ exists if and only if the set*

$$B = \{x \in Q : f \text{ is not continuous at } x\}$$

has measure 0.

Continuation of proof. Fix $\varepsilon > 0$. We want to find a partition P such that

$$\sum_R \left(\max_R f - \min_R f\right) \cdot v(R) = U(f,P) - L(f,P) < \varepsilon.$$

Since f is bounded, there exists $M \geq 0$ such that $|f(x) - f(y)| \leq M$ for all $x, y \in Q$. Fix now $\varepsilon', \varepsilon'' > 0$ (which we will fix later) and use the facts that

- B has measure 0
- f is continuous on $Q \setminus B$
- Q is compact

to find a partition P such that the subrectangles for which $\max_R(f) - \min_R(f) > \varepsilon''$ have total volume $< \varepsilon'$. We already constructed the partition last time, so we will omit the construction here. Then given P we have that

$$U(f,P) - L(f,P) = \sum_{R \text{ bad}} \left(\max_R - \min_R\right) v(R) + \sum_{\text{other } R} \left(\max_R - \min_R\right) v(R)$$

where the first summation is over the rectangles where we can't contain the variation within ε''. Therefore the term in parenthesis in the first summation is bounded by M, and in the second sum it is bounded by ε''.

$$\sum_{R \text{ bad}} \left(\max_R - \min_R\right) v(R) + \sum_{\text{other } R} \left(\max_R - \min_R\right) v(R) \leq M\varepsilon' + \varepsilon'' v(Q).$$

We can now set

$$\varepsilon' = \frac{\varepsilon}{2M} \qquad \varepsilon'' = \frac{\varepsilon}{2v(Q)}$$

so that

$$M\varepsilon' + \varepsilon'' v(Q) < \varepsilon.$$

□

Example 22.2. Is it true or false that the map $f : [0,1] \to \mathbb{R}$ defined by

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

is integrable because $[0,1] \cap \mathbb{Q}$ has measure zero? It is false, because f is nowhere continuous.

22.2. **Integration.**

22.2.1. *Overview.* We will focus on continuous maps, i.e. $f : Q \to \mathbb{R}$ continuous. The following results will be our main goal:

- if $Q \subset \mathbb{R}$, the fundamental theorem of calculus reduces computing $\int_Q f$ to finding antiderivatives;
- if $Q \subset \mathbb{R}^n$, Fubini's theorem reduces computing $\int_Q f$ to computing n 1-dimensional integrals.
- change of variables: we can rewrite an integral to make it easier to compute and find its antiderivatives.

22.2.2. *Fundamental theorem of calculus.* As a warmup, consider the function $f : [0,2] \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 1 & x \in [0,1] \\ 2 & x \in (1,2] \end{cases},$$

and define

$$F(x) = \int_{[0,x]} f.$$

Is F continuous? We can explicitly write

$$F(x) = \begin{cases} x & x \leq 1 \\ 1 + 2(x-1) & x \geq 1 \end{cases},$$

therefore F is continuous. In general we have the following

Slogan: integration makes functions smoother.

Theorem 22.3 (Fundamental Theorem of Calculus (FTC)). *Let $f : [a,b] \to \mathbb{R}$ be a continuous map. Then*

(i) the function

$$F(x) = \int_{[a,x]} f$$

is differentiable on (a,b) and $F'(x) = f(x)$;

(ii) if g is continuous on $[a,b]$ and $g'(x) = f(x)$ on (a,b), then

$$\int_{[a,b]} f = g(b) - g(a).$$

In this case g is called an ***antiderivative*** *of f.*

Remark. We can rewrite the above as

$$D\left(\int_a^x f\right)(x) = f(x) \qquad \int_a^b Dg = g(b) - g(a).$$

Remark. This is remarkable because it greatly simplifies the process of finding derivatives.

Example 22.4. We want to compute

$$\int_0^1 f, \qquad f(x) = x^2 - x^3.$$

For $g(x) = x^3/3 - x^4/4$ we have that $g' = f$ and therefore

$$\int_0^1 f = \frac{1}{3}x^3 - \frac{1}{4}x^4\Big|_0^1 = \frac{1}{3} - \frac{1}{4}$$

Proof of Theorem 22.3. Fix $c \in (a,b)$. We want to show that

$$\lim_{h\to 0} \frac{F(c+h)-F(c)}{h} = f(c).$$

Let now $h > 0$. Then

$$F(c+h) - F(c) = \int_{[a,c+h]} f - \int_{[a,c]} f = \int_{[a,c]} f + \int_{c,c+h} - \int_{[a,c]} f = \int_{[c,c+h]}$$

where the second equality comes from homework 7. We now define

$$m_h = \inf\{f(x) : c \le x \le c+h\}$$
$$M_h = \sup\{f(x) : c \le x \le c+h\}.$$

By homework 7 we find

$$m_h \cdot h \le \int_{[c,c+h]} f \le M_h \cdot h$$

and therefore

$$m_h \le \frac{1}{h}\int_{[c,c+h]} f \le M_h.$$

We can make the same argument for $h < 0$ by swapping c and $c+h$ in the above argument. By hypothesis, f is continuous, and therefore

$$\lim_{h\to 0} m_h = f(c) = \lim_{h\to 0} M_h$$

which implies

$$\lim_{h\to 0} \frac{F(c+h)-F(c)}{h} = f(c).$$

This proves (i). To prove (ii), we now know that

$$g'(x) = f(x) = F'(x).$$

Thus there exists C such that $F = g + C$. Since

$$0 = F(a) = g(a) + C$$

we find that $C = -g(a)$. Therefore for all x we have $F(x) = g(x) - g(a)$. For $x = b$ we get

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

□

Example 22.5. Using the chain rule, we get that defining

$$F(x) = \int_0^{x^3} \sin^3(x)$$

we can compute (exercise)

$$F'(x) = 3x^2 \sin^3(x^3).$$

22.3. **Fubini's theorem.** Having considered the 1-dimensional case, we now move on to higher dimensions.

Problem. We want to compute $\int_Q f$ where $Q = [a,b] \times [c,d] \subset \mathbb{R}^2$ and $f(x,y) = x^2 y$. We can do two different integrals in order, for example by first integrating along the x axis direction and then along the y axis. In this case

$$\begin{aligned}\int_Q f &= \int_{y=c}^{d} \left(\int_{x=a}^{b} x^2 y \right) \\ &= \int_c^d \left(\frac{b^3 - a^3}{3} y \right) \\ &= \frac{\left(b^3 - a^3\right)\left(d^2 - c^2\right)}{6}.\end{aligned}$$

23. March 26

23.1. **Fubini's theorem.** **Goal:** given $Q \subset \mathbb{R}^2$ and a bounded function $f : Q \to \mathbb{R}$ we want to compute $\int_Q f$ with 1-dimensional integrals. For $Q = [a,b] \times [c,d]$ and $y \in [c,d]$ consider $g_y : [a,b] \to \mathbb{R}$ such that $g_y(x) = f(x,y)$ and define

$$I(y) := \int_a^b g_y.$$

Our hope is that

$$\int_Q f = \int_c^d I(y) = \int_{y=c}^d \left(\int_{x=a}^b f(x,y) \right).$$

However, our hope is crushed by the fact that $I(y)$ may not exist for some y. For example, consider the function $f : [0,1] \times [0,1] \to \mathbb{R}$ defined as

$$f(x,y) = \begin{cases} 1 & y = \frac{1}{2}, x \in \mathbb{Q} \\ 0 & \text{else} \end{cases}.$$

Then

$$g_{1/2}(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$$

is not integrable and therefore $I(y)$ does not exist. However, f is discontinuous on a set of measure 0, and so f is integrable. Also, the set $\{f \neq 0\}$ has measure 0 so $\int_Q f = 0$. To fix this problem we need the following theorem:

Theorem 23.1 (Fubini)**.** *Let $Q = A \times B \in \mathbb{R}^k \times \mathbb{R}^m$ be a closed rectangle, and let $f : Q \to \mathbb{R}$ bounded. Define*

$$\underline{I}(y) = \underline{\int}_{x \in A} f(x,y) \qquad \overline{I}(y) = \overline{\int}_{x \in A} f(x,y),$$

which always exist since upper and lower itnegrals always exist. If f is integrable over Q then $\underline{I}(y), \overline{I}(y)$ are integrable and

$$\int_Q f = \int_{y \in B} \overline{I}(y) = \int_{y \in B} \underline{I}(y) = \int_{y \in B} \left(\underline{\int}_{x \in A} f(x,y) \right).$$

Example 23.2. Let f as before. What are $\overline{I}, \underline{I}$ in this case? We have that $\underline{I}(y) = 0$ for all y, and

$$\overline{I}(y) = \begin{cases} 1 & y = \frac{1}{2} \\ 0 & y \neq \frac{1}{2} \end{cases}.$$

This confirms Fubini's theorem since

$$0 = \int_Q f = \int_{y=0}^1 \underline{I}(y) = \int_{y=0}^1 \overline{I}(y).$$

Remark. If f is continuous then g_y is continuous for each y, and therefore

$$\underline{I}(y) = \int_A g = \overline{I}(y)$$

and so

$$\int_Q f = \int_{y\in B}\int_{x\in A} f(x,y)$$

and we can also write

$$\int_Q = \int_{x\in A}\int_{y\in B} f(x,y).$$

23.2. **Proof of Fubini's theorem.** We can start by visualizing the problem. We have a rectangle $Q \subset \mathbb{R}^k \times \mathbb{R}^m$ and we know that f is integrable over Q. This means that we can choose a partition P so that the difference between the upper sum and the lower sum $U(f,P) - L(f,P)$ is small. To show that $\underline{I}, \overline{I}$ are integrable, we want to find partitions P_B of B such that both $U(\underline{I},P_B) - L(\underline{I},P_B)$ and $U(\overline{I},P_B) - L(\overline{I},P_B)$ are small. We know that

$$L(\underline{I},P_B) \le L(\overline{I},P_B), U(\underline{I},P_B) \le U(\overline{I},P_B)$$

as well as the usual $L \le U$ intequalities. We want to show $L(f,P) \le L(\underline{I},P_B)$ and $U(\overline{I},P_B) \le U(f,P)$, so that we can sandwich the above equations to make everything small. This will be done by using the minimum principle, namely that mmin $\le$ average $\le$ max.

Proof of Fubini's theorem. Throughout the proof, R_A and R_B is going to indicate subrectangles of P sitting in A and B respectively. We are going to show

$$L(f,P) \le L(\underline{I},P_B)$$

for every partition $P = (P_A, P_B)$.

By definition,

$$\begin{aligned}
L(\underline{I},P_B) &= \sum_{R_B} \min_{R_B}(\underline{I}) \cdot v(R_B) \\
L(f,P) &= \sum_{R_A\times R_B} \min_{R_A\times R_B}(F) \cdot v(R_A\times R_B) \\
&= \sum_{R_B}\left(\sum_{R_A} \min_{R_A\times R_B}(f)\cdot v(R_A)\right)\cdot v(R_B)
\end{aligned}$$

where we used the fact that $v(R_A\times R_B) = v(R_A)v(R_B)$. We want to show

$$\min_{R_B}(\underline{I}) = \sum_{R_A}\min_{R_A\times R_B}(f)\cdot v(R_A)\sum_{R_A}\min_{R_A\times R_B}(f)\cdot v(R_A).$$

Fix R_B and $y_0 \in R_B$. Fix R_A. For all $x \in R_A$ we have that

$$\min_{R_A\times R_B}(f) \le f(x,y_0).$$

This implies that

$$\begin{aligned}\sum_{R_A} \min_{R_A \times R_B}(f) \cdot v(R_A) \leq& \sum_{R_A} \min_{R_A} f(x, y_0) \cdot v(R_A) \\ =& L(g_{y_0}, P_A) \\ \leq& \underline{\int}_{x \in A} f(x, y_0) \\ =:& \underline{I}(y_0).\end{aligned}$$

This holds for all $y_0 \in R_B$, and therefore

$$\min_{R_B}(\underline{I}) = \sum_{R_A} \min_{R_A \times R_B}(f) \cdot v(R_A) \sum_{R_A} \min_{R_A \times R_B}(f) \cdot v(R_A).$$

This proves

$$L(f, P) \leq L(\underline{I}, P_B)$$

and in a similar way one proves

$$U(\bar{I}, P_B) \leq U(f, P).$$

Therefor we conclude

$$\int_Q f = \sup L(f, P) \leq \sup L(\underline{I}, P_B) \leq \inf U(\bar{I}, P_B) \leq \inf U(f, P) = \int_Q f.$$

□

We can apply Fubini's theorem to the calculation of the volume of the pyramid. Let

$$C = \left\{(x, y, z) \in \mathbb{R}^3 : x, y, z \geq 0, x + y + y = 1\right\}.$$

Then $C \subset [0,1]^3 = Q$ and therefore

$$\operatorname{vol}(C) = \int_C 1 = \int_Q \chi_C$$

where

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

Therefore by Fubini's theorem

$$\begin{aligned}\int_Q \chi_C =& \int_{z=0}^{1} \int_{y=0}^{1-z} \int_{x=0}^{1-z-y} 1 \\ =& \int_{z=0}^{1} \int_{y=0}^{1-z} x \\ =& \int_{z=0}^{1} y - yz - \frac{1}{2} y^2 \\ =& \frac{1}{6}.\end{aligned}$$

24.1. **Change of variables.** The main techniques of integration we've seen so far are the fundamental theorem of calculus and Fubini's theorem. Today we are going to talk about a thrid technique, change of variables.

Definition 24.1. let $A \subset \mathbb{R}^n$ be open. We say $g : A \to \mathbb{R}^n$ is a **change of variables** if g is injective, C^1 and Dg is invertible for all $x \in A$(namely $\det Dg(x) \neq 0$ for all x).

Example 24.2. An example is polar coordinates, defined as a map

$$\begin{aligned} g : \mathbb{R}^2 &\to \mathbb{R}^2 \\ g(r, \theta) &= (r\cos\theta, r\sin\theta). \end{aligned}$$

Then g is injective on $(0, \infty) \times (0, 2\pi)$ and

$$Dg = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and therefore $\det Dg(r, \theta) = r \neq 0$. Thus $g : (0, \infty) \times (0, 2\pi)$ is a change of variables.

Example 24.3. Similarly we have spherical coordinates given by

$$\begin{aligned} g : \mathbb{R}^3 &\to \mathbb{R}^3 \\ g(\rho, \theta, \varphi) &= (\rho\sin\phi\cos\theta, \rho\sin\varphi\sin\theta, \rho\cos\theta). \end{aligned}$$

We see that g is injective on $(0, \infty) \times (0, 2\pi) \times (0, \pi)$ and $\det Dg(\rho, \theta, \varphi) = \rho^2 \sin\varphi \neq 0$ so that $g : (0, \infty) \times (0, 2\pi) \times (0, \pi)$ is a change of variables.

Theorem 24.4 (Change of variables). *Let $g : A \to \mathbb{R}^n$ be a change of variables and $B = g(A)$. If $f : B \to \mathbb{R}$ is continuous, then f is integrable over B if and only if $(f \circ g)|\det Dg|$ is integrable over A. Moreover,*

$$\int_B f = \int_A (f \circ g)|\det Dg|.$$

Remark. Note that we haven't defined integration on arbitrary open sets. In fact, integrals like

$$\int_{(0,\infty)} \frac{1}{x^2}$$

are undefined. We will get back to this later.

Remark. Recall that the determinant measures volume expansion of a linear map. For $f = 1$, $B \subset \mathbb{R}^n$ bounded and $g : A \to B$ we see that

$$\begin{aligned} vol(B) &= \int_B 1 \\ &= \int_A (f \circ g)|\det Dg| \\ &= \int_A |\det Dg| \end{aligned}$$

so that we can interpret $|\det Dg|$ as the infinitesimal volume expansion.

Example 24.5. Let

$$B = \left\{(x,y) : x,y > 0, x^2+y^2 < a^2\right\}.$$

We want integrate the function x^2y^2 over B. By Fubini's theorem,

$$\begin{aligned}\int_B x^2y^2 &= \int_{y=0}^{a}\int_{x=0}^{\sqrt{a^2-y^2}} x^2y^2 \\ &= \int_{y=0}^{a} \frac{1}{3}x^3y^2\bigg|_{x=0}^{x=\sqrt{a^2-y^2}} \\ &= \int_{y=0}^{a} \frac{1}{3}\left(a^2-y^2\right)^{\frac{3}{2}} y^2.\end{aligned}$$

This is hard to do, so it's best to use change of variables. Specifically, using polar coordinates change of variables we get

$$\begin{aligned}\int_B x^2y^2 &= \int_A r^2\cos^2\theta r^2\sin^2\theta \cdot r \\ &= \int_{r=0}^{a} r^5 \int_{\theta=0}^{\pi/2} \cos^2\theta\sin^2\theta\end{aligned}$$

which is easy to solve.

Example 24.6. Suppose B is the open disk instead, namely $B = \left\{(x,y) : x^2+y^2 < a^2\right\}$ (as opposed to its intersection with the first quadrant). Then g is not injective on $(0,a)\times[0,2\pi]$. However, if we let B' be equal to B with the positive axis removed, we can use the change of variables on $A = (0,a)\times(0,2\pi)$, and we know that

$$\int_B x^2y^2 = \int_{B'} x^2y^2 = \int_B \chi_{B'}x^2y^2 = 0$$

since the latter integrand is 0 outside a set of measure 0.

24.2. 1-**dimensional change of variable.**

Theorem 24.7. *Let $g : I = [a,b] \to J = [g(a), g(b)]$ and $J \to f\mathbb{R}$, with f continuous, $g \in C^1$ and $g' \neq 0$ on $[a,b]$. Then*

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

or equivalently

$$\int_J f = \int_I (f\circ g)|g'|.$$

Convention. We need to be careful about conventions to see why the absolute value is needed. If $c < d$ we write

$$\int_{[c,d]} f =: \int_c^d f = -\int_d^c f.$$

If $g' > 0$, then $g' = |g'|$ and $g(a) < g(b)$ and so

$$\int_J f = \int_{g(a)}^{g(b)} f = \int_a^b (f \circ g)|g'| = \int_I (f \circ g)|g'|.$$

If $g' < 0$ then $g = -|g'|$ and $g(b) < g(a)$ so that

$$\int_J f = -\int_{g(a)}^{g(b)} f = -\int_b^a (f \circ g)g' = \int_I (f \circ g)|g'|.$$

This exhausts all possible cases.

Proof of theorem 24.7. Define

$$F(y) = \int_{g(a)}^{y} f$$

so that

$$F'(y) = f(y)$$

by the first fundamental theorem of calculus. Moreover,

$$F \circ g(y) = \int_{g(a)}^{g(b)} f$$

and by the chain rule

$$\begin{aligned}(F \circ g)'(y) =& F'(g(y))g'(y) \\ =& (f \circ g)(y)g'(y)\end{aligned}$$

and by the second fundamental theorem of calculus we have

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

□

Exercise. Compute

$$\int_0^1 (2x^2 + 1)^{10} 4x.$$

24.3. **Defining integration on open sets.** We defined $\int_B f$ where B is bounded and f is bounded. We want to extend the definition to allow either to be unbounded. One option is to decompose B into compact sets and define

$$\int_B f := \sup_{C \text{ compact}} \int_C f$$

Another option is to decompose f into integrable functions using partitions of unity, which we will see later.

25.1. Partitions of unity.

Definition 25.1. Let $A \subset \mathbb{R}^n$ be an open rectangle, and let $f : A \to \mathbb{R}$. Define

$$X = \{x \in \mathbb{R}^n : f(x) \neq 0\}.$$

We define the **support** of f to be

$$\operatorname{supp}(f) := X \cup \operatorname{bd}(X).$$

Example 25.2. For $f(x) = x^2 - x$ we have $\operatorname{supp}(f) = \mathbb{R}$, and for

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \cap [0,1] \\ 0 & \text{else} \end{cases}$$

we have $X = \mathbb{Q} \cap [0,1]$, and $\operatorname{supp}(f) = [0,1]$.

A few facts:

- in general, $x \in \operatorname{supp}(f)$ if and only if there exists an open rectangle R such that $x \in R$ and $f\big|_R = 0$.
- $\operatorname{supp}(f)$ is always closed, since $\operatorname{supp}(f)^c = \operatorname{ext}(X)$ which is open by definition.
- $\operatorname{supp}(f)$ is the smalles closed set V such that $f\big|_{V^c} = 0$;
- we saw that f has **compact support** if $\operatorname{supp}(f)$ is compact.

Theorem 25.3 (Partitions of unity)**.** *Let $A \subset \mathbb{R}^n$ be an open rectangle, $\mathcal{O}$ an open cover, and $A = \bigcup_{U \in \mathcal{O}} U$. Then there exists a collection Φ of C^1 (or smooth) functions $\varphi \in \Phi$ defined on A such that*

(i) $0 \leq \varphi \leq 1$ for all $\varphi \in \Phi$

(ii) (local finiteness) for all $x \in A$ we have $\varphi(x) = 0$ for all but finitely many $\varphi \in \Phi$

(iii) for all $x \in A$

$$\sum_{\varphi \in \Phi} \varphi(x) = 1$$

(iv) for all $\varphi \in \Phi$ there exists $U \in \mathcal{O}$ such that $\operatorname{supp}(\varphi) \subset U$

(v) every $\varphi \in \Phi$ has compact support.

Remark. Properties (i) to (iii) define a partition of unity. Property (iv) can be stated as Φ being **subordinate** to $\mathcal{O}$, and property (v) can be stated as Φ having compact support.

Example 25.4. Let $A = \mathbb{R}$ and $\mathcal{O} = \{(-\infty, 2), (-2, \infty)\}$. Define

$$\varphi_1(x) = \begin{cases} 1 & x \leq -1 \\ \frac{1}{2}(1-x) & x \in [-1,1] \\ 0 & x \geq 1 \end{cases}$$

$$\varphi_2(x) = \begin{cases} 0 & x \leq -1 \\ \frac{1}{2}(1+x) & x \in [-1,1] \\ 1 & x \geq 1 \end{cases}.$$

Then $\Phi = \{\varphi_1, \varphi_2\}$ is almost a partition of unity, but φ_1, φ_2 are not C^1 and Φ does not have compact support. Partitions of unity are useful for "gluing" (they are a local-to-global tool). Suppose that given $f_i : U \to \mathbb{R}$ and we want to "glue" them to get a function defined on $\mathbb{R} = U_1 \cup U_2$. Define

$$f(x) = \varphi_1 f_1 + \varphi_2 f_2.$$

Then this function interpolates between -1 and 1.

25.1.1. *Application to integration.* Let $A \subset \mathbb{R}^n$ be open with $\mathcal{O}$ an open cover and Φ a POU subordinate to $\mathcal{O}$ with compact support and $f : A \to \mathbb{R}$ continuous. For $\varphi \in \Phi$ we know that $\operatorname{supp}(\varphi) \subset Q$ for some close rectangle, so that $\varphi f = 0$ outside of Q and is bounded on Q (by continuity). We define

$$\int_A \varphi f := \int_Q \varphi \cdot f$$

and we saw that f is **integrable (in the extended sense)** if if $\sum_i \int_A \varphi_i |f|$ exists and define

$$\int_A f = \sum_i \int_A \varphi_i f.$$

25.2. **Existence of partitions of unity.**

Lemma 25.4.1. *For any closed rectangle $Q \subset \mathbb{R}^n$ there exists a smooth $\psi_Q : \mathbb{R}^n \to \mathbb{R}$ such that $\psi_Q(x) > 0$ for all $x \in \operatorname{int} Q$ and $\psi_Q(x) = 0$ for all $x \notin \operatorname{int}(Q)$.*

Example 25.5. For $Q = [0,1] \subset R$, an example is

$$\psi(x) = \begin{cases} x(1-x) & x \in [0,1] \\ 0 & \text{else} \end{cases}.$$

For $Q = [a,b]$ define

$$\varphi_Q(x) = \psi\left(\frac{x-a}{b-a}\right)$$

and for $Q = [a_1, b_1] \times \cdots \times [a_n, b_n]$ define

$$\varphi_Q(x) = \psi\left(\frac{x_1 - a_1}{b_1 - a_1}\right) \cdots \varphi\left(\frac{x_n - a_n}{b_n - a_n}\right).$$

This example is called a **bump function**.

Theorem 25.6 (Technical theorem)**.** *Given $A \subset \mathbb{R}^n$ with $\mathcal{O}$ an open cover, there exist closed rectangles $Q_j \subset A$ such that*

(a) the interiors of the Q_i's cover A
(b) $Q_i \subset U$ for some $U \in \mathcal{O}$
(c) (local finiteness) for all $x \in A$ there exists a neighborhood that intersects only finitely many Q_i.

Given this theorem, we can prove Theorem 25.3:

Proof of Theorem 25.3 given Theorem 25.6. Take Q_i as in the technical theorem and ψ_{Q_i} as in the lemma. Let

$$\lambda(x) = \sum_i \varphi_{Q_i}(x) > 0$$

and define

$$\varphi_i = \frac{\psi_{Q_i}(x)}{\lambda(x)}.$$

Then $\Phi = \{\varphi_i\}$ do the job (check this!). □

Proof of Theorem 25.6. Let $C_i \subset A$ compact such that $C_i \subset \operatorname{int}(C_{i+1})$ and $A = \bigcup_i C_i$. Let $D_i = C_{i+1} \setminus \operatorname{int}(C_i)$. This is a compact set. For $x \in D_i$ choose $Q_x \subset A$ such that $Q_x \cap C_{i-1} = \emptyset$ and $Q_x \subset U$ for some $U \in \mathcal{O}$. Then

$$D_i \subset \bigcup_{x \in D_i} \operatorname{int}(Q_x).$$

Take a finite subcover (by compactness) and check that it satisfies the theorem. □

26.1. **Diffeomorphisms.** This week we will prove the Change of Variables theorem. Today we will take a closer look at change of variable functions.

Definition 26.1. Let $A, B \subset \mathbb{R}^n$ be open. If $g : A \to B$ is C^1 and bijective with g^{-1} being C^1 as well, we say that g is a **diffeomorphism**.

Example 26.2. The open unit square $(0,1)^2$ is diffeomorphic to the open circle with removed positive axis

$$\left\{(x,y) : 0 < x^2 + y^2 < 1, y = 0 \Longrightarrow x < 0\right\}.$$

The diffeomorphism is

$$g(r,t) = (r\cos 2\pi t, r\sin 2\pi t).$$

One can use algebraic topology to show that the open unit disk is not diffeomorphic to the open unit disk with the origin removed.

Remark. A diffeomorphism is the same as a change of variable. In fact, if $g : A \to B \subset \mathbb{R}^n$ is a diffeomorphism, then Dg is invertible by the chain rule. Moreover, if $g : A \to g(A) \subset \mathbb{R}^n$ is a change of variable, then $g(A)$ is open and g^{-1} is C^1 is by the inverse function theorem.

Remark. As an analogy, one can think that diffeomorphisms are to open subsets of $\mathbb{R}^n$ what linear isomorphisms are to vector spaces.

26.2. **Diffeomorphism behavior.** Our slogan is that *diffeomorphisms preserve your favorite property*.

Lemma 26.2.1. *If $g : A \to B$ is a diffeomorphism then*
- *(i) $U \subset A$ is open if and only if $g(U)$ is open*
- *(ii) $E \subset A$ has measure 0 if and only if $g(E)$ has measure 0*
- *(iii) for $K \subset A$ compact, we have*

$$\begin{aligned} g(\operatorname{int} K) &= \operatorname{int}(g(K)) \\ g(\operatorname{bd} K) &= \operatorname{bd}(g(K)). \end{aligned}$$

Proof. (i) follows from continuity of g, g^{-1} (recall that a function is continuous if and only if preimages of open sets are open). We now prove (ii). The key to the proof is that if $C \subset A$ is a cube of width w and $\|Dg(x)\| \le M$ for all $x \in C$, then $g(C)$ is contained in a cube of width $n \cdot M \cdot w$. In fact, by the multivariable mean value theorem we know that that for all $x, u \in C$ there exists c on the line between u and x such that

$$g(x) - g(a) = Dg(c)(x-u).$$

Therefore

$$\begin{aligned} \|g(x) - g(u)\| =& |Dg(c)(x-u)\| \\ \le& n\|Dg(c)\| \cdot \|x-u\| \\ \le& n \cdot M \cdot w. \end{aligned}$$

Suppose now that E has meausre 0. For more details, see Munkres' proof. □

Remark. The condition that g be C^1 is actually necessary. In fact, there exist continuous maps $g : (0,1) \to (0,1)^2$ such that the image is the whole square. These are called *Peano space filling curves*.

26.3. **Primitive diffeomorphism.**

Definition 26.3. A diffeomorphism $h = (h_1, \dots, h_n) : A \to B$ is **primitive** if $h_i(x) = x_i$ for some i.

Example 26.4. The diffeomorphism $h(x,y) = (3x + 7y, y)$ is primitive, whereas $h(x,y) = (y,x)$ is not.

Theorem 26.5 (Primitive diffeomorphism)**.** *Let $A, B \subset \mathbb{R}^n$ be open with $n \geq 2$. Let $g : A \to B$ be a diffeomorphism. Given $a \in A$ there exists a neighborhood $U_0 \ni a$ and diffeomorphism*

$$U_0 \xrightarrow{h_1} U_1 \xrightarrow{h_2} U_2 \to \cdots \xrightarrow{h_k} U_k \subset B$$

such that $g\big|_{U_0} = h_k \circ \cdots \circ h_1$ and h_j is primitive for each j.

Example 26.6. We can look at special cases.

(1) In the case where g is a translation, we can write

$$g(x) = x + c = (\cdots((x + c_1 e_1) + c_2 e_2) + \cdots + c_n e_n)$$

and therefore

$$g = T_{c_n e_n} \circ \cdots \circ T_{c_1 e_1}$$

where $T_{c_i e_i}$ is a translation by $c_i e_i$ and is certainly primitive.

(2) If $g = A \in M_n(\mathbb{R})$ is a linear isomorphism, we can used row operations (which we showed in the homework to be primitive) to decompose g.

27. April 6

27.1. **Diffeomorphisms. Recall.**

- Given open sets $A, B \subset \mathbb{R}^n$ a map $g : A \to B$ is a **diffeomorphism** if C^1 is a bijection with C^1 inverse.
- Moreover, the inverse function theorem says that, given a C^1 map $f : A \to B$, if $Df(a)$ is invertible then there exists open $U \ni a$ such that $f : U \to f(U)$ is a diffeomorphism (we say that f is locally a diffeomorphism).
- A diffeomorphism $g : \mathbb{R}^2 \mathbb{R}^2$ is **primitive** if either
$$g(x,y) = (g_1(x,y), y)$$
or
$$g(x,y) = (x, g_2(x,y)).$$

Example 27.1. Consider the function $f(x,y) = (x, x^2 y)$ on $A = (-2,2) \times (-1,1)$. This is not a diffeomorphism since it is not injective on the y axis, but it is a diffeomorphism if we restrict it to $A' = (0,2) \times (-1,1)$. The inverse is $h(x,y) = (x, y/x^2)$, which is also C^1. Here
$$f(A') = \{(x,y) : 0 < x < 2, f(x,-1) < y < f(x,1)\}.$$

Example 27.2. The map $f(x,y) = (x^2 - y^2, 2xy)$ on $A = (-1,3) \times (1,2)$ is a diffeomorphism, but it is not primitive. However, it is harder in this case to show that this is a diffeomorphism. It is also harder to write the image $f(A)$ in terms of the x, y coordinates (i.e. it is hard to compute
$$v(f(A)) = \int_Q \chi_{f(A)}$$
with Fubini).

Lemma 27.2.1 (Diffeomorphisms are locally primitive)**.** *Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ such that $g(0) = 0$ and $Dg(0) = \mathrm{Id}$. Then there exists $U \ni 0$ and primitive diffeomorphisms*
$$U \xrightarrow{g_1} g_1(U) \xrightarrow{g_2} g_2 g_1(U)$$
so that $g = g_2 \circ g_1$.

Proof. Given $g(x,y) = (g_1(x,y), g_2(x,y))$ we will try to write $g = k \circ h$ such that k preserves the x coordinate and h preserves the x coordinate. Then we want
$$h(x,y) = (g_1(x,y), y)$$
and so we need
$$k(x,y) = (x, g_2(h^{-1}(x,y))).$$
This is because
$$\begin{aligned} k \circ h(x,y) =& k(g_1(x,y), y) \\ =& \left(g_1(x,y), g_2\left(h^{-1}(g_1(x,y), y)\right)\right) \\ =& g(x,y). \end{aligned}$$

However, we still need to check that h is invertible near 0 and that K, h, h^{-1} are diffeomorphisms near 0. This follows from the inverse function theorem. In fact,

$$Dh(0) = \begin{pmatrix} D_1 g_1(0) & D_2 g_1(0) \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is invertible and therefore there exists $U_1 \ni 0$ such that $h : U_1 \to h(U_1)$ is a diffeomorphism. By the chain rule,

$$Dk(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

as well (check this). Therefore, there eixsts $U_2 \subset h(U_1)$ such that $k : U_2 \to K(U_2)$ is a diffeomorphism. □

27.2. **Change of variables theorem.**

Theorem 27.3 (Change of variables)**.** *If $g : A \to B$ is a diffeomorphism, for every $f : B \to \mathbb{R}$. If $\int_B f$ exists then*

$$\int_A (f \circ g) |\det Dg|$$

exists and

$$\int_B f = \int_A f \circ g |\det Dg|.$$

The proof will be divided in 5 steps. We will refer to Munkres for details.

Proof. Step 1. We want to show that the theorem is true for composition. Namely, suppose we have diffeomorphisms

$$U \xrightarrow{g} V \xrightarrow{h} W$$

such that the theorem holds for g, h. Then we want to show that if $f : W \to \mathbb{R}$ is continuous then $\int_W f$ exists and

$$\int_W f = \int_U f \circ h \circ g |\det D(h \circ g)|.$$

This follows from composition, since

$$\begin{aligned} \int_W f &= \int_V f \circ h |\det Dh| \\ &= \int_U f \circ g \circ h |\det Dh \circ g| \cdot |\det Dg| \\ &= \int_U f \circ g \circ h |\det D(h \circ g)| \end{aligned}$$

where the last step follows from the chain rule.

Step 2. This is a local to global statement. Given $g : A \to B$ suppose that for all $x \in A$ there exists a neighborhood U_x such that the theorem is true for $g : U_x \to g(U_x) =: V_x$. We want to show that the theorem is then true for $g : A \to B$. In fact, $\{V_x\}_{x \in U}$ is a cover of B. We can choose a partition of unity $\{\varphi_i : B \to \mathbb{R}\}$ with compact support $\operatorname{supp}(\varphi_i) \subset V_i$ so that

$$\begin{aligned}\int_B f &= \sum_i \int_{V_i} \varphi_i f \\ &= \sum_i \int_{U_i} (\varphi_i \circ g)(f \circ g)|\det Dg| \\ &= \int_A f \circ g |\det Dg|.\end{aligned}$$

Step 3. We proved the theorem in dimension 1 for closed intervals, so by Step 2 we know it holds in the general case.

Step 4. We know that every diffeomorphism is locally a composition of primitive diffeomorphisms. Then by steps 1 and 2 it suffices to show this for primitive diffeomorphisms.

Step 5. To prove this for primitive diffeomorphisms, we use Fubini's theorem. We leave the proof to Munkres.

□

Application. This can be used to make Archimedes' exhaustion principle rigorous. Let D be the circle with radius r, inscribed in the square $Q = [-r, r]^2$. Therefore

$$\begin{aligned}v(D) &= \int_Q \chi_D \\ &= 2\int_{x=-r}^{r} \int_{y=0}^{r} \chi_D(x,y) \\ &= 2\int_{-r}^{r} \int_{y=0}^{\sqrt{r^2-x^2}} 1 \\ &= 2\int_{x=-r}^{r} \sqrt{r^2-x^2}.\end{aligned}$$

Now we introduce the change of variables $g(\theta) = r\sin\theta$ so that the above becomes

$$2\int_{\theta=-\pi/2}^{\pi/2} r^2\cos^2\theta = \pi r^2.$$

28. April 9

28.1. k-**dimensional volumes in** $\mathbb{R}^n$ **(forms). Recall.** The determinant

$$\det : \underbrace{\mathbb{R}^k \times \mathbb{R}^k}_{k \text{ times}} \to \mathbb{R}$$

is the only function which is multilinear and alternating and such that

$$\det(e_1, \dots, e_k) = 1.$$

For $A \in M_k(\mathbb{R})$ we view $\det A$ as a function of the columns. For a parallelogram P with sides $(a,c),(b,d) \in \mathbb{R}^2$, its signed area is given by the determinant:

$$\text{(signed) area} = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

This actually follows from the change of variables theorem: since A is a diffeomorphism, we can set $P = A(Q)$ where Q is the unit square, and therefore

$$\begin{aligned} v(P) &= \int_P 1 \\ &= \int_Q |\det A| \\ &= |\det A|. \end{aligned}$$

Definition 28.1. A k-**form on** $\mathbb{R}^n$ is a multilinear, alternating function

$$\varphi : \underbrace{\mathbb{R}^n \times \mathbb{R}^n}_{k \text{ times}} \to \mathbb{R}.$$

The set of k-forms on $\mathbb{R}^n$, denoted by $\bigwedge^k(\mathbb{R}^n)$, is a vector space.

Example 28.2. Fix $1 \leq i_1, \dots i_k \leq n$. Define

$$\varphi := dx_{i_1} \wedge \cdots \wedge dx_{i_k} \in \bigwedge^k \mathbb{R}^n.$$

This form is defined as follows. Given vectors $v_1, \dots, v_k \in \mathbb{R}^n$ we organize them in a matrix as

$$\begin{pmatrix} v_{11} & \cdots & v_{1k} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nn} \end{pmatrix}$$

and define

$$\varphi(v_1, \dots, v_k) = \det \begin{pmatrix} v_{i_1,1} & \cdots & v_{i_1,k} \\ \vdots & \ddots & \vdots \\ v_{i_k,1} & \cdots & v_{i_k,k}. \end{pmatrix}$$

This map is multilinear and alternating because det is. Such form is called an **elementary** k**-form**. Note that

$$\begin{aligned} dx_i : \mathbb{R}^n &\to \mathbb{R} \\ v = \sum_{i=1}^{n} c_i e_i &\mapsto c_i, \end{aligned}$$

namely dx_i is the 1-form that picks out the ith coordinate.

To see how we can interpret these forms as k-dimensional volumes on $\mathbb{R}^n$, we see that in the case of elementary forms we are measuring signed volume of projections of parallelogram. For example, the form $dx_1 \wedge dx_2$ on $\mathbb{R}^3$ takes two vectors $v_1, v_2 \in \mathbb{R}^3$ and returns the signed area of the projection of the parallelogram on the xy plane.

Note. On $\mathbb{R}^n$ we see that $dx_i \wedge dx_j = -dx_j \wedge dx_i$ since the determinant is alternating. Moreover,

$$\begin{aligned} dx_1 \wedge dx_2 \wedge dx_1 &= 0 \\ dx_2 \wedge dx_3 \wedge dx_1 &= dx_1 \wedge dx_2 \wedge dx_3. \end{aligned}$$

Definition 28.3. If $1 \leq i_1 < \cdots < i_k \leq n$ we call

$$dx_{i_1} \wedge \cdots dx_{i_k}$$

an **elementary** k**-form**. For a permutation

$$\sigma : \{1, \ldots, k\} \to \{1, \ldots, k\}$$

we set $j_\ell = i_{\sigma(\ell)}$ and define

$$dx_{j_1} \wedge \cdots \wedge dx_{j_k} = \operatorname{sign}(\sigma) dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

In particular these forms are linearly dependent in $\bigwedge^k(\mathbb{R}^n)$.

Theorem 28.4. *The elementary k-forms are a basis for $\bigwedge^k(\mathbb{R}^n)$.*

Example 28.5. Every $\varphi \in \bigwedge^2(\mathbb{R}^3)$ can be written uniquely as

$$\varphi = a \underbrace{dx_1 \wedge dx_2}_{\varphi_{12}} + b \underbrace{dx_2 \wedge dx_3}_{\varphi_{23}} + c \underbrace{dx_1 \wedge dx_3}_{\varphi_{13}}.$$

Corollary 28.5.1. *It follows that*

$$\dim \bigwedge^k(\mathbb{R}^n) = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Proof of Theorem 28.4. We will only consider the case $\bigwedge^2(\mathbb{R}^3)$. The general case differs only in notation. We want to show that $\varphi_{12}, \varphi_{23}, \varphi_{13}$ are linearly independent and span. The key is that 2-forms on $\mathbb{R}^3$ are determined by their values on pairs of distinct basis vectors of $\mathbb{R}^3$.

Note. If $i_1 < i_2, j_1 < j_2$ then

$$\varphi_{i_1, i_2}(e_{j_1}, e_{j_2}) = \begin{cases} 1 & i_1 = j_1, i_2 = j_2 \\ 0 & \text{else} \end{cases}.$$

Step 1. We now prove linear independence. Suppose $0 = a\varphi_{12} + b\varphi_{13} + c\varphi_{23}$. Evaluating this on (e_i, e_j) shows that $a = b = c = 0$ (for example, $a\varphi_{12}(e_1, e_2) = a$).

Step 2. We now prove that they span. Fix $\varphi \in \bigwedge^2 \mathbb{R}^3$. We want to write

$$\varphi = \sum_{i<j} c_{ij}\varphi_{ij}.$$

Fix $a, b \in \mathbb{R}^3$. Then

$$\begin{aligned}\varphi(a,b) &= \varphi\left(\sum a_i e_i, \sum b_j e_j\right)\\ &= \sum_{1\le i,j\le 3} a_i b_j \varphi(e_i, e_j)\\ &= (a_1b_2 - a_2b_1)\,\varphi(e_1,e_2) + (a_1b_3 - b_3a_1)\varphi(e_1,e_3) + (a_2b_3 - a_3b_2)\varphi(e_2,e_3)\end{aligned}$$

and so

$$\varphi = \varphi(e_1,e_2)\varphi_{12} + \varphi(e_1,e_3)\varphi_{13} + \varphi(e_2,e_3)\varphi_{23}.$$

□

28.2. **Wedge product of forms.** The product

$$\wedge : \bigwedge^k(\mathbb{R}^n) \times \bigwedge^\ell(\mathbb{R}^n) \to \bigwedge^{k+\ell}(\mathbb{R}^n)$$

is implicit in our notation for elementary forms. We can extend this operation from elementary forms to arbitrary forms by forcing $\wedge$ to be distributive and associative.

Remark. If $k > n$ we have $\bigwedge^k(\mathbb{R}^n) = 0$.

Example 28.6. We see that

$$\begin{aligned}&(7dx_1 \wedge dx_2 - 2dx_1 \wedge dx_3) \wedge (dx_2 + dx_3)\\ =&7dx_1 \wedge dx_2 \wedge dx_2 + 7dx_1 \wedge dx_2 \wedge dx_3 - 2dx_1 \wedge dx_3 \wedge dx_2 - 2dx_1 \wedge dx_3 \wedge dx_3\\ =&9dx_1 \wedge dx_2 \wedge dx_3.\end{aligned}$$

29. April 11

29.1. **Differential k-forms.** A way of writing the fundamental theorem of calculus is that given a function $f:[a,b]\to\mathbb{R}$ such that $f=Dg$ for some $g:[a,b]\to\mathbb{R}$ then

$$\int_{[a,b]} f = \int_{\partial[a,b]} g = g(b)-g(a)$$

where $\partial[a,b]=b\cup a$ is the oriented boundary of $[a,b]$. This is a particular case of Stokes' theorem, which asserts that given a form ω such that $\eta = d\omega$ we have

$$\int_Q \eta = \int_{\partial Q}\omega.$$

Today we will see that a differential k-form is something that computes "volumes" of k-manifolds in $\mathbb{R}^n$.

Definition 29.1. Let $U\subset\mathbb{R}^n$ be open. A **differential k-form** on U is written

$$\omega = \sum f_{i_1,\dots,i_k} dx_{i_1}\wedge\cdots\wedge dx_{i_k},$$

where $f_{i_1,\dots,i_k}:U\to\mathbb{R}$ are smooth functions. By adopting the convenient notation $I=(i_1,\dots,i_k)$ we can write

$$\omega=\sum f_I dx_I.$$

Note that for each $p\in U$

$$\omega(p)=\sum f_I(p)dx_I\in\bigwedge^k(\mathbb{R}^n).$$

Therefore given vectors $v_1,\dots,v_k\in\mathbb{R}^n$ we can evaluate $\omega(p)(v_1,\dots,v_k)\in\mathbb{R}$. Thus we view ω as a smooth choice of k-form at each $p\in U$, i.e. as a function $\omega: U\to\bigwedge^k(\mathbb{R}^n)$.

Convention. A 0-form is a function $f:U\to\mathbb{R}$. The set of differential k-forms on U is denoted $\Omega^k(U)$ and it is an infinite dimensional vector space.

Example 29.2. On $\mathbb{R}^3$, a 0-form is a function $f:\mathbb{R}^3\to\mathbb{R}$. A 1-form ω is expressed as

$$\omega = f_1dx+f_2dy+f_3dz.$$

If we define

$$F=(f_1,f_2,f_3):\mathbb{R}^3\to\mathbb{R}^3$$

we can view F as a vector field, namely a function that attaches a vector to every point of $\mathbb{R}^3$. For $v=v_1e_1+v_2e_2+v_3e_3\in\mathbb{R}^3$ we have

$$\begin{aligned}\omega(p)(v) &= (f_1(p)dx+f_2(p)dy+f_3(p)dz)(v_1e_1+v_2e_2+v_3e_3) = f_1(p)v_1+f_2(p)v_2+f_3(p)v_3\\ &= F(p)\cdot v\end{aligned}$$

where $\cdot$ is the dot product.

Remark. In physics 1-forms arise as the work of a "force" vector field F. In this case, $F(p)\cdot v$ is the infinitesimal work on a particle at p moving with velocity v. Along a path $\gamma:[a,b]\to\mathbb{R}^3$ the total work is then equal to

$$\int_{t=a}^{b} F(\gamma(t))\cdot\gamma'(t).$$

2−forms can be expressed as

$$\omega = f_1 dx\wedge dy + f_2 dx\wedge dz + f_3 dy\wedge dz.$$

Given $F=(f_1,f_2,f_3):\mathbb{R}^3\to\mathbb{R}^3$ we can write

$$\omega(p)(u,v) = \det(F(p),u,v) = F(p)\cdot(u\times v),$$

where $\times$ is the cross-product on $\mathbb{R}^3$. You will prove this in the homework. In physics, 2-forms arise as flux. The quantity $F(p)\cdot(u\times v)$ is the infinitesimal emasure of the amount of fluid flow through the parallelogram determined by u,v at p. For example, if the flux is parallel to the area, the flux is 0, and it is at a maximum when the parallelogram is perpendicular.

3-forms on $\mathbb{R}^3$ can be written as $\omega = f\,dx\wedge dy\wedge dz$. This is a volume form.

Example 29.3. On $\mathbb{R}^2$, given $\omega = dx$ and $\eta = xdy$, calculate the value of ω,η on v_1,v_2,v_3. We see that $\omega(0,0)(v_1)=0$, and $\omega(2,2)=dx$ so that $\omega(2,2)(v_2)=-1$ and $\omega(2,2)(v_3)=1$.

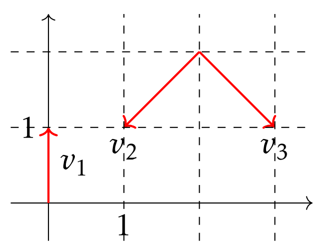

FIGURE 1.

Similarly $\eta(0,0)=0$ and $\eta(2,2)=2y$ so that $\eta(2,2)(v_2)=\eta(2,2)(v_3)=-2$.

29.2. **Exterior derivative.** Let $U\subset\mathbb{R}^n$ be open. We want to define an operator

$$d:\Omega^k(U)\to\Omega^{k+1}(U)$$

which allows us to differentiate forms. We define

$$d:\Omega^0(U)\to\Omega^1(U)$$

$$f\mapsto df=\sum_{i=1}^{n} D_i f\,dx_i.$$

Then d is linear, as

$$d(af+bg)=adf+bdg.$$

Also, the product rule holds, i.e.

$$d(fg)=fdg+gdf.$$

We want to extend this uniquely to a linear map

$$d : \Omega^k(U) \to \Omega^{k+1}(U)$$

by instisting that $d(dx_i) = 0$ and for a k-form α and a ℓ-form β we have

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta.$$

This can be done by defining, for any k-form

$$\omega = \sum_I f_I \, dx_I,$$

the **exterior derivative**

$$d\omega = \sum_I \sum_{i=1}^n D_i f_I \, dx_i \wedge dx_I.$$

Example 29.4. In $\mathbb{R}^3$, a function $f \in \Omega^0(\mathbb{R})$ gives us a vector field

$$df = D_1 f \, dx + D_2 f \, dy + D_3 f \, dz$$

also denoted ∇f (the gradient).

30.1. **Forms and integration.** The main thing we are going to see is how forms are related to integration. What we are aiming for is Stokes' theorem:

Theorem 30.1 (Stokes)**.** *If η is a k-form on $\mathbb{R}^n$ and $\eta = d\omega$ for some $(k-1)$-form ω, then for a k-dimensional rectangle Q*

$$\int_Q \eta = \int_{\partial Q} \omega.$$

In order to prove this we need to do three things:

(1) Define what it means to integrate a form over a rectangle, i.e. defining
$$\int_Q \eta;$$
(2) Discuss exterior derivatives further;
(3) Define ∂Q.

Suppose η is a k-form on $\mathbb{R}^k$. Then

$$\eta = f\,dx_1 \wedge \cdots \wedge dx_k$$

for a unique function $f : \mathbb{R}^k \to \mathbb{R}$. In fact, recall that

$$\dim \bigwedge^k(\mathbb{R}^k) = 1$$

and this space is spanned by

$$\begin{aligned} dx_1 \wedge \cdots \wedge dx_k : \mathbb{R}^k \times \cdots \times \mathbb{R}^k &\to \mathbb{R} \\ (v_1, \ldots, v_k) &\mapsto \det\left(v_1 \middle| \cdots \middle| v_k\right). \end{aligned}$$

For a closed rectangle $Q \subset \mathbb{R}^k$ we then define the **integral of** η **over** Q as

$$\int_Q \eta = \int_Q f.$$

We also write

$$\int_Q f\,dx_1 \wedge \cdots \wedge dx_n.$$

We also want to be able to integrate a k-form on a k-dimensional subspace of $\mathbb{R}^n$. To do that we will need pullbacks.

30.1.1. *Pullbacks.* Let $V \subset \mathbb{R}^m, U \subset \mathbb{R}^n$ be open sets. For a map $g : V \to U$ and a k-form $\omega \in \Omega^k(U)$, the **pullback** of ω along g is a k-form $g^*\omega \in \Omega^k(V)$. For $q \in V$ and $u_1, \ldots, u_k \in \mathbb{R}^m$ this is defined as

$$g^*\omega(q)(u_1, \ldots, u_k) = \omega(g(q))(Dg(q)(u_1), \ldots, Dg(q)(u_k)).$$

Remark. A pullback defines a *linear* map

$$g^* : \Omega^k(U) \to \Omega^k(V).$$

Convention. For a function $f \in \Omega^0(U)$ we write $g^*f = f \circ g$.

Example 30.2. Let $U, V = \mathbb{R}^n$, and let $\omega = dx_1 \wedge \cdots \wedge dx_n$ (the volume form). For a map $g : V \to U$ we know that

$$g^*\omega = \varphi dx_1 \wedge \cdots \wedge dx_n$$

for some unique φ. What is φ? For $q \in V$ we then have

$$g^*\omega(q) = \varphi(q) dx_1 \wedge \cdots \wedge dx_n.$$

We now compute

$$\begin{aligned} g^*\omega(q)(e_1, \dots, e_n) &= \omega(g(q))\,(Dg(q)e_1, \dots, Dg(q)e_n) \\ &= (dx_1 \wedge \cdots \wedge dx_n)\,(D_1 g(q), \dots, D_n g(q)) \\ &= \det\left(D_1 g(q) \middle| \cdots \middle| D_n g(p)\right) \\ &= \det Dg(q). \end{aligned}$$

Thus $\varphi = \det Dg$, i.e.

$$g^*(dx_1 \wedge \cdots \wedge dx_n) = \det Dg \cdot dx_1 \wedge \cdots \wedge dx_n.$$

30.1.2. *Properties of pullbacks.* We won't prove them now.

(i) $g^*(\eta \wedge \omega) = g^*\eta \wedge g^*\omega$;
(ii) $g^*(d\omega) = d(g^*\omega)$.

Definition 30.3. Given a k-form $\omega \in \Omega^k(\mathbb{R}^n)$ and a map $C : [0,1]^k \to \mathbb{R}^n$ we define

$$\int_C \omega = \int_{[0,1]^k} c^*\omega.$$

If

$$c^*\omega = \varphi dx_1 \wedge \cdots \wedge dx_n$$

then

$$\int_{[0,1]^k} c^*\omega = \int_{[0,1]^k} \varphi.$$

Example 30.4. Let $A = \mathbb{R}^2 \setminus \{0\}$. Consider the form

$$\eta = \frac{x}{x^2 + y^2} dx + \frac{y}{x^2 + y^2} dy.$$

We see that

$$\eta(1,0) = dx \quad \eta(1,1) = \frac{1}{2}(dx + dy) \quad \eta(2,0) = \frac{1}{2} dx.$$

Recall from last time that for $v \in \mathbb{R}^2$ we can write

$$\eta(x,y)(v) = F(x,y) \cdot$$

where the vector field F is

$$F(x,y) = \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right).$$

Thus for every circle center at the origin, F will be perpendicular to the circle and will grow as it approaches the origin. Consider not the curve

$$\begin{aligned} c : [0,1] \to & A \\ c(t) = & (1+2t, 2+4t). \end{aligned}$$

We compute

$$\int_C \eta = \int_{[0,1]} c^*\eta.$$

Using the two properties listed above, the above becomes easier to compute. In fact,

$$\begin{aligned} c^*\eta =& c^*\left(\frac{x}{x^2+y^2}dx + \frac{y}{x^2+y^2}dy\right) \\ =& c^*\left(\frac{x}{x^2+y^2}dx\right) + c^*\left(\frac{y}{x^2+y^2}\right). \end{aligned}$$

By setting

$$f(x,y) = \frac{x}{x^2+y^2}$$

we see

$$\begin{aligned} c^*(f(x,y)dx) =& (c^*f)(x,y)c^*dx \\ =& (f \circ c)(x,y) \cdot d(c^*x) \\ =& \frac{1+2t}{(1+2t)^2 + (2+4dt^2)} d(1+2t) \\ =& \frac{1+2t}{5(1+2t^2)} \cdot 2dt \end{aligned}$$

and doing a similar computation for dy we see that

$$c^*\eta = \frac{2}{1+2t}dt$$

and therefore

$$\int_C \eta = \int_0^1 \frac{2}{1+2t}dt = \log 3.$$

Compare this with the curve

$$\begin{aligned} d : [0,1] \to & A \\ d(t) = & (\cos 2\pi t, \sin 2\pi t). \end{aligned}$$

In this case,

$$\begin{aligned}
d^*\eta =& \frac{\cos 2\pi t}{1} d(\cos 2\pi t) + \frac{\sin 2\pi t}{1} d(\sin 2\pi t) \\
=& \cos 2\pi t \cdot (-\sin 2\pi t) \cdot 2\pi dx + \sin 2\pi t (\cos 2\pi t) \cdot 2\pi t \\
=& 0.
\end{aligned}$$

This should be evident since the vector field determined by η is always orthogonal to the curve d.

Example 30.5. Consider the form

$$\omega = \frac{-y}{x^2+y^2} dx + \frac{x}{x^2+y^2} dy.$$

If we consider thsi as a vector field, in this case the field is tangent at each point to circles centered at the origin. In particular

$$\begin{aligned}
\int_c \omega =& 0 \\
\int_d \omega =& 2\pi.
\end{aligned}$$

31.1. **Stokes' theorem. Recall.** Let $U \subset \mathbb{R}^n$ be open, and let

$$\omega = \sum_I f_I dx_I$$

be a differential k-form. Its exterior derivative is denoted $d\omega$. Let $c : [0,1]^k \to U$ be a k-cube. We define a k-chain to be a formal sum

$$z = \sum_i a_i c_i$$

for some $a_i \in \mathbb{Z}$. We define the boundary of z to be

$$\partial z = \sum_{i=1}^{k} \sum_{\alpha=0,1} (-1)^{i+\alpha} c_{i,\alpha}.$$

Today we will prove the following:

Theorem 31.1 (Stokes' Theorem)**.** *If $\omega = d\eta$ is a k-form and z is a k-chain, then*

$$\int_z \omega = \int_{\partial z} \eta.$$

Remark. The above can be written as

$$\int_z d\eta = \int_{\partial z} \eta,$$

which suggests that in a way "∂ and d are dual to each other."

Example 31.2. Let $k = 1, U = \mathbb{R}, \omega = d\eta$. We write $\omega = f(x)dx, c : [0,1] \to [a,b]$ so that

$$\int_C \omega = \int_{[0,1]} c^* \omega = \int_a^b f(t)dt$$

and

$$\int_{\partial c} = \eta(b) - \eta(a).$$

Thus Stokes' theorem gives us

$$\int_a^b \eta'(t)dt = \eta(b) - \eta(a),$$

which is the fundamental theorem of calculus.

Example 31.3. Let $k = 2, U = \mathbb{R}^2, \omega = dx \wedge dy$. Note that $\omega = d(xdy)$. Therefore, by Stokes's theorem

$$\int_C \omega = \int_{\partial C} xdy.$$

Therefore Stokes' theorem tells us that to calculate the area of C we just integrate the vertical vector field xdy over the boundary of C. For a simple example, consider the rectangle with sides of length a, b with the bottom left corner centered at the origin. Then the integral over the horizontal edges is zero since the field and the sides are perpendicular, and moreover the integral over the left vertical side is 0 since the field is 0. Thus

$$\operatorname{area}(C) = \int_0^b a\,dt = ab.$$

Example 31.4. We will now see a case in which Stokes' theorem does *not* hold. Consider $k = 1, U = \mathbb{R}^2 \setminus \{0\}$, and define

$$\omega = \frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy.$$

In polar coordinates we can write $\omega = d\theta$. We want to show that there does not exist any form η such that $\omega = d\eta$. Consider $c : [0,1] \to \mathbb{R}^2$ which loops around the origin, i.e.

$$c(t) = (\cos 2\pi t, \sin 2\pi t).$$

If $\omega = d\eta$ it would follow from Stokes' theorem that

$$2\pi = \int_C \omega = \int_{\partial C} \eta = 0,$$

where the last equality holds since the boundary is empty.

Proof of Theorem 31.1. Let's fix the notation first. Let ω be a k-form, and η a $(k-1)$-form with $d\eta = \omega$. For simplicity, let $k = 3$, and let $I^3 : [0,1]^3 \xrightarrow{Id} [0,1]^3$. Let's also assume that $\eta = f\,dx_1 \wedge dx_2$. We want to compare

$$\underbrace{\int_c \omega}_{(1)} \qquad \text{and} \qquad \underbrace{\int_{\partial c} \eta}_{(2)}.$$

We write

$$\begin{aligned}
(1) &= \int_{[0,1]^k} \omega \\
&= \int_{[0,1]^3} D_3 f\, dx_3 \wedge dx_1 \wedge dx_2 \\
&= \int_{x_1=0}^1 \int_{x_2=0}^1 f(x_1,x_2,1) - f(x_1,x_2,0).
\end{aligned}$$

For (2), we start by computing

$$\partial I^3 = \sum_{i=1}^3 \sum_{\alpha=0,1} (-1)^{i+\alpha} I^3_{(i,\alpha)}.$$

Therefore

$$\begin{aligned}\int_{\partial I^3} \eta &= \sum_{i=1}^{3} \sum_{\alpha=0,1} (-1)^{i+\alpha} \int_{I^3_{(i,\alpha)}} \eta \\ &= \sum_{i=1}^{3} \sum_{\alpha=0,1} (-1)^{i+\alpha} \int_{[0,1]^2} (I^3_{(i,\alpha)})^* \eta.\end{aligned}$$

We know that

$$\begin{aligned}I^3_{1,\alpha}(x_1, x_2) =&(\alpha, x_1, x_2)\\ I^3_{2,\alpha}(x_1, x_2) =&(x_1, \alpha, x_2)\\ I^3_{3,\alpha}(x_1, x_2) =&(x_1, x_2, \alpha).\end{aligned}$$

With this we see that

$$\begin{aligned}(I^3_{1,\alpha})^* \eta =&(I^3_{1,\alpha})^* f \, dx_1 \wedge dx_2\\ =&f \circ I^3_{1,\alpha} d\alpha \wedge dx_2 = 0\end{aligned}$$

and similarly $(I^3_{2,\alpha})^* \eta = 0$. The only nonvanishing components are

$$(I^3_{3,\alpha})^* f \, dx_1 \wedge dx_2 = f \circ I^3_{3,\alpha} dx_1 \wedge dx_2.$$

Therefore

$$\begin{aligned}\sum_{i=1}^{3} \sum_{\alpha=0,1} (-1)^{i+\alpha} \int_{[0,1]^2} (I^3_{(i,\alpha)})^* \eta &= \sum_{\alpha=0,1} (-1)^{3+\alpha} \int_{[0,1]^2} (I^3_{3,\alpha})^* \eta \\ &= \sum_{\alpha=0,1} (-1)^{3+\alpha} \int_{[0,1]^2} f(x_1, x_2, \alpha) dx_1 \wedge dx_2 \\ &= -\int_{x=0}^{1} \int_{x_2=0}^{1} f(x_1, x_2, 0) + \int_{x=0}^{1} \int_{x_2=0}^{1} f(x_1, x_2, 1),\end{aligned}$$

which agrees with what we calculated for (1). $\square$

32.1. Winding numbers. For today, we will fix the following notation:

- $A = \mathbb{R}^2 \setminus \{0\}$, $B = A \setminus$ positive x axis
- $c : [0,1] \to A, c(0) = c(1)$ is a closed curve
- we fix a 1-form
$$\omega = \frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy$$
on A.

From the homework you know that

- $d\omega = 0$ (i.e. ω is **closed**)
- if $\theta : B \to (0, 2\pi)$ is defined implicitly by $(x,y) = (r\cos\theta, r\sin\theta)$ then $\omega = d\theta$ on B, i.e. ω is **exact** on B
- ω is *not* exact on A, i.e. there does not exist $f : A \to \mathbb{R}$ such that $\omega = df$ (you will often see ω denoted by $d\theta$, but it is an abuse of notation).

Definition 32.1. For a closed curve c in A, the **winding number of c around** 0 is
$$\mathrm{wind}(c) = \frac{1}{2\pi}\int_c \omega.$$

Example 32.2. From homework 10, if
$$\begin{aligned} c_{R,n} : [0,1] &\to A \\ t &\mapsto (R\cos 2\pi nt, R\sin 2\pi nt) \end{aligned}$$
then $\mathrm{wind}(c_{R,n}) = n$.

Example 32.3. If $c = \partial d$ for some $d : [0,1]^2 \to A$, then
$$\mathrm{wind}(c) = \int_{\partial d}\omega = \int_d d\omega = 0$$
since ω is closed.

Question: What values does $\mathrm{wind}(c)$ take?
Claim. We claim that $\mathrm{wind}(c)$ is always an integer. This follows from 2 observations:

(1) If c_1, c_2 are closed curves and $b : [0,1]^2 \to A$ is such that $\partial b = c_1 - c_2$ then $\mathrm{wind}(c_1) = \mathrm{wind}(c_2)$ cecause
$$\begin{aligned} \mathrm{wind}(c_1) - \mathrm{wind}(c_2) &= \int_{c_1}\omega - \int_{c_2}\omega \\ &= \int_{\partial b}\omega \\ &= \int_b d\omega = 0. \end{aligned}$$

(2) From homework 10, for any c there exists n and a cube $b : [0,1]^2 \to A$ such that
$$\partial b = c - c_{1,n}.$$

This implies that

$$\mathrm{wind}(c) = \mathrm{wind}(c_{1,n}).$$

Remark. The fact that ω is closed but not exact makes ω interesting! If $\omega = df$ were exact on A then for all closed curves c we would have that

$$\int_c \omega = \int_{\partial c} f = f(c(1)) - f(c(0)) = 0.$$

In particular, $\mathrm{wind}(c)$ is not interesting on B.

32.2. **Fundamental theorem of algebra.** Winding numbers can be used to prove the fundamental theorem of algebra.

Theorem 32.4 (Fundamental theorem of algebra)**.** *Let*

$$f = z^n + a_1 z^{n-1} + \cdots + a_n \in \mathrm{Poly}(\mathbb{C})$$

with $n \geq 1$. Then f has a root, i.e. $f(\alpha) = 0$ for some $\alpha \in \mathbb{C}$.

Remark. Recall that this was crucial last semester for finding eigenvalues.

Observations/intuition: For z large, $f(z) \approx z^n$. Another way of saying this is that given $r > 0$, there exists $R > 0$ so that $|z| \geq R$ and $|f(z)| \geq r > 0$. In particular, if $|z| \geq R$ then $f(z) \neq 0$. For $p(z) = z^n$, consider the circle $c_{r,1}$. Then

$$(R\cos 2\pi t + iR\sin 2\pi t)^n = R^n \cos 2\pi nt + iR^n \sin 2\pi nt$$

(which can be seen by writing the circle as $Re^{i2\pi t}$). Thus $p(z)$ makes the circle wind around n times. Then if $f \approx p$, at least at large z, we expect a similar behavior for f.
Main idea. By contradiction, suppose f has no root. Consider $f \circ c_{r,1}$, with $r \gg 0$. Can 0 escape from the maze? That is, does this curve have winding number 0? By the observations we made earlier, we don't expect this to be the case. In fact, we said that that $f(z) \approx z^n$ for z large and therefore $f \circ c_{r,1} \approx p \circ c_{r,1} = c_{r^n,n}$, so that we would expect

$$\mathrm{wind}(f \circ c_{r,1}) = \mathrm{wind}(c_{r^n,n}) = n.$$

On the other hand, suppose that $\partial d = c$. Then

$$f \circ c_{r,1} = \partial(f \circ d)$$

and therefore $\mathrm{wind}(f \circ c_{r,1}) = 0$. What gives?

Proof of Theorem 32.4. We will explain why $\mathrm{wind}(f \circ c_{r,1}) = n$ for $r \gg 0$.
Claim. There exists $r \gg 0$ and $b : [0,1]^2 \to A$ such that $\partial b = c_{r^n,n} - f \circ c_{r,1}$. Consider the "straight line interpolation"

$$b(s,t) = s \cdot c_{r^n,n}(t) + (1-s) f \circ c_{r,1}(t).$$

Why is the image of b never 0? (If it were, the image of b would not be in A, which we require.) Weite $z_t = c_{r,1}(t)$. Note that $c_{r^n,n}(t) = z_t^n$ and so

$$\begin{aligned} b(s,t) &= s z_t^n + (1-s)\left[z_t^n + a_1 z_t^{n-1} + \cdots + a_n\right] \\ &= z^n + (1-s)\left[a_1 z_t^{n-1} + \cdots + a_n\right] \end{aligned}$$

which is nonzero for r large enough. ∎

33.1. **Manifolds and Stokes' Theorem.** For now we limited ourselves to integration on chains when discussing Stokes' theorem. However, it can be extended to manifolds. To do so, we need to introduce the concept of a manifold with boundary.
Informal definition. A k**-dimensional manifold with boundary in** $\mathbb{R}^n$ is a subset $M \subset \mathbb{R}^n$ such that locally M either looks like $\mathbb{R}^k$ or $\mathbb{R}^{k-1} \times [0, \infty]$. Namely, for all $x \in M$ there exists an open neighborhood $x \in U \subset \mathbb{R}^n$ such that $U \cap M$ is eiter diffeomorphic to $\mathbb{R}^k$ or $\mathbb{R}^{k-1} \times [0, \infty]$. The points of the second type are called the **boundary of** M, and are denoted ∂M.

Example 33.1. A disk in $\mathbb{R}^2$ is a 2-manifold with boundary, and the boundary is the circle. A line segment is a 1-manifold with boundary two points.

Example 33.2. Consider the set

$$A = \left\{ x \in \mathbb{R}^2 : r_1 \leq |x| \leq r_2 \right\}.$$

This is a 2-manifold with boundary

$$\partial A = \{|x| = r_1\} \cup \{|x| = r_2\}.$$

However, the set

$$A' = \left\{ x \in \mathbb{R}^2 : r_1 \leq |x| < r_2 \right\}$$

is a 2-manifold with boundary

$$\partial A' = \{|x| = r_1\}.$$

Remark. We now have three different notions of boundary: on subsets of $\mathbb{R}^n$, on k-chains, and on manifolds.

Having formulated such definitions, we can now state Stokes' theorem for manifolds:

Theorem 33.3 (Stokes' theorem for manifolds)**.** *If $M \subset \mathbb{R}^n$ is a k-manifold and $d\eta$ is a k-form, then*

$$\int_{\partial M} \eta = \int_M d\eta.$$

We don't really know how to integrate forms on manifolds, so for now we will use the following working definition:
Working definition. Suppose $M = \mathrm{Im}(c)$ for some $c : [0,1]^k \to \mathbb{R}^n$ (as a parametrization). Then for $\omega \in \Omega^k(\mathbb{R}^n)$ we define

$$\int_M \omega = \int_c \omega.$$

33.2. **Green's theorem.** This is one of the main applications of Stokes' theorem:

Theorem 33.4 (Green's theorem)**.** *Let $M \subset \mathbb{R}^2$ be a compact 2-manifold with boundary, and let*

$$F = (P, Q) : \mathbb{R}^2 \to \mathbb{R}^2$$

be a smooth vector field. Then

$$\int_{\partial M} P dx + Q dy = \int_M \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx \wedge dy.$$

33.2.1. *Application.* We can apply Green's theorem to calculating areas in $\mathbb{R}^2$. In fact, the area of M is

$$\operatorname{area}(M) = \int_M 1 dx \wedge dy.$$

We can compute this using Green's theorem as soon as we find (P, Q) such that

$$\frac{\partial P}{\partial x} - \frac{\partial Q}{\partial y} = 1.$$

Many fields work, such as for example

$$\begin{aligned} P &= 0 & \qquad\qquad P &= -\frac{y}{2} \\ Q &= x & Q &= \frac{x}{2}. \end{aligned}$$

Thus for example

$$\operatorname{area}(M) = \frac{1}{2} \int_{\partial M} -y dx + x dy.$$

Example 33.5. Consider M being the ellipse

$$M = \left\{ (x, y) : \frac{x^2}{a^2} + \frac{y^2}{b^2} \le 1 \right\}$$

with boundary

$$\partial M = \left\{ (x, y) : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \right\}.$$

We can parametrize ∂M as

$$c(t) = (a \cos t, b \sin t).$$

Thus

$$\begin{aligned}
\operatorname{area}(M) =& \frac{1}{2}\int_c -y dx + x dy \\
=& \frac{1}{2}\int_0^{2\pi} c^*(y dx) + c^*(x dy) \\
=& \frac{1}{2}\int_0^{2\pi} (-b\sin t)(-a\sin t)dt + (a\cos t)(b\sin t)dt \\
=& \frac{1}{2}(ab)\int_0^{2\pi} 1 dt \\
=& \pi ab.
\end{aligned}$$

Setting $a = b = r$ we recover the area of the circle.

33.3. **Divergence theorem.** The divergence theorem is a 3-dimensional version of Green's theorem. The sample problem is to compute the surface area of a sphere of radius r.

Definition 33.6. For a 2-cube $c : [0,1]^2 \to \mathbb{R}^3$ the **unit normal vector** is

$$n(s,t) = \frac{D_1c(s,t) \times D_2c(s,t)}{\|D_1c(s,t) \times D_2c(s,t)\|}$$

Theorem 33.7 (Divergence theorem)**.** *Let $M \subset \mathbb{R}^3$ be a compact 3-manifold with boundary with normal vector n on ∂M. Let*

$$F = (F_1, F_2, F_3) : \mathbb{R}^3 \to \mathbb{R}^3$$

be a smoot vector field. Then

$$\int_M \operatorname{div} F dV = \int_{\partial M} \langle F, n\rangle dA,$$

where

$$\begin{aligned}
\operatorname{div} F =& \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} \\
dV =& dx \wedge dy \wedge dz \text{ (“volume element”)}
\end{aligned}$$

and dA is a 2-form on ∂M called the “area element” such that for $p \in \partial M$ and $u, v \in T_p\partial M$ we have

$$dA(p)(u,v) = \langle u \times v, n(p)\rangle = \det(u, v, n(p))$$

Example 33.8. If $c : [0,1]^2 \to \mathbb{R}^3$ parametrizes ∂M then $c^*(dA)$ is given by

$$c^*(dA) = |D_1c(s,t) \times D_2c(s,t)| ds \wedge dt.$$

Proof. Assume that $M = \operatorname{Im}(b)$ with $b : [0,1]^3 \to \mathbb{R}^3$ and $\partial M = \operatorname{Im}(c)$ with $c : [0,1]^2 \to \mathbb{R}^3$. We define

$$\omega = F_1 dy \wedge dz + F_2 dz \wedge dx + F_3 dx \wedge dy$$

so that

$$d\omega = \operatorname{div} F dx \wedge dy \wedge dz.$$

By Stokes' theorem

$$\int_M \operatorname{div} F dV = \int_M d\omega = \int_{\partial M} \omega$$

and thus we want to show that

$$\int_{\partial M} \omega = \int_{\partial M} \langle F, n \rangle dA.$$

To do so we will show that

$$c^* \omega = c^* (\langle F, n \rangle dA).$$

We can rewrite the right hand side as

$$\langle F(c(s,t)), n(s,t) \rangle |D_1 c(s,t) \times D_2 c(s,t)| ds \wedge dt$$

and for the left hand side we know that

$$c^* \omega = \varphi ds \wedge dt$$

for some φ. We want to evaluate the above at the basis vectors e_1, e_2:

$$\begin{aligned}
(c^*\omega)(s,t)(e_1, e_2) =& \omega(c(s,t)) (D_1 c(s,t), D_2 c(s,t)) \\
=& \langle F(c(s,t)), (D_1 c(s,t), D_2 c(s,t)) \\
=& \langle F(c(s,t)), n(s,t) \rangle |D_1 c(s,t) \times D_2 c(s,t)|.
\end{aligned}$$

□

33.3.1. *Applications.* As we said earlier, we can apply the above to computing the areao of a sphere. For $F = (x, y, z)$ we see $\operatorname{div} F = 3$ and thus

$$\begin{aligned}
3v(B) =& \int_B \operatorname{div} F dV \\
=& \int_S \langle F, n \rangle dA \\
=& \int_S r dA \\
=& r \cdot v(S)
\end{aligned}$$

and thus the volume is equal to

$$\frac{4}{3} \pi r^3.$$

34. April 25–Last Class!

34.1. **States by area.** How can you order the states of Vermont, New Hampshire, and Massachussetts by area? Today we are going to develop some tools for this. How are we to approach this problem, given what we know? One option is to use Green's theorem. Imagine that our state is given by S, with boundary $C = \partial S$. Then Green's theorem says that

$$\text{area}(S) = \int_S dx \wedge dy = \int_C x dy.$$

It is not necessay to use the xdy as our 1-form; any form η such that $d\eta = dx \wedge dy$ works. This approach is probably hopeless, since we need to parametrize the boundary of the state with a curve. What else can we try? We could try Archimedes' method of exhaustion, but that would not be ideal either–it would be exhausting. We can cut the paper and weight it. We could also try Fubini's theorem, but this too would bee a nightmare. Let's try to rectify the situation.

34.2. **Rolling wheels and sweeping areas.** We will introduce a physical tool to compute areas, called the **planimeter**. We will start with a simple version, consisting of a rod and a wheel in its middle. We can use the rod to sweep areas in the plane, either vertically, horizontally (in which case it does not sweep any area) or in between. A first observation is that the area swept by the planimeter is governed by the rolling of the wheel (that is, by how much the wheel rolls or slides). We thus introduce a vector $n(t)$ which we define as the unit vector normal to the rod. We also consider the vector $\gamma'(t)$, which is the derivative of the path $\gamma(t)$ of the wheel. Therefore the total rolling of the wheel is given by

$$\int_a^b \langle \gamma'(t), n(t) \rangle dt.$$

How about the area? For an infinitesimal distance, the area of a parallelogram is equal to

$$\text{area} = \ell \cdot |\gamma'(t)| \cdot \cos\theta = \ell \langle \gamma'(t), n(t) \rangle$$

where ℓ is the length of the rod. Therefore the total area is given by

$$\int_a^b \ell \langle \gamma'(t), n(t) \rangle dt.$$

This area can be negative (we are talking about signed area).

34.3. **The planimeter.** An actual planimeter is a bit more complex. We have a base which is fixed, as well as two arms: the *tracer arm*, where the wheel is, and the *polar arm*. The amazing fact about this is that the total rolling of the wheel as the tracer arm traverses $C = \partial S$ is proportional to the area of S. An idea of the proof is to use Green's theorem.
Proof idea. We define a vector field $n = (n_1, n_2)$ where n is the normal vector to the end of the tracer arm. We define the 1-form $\omega = n_1 dx + n_2 dy$. If $C : [0,1] \to \mathbb{R}^2$ is the boundary of S we see that

$$\begin{aligned} C^*\omega(t) =& n_1(C(t))C_1'(t)dt + n_2(C(t))C_2'(t)dt \\ =& \langle n(C(t)), C'(t) \rangle. \end{aligned}$$

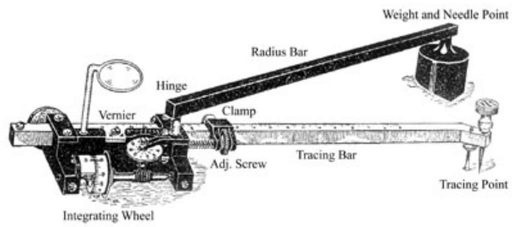

Figure 2. A planimeter

Moreover,

$$d\omega = K dx \wedge dy$$

for some constant K. (A way of seeing this is by imagining the planimeter going aroud a circle; the closer the tracer is to the origin, the more outward the normal vector will point.) By Green's theorem

$$K \cdot \text{area}(S) = \int_S d\omega = \int_C \omega = \int_0^1 \langle n(C(t)), C'(t) \rangle$$

which is equal to the rolling of the wheel.

Index