21-235 Recitations

Thomas Lam

2022 Fall

Contents

1 Repeated Integration I

1.1 The Lebesgue Measure and Measurability

The Story: Mathematicians wanted a way to rigorously define the intuitive notion of "area". In a perfect world, we'd hope that "area" satisfies a number of properties, such as:

1. (Universality) All sets have an "area".
2. (Additivity) If E, F area disjoint, then the area of $E \cup F$ is the sum of the areas of E and F.

Unfortunately, it turns out that there is no "perfect" way to define area. Every possible definition has a weird problem. For example, one could try using the *Lebesgue Outer Measure* to define area.

Definition 1.1 (Lebesgue Outer Measure)

The *Lebesgue Outer Measure*, denoted by $\mathcal{L}_o^N$, is defined by

$$\mathcal{L}_o^N(E) := \inf \left\{ \sum_{n=1}^{\infty} \operatorname{meas} R_n : R_n \text{ is a rectangle, and } \bigcup_{n=1}^{\infty} R_n \supseteq E \right\}$$

for *all* sets $E \subseteq \mathbb{R}^N$.

This comes really close to working. We get a lot of nice "area-like" properties, like $\mathcal{L}_o^N(\emptyset) = 0$ and $\mathcal{L}_o^N(E) \leq \mathcal{L}_o^N(F)$ whenever $E \subseteq F$. **Unfortunately, *additivity* breaks.** A horrific construction, called the *Vitali Set*, gives an example of disjoint $E, F \subseteq \mathbb{R}$ for which $\mathcal{L}_o^1(E \cup F) \neq \mathcal{L}_o^1(E) + \mathcal{L}_o^1(F)$.

This is really bad. If we don't have additivity then we really can't do much with "area". So **we sacrifice universality** in order to get additivity back. That is, we *restrict* $\mathcal{L}_o^N$ to only work on "nice enough" sets.

Definition 1.2 (Lebesgue Measure)

The *Lebesgue Measure*, denoted by $\mathcal{L}^N$, is the restriction of $\mathcal{L}_o^N$ to a family (...a *sigma-algebra*) of sets of $\mathbb{R}^N$ called *Lebesgue measurable sets*.

This way we can write things like $\mathcal{L}^N(E \cup F) = \mathcal{L}^N(E) + \mathcal{L}^N(F)$ (for disjoint E, F) with a peace of mind, because the sets we're feeding into $\mathcal{L}^N$ have to be measurable in order for this to work anyway.

But what does it mean for a set to be measurable?

Definition 1.3 (Measurable)

A set $E \subseteq \mathbb{R}^N$ is *Lebesgue measurable* if for all half-spaces A, we have that

$$\mathcal{L}_o^N(E) = \mathcal{L}_o^N(E \cap A) + \mathcal{L}_o^N(E \setminus A).$$

This is known as the *Catheodory Cutting Condition.*

This is kind of a mess to work with so here are some properties, as well as some "properties".

1.1.1 Something something outer regularity idk i forgot everything from 21-720 pls dont fire me

E is measurable iff for all $\varepsilon > 0$ there exists an open set $U \supseteq E$ such that $\mathcal{L}_o^N(U \setminus E) < \varepsilon$. (Search terms for the curious: "Outer regularity", "Radon Measure")

Of course, this immediately implies that all open sets are measurable. To wit...

1.1.2 Some Niceish Sets

- All open sets are measurable.
- Heck, all closed sets are measurable too.
- If you have a countable family of measurable sets, then their union is measurable. And their intersection.

Huh, a ton of things are measurable! In fact...

1.1.3 The "Officer I'm Not Trying To Break Math I Swear" Condition

Theorem 1.1 (not actually a theorem but it really should be tbh)

If you ever define a set $E \subseteq \mathbb{R}^N$ and you're *not* actively trying to break math, then E must be measurable.
More precisely, if you're defining E in an explicit, *constructive* manner (i.e. not non-constructively), then E has to be measurable.

Examples:

- Let $E = B(0,1) \subseteq \mathbb{R}^N$. I am not trying to break math. Therefore E is measurable.
- Let $E = \mathbb{Q}^N$. I am still not actively trying to break math. So E is measurable.
- Let
$$E = \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} (\log(m^n + \pi^e), ((m+n)!)!!] \subseteq \mathbb{R}.$$
Even here, I am not engaging in the sacrilegious breaking of math. (Specifically, I'm giving a very explicit construction of the set E. You can pick any $x \in \mathbb{R}$ and I can tell you with confidence whether or not x is in it.) Thus E is measurable.

What do I mean by breaking math / non-constructive shenanigans? It's the usage of the **Axiom of Choice**. As long as you're not using the Axiom of Choice, any set you define is guaranteed to be measurable. In fact, we have the following incredibly silly property:

If you reject the Axiom of Choice, then every set is measurable!

One last example: Let $\mathcal{B}(\mathbb{R}^N)$ be the *Borel Sigma-Algebra* on $\mathbb{R}^N$, which is basically the family of all sets you can make by applying countable union, countable intersection, and set difference operations on *open sets*. The sets in $\mathcal{B}(\mathbb{R}^N)$ are called *Borel sets*.

Unsurprisingly, all Borel sets are measurable.

(...though there exist measurable sets that are not Borel. See https://www.math3ma.com/blog/lebesgue-but-not-borel.)

1.2 Measurable Functions

Definition 1.4

A function $f : E \to \mathbb{R}$ is *measurable* if $\{x \in E : f(x) > a\}$ is measurable for all $a \in \mathbb{R}$.

Exercise: It turns out that the following is an equivalent condition: $f^{-1}(F)$ is measurable for all Borel sets $F \in \mathcal{B}(\mathbb{R})$. Can you imagine why that is?

I don't have a great intuitive explanation for why we take this definition, but hopefully you buy that measurable functions are functions that aren't horrifying. In fact:

Theorem 1.2 (not actually a theorem but it really should be tbh v2)

If you ever define a function $f : E \subseteq \mathbb{R}$ and you're *not* actively trying to break math, then f must be measurable.
More precisely, if you're defining f in an explicit, *constructive* manner (i.e. not non-constructively), then f has to be measurable.

Examples:

- All continuous functions on measurable domains are measurable.
- The slightly horrifying function

$$f(x) := \sum_{n=1}^{\infty} \frac{1}{2^n} \sin\left(e^{x-\pi\cdot\sqrt{n}} - 17 - 9001x\right) \cdot 1_{[-n,n]}$$

  is measurable.

1.3 Swapping Integrals: A Case Study

The story here is that we'd like to be able to swap integrals like

$$\int_E \int_F \cdots \, dx \, dy = \int_F \int_E \cdots \, dy \, dx.$$

But is this allowed? Not always!

The construction of this counter-example is inspired by the Grandi Series. That is,

$$1 - 1 + 1 - 1 + 1 - 1 + \ldots$$

is not a well-defined sum. Some middle-schoolers think its 0 because you can write it like $(1-1)+(1-1)+\ldots$. Other middle-schoolers think it's 1 because you can write it like $1+(-1+1)+(-1+1)+\ldots$. And that's exactly what we're going to exploit.

Let:

$$E := \bigcup_{m,n\in\mathbb{N}, m=n} (m-1, m) \times (n-1, n)$$

$$F := \bigcup_{m,n\in\mathbb{N}, m=n+1} (m-1, m) \times (n-1, n)$$

Then we take $f : \mathbb{R}^2 \to \mathbb{R}$ via:

$$f(x,y) = \begin{cases} 1, & x \in E \\ -1, & x \in F \\ 0, & \text{otherwise} \end{cases}$$

By the draw-a-picture theorem, we see that:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx = 0$$

Swapping integrals has failed! We will see that:

- Tonelli's Theorem indeed cannot be applied to this function f because it is sometimes negative, and
- Fubini's Theorem indeed cannot be applied either because f is not integrable.

1.4 Tonelli's Theorem

Usually when people swap integrals, they cite Fubini's Theorem. I think that Tonelli does not get enough love! It can be useful in some situations that Fubini cannot be used in. Really, all Tonelli needs is that your function is non-negative.

Theorem 1.3 (Tonelli)

Let $E \subseteq \mathbb{R}^N, F \subseteq \mathbb{R}^M$ be measurable. If $f : E \times F \to \mathbb{R}$ is measurable and **non-negative**, then

$$\int_E \int_F f(x,y)\,dy\,dx = \int_F \int_E f(x,y)\,dx\,dy = \int_{E\times F} f(x,y)d(x,y).$$

Example 1.1: Let $E = \{(x,y) \in \mathbb{R}^2 : 0 < x \le 1, 0 \le y < x+1\}$. Compute

$$\int_E xy\,d(x,y).$$

Solution. This solution will be *very* pedantic and rigorous, just so you can see what such rigorous arguments may look like. I'll wave my hands more after this.

The first thing to check before we do anything is that the integrand is measurable. If it isn't, all bets are off. Fortunately, it's continuous, and E is measurable because you can show that it's Borel (this is the one time I'll wave my hands here; this isn't hard just annoying) so we're chilling.

To properly apply Tonelli, we want the domain to be of the form $F \times G$. E isn't quite of this form. The prank here is to remember what $\int_E$ means...

$$\int_E xy\, d(x,y) = \int xy \cdot 1_E(x,y)\, d(x,y) = \int_{\mathbb{R}\times\mathbb{R}} xy \cdot 1_E(x,y)\, d(x,y)$$

Each of these equalities is essentially by definition. Now, since the integrand is non-negative everywhere, we can apply Tonelli!

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} xy \cdot 1_E(x,y)\, dy\ dx$$

Note that because of the shape of E, it's better to integrate in y first.

We now split the inner integral into two parts.

$$= \int_{\mathbb{R}} \int_{[0,x+1)} xy \cdot 1_E(x,y)\, dy + \int_{\mathbb{R}\setminus[0,x+1)} xy \cdot 1_E(x,y)\, dy\, dx$$

Since the integrand in the second integral is always 0 (because of the indicator function), we see that the second integral simply evaluates to 0.

$$= \int_{\mathbb{R}} \int_{[0,x+1)} xy \cdot 1_E(x,y)\, dy\, dx$$

Now we want to apply the Fundamental Theorem of Calculus (FTC). Unfortunately, the indicator function is kinda in the way now. At this point we just argue that for all (x,y) with $x \in \mathbb{R}$ and $0 \le y < x+1$, we have that $1_E(x,y) = 1_{(0,1]}(x)$. This lets us write

$$= \int_{\mathbb{R}} \int_{[0,x+1)} xy \cdot 1_{(0,1]}(x)\, dy\, dx,$$

and finally we may apply the FTC via $\frac{\partial}{\partial y}\frac{1}{2}x1_{(0,1]}(x)y^2 = x1_{(0,1]}(x)y$ to get

$$= \int_{\mathbb{R}} \frac{1}{2}x(x+1)^2 \cdot 1_{(0,1]}\, dx.$$

This integral is just

$$= \int_0^1 \frac{1}{2}x(x+1)^2\, dx$$

by definition, and we may apply the FTC once more to get... **** it who cares just apply Mathematica to obtain $\boxed{\dfrac{17}{24}}$. ■

Here's something more exciting.

Example 1.2 (Integral representation for log / Frullani Integral): Prove that

$$\log a = \int_0^\infty \frac{e^{-x} - e^{-ax}}{x}\, dx.$$

Proof. The first step is magical. We write:

$$\int_0^\infty \frac{e^{-x} - e^{-ax}}{x}\, dx = \int_0^\infty \int_1^a e^{-xy}\, dy\, dx$$

Since $e^{-xy} \geq 0$ for all x and y, we can swap using Tonelli:

$$= \int_1^a \int_0^\infty e^{-xy}\, dx\, dy$$

Now we just compute using the FTC:

$$= \int_1^a \frac{1}{y}\, dy = \log(a) - \log(1) = \log a$$

□

1.5 Fubini's Theorem

This is the famous big-shot theorem. It can be used on functions that take both positive and negative values. The condition Fubini needs is *integrability.*

Theorem 1.4 (Fubini)

Let $E \subseteq \mathbb{R}^N, F \subseteq \mathbb{R}^M$ be measurable. If $f : E \times F \to \mathbb{R}$ is measurable and is **integrable**, i.e.

$$\int_{E \times F} |f(x,y)| d(x,y) < +\infty$$

then

$$\int_E \int_F f(x,y)\, dy\, dx = \int_F \int_E f(x,y)\, dx\, dy = \int_{E \times F} f(x,y) d(x,y).$$

In practice, Fubini is slightly harder to use than Tonelli because "f is integrable" takes a bit more work than "lol $f \geq 0$ ez". That's why you should first check that your integrand is non-negative before going crazy with Fubini.

Let's start with something simple.

Example 1.3: Evaluate

$$\int_{(123,456) \times (789,101112)} e^x - e^y\, d(x,y).$$

Solution. I'm not actually going to evaluate this lol.

First we check if we can use Tonelli... Aw man, this function seems negative at several places.

Can we use Fubini? Indeed, we can argue as follows: Everything is measurable. The integrand is continuous, and the domain is bounded. Thus the integrand is bounded (why?) by some M. Since it's bounded on a domain of finite measure, we see that it is integrable because

$$\begin{aligned}\int_{(123,456)\times(789,101112)} |e^x - e^y| \, d(x,y) &\leq \int_{(123,456)\times(789,101112)} M \, d(x,y) \\ &= M\mathcal{L}^2((123,456)\times(789,101112)) < \infty.\end{aligned}$$

Thus we may apply Fubini to write

$$\int_{(123,456)\times(789,101112)} e^x - e^y \, d(x,y) = \int_{123}^{456} \int_{789}^{101112} e^x - e^x \, dy \, dx.$$

Now you stuff this into Mathematica. ■

Let's try something slightly more involved.

Example 1.4: Integrate:

$$\int_{[-1,1]^3} \frac{1}{\sqrt[3]{xyz}} \, d(x,y,z)$$

Solution. Everything is measurable, yay. The integrand is sometimes positive and sometimes negative, so Tonelli is out of the question. We must try and apply Fubini.

We need to prove that $\int_{[-1,1]^3} \left|\frac{1}{\sqrt[3]{xyz}}\right| \, d(x,y,z) < \infty$. How, you ask? Using Tonelli, of course! We have by Tonelli that

$$\begin{aligned}\int_{[-1,1]^3} \left|\frac{1}{\sqrt[3]{xyz}}\right| d(x,y,z) &= \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} \left|\frac{1}{\sqrt[3]{xyz}}\right| \, d(x,y,z) \\ &= \left(\int_{-1}^{1} \frac{1}{|\sqrt[3]{x}|} \, dx\right)\left(\int_{-1}^{1} \frac{1}{|\sqrt[3]{y}|} \, dy\right)\left(\int_{-1}^{1} \frac{1}{|\sqrt[3]{z}|} \, dz\right) < \infty\end{aligned}$$

Thus we are justified in applying Fubini:

$$\begin{aligned}\int_{[-1,1]^3} \frac{1}{\sqrt[3]{xyz}} \, d(x,y,z) &= \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} \frac{1}{\sqrt[3]{xyz}} \, d(x,y,z) \\ &= \left(\int_{-1}^{1} \frac{1}{\sqrt[3]{x}} \, dx\right)\left(\int_{-1}^{1} \frac{1}{\sqrt[3]{y}} \, dy\right)\left(\int_{-1}^{1} \frac{1}{\sqrt[3]{z}} \, dz\right) = 0\end{aligned}$$

■

2 Repeated Integration II

2.1 Change of Variables

Now is the time to be a *tad* bit more careful with the terminology "measurable". Although this is typically interpreted as *Lebesgue Measurable*, it's actually a bit vague.

Recall the definition of a Lebesgue Measurable (LM) function.

Definition 2.1 (Lebesgue Measurable)

A function $f : E \to \mathbb{R}$ is ***Lebesgue*** *measurable* if $\{x \in E : f(x) > a\}$ is Lebesgue measurable for all $a \in \mathbb{R}$.
(Equivalently, $f^{-1}(F)$ is Lebesgue measurable for all Borel F.)

The definition of a *Borel function* is essentially the same.

Definition 2.2 (Borel Measurable)

A function $f : E \to \mathbb{R}$ is ***Borel*** *(measurable)* if $\{x \in E : f(x) > a\}$ is Borel for all $a \in \mathbb{R}$.
(Equivalently, $f^{-1}(F)$ is Borel for all Borel F.)

Recall that a Borel set is a set that can essentially be written in terms of open sets. (Concretely, it is the *sigma-algebra* that is *generated* by the topology of open sets!)

Some notes:

- Since every Borel set is LM, we have that every Borel function is a LM function. (We do not have the converse :c)
- Every continuous $f : E \to \mathbb{R}$ is Borel, provided that E is Borel.

We are now ready to state the Change of Variables Theorem in N dimensions.

Theorem 2.1 (Change of Variables)

Let U be open and $g : U \to \mathbb{R}^N$ be a continuous function, such that:

1. ("Differomorphism a.e.") there is $F \subseteq U$ with $\mathcal{L}^N(U \setminus F) = \mathcal{L}^N(g(U \setminus F)) = 0$ such that g is differentiable over F, and
2. ("Injection a.e.") there is $G \subseteq U$ with $\mathcal{L}^N(g(U \setminus G)) = 0$ such that g is an injection over G.

Let $E \subseteq U$ be Borel, and let $f : g(E) \to \mathbb{R}$ be Borel such that either $f \geq 0$ or f is integrable. Then

$$\int_{g(E)} f(y)\, dy = \int_E f(g(x)) \cdot |\det J_g(x)|\, dx.$$

This Theorem is the "most applicable / general" one. But also it hurts my eyes to read. Here is a version that's less general but also a lot simpler.

Theorem 2.2 (Change of Variables: Simplified Version)

Let $f : E \to \mathbb{R}$ be Borel such that either $f \geq 0$ or f is integrable.
Let U be open and $g : U \to \mathbb{R}^N$ be a differentiable injection such that $E \subseteq g(U)$ (i.e. the range of g covers E).
Then

$$\int_E f(y)\, dy = \int_{g^{-1}(E)} f(g(x)) \cdot |\det J_g(x)|\, dx.$$

What's the intuition for Change of Variables? In particular, why do we have this crazy $|\det J_g(x)|$ factor (the Jacobian determinant)?

Well, you can't just change the variables and simplify things for free, that'd be too easy. There's gotta be a catch. Indeed, a change of variables does some weird morphing of the domain space. Some parts of the domain by expand or shrink under the change of variables. The expanding and shrinking behavior is captured by the Jacobian, naturally. Since the determinant essentially "measures how much a linear transformation expands or shrinks things", we have that the Jacobian determinant should, intuitively, measure how much a change of variables is expanding or shrinking space near a particular point. Multiplying by this factor "counteracts" this effect, in a sense.

Why do we need f Borel instead of just LM? That's because the composition of LM functions may not be LM. But if f is Borel and g is LM, then f^{-1} sends Borel sets to Borel sets and then g^{-1} sends *those* Borel sets to LM sets, so that the composition just sends Borel sets to LM sets. This chain is broken if we only had that f was LM!

2.2 Simple Example: Scaling

Let $f : E \to \mathbb{R}$ for $E \subseteq \mathbb{R}^N$. Let's assume for ease that f is integrable, though this won't be necessary for this example (and the reason. The goal here is to find a relationship between the quantities

$$\int_E f(x)\,dx \qquad \int_{rE} f(x/r)\,dx$$

for $r > 0$. Here, $rE = \{r\vec{x} : \vec{x} \in E\}$.

Of course, for $N = 1$, you probably know from u-substitution that

$$\int_a^b f(x)\,dx = r\int_{ra}^{rb} f(x/r)\,dx.$$

We claim that this generalizes to N dimensions. Naturally, we wish to change variables to transform $\int_E f(x)\,dx$ into something else. Naturally, we pick $g : \mathbb{R} \to \mathbb{R}$ with

$$g(x) = x/r.$$

This is indeed a differentiable injection whose range covers E. So we may apply the change of variables theorem to obtain

$$\int_E f(x)\,dx = \int_{g^{-1}(E)} f(g(x)) \cdot |\det J_g(x)|\,dx = \int_{rE} f(x/r) \cdot |\det J_g(x)|\,dx.$$

It remains to evaluate $\det J_g(x)$. To figure that out, it might be clearer to write g as

$$g(x_1, x_2, \cdots, x_N) = (x_1/r, x_2/r, \cdots, x_N/r).$$

Now it's clear that $\frac{\partial g_i}{\partial x_i}(x) = \frac{1}{r}$ for all i, and $\frac{\partial g_i}{\partial x_j}(x) = 0$ for all i, j with $i \neq j$. Thus $J_g(x)$ is just $\frac{1}{r}I_N$ where I_N is the $N \times N$ identity matrix. In particular, $|\det J_g(x)| = \frac{1}{|r^N|} = \frac{1}{r^N}$. Thus we have the relationship

$$\int_E f(x)\,dx = \frac{1}{r^N}\int_{rE} f(x/r)\,dx.$$

Yay!

Exercise: Can this result be extended to hold for all LM f?

2.3 Polar Coordinates: The 2D Case

Sometimes a function's value at a point (x, y) is better expressed in terms of "radius and angle" instead of "x-coordinate and y-coordinate", especially if we are integrating over $\mathbb{R}^2$ or the ball $B((0, 0), R)$.

(I'm not going to explain what polar coordinates are, if you're unfamiliar then you should probably look it up.)

Corollary 2.1 (Polar COV)

Let $0 < R \leq \infty$ and let $f : B((0,0), R) \to \mathbb{R}$ be Borel and either non-negative or integrable. Then

$$\int_{B((0,0),\mathbb{R}} f(x,y)\, d(x,y) = \int_0^{2\pi} \int_0^R r f(r\cos\theta, r\sin\theta)\, dr\, d\theta.$$

(Note that this works for $\int_{\mathbb{R}^2} f\, d(x,y)$ too, by taking "$R = +\infty$".)

Motto: "The price to pay in order to change to polar coordinates is a factor of r."

Proof. We are tempted to choose the change of variables $g : [0,R) \times [0, 2\pi) \to B((0,0), R)$ defined as

$$g(r,\theta) := (r\cos\theta, r\sin\theta).$$

However, this isn't actually injective because g maps $(0,\theta)$ to $(0,0)$ for all θ. Hence this doesn't satisfy the conditions of the simplified COV theorem (do note, though, that this actually satisfies the original COV theorem that I wrote, since g is an injection almost everywhere).

If this concerns you, the fix is pretty easy: Delete $(0,0)$ from the domain of integration, and restrict g to only take positive r so that $g : (0,R) \times [0,2\pi) \to B((0,0),R) \setminus \{(0,0)\}$. Tada! (EDIT: Actually this domain isn't open. To fix this, throw out 0 so thhat $g : (0,R) \times (0,2\pi) \to B((0,0),R) \setminus \{(0,0)\}$. This is ok because this throws out a set of measure zero. See next subsection.)

Now g is a differentiable injection and its range covers the domain over which we integrate f, so the simplified COV theorem applies.

What's the Jacobian determinant, i.e. the "price" we pay to make this change of variables? It's not too hard to compute:

$$\det J_g(r,\theta) = \begin{vmatrix} \frac{\partial g_1}{\partial r}(r,\theta) & \frac{\partial g_1}{\partial \theta}(r,\theta) \\ \frac{\partial g_2}{\partial r}(r,\theta) & \frac{\partial g_2}{\partial \theta}(r,\theta) \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial r} r\cos\theta & \frac{\partial}{\partial \theta} r\cos\theta \\ \frac{\partial}{\partial r} r\sin\theta & \frac{\partial}{\partial \theta} r\sin\theta \end{vmatrix}$$
$$= \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r(\cos^2\theta + \sin^2\theta) = r$$

Since $r > 0$ we simply have $|\det J_g(r,\theta)| = r$. This is the claimed Jacobian determinant.

(Also, Tonelli/Fubini was implicitly applied to write the RHS as two integrals instead of one.) $\square$

2.4 Quick Addendum: Yeeting Null Sets

Somehow I forgot to consider that measure theory is hard and that this was not something known, so very quickly I wanted to add this note.

In the previous section, I did some weird tricks in which I threw away sets of measure zero as if that was totally ok. Fortunately it *is* ok.

This is because if E_0 is such that $\mathcal{L}^N(E_0) = 0$, then $\int_{E_0} f(x)\,dx = 0$ no matter what f is (well, you should assume it's measurable, but other than that...).

So, if $E \subseteq \mathbb{R}^N$ is some LM set, and $E_0 \subseteq E$ is such that $\mathcal{L}^N(E_0) = 0$, then we can do this trick:

$$\int_E f(x)\,dx = \int_{E\setminus E_0} f(x)\,dx + \int_{E_0} f(x)\,dx = \int_{E\setminus E_0} f(x)\,dx$$

In other words, we can yeet sets of measure zero and you can't stop me.

Concrete example: Consider integrating

$$\int_{[0,1]^2} x^2 + y\,d(x,y)$$

for some reason. Let's say you're allergic to closed sets or something. Fortunately, $\partial[0,1]^2$ has Lebesgue measure zero, meaning you can yeet it to write

$$\int_{[0,1]^2} x^2 + y\,d(x,y) = \int_{(0,1)^2} x^2 + y\,d(x,y)$$

and nobody can stop you.

Ok addendum over.

2.5 Gaussian Integral

We now look at an application of polar integration.

Theorem 2.3

$$\int_{\mathbb{R}} e^{-x^2}\,dx = \sqrt{\pi}$$

Proof. The prank here is to instead evaluate $\left(\int_{\mathbb{R}} e^{-x^2}\,dx\right)^2$. We can manipulate this as follows:

$$\left(\int_{\mathbb{R}} e^{-x^2}\,dx\right)^2 = \left(\int_{\mathbb{R}} e^{-x^2}\,dx\right)\left(\int_{\mathbb{R}} e^{-x^2}\,dx\right)$$

$$= \left(\int_{\mathbb{R}} e^{-x^2}\, dx\right)\left(\int_{\mathbb{R}} e^{-y^2}\, dy\right)$$

$\left(\int_{\mathbb{R}} e^{-x^2}\, dx\right)$ is just a number, so we can shove it into the y-integral:

$$= \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-x^2}\, dx\right) e^{-y^2}\, dy\right)$$

For each y, we have that e^{-y^2} is just a number, so we can shove it into the x-integral:

$$= \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-x^2-y^2}\, dx\right)\, dy\right)$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-x^2-y^2}\, dx\, dy$$

Now we apply Tonelli:

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

We may now change variables to polar coordinates. We have $x^2 + y^2 = r^2$, so the integral just becomes:

$$= \int_0^{2\pi} \int_0^{\infty} re^{-r^2}\, dr\, d\theta$$

Remember that we get an extra factor of r when changing to polar... and this is exactly what makes this integral solvable! By changing variables again with $u = r^2$, this becomes

$$= \int_0^{2\pi} \frac{1}{2} \int_0^{\infty} e^{-u}\, du\, d\theta$$

$$= \int_0^{2\pi} \frac{1}{2}\, d\theta = \pi$$

This is the answer *squared*, so taking the square root of this gives the desired answer to the integral! $\square$

2.6 Measure of the N-Ball

Here is an application of the Gaussian integral and Tonelli.

Theorem 2.4

$$\mathcal{L}^N(B_N(0,1)) = \frac{\pi^{N/2}}{\Gamma\left(\frac{N}{2}+1\right)}$$

Proof. We're going to use a useful, general technique/prank here, so pay attention! We start with the Gaussian integral on $\mathbb{R}^N$. That is,

$$\int_{\mathbb{R}^N} e^{-\|x\|^2}\,dx = \prod_{i=1}^N \int_{\mathbb{R}} e^{-x_i^2}\,dx_i = \pi^{N/2}$$

by Tonelli applied a billion times.

Here comes to prank. Let's rewrite the integrand by introducing an indicator function and a second integral!

$$\pi^{N/2} = \int_{\mathbb{R}^N} e^{-\|x\|^2}\,dx = \int_{\mathbb{R}^N}\int_0^\infty 1_{[0,e^{-\|x\|^2}]}(t)\,dt\,dx$$

Make sure you see why this works. We now apply Tonelli to swap:

$$= \int_0^\infty \int_{\mathbb{R}^N} 1_{[0,e^{-\|x\|^2}]}(t)\,dx\,dt$$

To simplify this thing, we now want to rewrite the indicator function to be a function of x instead of t. To do this, we reason that for all $0 \le t \le \infty$, we have that $0 \le t \le e^{-\|x\|^2}$ holds iff $\log t \le -\|x\|^2$ iff $\|x\|^2 \le -\log t$ iff $\|x\| \le \sqrt{-\log t}$ iff $x \in B_N(0, \sqrt{-\log t})$, and in fact this is valid only for $t \in [0,1]$. Thus $1_{[0,e^{-\|x\|^2}]}(t) = 1_{B_N(0,\sqrt{-\log t})}(x) \cdot 1_{[0,1]}(t)$, so in addition to replacing the indicator, we also replace the $\int_0^\infty dt$ with $\int_0^1 dt$

$$\pi^{N/2} = \int_0^1 \int_{\mathbb{R}^N} 1_{B_N(0,\sqrt{-\log t})}(x)\,dx\,dt = \int_0^1 \mathcal{L}^N(B_N(0,\sqrt{-\log t}))\,dt$$

By a scaling argument, we know that $\mathcal{L}^N(B_N(0,\sqrt{-\log t})) = \left(\sqrt{-\log t}\right)^N \mathcal{L}^N(B_N(0,1))$. Thus

$$\pi^{N/2} = \mathcal{L}^N(B_N(0,1)) \int_0^1 (-\log t)^{N/2}\,dt.$$

Lastly we change variables with $s = -\log t$. Then $ds = (-1/t)\,dt$ i.e. $dt = -t\,ds = -e^{-s}\,ds$ to get

$$\pi^{N/2} = \mathcal{L}^N(B_N(0,1)) \int_\infty^0 -e^{-s}s^{N/2}\,ds = \mathcal{L}^N(B_N(0,1)) \int_0^\infty e^{-s}s^{N/2}\,ds = \Gamma\left(\frac{N}{2}+1\right).$$

Rearranging, we conclude that $\boxed{\mathcal{L}^N(B_N(0,1)) = \dfrac{\pi^{N/2}}{\Gamma\left(\frac{N}{2}+1\right)}}$. $\square$

3 Spherical Coordinates

We started with polar coordinates, given by the COV

$$g\left(\begin{bmatrix} r \\ \theta_1 \end{bmatrix}\right) = \begin{bmatrix} r\cos\theta_1 \\ r\sin\theta_1 \end{bmatrix}.$$

Spherical coordinates generalizes to N coordinates, with the COV

$$g\left(\begin{bmatrix} r \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{N-2} \\ \theta_{N-1} \end{bmatrix}\right) = \begin{bmatrix} r\cos\theta_1 \\ r\sin\theta_1\cos\theta_2 \\ r\sin\theta_1\sin\theta_2\cos\theta_3 \\ \vdots \\ r\sin\theta_1\sin\theta_2\sin\theta_3\dots\sin\theta_{N-2}\cos\theta_{N-1} \\ r\sin\theta_1\sin\theta_2\sin\theta_3\dots\sin\theta_{N-2}\sin\theta_{N-1} \end{bmatrix},$$

where we have $0 \le \theta_1, \theta_2, \cdots, \theta_{N-2} \le \pi$ and $0 \le \theta_{N-1} \le 2\pi$.

The way this works out is inductively (draw a picture for $N = 3$ to follow along!).

- θ_1 is the angle between x and the positive x_1 axis, so that $x_1 = r\cos\theta_1$. (Note that we have $0 \le \theta_1 \le \pi$; if you think about it, it makes no sense for two vectors to form an angle of $270°$...)
- Then, we remove/"zero-out" the x_1 component from x by projecting it unto the $N-1$-dimensional subspace formed by $x_2, x_3, \cdots, x_N$. This doesn't change the $x_2, x_3, \cdots, x_N$ coordinates, but it does change the magnitude from r to $r\sin\theta_1$.
- Then we keep going: The angle between this projected vector and x_2 we shall call θ_2, so that $x_2 = (\text{new magnitude})\cos\theta_2 = r\sin\theta_1\cos\theta_2$. (Again, $0 \le \theta_2 \le \pi$.)
- Now you project unto the $N-2$ subspace formed by $x_3, \cdots, x_N$ and etc. etc. etc.
- After enough projections, we project unto a 2D plane. After that, we just need one last angle θ_{N-1} to specify where we are in the plane, given that our projection has length $r\sin\theta_1\sin\theta_2\sin\theta_3\dots\sin\theta_{N-2}$. That's why the last two components of this coordinate system end in a $\cos\theta_{N-1}$ and $\sin\theta_{N-1}$, like in 2D polar coordinates. This is also why we don't need to restrict θ_{N-1} to $[0, \pi]$ like the other angles.

If you rough it out, the Jacobian determinant is given by

$$\det J_g(x) = r^{N-1}\sin^{N-2}\theta_1\sin^{N-3}\theta_2\dots\sin\theta_{N-2}.$$

Example 3.1 (Hemisphere Center of Mass): Compute

$$\frac{1}{\frac{2}{3}\pi}\int_E z\, d(x,y,z),$$

where $E = \{(x,y,z) \in B_3(0,1) : z > 0\}$ is the upper unit hemisphere.

Solution. We use spherical coordinates with the COV

$$g\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) \stackrel{?}{=} \begin{bmatrix} r\cos\theta \\ r\sin\theta\cos\varphi \\ r\sin\theta\sin\varphi \end{bmatrix}.$$

But the astute reader may observe that the condition $z > 0$ would translate to $r\sin\theta\sin\varphi > 0$, which is ugly to deal with, so let's switch the order around real quick so we don't have to deal with that:

$$g\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} r\sin\theta\sin\varphi \\ r\sin\theta\cos\varphi \\ r\cos\theta \end{bmatrix}.$$

Now $z > 0$ iff $r\cos\theta > 0$, which is much easier to deal with; in fact, this occurs iff $-\pi/2 < \theta < \pi/2$. This restriction, combined with the restrictions $0 < r < 1$, $0 \le \theta \le \pi$, and $0 \le \varphi \le 2\pi$, tells us what our new integral(s) look like:

$$\frac{1}{\frac{2}{3}\pi}\int_E z\, d(x,y,z) = \frac{1}{\frac{2}{3}\pi}\int_0^1 \int_0^{2\pi} \int_0^{\pi/2} r\cos\theta \cdot \left(r^2\sin\theta\right)\, d\theta\, d\varphi\, dr.$$

Remark 1: Note the Jacobian determinant of $r^2\sin\theta$.

Remark 2: This should make a lot of sense: To specify what angle you're facing, among the points on the surface of a hemisphere, you first specify a **pitch** ("up and down-ness"; imagine nodding "yes") within a range of 180 degrees, and then you specifiy a **yaw** ("left and right-ness"; imagine shaking your head ala "no") between 0 and 360 degrees!

Using Tonelli spam, we may evaluate these new integrals:

$$= \frac{1}{\frac{2}{3}\pi}\left(\int_0^1 r^3\, dr\right)\left(\int_0^{2\pi} 1\, d\varphi\right)\left(\int_0^{\pi/2} \sin\theta\cos\theta\, d\theta\right) = \frac{1}{\frac{2}{3}\pi}(1/4)(2\pi)(1/2) = \boxed{\frac{3}{8}}.$$

■

3.1 Integrals depending only on magnitude

At some point (perhaps in a certain advanced real analysis course?), you might become very interested in integrating functions that depend only on magnitude, such as

$$\int_{\mathbb{R}^3} \frac{1}{1+x^2+y^2+z^2}\, d(x,y,z).$$

The general form for such a problem would be $\int_{a\le\|x\|\le b} f(\|x\|)\,dx$, for $0 \le a \le b \le \infty$ and an integrable function $f : [0,\infty) \to \mathbb{R}$. Naturally, we'd like to tame this using spherical coordinates. We have:

$$\int_{a\le\|x\|\le b} f(\|x\|)\,dx$$

$$= \int_{[0,\pi]^{N-2}\times[0,2\pi]} \int_a^b f(r)\cdot r^{N-1} \sin^{N-2}\theta_1 \sin^{N-3}\theta_2 \dots \sin\theta_{N-2}\, dr\, d(\theta_1,\cdots,\theta_{N-1})$$

Ewww!!! Blegh! Let's rearrange this into:

$$= \int_a^b f(r)\cdot r^{N-1}\,dr \cdot \left(\int_{[0,\pi]^{N-2}\times[0,2\pi]} \sin^{N-2}\theta_1 \sin^{N-3}\theta_2 \dots \sin\theta_{N-2}\, d(\theta_1,\cdots,\theta_{N-1})\right)$$

The good news is that that second term does not depend on f! In fact, it depends only on N, so we may call this a constant α_N.

$$\boxed{\int_{a\le\|x\|\le b} f(\|x\|)\,dx = \alpha_N \int_a^b r^{N-1} f(r)\,dr} \qquad (*)$$

Even if we don't really know what α_N is, this is already really useful for ascertaining *convergence.*

Example 3.2: For what exponents $a \in \mathbb{R}$ does the integral

$$\int_{B_N(0,1)} \frac{1}{\|x\|^a}\,dx$$

converge? For what exponents $b \in \mathbb{R}$ does the integral

$$\int_{\mathbb{R}^N\setminus B_N(0,1)} \frac{1}{\|x\|^b}\,dx$$

converge?

Solution. Using our cute formula, we have that

$$\int_{B_N(0,1)} \frac{1}{\|x\|^a}\,dx = \alpha_N \int_0^1 \frac{r^{N-1}}{r^a}\,dr = \alpha_N \int_0^1 \frac{1}{r^{a-N+1}}\,dr,$$

and classically this converges exactly when $a - N + 1 < 1$, i.e. $a < N$.

Similarly, the integral $\int_{\mathbb{R}^N \setminus B_N(0,1)} \frac{1}{\|x\|^b}\, dx$ exactly when $b > N$. ■

But, let's say you were really, *really* curious about what α_N is. How would we go about finding it?

The key idea is that our formula $(*)$ works for any $0 \leq a \leq b \leq \infty$ and any f that we plug into it. Why don't we try taking $a = 0$, $b = 1$, and $f = 1$? Then we'd get

$$\int_{B_N(0,1)} 1\, dx = \alpha_N \int_0^1 r^{N-1}\, dr,$$

or

$$\mathcal{L}^N(B_N(0,1)) = \alpha_N \cdot \frac{1}{N}.$$

Thus $\boxed{\alpha_N = N \cdot \mathcal{L}^N(B_N(0,1))}$.

4 Cantor Ruins the Day

If a measurable set (in the Lebesgue sense) has positive measure, then it must be uncountable (why?). But the converse is not true! There is a set that is uncountable and yet has zero measure, called the *Cantor Set*.

4.1 Ternary Construction

Definition 4.1

The Cantor set $\mathcal{C}$ is the set of all real numbers in $[0,1]$ that has a ternary representation using only 0's and 2's.

- I write "a" ternary representation because some numbers have multiple! (Think about how $0.\overline{9} = 1$...)
- $2/3 \in \mathcal{C}$ because $2/3 = 0.2_3$ has no 1's.
- $1/3 \in \mathcal{C}$ because $1/3 = 0.1_3 = 0.02222\ldots$ has no 1's.
- $1/2 \notin \mathcal{C}$. This is because $1/3 < 1/2 < 2/3$ which forces the first digit of $1/2$ to be a 1.

In general, suppose that the first $n-1$ digits of x are known to be $d_1, d_2, \cdots, d_{n-1}$. Then x is somewhere in the range $\left[0.d_1d_2\ldots d_{n-1}, 0.d_1d_2\ldots d_{m-1} + \frac{1}{3^{n-1}}\right]$. Moreover, $d_n = 1$ is *forced* exactly when

$$x \in \left(0.d_1d_2\ldots d_{m-1} + \frac{1}{3^n}, 0.d_1d_2\ldots d_{m-1} + \frac{2}{3^n}\right).$$

Taking the union over all possible digits $d_1, \cdots, d_{n-1}$, we see that for $x \in [0,1]$, its nth digit in every ternary representation of x is forced to be 1 exactly when

$$x \in \bigcup_{d_1, d_2, \cdots, d_{n-1}} \left(0.d_1d_2\ldots d_{n-1} + \frac{1}{3^n}, 0.d_1d_2\ldots d_{n-1} + \frac{2}{3^n}\right)$$

$$= \bigcup_{k=0}^{3^{n-1}-1} \left(\frac{k}{3^{n-1}} + \frac{1}{3^n}, \frac{k}{3^{n-1}} + \frac{2}{3^n}\right)$$

Taking the union over all n, we see that *some* digit of x is forced to be 1 (i.e. $x \in [0,1] \setminus \mathcal{C}$) exactly when

$$x \in \bigcup_{n=1}^{\infty} \bigcup_{k=0}^{3^{n-1}-1} \left(\frac{k}{3^{n-1}} + \frac{1}{3^n}, \frac{k}{3^{n-1}} + \frac{2}{3^n}\right).$$

It follows that

$$\boxed{\mathcal{C} = [0,1] \setminus \bigcup_{n=1}^{\infty} \bigcup_{k=0}^{3^{n-1}-1} \left(\frac{k}{3^{n-1}} + \frac{1}{3^n}, \frac{k}{3^{n-1}} + \frac{2}{3^n} \right)}.$$

4.2 "Middle Thirds" Construction

(Just google this lol)

4.3 Properties

1. $\mathcal{C}$ is compact. (It is closed by the explicit construction.)
2. $\mathcal{C}$ is Borel (...thus it is Lebesgue measurable).
3. $\mathcal{C}^\circ = \emptyset$
4. $\operatorname{acc} \mathcal{C} = \mathcal{C}$. (Sketch: Take $x \in \mathcal{C}$. Flip some digit from a 0 to 2 or vice versa. If the digit is "sufficiently far" then the change is $< \varepsilon$.)
5. $\mathcal{C}$ is self-similar, and is thus a fractal. In particular:

$$\mathcal{C}/3 \subseteq \mathcal{C}$$

$$\frac{2}{3} + \mathcal{C}/3 \subseteq \mathcal{C}$$

$$\mathcal{C} = \left(\frac{\mathcal{C}}{3}\right) \cup \left(\frac{2}{3} + \frac{\mathcal{C}}{3}\right)$$

   (Motto: "x has no 1's iff the first digit is either 0 or 2, and the rest of the digits have no 1's")

6. $\mathcal{L}^1(\mathcal{C}) = 0$

   Proof. From the previous property, we have that

$$\mathcal{L}^1(\mathcal{C}) = \mathcal{L}^1(\mathcal{C})/3 + \mathcal{L}^1(\mathcal{C})/3.$$

   The only solutions to this are $\mathcal{L}^1(\mathcal{C}) = 0$ and $\mathcal{L}^1(\mathcal{C}) = +\infty$. We can easily toss the latter. □

7. $\mathcal{C}$ is uncountably infinite. (There are uncountably many ternary strings with only 0's and 2's!)

4.4 The Devil's Staircase

Also called the "Cantor Function", but it doesn't sound as cool.

Definition 4.2 (Devil's Staircase)

Consider the function $f : [0,1] \setminus \mathcal{C} \to [0,1]$ defined as follows: For $x \in [0,1] \setminus \mathcal{C}$, let d_n be the first digit that is 1, so that $1, 2, \cdots, d_{n-1} \in \{0,2\}$ and $d_n = 1$. Then:

$$f(x) := \sum_{k=1}^{n-1} \frac{d'_k}{2^k} + \frac{1}{2^n},$$

where

$$d'_k = \begin{cases} 0, & d_k = 0 \\ 1, & d_k = 2 \end{cases}. \qquad (*)$$

Essentially, we replace every 2 with a 1, and then read the first n digits as binary. For example,

$$(0.0200202021020020\ldots)_3 \text{ maps to}$$
$$(0.0100101011)_2.$$

It can be shown that f is increasing and continuous (and also, if one writes $[0,1] \setminus \mathcal{C}$ as a countable union of disjoint open intervals, then f is constant on each interval!). It follows that f has a unique continuous extension to all of $[0,1]$, and we call this the *Devil's Staircase*. An explicit formula for $f(x)$ when $x \in \mathcal{C}$ is given as

$$f(x = (0.d_1 d_2 d_3 d_4 \ldots)_3) := (0.d'_1 d'_2 d'_3 d'_4 \ldots)_3,$$

where d'_k is defined as in $(*)$.

Some properties:

1. f is continuous and increasing.
2. f is constant over every $\left(\frac{k}{3^{n-1}} + \frac{1}{3^n}, \frac{k}{3^{n-1}} + \frac{2}{3^n}\right)$.
3. $f'(x) = 0$ for almost every $x \in [0,1]$.
4. As a corollary,
$$1 = f(1) - f(0) \neq \int_0^1 f'(x)\, dx = 0.$$
Thus the Fundamental Theorem of Calculus has been broken. The necessary (and sufficient!) fix is to enforce that f is not only continuous, but *absolutely continuous*. At some point, lecture will focus a lot on absolute continuity, and it's very important!

5. (For the probability nerds) Let X be a random variable whose CDF is given by f. Then the CDF of X is continuous but X is not a continuous random variable, because X does not admit a density.

6. (For the time travelers) $f \in C^{0,\log_2 3}([0,1])$ (i.e. f is *Hölder continuous* with exponent $\log_2 3$). It follows that f is uniformly continuous. At some point we will probably talk about what all this means in recitation.

5 We Stan Domination

Theorem 5.1 (Lebesgue Domination)

Suppose $f_n : E \to \mathbb{R}$ is such that $f_n \to f$ almost everywhere, and moreover $|f_n| \leq g$ for $g \in L^1(E)$ (i.e. $\int_E g\,dx < \infty$). Then

$$\lim_{n\to\infty} \int_E f_n\,dx = \int_E \lim_{n\to\infty} f_n\,dx.$$

This is the most important theorem in all of real analysis.

5.1 Some Example Applications

Example 5.1: Compute the limit

$$\lim_{n\to\infty} \int_0^2 (2x - x^2)^n\,dx.$$

Solution. Let $f_n(x) = (2x - x^2)^n\,dx$. We recognize that $0 < 2x - x^2 < 1$ for all $x \in (0, 1)$, thus

$$\lim_{n\to\infty} f_n(x) = 0$$

for all $x \in (0, 2)$. Thus $f_n \to f$ pointwise, where $f \equiv 0$. We now must show that the functions $\{f_n\}_n$ are dominated by some integrable function g. Indeed, we can just take $g(x) = 1$, which is integrable over $(0, 2)$. So we may swap the limit and integral to get

$$\lim_{n\to\infty} \int_0^2 (2x - x^2)^n\,dx = \int_0^2 \lim_{n\to\infty} (2x - x^2)^n\,dx = \int_0^2 0\,dx = 0.$$

■

Example 5.2: Compute the limit

$$\lim_{n\to\infty} n \int_0^1 \frac{\cos(x)}{1 + n^2x^2}.$$

Solution. We first change variables with $u = nx$ to rewrite the integral as

$$n \int_0^1 \frac{\cos(x)}{1 + n^2x^2} = \int_0^n \frac{\cos(u/n)}{1 + u^2}\,du = \int_0^\infty \frac{\cos(u/n)}{1 + u^2} \cdot 1_{[0,n]}(u)\,du.$$

Letting $f_n(u) := \frac{\cos(u/n)}{1+u^2} \cdot 1_{[0,n](u)}$, it is clear that $f_n \to f$ pointwise, where $f(u) := \frac{1}{1+u^2}$.

Moreover, the $\{f_n\}_n$ are dominated by $g(u) := f(u)$. It follows that

$$\lim_{n\to\infty}\int_0^\infty \frac{\cos(u/n)}{1+u^2}\cdot 1_{[0,n]}(u)\,du = \int_0^\infty \frac{1}{1+u^2}\,du = \boxed{\frac{\pi}{2}}.$$

■

5.2 Regularity of the Gamma function

Recall the definition of the Gamma function:

$$\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}\,dt \qquad \forall x > 0$$

Theorem 5.2

Γ is continuous.

Proof. Note: This example ended up being more convoluted than I thought because there is some fussiness involved with $t > 1$ vs. $t < 1$...

Fix $x_0 > 0$. We would like to show that $\lim_{x\to 0}\Gamma(x) = \Gamma(x_0)$. It suffices to prove that for any $x_n \to x_0$ we have $\lim_{n\to\infty}\Gamma(x_n) = \Gamma(x_0)$. That is,

$$\lim_{n\to\infty}\int_0^\infty t^{x_n-1}e^{-t}\,dt \stackrel{?}{=} \int_0^\infty t^{x_0-1}e^{-t}\,dt.$$

We want to try and apply dominated convergence! Let $f_n(t) := t^{x_n-1}e^{-t}$ and $f(t) := t^{x_0-1}e^{-t}$. It is clear that $f_n \to f$ pointwise in $(0,\infty)$. We now just need to dominate $\{f_n\}_n$.

Since $x_n \to x_0$, we have in particular that $\{x_n\}_n$ is bounded from above by some $M > 0$ and bounded from below by some $\delta > 0$ (why?). We claim that the function

$$g(t) := \begin{cases} t^{M-1}e^{-t}, & t \geq 1 \\ t^{\delta-1}e^{-t}, & 0 < t < 1 \end{cases}$$

dominates!

By taking cases on whether whether $t \geq 1$ or $t < 1$, we can show that $|f_n| \leq g$ for all n (note that $x \mapsto t^x$ is *decreasing* if $t < 1$, and *increasing* if $t > 1$), so we need only show that g is integrable.

- On one hand, we have that

  $$\int_0^1 t^{\delta-1}e^{-t}\,dt \leq \int_0^1 t^{\delta-1}\,dt < \infty$$

  because $\delta - 1 > -1$.

- On the other hand, you can show that

  $$\int_1^\infty t^{M-1}e^{-t}\,dt < \infty,$$

  which should be believable since this is less than $\Gamma(M)$, which should be finite. If you want an actual proof, try using the fact that $t^{M-1}e^{-t} \le e^{-t/2}$ for all t large enough (say, $t \ge T$).

This concludes the proof of continuity! □

Next let's prove that the Gamma function is differentiable.

Theorem 5.3

Γ is differentiable, with

$$\Gamma'(x) = \int_0^\infty \log(t)t^{x-1}e^{-t}\,dt.$$

Proof. Take $x_0 > 0$. We wish to prove that the limit

$$\lim_{x\to x_0}\frac{\Gamma(x)-\Gamma(x_0)}{x-x_0} = \int_0^\infty \frac{t^{x-1}-t^{x_0-1}}{x-x_0}e^{-t}\,dt$$

exists, and is equal to the claimed derivative.

It suffices to prove that

$$\lim_{n\to\infty}\int_0^\infty \frac{t^{x_n-1}-t^{x_0-1}}{x_n-x_0}e^{-t}\,dt = \int_0^\infty \log(t)t^{x_0-1}e^{-t}\,dt$$

for an arbitrary sequence $x_n \to x_0$.

It's clear by definition of derivative that we have convergence of the integrands pointwise. So, we just need to dominate the integrands.

By the MVT (applied to the function $x \mapsto t^{x-1}$, whose derivative is $\log(t)t^{x-1}$) there exists c_n between x_n and x_0 such that

$$\frac{t^{x_n-1}-t^{x_0-1}}{x-x_0}e^{-t} = \log(t)t^{c_n-1}e^{-t}.$$

Since $x_n \to x_0$, we must have $c_n \to x_0$, so in particular $\{c_n\}$ is bounded so that we can obtain bounds $\delta \le c_n \le M$ for some $M, \delta > 0$. It follows that when $t \ge 1$ we have the upper bound

$$|\log(t)t^{c_n-1}e^{-t}| \le \log(t)t^{M-1}e^{-t},$$

and when $0 < t < 1$ we have the upper bound

$$|\log(t)t^{c_n-1}e^{-t}| \leq |\log(t)|t^{\delta-1}e^{-t}.$$

(Remember that we have absolute values because we want to write $|f_n| \leq g$!)

We have finally arrived at an upper bound

$$\left|\frac{t^{x_n-1}-t^{x_0-1}}{x_n-x_0}e^{-t}\right| = |\log(t)t^{c_n-1}e^{-t}| \leq g(t) := \begin{cases} |\log(t)|t^{M-1}e^{-t}, & t \geq 1 \\ |\log(t)|t^{\delta-1}e^{-t}, & 0 < t < 1 \end{cases}$$

that is independent of n, and it remains to show that this is integrable.

Similarly to the previous proof, we can abuse the asymptotics pertaining to the fact that e^{-t} decays much faster that t^k and $\log t$ to show that $\int_1^\infty |\log(t)|t^{M-1}e^{-t}\,dt < \infty$. Then, we may make the bound $\int_0^1 |\log(t)|t^{\delta-1}e^{-t}\,dt \leq \int_0^1 |\log(t)|t^{\delta-1}\,dt$, and then, perhaps by integration by parts, we may show that this is finite. □

5.3 wtf

Example 5.3: Evaluate the integral

$$\int_0^1 \frac{x^{17}-1}{\log x}\,dx.$$

Solution. Define the function $I : \mathbb{R}^+ \to \mathbb{R}$ via

$$I(t) := \int_0^1 \frac{x^t-1}{\log x}\,dx.$$

We claim that $I(t)$ is differentiable for all $t > 0$. To see this, let $t_0 > 0$ and let $h_n \to 0$ be an arbitrary sequence. Then

$$\frac{I(t_0+h_n)-I(t_0)}{h_n} = \int_0^1 \frac{x^{t_0}}{\log x}\cdot\frac{x^{h_n}-1}{h_n}\,dx.$$

By the MVT, there exists c_n between h_n and 0 for which

$$\int_0^1 \frac{x^{t_0}}{\log x}\cdot(\log(x)x^{c_n})\,dx = \int_0^1 x^{t_0+c_n}\,dx,$$

and from here it is not hard to find a dominating function for the integrand (say, $g(x) = x^{t_0+\inf_n c_n}$. Or... even $g(x) = 1$ if you squint). So by domination it follows that

$$I'(t_0) = \lim_{n\to\infty}\frac{I(t_0+h_n)-I(t_0)}{h_n} = \int_0^1 \lim_{n\to\infty} x^{t_0+c_n}\,dx = \int_0^1 x^{t_0}\,dx = \frac{1}{1+t_0}.$$

As t_0 was arbitrary, we deduce that

$$I'(t) = \frac{1}{1+t}$$

for all $t > 0$. I is sufficiently regular, so we may integrate and apply the FTC to obtain

$$I(t) = I(0) + \log(1+t) = \log(1+t).$$

Thus the answer is $I(17) = \boxed{\log(18)}$. ■

5.4 thomas go to sleep it's 3 am

Example 5.4: Evaluate the integral

$$\int_{-\infty}^{\infty} e^{-x^2} \cos(2x)\, dx.$$

Solution. With no motivation whatsoever, let's define

$$I(t) = \int_{-\infty}^{\infty} e^{-x^2} \cos(2tx)\, dx.$$

We now compute the derivative of I. We have

$$\begin{aligned} \frac{I(t+h_n) - I(h_n)}{h_n} &= \int_{-\infty}^{\infty} e^{-x^2} \cdot \frac{\cos(2xt + 2xh_n) - \cos(2xt)}{h_n}\, dx \\ &= \int_{-\infty}^{\infty} 2xe^{-x^2} \cdot \frac{\cos(2xt + 2xh_n) - \cos(2xt)}{2xh_n}\, dx. \end{aligned}$$

By the MVT, we have for some c_n between $2xt$ and $2xh_n$ that this is equal to

$$= \int_{-\infty}^{\infty} 2xe^{-x^2} \cdot (-\sin(c_n))\, dx.$$

Note that by choice of c_n, we have that $c_n \to 2xt$.

We now claim that the integrands are dominated. Indeed, we have that

$$\left| 2xe^{-x^2} \cdot (-\sin(c_n)) \right| \leq 2|x|e^{-x^2},$$

and since $2|x|e^{-x^2}$ is integrable (Hint: Integrate over $x > 0$, and take $u = x^2$), we may take this as the dominating function. Thus we can pass the limit through to get

$$I'(t) = \lim_{n\to\infty} \int_{-\infty}^{\infty} 2xe^{-x^2} \cdot (-\sin(c_n))\, dx = \int_{-\infty}^{\infty} 2xe^{-x^2} \cdot (-\sin(2xt))\, dx.$$

Let us now integrate by parts.

$$I'(t) = \left(-e^{-x^2} \cdot (-\sin(2xt)) \Big|_{x=-\infty}^{\infty} \right) - \int_{-\infty}^{\infty} e^{-x^2} 2t \cos(2xt)\, dx$$
$$= \int_{-\infty}^{\infty} -2te^{-x^2} \cos(2xt) = -2tI(t).$$

Since $I(0) = \sqrt{\pi}$, we have obtained the *initial value problem*

$$\begin{cases} I'(t) = -2tI(t) \\ I(0) = \sqrt{\pi} \end{cases}.$$

To solve this, we note that

$$I'(t) + 2tI(t) = 0,$$

so that

$$I'(t)e^{t^2} + 2tI(t)e^{t^2} = 0.$$

By magic, this may be written as

$$\frac{d}{dt}\left[I(t)e^{t^2} \right] = 0,$$

and thus by applying the FTC we may obtain

$$I(t)e^{t^2} - I(0)e^{0^2} = 0.$$

This gives the form $I(t) = \sqrt{\pi}e^{-t^2}$. Our answer is thus $I(1) = \boxed{\dfrac{\sqrt{\pi}}{e}}$. ■

The following exercise may be good practice for whoever is reading this.

Example 5.5: Let

$$I(t) := \int_0^{\infty} \frac{\sin x}{x} e^{-tx}\, dx.$$

1. (Maybe a bit tricky?) Find the set E of all $t \in \mathbb{R}$ for which $I(t)$ is well-defined. (In particular, does the integral converge for $t = 0$...?)
2. Show that $I(t)$ is differentiable in E, and compute said derivative.
3. Using tricks, find a closed form for $I'(t)$.
4. Deduce the value of the integral

$$\int_0^{\infty} \frac{\sin x}{x} e^{-x}\, dx.$$

6 Continuity and Friends!

6.1 Meet the Family

There are a bunch of notions of continuity!

Definition 6.1 (Continuous)

Let $f : (X, d_X) \to (Y, d_Y)$.
f is **continuous** if for each x_0, we have for all $\varepsilon > 0$ that there exists $\delta > 0$ for which

$$d_Y(f(x), f(x_0)) < \varepsilon$$

for all x with $d_X(x, x_0) < \delta$.
We write $f \in C^0(X; Y)$.

Definition 6.2 (Uniform)

Let $f : (X, d_X) \to (Y, d_Y)$.
f is **uniformly continuous** if for all $\varepsilon > 0$ there exists $\delta > 0$ for which

$$d_Y(f(x), f(y)) < \varepsilon$$

for all x, y with $d_X(x, x_0) < \delta$.

Motto: "There is no dependence on *where* you are. We're "just as continuous" everywhere. For each $\varepsilon > 0$, the *same* $\delta > 0$ works *everywhere*."

Definition 6.3 (Lipschitz)

Let $f : (X, d_X) \to (Y, d_Y)$.
f is *Lipschitz continuous* if there exists $L > 0$ for which

$$d_Y(f(x), f(y)) \leq L \cdot d_X(x, y)$$

for all x, y.
We write $f \in C^{0,1}(X; Y)$.
On the real line, this looks like

$$|f(x) - f(y)| \leq L|x - y|.$$

You can try and visualize this if you want (imagine drawing an "X" through each point on the "graph"...), though in general I consider this to be a nice property that gives "nice ways to estimate differences". Often used in differential equations.

Definition 6.4 (Hölder)

Let $f : (X, d_X) \to (Y, d_Y)$. Let $\alpha > 0$.
f is *α-Hölder continuous* if there exists $H > 0$ for which

$$d_Y(f(x) - f(y)) \leq H \cdot d_X(x, y)^\alpha$$

for all x, y.
We write $f \in C^{0,\alpha}(X; Y)$.
On the real line, this looks like

$$|f(x) - f(y)| \leq H|x - y|^\alpha.$$

Motto: Hölder got jealous and copied Lipschitz.

I defined it on metric spaces, but you pretty much only see Hölder continuity used on $\mathbb{R}^N$.

The next notion of continuity is only really sane on $\mathbb{R}$.

Definition 6.5 (Absolutely Continuous)

Let $I \subseteq \mathbb{R}$ be an interval, $f : I \to \mathbb{R}$. f is *absolutely continuous* if for any $\varepsilon > 0$ we can find $\delta > 0$ such that the following holds:
Whenever we have pairwise disjoint intervals $\{(a_i, b_i)\}_{i=1}^n$ for some n, whose lengths sum to at most δ (i.e. $\sum_{i=1}^n b_i - a_i < \delta$), we have that

$$\sum_{i=1}^n |f(b_i) - f(a_i)| < \varepsilon.$$

This is incredibly important, and a big portion of this course will be spent studying this notion of continuity.

6.2 Hierarchy

Theorem 6.1

Let $0 < \alpha < 1$, we have the following hierarchy:

$$\begin{tikzcd}
 & \text{AC} \arrow[dr] & & \\
\text{Lipschitz} \arrow[ur] \arrow[dr, dashed] & & \text{Uniform} \arrow[r] & \text{Continuous} \\
 & \alpha\text{-Holder} \arrow[ur] & &
\end{tikzcd}$$

- Putting AC on the diagram only make sense when $f : I \to \mathbb{R}$. In any case we have the implication Lipschitz $\implies$ Uniform.
- The dashed arrow holds only when f is bounded. (If f need not be bounded, then $f(x) = x$ for $x \in \mathbb{R}$ is a counterexample!)

Proof. I will assume that $f : I \to \mathbb{R}$ because I'm too lazy to write metric stuff.

Lipschitz implies AC

Suppose $|f(x) - f(y)| \leq L|x - y|$. Fix $\varepsilon > 0$. Take $\delta = \varepsilon/L$.

Then for any $\{(a_i, b_i)\}_{i=1}^n$ with $\sum_{i=1}^n b_i - a_i < \delta$, we have that

$$\sum_{i=1}^n |f(b_i) - f(a_i)| \leq \sum_{i=1}^n L|b_i - a_i| \leq L \cdot \delta = \varepsilon.$$

AC implies Uniform

Take $n = 1$ lmfao.

Lipschitz implies Uniform

This isn't immediate from the first two implications if we're in a general metric space, but no matter the case, this is easy.

Lipschitz implies Hölder when f is bounded

Suppose f is Lipschitz with Lipschitz constant L, so $|f(x) - f(y)| \leq L|x - y|$.

Take $x, y \in I$.

- If $|x-y|<1$, then
$$|f(x)-f(y)| \leq L|x-y| \leq L|x-y|^{\alpha}$$
because $\alpha \in (0,1)$.
- If $|x-y| \geq 1$, then
$$|f(x)-f(y)| \leq 2M \leq 2M|x-y|.$$

So f is α-Hölder continuous with Hölder constant $H := \max(2M, L)$.

Hölder implies Uniform

This is easy.

Uniform implies Continuous

I mean, duh. $\square$

The next natural question to ask is if all these implications are strict. I won't do *all* of them, here are some of the more interesting/instructive ones.

6.2.1 Function that is Continuous but not Uniformly Continuous

$f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$ works. Notice that since x^2 gets "steeper and steeper", we can reason that "the same δ cannot work everywhere".

In general, any $f : \mathbb{R} \to \mathbb{R}$ that is uniformly continuous must be *sublinear*.

Another example is given $f : (-1,0) \cup (0,1) \to \mathbb{R}$ with
$$f(x) := \sin\left(\frac{1}{x}\right).$$
This is technically continuous, but the high oscillatory nature near $x = 0$ is what prevents f from being uniformly continuous.

6.2.2 Function that is Uniformly Continuous but not Absolutely Continuous

The Devil's Staircase works.

On one hand, the Devil's staircase is uniformly continuous because it is continuous on a compact set (a property that we will prove). On the other hand, it can be reasoned that the Devil's staircase could not be AC. This is because it fails the Fundamental Theorem of Calculus (which is in a sense equivalent to AC, as you will see later in the semester!).

6.2.3 Function that is α-Hölder but not Absolutely Continuous

The Devil's Staircase again!

It's not just uniformly continuous. It's also Hölder continuous of exponent $\log_3 2$.

6.2.4 Function that is Absolutely Continuous but not Lipschitz

$f(x) := \sqrt{x}$ on $[0, 1]$ works.

Proof that f is AC

We use the following inequality: For $0 \leq z \leq x \leq y$, we have $\sqrt{x} - \sqrt{y} \leq \sqrt{x - z} - \sqrt{y - z}$. (Proof: Square both sides and move things around until you believe me.)

Intuitively, if we "shift" the interval (x, y) to the left, then the value of $\sqrt{y - x}$ can only increase.

Now for any $\{(a_i, b_i)\}_{i=1}^n$ pairwise disjoint, we "shift" each interval to the left until all the n intervals are next to each other, starting at 0. This gives

$$\sum_{i=1}^n \sqrt{b_i} - \sqrt{a_i} \leq \sqrt{\sum_{i=1}^n b_i - a_i} - \sqrt{0}.$$

From here, the finish is easy. When $\varepsilon > 0$, then we can take $\delta = \sqrt{\varepsilon}$, so that if $\sum_{i=1}^n b_i - a_i < \delta$ then

$$\sum_{i=1}^n \sqrt{b_i} - \sqrt{a_i} \leq \sqrt{\delta} - \sqrt{0} = \varepsilon.$$

Proof that f is not Lipschitz

If it were, then there is $L > 0$ such that $\sqrt{x} \leq Lx$ for all $x \in [0, 1]$. Dividing, it follows that $\frac{1}{\sqrt{x}} \leq L$ for all $x \in (0, 1]$. Sending $x \to 0^+$ gives a contradiction.

6.2.5 Function that is Uniformly Continuous but not α-Hölder, or α-Hölder but not Lipschitz

A general statement that can be made here: Let $\beta \in (0, 1)$, and define $f(x) := x^\beta$ for all $x \in [0, 1]$. Then:

- f is uniformly continuous. (Even better: It's absolutely continuous!)

- f is α-Hölder continuous exactly when $0 < \alpha \leq \beta$.

The proof methodology is similar to the methods used in the previous two counterexamples.

6.3 Properties of Lipschitz Continuity

Here is a natural criterion for ascertaining that a differentiable function is Lipschitz.

Theorem 6.2

Let $E \subseteq \mathbb{R}^N$ be convex, and $f : E \to \mathbb{R}$ be differentiable. Suppose there exists $M > 0$ such that $\left|\frac{\partial f}{\partial x_i}(x)\right| \leq M$ for all $x \in E$ and all $1 \leq i \leq N$. Then f is Lipschitz.

Proof. The uniform bound on the partial derivatives implies that there exists some constant $L > 0$ such that $\|\nabla u(x)\| \leq L$ for all $x \in E$.

Now take any $x, y \in E$. Let $g : [0, 1] \to \mathbb{R}$ be defined as $g(t) := f((1-t)x + ty)$. (Here we need the convexity of E for this to be well-defined.)

Then, by applying the Mean Value Theorem, we get

$$g(1) - g(0) = g'(c)$$

for some $c \in (0, 1)$. It follows that

$$f(y) - f(x) = \nabla f((1-c)x + cy) \cdot (y - x),$$

and now by Cauchy Schwarz we obtain the bound

$$\|f(y) - f(x)\| \leq \|\nabla f((1-c)x + cy)\| \cdot \|y - x\| \leq L\|y - x\|.$$

□

In fact...

Theorem 6.3

Let $K \subseteq \mathbb{R}^N$ be compact and convex, and $f : K \to \mathbb{R}$ be of class C^1. Then f is Lipschitz.

Proof. The function $\nabla f : K \to \mathbb{R}^N$ is continuous, and in particular $\|\nabla f(\cdot)\| : K \to \mathbb{R}$ is continuous. Thus by the extreme value theorem, there exists $L > 0$ such that $\|\nabla f(x)\| \leq L$ for all $x \in K$. We may conclude that f is Lipschitz by the previous theorem. □

Remarks:

- You can generalize both theorems to account for a range of $\mathbb{R}^M$ instead of $\mathbb{R}$. It's just tedious.
- In both theorems, it is possible to remove the requirement that the domain be convex. We might cover how to do that *next* semester.

6.4 Properties of Hölder Continuity

Y'know how we keep requiring that $\alpha \in (0,1)$? Well here's why.

Theorem 6.4

Let $I \subseteq \mathbb{R}$ be an interval. Any $f : I \to \mathbb{R}$ that is α-Hölder continuous, where $\alpha > 1$, is **constant**.

Proof. Fun exercise. (I can think of two nice ways to go about this...) □

6.5 Properties of Uniform Continuity

Theorem 6.5

Let K be compact and $f : K \to \mathbb{R}$ be continuous. Then f is uniformly continuous (!).

Proof. Heads up: We take a bunch of weird "$\cdots/2$" 's in order to make things work here.

Fix $\varepsilon > 0$. Then for each $x \in K$ there exists $\delta_x > 0$ for which

$$d_Y(f(x), f(y)) < \varepsilon/2$$

for all $y \in B(x, \delta_x)$.

Since $\{B(x, \frac{\delta_x}{2})\}_{x \in K}$ covers K, we have by compactness that there exists $x_1, x_2, \cdots, x_n$ for which $\{B(x_k, \frac{\delta_{x_k}}{2}\}_{1 \le k \le n}$ covers K.

Now we take $\delta = \min\limits_{1 \le k \le n} \dfrac{\delta_{x_k}}{2}$. We claim this works.

Indeed, consider $x, y \in K$ for which $d_X(x,y) < \delta$. Find k for which $x \in B(x_k, \delta_{x_k}/2)$. Then

$$d_X(y, x_k) \le d_X(y,x) + d_X(x, x_k) \le \delta + \delta_{x_k}/2 \le \delta_{x_k}/2 + \delta_{x_k}/2 = \delta_{x_k}.$$

Since $d_X(x, x_k) < \delta \le \delta_{x_k}$ it follows that both $d_Y(f(x), f(x_k)) < \varepsilon/2$ and $d_Y(f(x_k), f(y)) < \varepsilon/2$, hence $d_Y(f(x), f(y)) < \varepsilon$. □

(*Remark:* Jacob mentioned a less messy proof, but it requires a notion of sequential compactness. We'll talk about it next semester.)

The next theorem is arguably the most important property of uniformly continuous functions, but unfortunately we do not currently have the technology necessary to prove it or exploit it in full.

Theorem 6.6 (Extension of Uniformly Continuous Functions)

Let $E \subseteq \mathbb{R}^N$ and $f : E \to \mathbb{R}^M$ be uniformly continuous. Then f can be **continuously extended** to all accumulation points of E!
That is, we can define a value $f(x)$ at each point $x \in \overline{E} \setminus E$ so that f is still continuous!
Also this extension is unique.

Proof. Hard. We'll talk about it next semester, probably! □

Remark: The same is true for $f : E \to Y$, where $E \subseteq X$, X, Y are metric spaces, and Y is **complete**. Again, we'll cover that next semester.

Theorem 6.7 (Sublinear Growth)

Let $f : \mathbb{R} \to \mathbb{R}$ be uniformly continuous. Then there exists $a, b \in \mathbb{R}$ for which

$$|f(x)| \leq a|x| + b$$

for all $x \in \mathbb{R}$.

Proof. Exercise. ;) □

7 Absolute Continuity

Absolute continuity is the most important continuity (...in my opinion)!

7.1 Closure properties

Theorem 7.1 (AC closed under linear combinations or whatever)

Suppose $f, g : I \to \mathbb{R}$ are AC. Then $f + g$ is AC. Also tf is AC for any $t \in \mathbb{R}$.

Proof. tRiViAl ☐

But it is not true that the product of AC functions is AC. Consider, e.g. $f(x) = x$ and $g(x) = x$ over $\mathbb{R}$. These are both AC, but $(fg)(x) = x^2$ is not.

Theorem 7.2 (Products)

Suppose $f, g : I \to \mathbb{R}$ are AC and **bounded**. Then fg is AC.

Proof. We do pranks.

Let's instead prove that if $h : I \to \mathbb{R}$ is AC and bounded, then h^2 is AC.

Let $M = \sup_I |h| < \infty$. Fix $\varepsilon > 0$. Find $\delta > 0$ such that if $\sum b_i - a_i < \delta$ then $\sum |h(b_i) - h(a_i)| < \varepsilon/(2M)$. Then for all $\{(a_i, b_i)\}$ with $\sum b_i - a_i < \delta$ we may write

$$\sum_{i=1}^{n} |h(b_i)^2 - h(a_i)^2| \le \sum_{i=1}^{n} |h(b_i) - h(a_i)| \cdot |h(b_i) + h(a_i)| \le 2M \sum_{i=1}^{n} |h(b_i) - h(a_i)| \le 2M \cdot \frac{\varepsilon}{2M} = \varepsilon.$$

So h^2 is indeed AC.

To finish the prank, apply the intermediate result on $f + g$ to see that $f^2 + 2fg + g^2$ is AC. But applying the result on f and g independently shows that f^2 and g^2 are AC. Hence $f^2 + 2fg + g^2 - f^2 - g^2$ is AC, so $2fg$ is AC. Hence fg is AC. ☐

We're now interested in whether quotients are AC. Obviously, not necessarily (e.g. $f(x) = 1$, $g(x) = x$ are AC over $(0, 1)$, but not $(f/g)(x) = 1/x$). Here we should simply ensure that the denominator stays away from 0.

Theorem 7.3

Suppose $f, g : I \to \mathbb{R}$ are AC, such that f is bounded and 0 is not an accumulation point of the range of g (i.e. $g(I)$). Then f/g is AC.

Proof. Since g is continuous, we may argue that $|g(x)| \geq a$ for some $a > 0$ (why?). Hence $1/g$ is bounded and so by the previous theorem it suffices to show that $1/g$ is AC. But upon writing

$$\sum_{i=1}^{n} \left| \frac{1}{g(b_i)} - \frac{1}{g(a_i)} \right| = \sum_{i=1}^{n} \frac{|g(b_i) - g(a_i)|}{|g(a_i)g(b_i)|} \leq \sum_{i=1}^{n} \frac{1}{a^2} |g(b_i) - g(a_i)|,$$

we see that this is clear. □

As a corollary, we have all of the following properties for functions on *compact* intervals.

Corollary 7.1

Let $f, g : [a, b] \to \mathbb{R}$ be AC. Then

- tf is AC,
- $f \pm g$ is AC,
- fg is AC, and
- f/g is AC provided that g has no zeroes.

The next question is, what about composition? It is not true that the composition of AC functions is AC. One weird example is given by $f(x) = x^2|\sin(1/x)|$ and $g(x) = \sqrt{x}$, both over $[0, 1]$. These are both AC (why?), but the function $x|\sin(1/\sqrt{x})|$ is not. Hence a stronger assumption is necessary.

Theorem 7.4 (Composition)

Let $f : \mathbb{R} \to \mathbb{R}$ and $g : I \to \mathbb{R}$ be such that f is Lipschitz and g is AC. Then $f \circ g$ is AC.

Proof. Let L be a Lipschitz constant for f, and fix $\varepsilon > 0$. Take δ such whenever $\sum b_i - a_i < \delta$ (pairwise disjoint blah blah) we have $\sum |g(b_i) - g(a_i)| < \varepsilon/L$. Then for all pairwise disjoint intervals $\{(a_i, b_i)\}$ with $\sum b_i - a_i < \delta/L$, we have that

$$\sum_{i=1}^{n} |f(g(b_i)) - f(g(a_i))| \leq \sum_{i=1}^{n} L|g(b_i) - g(a_i)| < L \cdot \varepsilon/L = \varepsilon.$$

□

7.2 Other Properties

Do the intervals need to be pairwise disjoint? Then answer is yes. Consider $f(x) = \sqrt{x}$ over $[0, 1]$. This is AC. But note that given any $\delta > 0$, the n intervals $(0, \delta/n), (0, \delta/n), \cdots, (0, \delta/n)$

have total length at most δ, and $\sum_{i=1}^n \sqrt{\delta/n} - \sqrt{0} = \sqrt{n\delta}$, which can exceed any $\varepsilon > 0$ by choosing n large enough.

Fortunately, there are some valid ways to fudge the definition of AC.

Theorem 7.5

In the definition of AC, we make allow taking $n = \infty$.

Proof. Suppose f is AC. Fix $\varepsilon > 0$, and take a δ that witnesses the absolute continuity of f suppose that pairwise disjoint blah blah $\sum_{i=1}^\infty b_i - a_i < \delta$. Then for all n, we have $\sum_{i=1}^n b_i - a_i < \delta$, so

$$\sum_{i=1}^n |f(b_i) - f(a_i)| < \varepsilon.$$

As n was arbitrary, we can send $n \to +\infty$ to obtain $\sum_{i=1}^n |f(b_i) - f(a_i)| \leq \varepsilon$, which is good enough. □

This next "property" is really boring. I had no plans to do it in recitation and honestly I have no idea why I even decided to write it. what is wrong with me.. whatever i already wrote it and im too lazy to erase it so here you go

Theorem 7.6

Call a function $f : I \to \mathbb{R}$ *clearly continuous* (not an actual term) if for all $\varepsilon > 0$ we can find $\delta > 0$ such that whenever blah blah pairwise disjoint $\sum_{i=1}^n b_i - a_i < \delta$, we have

$$\left| \sum_{i=1}^n f(b_i) - f(a_i) \right| < \varepsilon.$$

(The difference is where the absolute value bars are...)
A function f is clearly continuous if and only if it is absolutely continuous.

Proof. The triangle inequality shows that if f is absolutely continuous then it is clearly continuous. For the converse, suppose that f is clearly continuous.

Fix $\varepsilon > 0$. Take δ such that whenever blah blah pairwise disjoint $\sum b_i - a_i < \delta$, we have that $|\sum f(b_i) - f(a_i)| < \varepsilon/2$.

Now consider pairwise disjoint intervals $\{(a_i, b_i)\}_{i=1}^n$. Split it up into two sets of intervals $\{(a_i', b_i') : 1 \leq i \leq n'\} \sqcup \{(a_i'', b_i'') : 1 \leq i \leq n''\}$, such that $f(b_i') > f(a_i')$ and $f(b_i'') \leq f(a_i'')$. Then since $\sum b_i' - a_i' < \delta$ and $\sum b_i'' - a_i'' < \delta$, we have

$$\sum_{i=1}^{n'} f(b_i') - f(a_i') < \varepsilon/2,$$

$$\sum_{i=1}^{n'} f(a_i') - f(b_i') < \varepsilon/2.$$

Since these summands are all non-negative, they are equal to their absolute values, so

$$\sum_{i=1}^{n} |f(b_i) - f(a_i)| = \sum_{i=1}^{n'} |f(b_i') - f(a_i')| + \sum_{i=1}^{n'} |f(b_i'') - f(a_i'')|$$

$$= \sum_{i=1}^{n'} f(b_i') - f(a_i') + \sum_{i=1}^{n'} f(a_i'') - f(b_i'') < \varepsilon.$$

Hence f is absolutely continuous. □

Ok one last thing. This characterization is actually pretty cute. I'm not going to prove it because it's really hard, and hence I won't let you cite it (though I sense that you'll have a tough time using it...). However, it might give you a better sense of how to tell whether something is AC, from an intuitive standpoint.

Theorem 7.7

A function $f : I \to \mathbb{R}$ is AC if and only if the following conditions hold:

1. f is continuous
2. f has bounded pointwise variation
3. f satisfies the *Lusin-N* property. That is, f sends measure-zero sets to measure-zero sets, i.e. for any $E \subseteq I$ with $\mathcal{L}^1(E) = 0$, we have that $\mathcal{L}^1(f(E)) = 0$.

Proof. idk □

8 Manifolds

Lemma 8.1 (Totally Not Suspicious Linear Algebra Warmup)

Let $P \leq N$, A be an $M \times N$ matrix, B be an $N \times P$ matrix, and $AB = C$. If $\operatorname{rank} C = P$. Then $\operatorname{rank} A \geq P$.

Proof. Since $P \leq N$ and C is full rank, we have that the columns of C are linearly independent. Viewing B as a matrix of vectors $[v_1, v_2, \cdots, v_P]$ in $\mathbb{R}^N$, we see that $\{Av_1, Av_2, \cdots, Av_P\}$ are precisely the columns of C, and thus are linearly independent.

It follows that the row space of A has dimension at least P, which is what we wanted to show. □

8.1 Stupid Example

Example 8.1: Show that $\{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 = x_2\}$ is a C^∞ manifold of rank 2.

Proof. We take $\varphi(y_1, y_2) := (y_1, y_1, y_2)$. An inverse is given by $(x_1, x_2, x_3) \mapsto (x_1, x_3)$ which is continuous (we also could have used (x_2, x_3). It's the same map when restricted to $\varphi(\mathbb{R}^2)$). So φ is a homeomorphism.

Now

$$D\varphi(y_1, y_2) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which is full rank because there exists a 2×2 matrix which is invertible. □

8.2 The Torus

8.2.1 That Weird Trick That Edward Showed Me Like Half an Hour Before Recitation

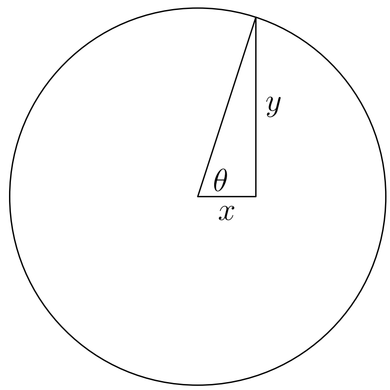
I totally sped through this in recitation so let's take the time to understand this essential component: Given the values of $\cos\theta$ and $\sin\theta$, how can we recover θ in a nice way? In other words, can we find a nice expression for the inverse of $f : \theta \mapsto (\cos\theta, \sin\theta)$ over $\theta \in (-\pi, \pi)$?

One alright way to do this is via

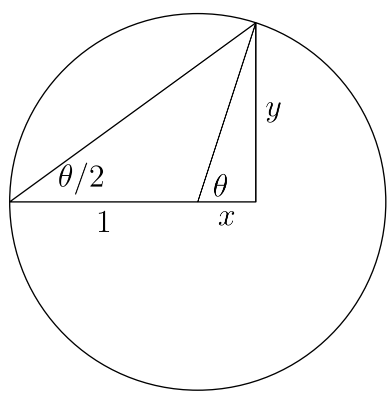
$$f^{-1}(x,y) := \begin{cases} \cos^{-1}(x), & y > 0 \\ 0, & y = 0 \\ -\cos^{-1}(x), & y < 0 \end{cases}.$$

This *isn't* continuous if you interpret as a function on like all of $\mathbb{R}^2$ or something, but it is if you're restricting it to the unit circle.

A cleaner but more black-magic-y way uses high school geometry. Draw the x and y-segments...



KEY POINT: We'd like to do $\tan^{-1}(y/x)$ to try and recover the angle θ. But **this doesn't work** because $\tan^{-1}$ only spits out values in the interval $(-\pi/2, \pi/2)$, whereas θ lives in $(-\pi, \pi)$. The trick here is to instead use the $\tan^{-1}$ to compute $\theta/2$. Where is $\theta/2$ in the diagram? Right here!



Now $\tan^{-1}\left(\frac{y}{1+x}\right)$ spits out $\theta/2$, and so $\boxed{2\tan^{-1}\left(\dfrac{y}{1+x}\right)}$ spits out θ. No piecewise stuff needed! It just uses more brain than I currently possess.

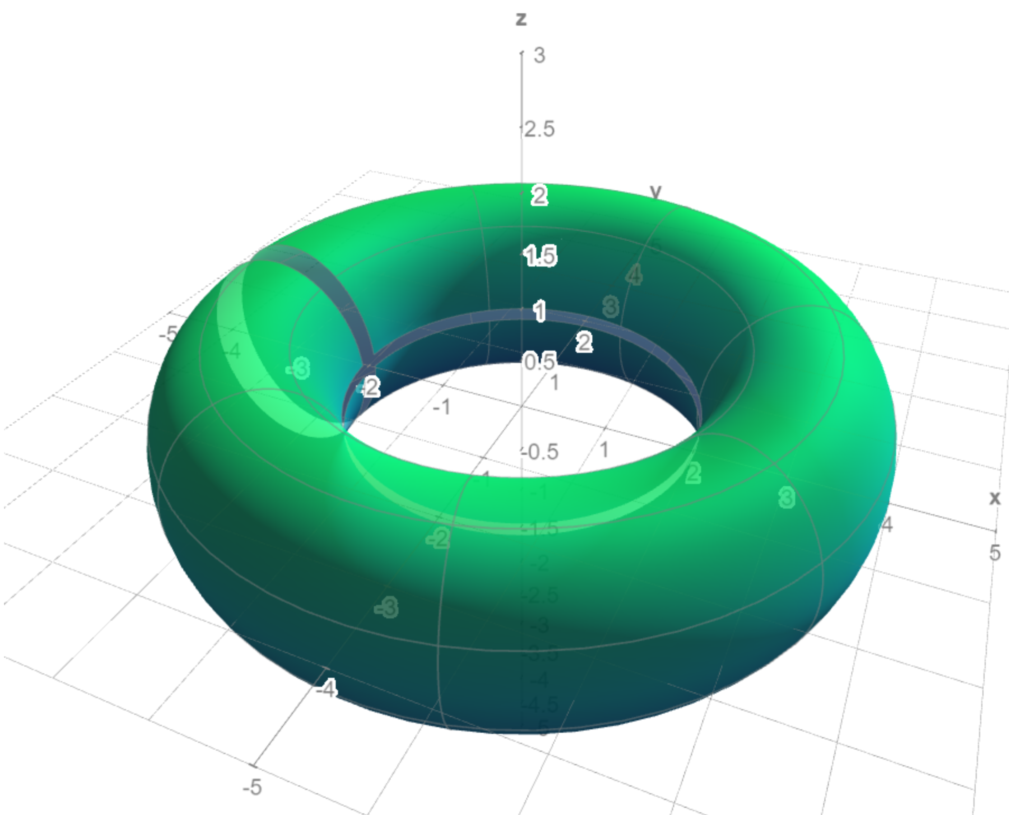
Compare the two methodologies here: https://www.math3d.org/7QVsHD1cL

8.2.2 ok let the pain begin

Example 8.2: Let $0 < R_2 < R_1$. Let

$$T := \left\{ (x,y,z) \in \mathbb{R}^3 : \left\| (x,y,z) - \frac{R_2(x,y,0)}{\sqrt{x^2+y^2}} \right\| = R_1 \right\}.$$

Let's make two "cuts" into T to get $\tilde{T}$. In particular let $\tilde{T} = T \setminus (\{(x,0,z) : x < 0\} \cup \{(x,y,0) : \|(x,y,0)\| = R_1\})$. Prove that $\tilde{T}$ is a C^∞ manifold of dimension 2.



Notice the two cuts: one along a small circle on the left, and one along the inner ring.

(*You can play with this here:* https://www.math3d.org/BsJCbfpqk)

Proof. It's a good idea to do this in "steps" so that it's easier to digest.

Step 1: First Curl

Let's start with a sheet of paper $U := (-\pi, \pi) \times (-\pi, \pi)$. We'll now curl it up like so:

$$\varphi_1 : (\alpha, \beta) \mapsto (\cos\beta, \alpha, \sin\beta)$$

Now the curled-up paper $\varphi_1(U)$ looks like a cylinder whose circular faces are aligned with the y-axis. There's also a slight slit in the cylinder along the "leftmost segment" of the curvy side.

Let's show that this is a homeomorphism. Clearly φ_1 is continuous, so we need to construct an inverse and show that it is continuous over $\varphi_1(U)$. By that weird trick, one such inverse is given by

$$(x', y', z') \mapsto \left(y', 2\tan^{-1}\left(\frac{z'}{1+x'}\right)\right),$$

and this is clearly continuous.

Step 2: Reposition and Dilate (to prep for the next step)

I now want to consider a transformation so that the middle circle of the cylinder $\varphi_1(U)$ becomes the circle at the "angle 0" cross-section of the torus, i.e. $\{(x, 0, z) : (x - R_1)^2 + y^2 = R_2^2\}$.

I need to start by stretching the circle so that its radius becomes R_2. This is given by the map $(x, y, z) \mapsto (R_2x, y, R_2z)$ (the circle is in the xz-plane, so I don't want to stretch in the y direction).

Now that the circle is the correct size, I just need to shift it over by R_1. This is given by the map $(x, y, z) \mapsto (R_1 + x, y, z)$.

Combining these two maps, I get the affine transformation

$$\varphi_2 : (x, y, z) \mapsto (R_1 + R_2x, y, R_2z).$$

It's clear that this map is infinitely nice, being a homeomorphism and everything, and particularly its 100% nice when restricted to the cylinder $\varphi_1(U)$. Yay.

Step 3: Second Curl

The intuition now is that, given a point (x, y, z) on the repositioned cylinder $(\varphi_2 \circ \varphi_1)(U)$, we can view $(x, 0, z)$ as "a point on a circle" and the y-component as "how much to rotate". (If you trace back, recall that y is essentially just α, so it lives in the interval $(-\pi, \pi)$.)

The way we do the second-curling is then clear: We send a point $(x, y, z) \in (\varphi_2 \circ \varphi_1)(U)$ to the point $(x, 0, z)$ rotated y-radians about the z-axis. If you squint, this is literally just given by

$$\varphi_3 : (x, y, z) \mapsto (x \cos y, x \sin y, z).$$

Tada! This has all the "slits" or "cuts" in the right places too.

For the last time, let's show that this is a homeomorphism. Clearly φ_3 is continuous. For the inverse, suppose we are given the values of $x' = x \cos y$, $y' = x \sin y$, and $z' = z$. We need to recover x, y, z.

Well, z is just z'. To get x, we can compute $\sqrt{(x')^2+(y')^2}$. Lastly to get y, we can use that weird trick again because we know the values of $\cos y = \frac{x'}{\sqrt{(x')^2+(y')^2}}$ and $\sin y = \frac{y'}{\sqrt{(x')^2+(y')^2}}$, so an inverse is given by

$$(x', y', z') \mapsto \left(\sqrt{(x')^2+(y')^2}, 2\tan^{-1}\left(\frac{y'}{\sqrt{(x')^2+(y')^2}+x'}\right), z'\right).$$

This is continuous, so we're done.

Yay we win

Since $\varphi_1, \varphi_2, \varphi_3$ are all homeomorphisms, their composition $\varphi := \varphi_3 \circ \varphi_2 \circ \varphi_1$ is a parametrization $\varphi : U \mapsto \mathbb{R}^3$ for $\tilde{T}$, and must be a homeomorpism because the composition of homeomorphisms is a homeomorphism (why?). Great!

Gathering everything up, the formula for φ is given by

$$\boxed{\varphi(\alpha, \beta) = ((R_1 + R_2\cos\beta)\cos\alpha, (R_1 + R_2\cos\beta)\sin\alpha, R_2\sin\beta)}$$

To actually complete the proof that $\tilde{T} = \varphi(U)$ is a manifold, we now need to show that the Jacobian is rank 2. The Jacobian is

$$D\varphi(r, \theta) = \begin{bmatrix} -(R_1 + R_2\cos\beta)\sin\alpha & (R_1 + R_2\cos\beta)\cos\alpha & 0 \\ -R_2\sin\beta\cos\alpha & -R_2\sin\beta\sin\alpha & R_2\cos\beta \end{bmatrix}.$$

There are two cases.

- If $\cos\beta \neq 0$, then for this to be rank 2 we just need something non-zero in the top row. Since $0 < R_2 < R_1$, the $R_1 + R_2\cos\beta$ and $-R_2$ factors are always non-zero. And, $\sin\alpha$ and $\cos\alpha$ can't both be 0, so we're good here.
- If $\cos\beta = 0$, then we need the first 2×2 submatrix to be invertible, which is

  $$\begin{bmatrix} -R_1\sin\alpha & R_1\cos\alpha \\ \pm R_2\cos\alpha & \pm R_2\sin\alpha \end{bmatrix}.$$

  This is invertible because its determinant is $\pm R_1R_2$, which is non-zero!!!

Lastly, φ is C^∞ because the composition and product of smooth maps is smooth. $\square$

8.3 The Surface of the N-Ball

QUESTION: Why not use those spherical coordinates that we learned a number of recitations ago?

ANSWER: There is a pretty severe issue of uniqueness. Back then, it was fine for changing variables because the spherical coordinates are an injection almost everywhere. The "problematic" points are just negligible.

Here though, we're going to try and parametrize all of $\partial B_N(0,1)$ with one chart. It turns out that this is impossible, which is why we're going to poke a hole at it at the north pole. That is, we'll parametrize $B_N(0,1) \setminus (0,0,\cdots,1)$. Doing this is still a pretty useful exercise.

Now it's possible to parameterize this punctured surface with only one chart! But the spherical coordinates still won't cut it. We'll use a different way to parameterize it called a **stereographic projection** (Google this term for some dank pics!). It's really cool.

Though, even once we get a parameterization, how are we ever going to prove that it's Jacobian is full rank? That sounds like a recipe for a disaster. This is where the following theorem that I came up with on Sunday comes into play.

Theorem 8.1

Let $M \leq N$, $U \subseteq \mathbb{R}^M$, $\varphi : U \to \mathbb{R}^N$. Suppose that φ is a bijection, and both φ and φ^{-1} are differentiable. Then $D\varphi$ has full rank, and moreover $\varphi(U)$ is a differentiable manifold of dimension M.

Proof. Let $V = \varphi(U)$. We know that $\varphi^{-1} \circ \varphi = \mathrm{id}_U$, thus we have that $D(\varphi^{-1} \circ \varphi) = I_{M\times M}$. By the chain rule,

$$(D\varphi)((D\varphi^{-1}) \circ \varphi) = I_{M\times M}.$$

(You're probably unfamiliar with this form of chain rule. It actually follows from the normal chain rule you're used to. If you need convincing, expand it out!)

We have that $D\varphi$ is $M \times N$. **By the warm-up (!!!)**, we see that $D\varphi$ has rank at least M. But $M \leq N$, so $\operatorname{rank} D\varphi \leq \min(M,N) = M$. So in fact $D\varphi$ has rank M, and is thus full rank.

Since clearly φ is a homeomorphism (differentiability implies continuity for both φ and φ^{-1}) we are done. $\square$

At last, we can get to the point.

Example 8.3: Prove that the surface of the unit N-ball, minus the point $(0, 0, \cdots, 1)$, is a C^∞ manifold.

In what follows, I will use Leoni's "abuse of notation" to denote each point in $\mathbb{R}^N$ by the tuple (x', x_N), where $x' \in \mathbb{R}^{N-1}$ and $x_N \in \mathbb{R}$. For example, $(0, 1)$ denotes the point at the north pole that we are excluding.

Proof. Take $y := (y_1, \cdots, y_{N-1})$. We get a corresponding point as follows: Consider $(y, 0) \in \mathbb{R}^N$ and connect it to $(0, 1)$ with a line. The line is parametrized by

$$t(y, 0) + (1 - t)(0, 1) = (ty, 1 - t),$$

and this intersects the surface when t is such that $t^2\|y\|^2 + (1 - t)^2 = 1$. Solving, we get $t = \frac{2}{1+\|y\|^2}$. At this value of t we have

$$(ty, 1 - t) = \left(\frac{2y}{1 + \|y\|^2}, \frac{\|y\|^2 - 1}{\|y\|^2 + 1}\right).$$

We take

$$\boxed{\varphi(y) := \left(\frac{2y}{1 + \|y\|^2}, \frac{\|y\|^2 - 1}{\|y\|^2 + 1}\right)}.$$

To see that this is a homeomorphism, we must construct an explicit inverse. For $x = (x', x_N) \in \mathbb{R}^N$, we send this to the $\{x_N = 0\}$ plane using the same projection idea. The line connecting x with $(0, 1)$ is parametrized as $t(x', x_N) + (1 - t)(0, 1) = (tx', tx_N + 1 - t)$, and this is exactly at the $\{x_N = 0\}$ plane when $tx_N + 1 - t = 0$, or $t = \frac{1}{1-x_N}$. So we claim that an inverse map is given by

$$\varphi^{-1} : (x', x_N) \mapsto \frac{1}{1 - x_N} x'.$$

If we've followed our intuition correctly then this *better* be correct. To verify, we compose it with φ to see that

$$\varphi^{-1}\left(\frac{2y}{1 + \|y\|^2}, \frac{\|y\|^2 - 1}{\|y\|^2 + 1}\right) = \frac{1}{1 - \frac{\|y\|^2-1}{\|y\|^2+1}}\left(\frac{2y}{1 + \|y\|^2}\right) = y.$$

Yay! It's clear φ^{-1} is continuous, so φ is a homeomorphism. Next, φ^{-1} is differentiable, so $D\varphi$ is full rank. Lastly, φ itself is C^∞ because if f, g are smooth (with g non-zero) then so is f/g, So we're done. $\square$

9 Surface Integrals

9.1 Other Notations for the Surface Integral

Leoni uses $d\mathcal{H}^k$. This is good because it's like, literally the most correct possible notation, being literally an integral with respect to the Hausdorff measure.

Most commonly, you see dS being used. This is usually paired with using like dV instead of $d(x, y, z)$. Blegh.

Other notations include $d\Sigma, d\sigma$, and dA.

9.2 Generic Example

Before we do anything, here is a very useful tool for computing Jacobians by hand:

Theorem 9.1 (Cauchy-Binet Formula)

Let A be an $N \times M$ matrix, where $M \leq N$. Then the determinant of $A^T A$ is the sum of the squares of the determinants of all $M \times M$ submatrices of A.
For example, if

$$A = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix},$$

then

$$\det(A^T A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix}^2 + \begin{vmatrix} a & b \\ e & f \end{vmatrix}^2 + \begin{vmatrix} c & d \\ e & f \end{vmatrix}^2$$

Proof. idk □

Example 9.1: Let $f : \mathbb{R}^3 \to \mathbb{R}$ be defined as $f(x, y, z) = x^2 + yz$. Compute the surface integral $\int_C f \, d\mathcal{H}^2$, where C is the curved surface of the upside-down circular cone with base $B_2(0, 1) \times \{1\}$ and vertex $(0, 0, 0)$.

Solution. Let's parameterize C with the chart $\varphi : (0, 1) \times (0, 2\pi)$ given by $\varphi(r, \theta) := (r \cos\theta, r \sin\theta, r)$. Then

$$\int_C f \, d\mathcal{H}^2 = \int_0^1 \int_0^{2\pi} f(r\cos\theta, r\sin\theta, r) |||\varphi(r, \theta)||| \, d\theta \, dr.$$

We now compute the Jacobian $|||\varphi(r,\theta)|||$. First we note that

$$D\varphi(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \\ 1 & 0 \end{pmatrix},$$

so by the Cauchy-Binet formula we see that

$$\begin{aligned}\det(D\varphi(y)^T D\varphi(y)) &= \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}^2 + \begin{vmatrix} \cos\theta & -r\sin\theta \\ 1 & 0 \end{vmatrix}^2 + \begin{vmatrix} \sin\theta & r\cos\theta \\ 1 & 0 \end{vmatrix}^2 \\ &= r^2 + r^2\sin^2\theta + r^2\cos^2\theta = 2r^2,\end{aligned}$$

thus

$$|||\varphi(r,\theta)||| = \sqrt{\det(D\varphi(y)^T D\varphi(y))} = \sqrt{2}r.$$

It follows that

$$\int_C f\, d\mathcal{H}^2 = \int_0^1 \int_0^{2\pi} (r^2\cos^2\theta + r^2\sin\theta) \cdot \sqrt{2}r\, d\theta\, dr = \sqrt{2}\int_0^1 \pi r^3\, dr = \boxed{\frac{\sqrt{2}\pi}{4}}.$$

■

9.3 Integrating over the Surface of a Graph

Lemma 9.1 (Sylvester's Determinant Identity)

Let A be $m \times n$ and B be $n \times m$. Then

$$\det(I_m + AB) = \det(I_n + BA).$$

Proof. Consider the block-form $(m+n) \times (m+n)$ matrix

$$\begin{pmatrix} I_m & A \\ -B & I_n \end{pmatrix}.$$

We evaluate the determinant of this matrix in two different ways. Using "row reduction", we have on one hand that

$$\det\begin{pmatrix} I_m & A \\ -B & I_n \end{pmatrix} = \det\begin{pmatrix} I_m & A \\ -B + BI_m & I_n + BA \end{pmatrix} = \det\begin{pmatrix} I_m & A \\ 0 & I_n + BA \end{pmatrix} = \det(I_n + BA).$$

On the other hand, we can use "column reduction" to get that

$$\det\begin{pmatrix} I_m & A \\ -B & I_n \end{pmatrix} = \det\begin{pmatrix} I_m + AB & A \\ -B + I_nB & I_n \end{pmatrix} = \det\begin{pmatrix} I_m + AB & A \\ 0 & I_n \end{pmatrix} = \det(I_m + AB).$$

□

Theorem 9.2 (Integration over a Graph)

Let $U \subseteq \mathbb{R}^N$ be open, and $g : U \to \mathbb{R}$ differentible and of class C^k for some $k \geq 0$. Then the set $M := \{(x, g(x)) : x \in U\}$ is a manifold of dimension N and of class C^k, and for $f : M \to \mathbb{R}$ $\mathcal{H}^N$-measurable such that either f is integrable or has a sign, we have the formula

$$\int_M f\, \mathcal{H}^N = \int_U f(x, g(x))\sqrt{1 + \|\nabla f(x)\|^2}\, dx.$$

Proof. A parameterization for M is easily given by $\varphi : x \mapsto (x, g(x))$. This is continuous, and in particular of class C^k, and an inverse is given by $(x', y') \mapsto x'$ which is obviously continuous. Moreover

$$D\varphi = \begin{pmatrix} I_N \\ \nabla g^T \end{pmatrix}$$

which has full rank because the identity submatrix is invertible, which confirms that M is a manifold of dimension N.

Now by the surface integral formula,

$$\int_M f\, \mathcal{H}^N = \int_U f(\varphi(x))\sqrt{\det(D\varphi(x)^T D\varphi(x))}\, dx$$

$$= \int_U f(x, g(x))\sqrt{\det\left[\begin{pmatrix} I_N & \nabla g(x) \end{pmatrix}\begin{pmatrix} I_N \\ \nabla g(x)^T \end{pmatrix}\right]}$$

$$= \int_U f(x, g(x))\sqrt{\det\left(I_N + \nabla g(x)\nabla g(x)^T\right)}.$$

Now we can apply Sylvester's Determinant Identity (!!!) to write this as

$$= \int_U f(x, g(x))\sqrt{\det(I_1 + \nabla g(x)^T \nabla g(x))}$$

$$= \int_U f(x, g(x))\sqrt{1 + \|\nabla g(x)\|^2}).$$

□

Example 9.2: Solve Example 9.1 again!

Solution. The cone C is the graph of $g(x, y) := \|(x, y)\|$ over $B(0, 1)$ (and this is smooth over $B(0, 1) \setminus \{0\}$ so this is fine). Thus

$$\int_C f\, d\mathcal{H}^2 = \int_{B(0,1)} f(x, y, \|(x, y)\|)\sqrt{1 + \|\nabla g(x, y)\|^2}\, d(x, y).$$

If you work it out, $\nabla g(x,y) = \frac{(x,y)}{\|(x,y)\|}$, so in fact, $\|\nabla g(x,y)\| = 1$, hence

$$\int_C f\, d\mathcal{H}^2 = \sqrt{2} \int_{B(0,1)} x^2 + y\|(x,y)\|\, d(x,y).$$

Changing to polar coordinates, this is

$$= \sqrt{2} \int_0^1 \int_0^{2\pi} r^2 \cos^2\theta + r^2 \sin\theta\, d\theta\, dr,$$

which is what we had before. ■

9.4 Spherical Coordinates Revisited

The circumference of a circle is $2\pi r$, and its area is πr^2. The surface area of a sphere is $4\pi r^2$, and its volume is $\frac{4}{3}\pi r^3$. Why is it the case that the derivative of N-dimensional volume of the N-ball gives its surface measure? Is it a coincidence?

Here we will demystify this and, in fact, prove something more general.

Theorem 9.3 (Spherical Coordinates)

Let $f : B_N(0,R) \to \mathbb{R}$ be either integrable or signed. Then

$$\int_{B_N(0,R)} f\, dx = \int_0^R \int_{\partial B_N(0,r)} f\, d\mathcal{H}^{N-1}\, dr.$$

We start with a lemma.

Lemma 9.2

Let $U \subseteq \mathbb{R}^M$, and $\varphi : U \to \partial B_N(0,1)$ be any differentiable function. Then $\frac{\partial \varphi}{\partial y_i}(y) \cdot \varphi(y) = 0$ for all $i = 1, \cdots, M$. In particular, $D\varphi(y)^T \varphi(y) = 0$.

Proof. Since $\|\varphi(y)\|^2 = 1$ for all $y \in U$, and $\nabla \|\cdot\|^2 = 2(\cdot)$, we have that

$$0 = \frac{\partial}{\partial y_i} \|\varphi(y)\|^2 = 2(\varphi(y)) \cdot \frac{\partial \varphi}{\partial y_i}(y)$$

for all i. □

We now turn to the proof of the theorem.

Proof.

Let $\varphi : U \to \partial B_N(0,1)$ be a parameterization whose range covers $\partial B_N(0,1)$ up to a set of measure zero. (For example, one can take φ to be the spherical change of variables, or even the stereographic projection.)

We now may consider the change of variables $g : (0,R) \times U$ given by $g(r,y) = r\varphi(y)$. Then

$$\int_{B_N(0,R)} f\,dx = \int_0^R \int_U f(r\varphi(y))|\det Dg(r,y)|\,dy.$$

Let's compute the Jacobian determinant. Since $\frac{\partial g}{\partial r}(r,y) = \varphi(y)$ we have that

$$Dg(r,y) = \begin{pmatrix} \varphi(y) & rD\varphi(y) \end{pmatrix}.$$

Computing the determinant of this would be a disaster. But we can do this weird trick (view r as a *constant or scalar* in these computations):

$$\begin{aligned}
(\det Dg)^2 = (\det Dg)(\det Dg) &= (\det(Dg)^T)(\det(Dg)) \\
&= \det\left(Dg^T Dg\right) = \det\left[\begin{pmatrix} \varphi^T \\ rD\varphi^T \end{pmatrix}\begin{pmatrix} \varphi & rD\varphi \end{pmatrix}\right] \\
&= \det\begin{pmatrix} \|\varphi\|^2 & r\varphi^T(D\varphi) \\ r(D\varphi)^T\varphi & r^2(D\varphi)^T D\varphi \end{pmatrix} \\
&= \det\begin{pmatrix} 1 & 0 \\ 0 & r^2(D\varphi)^T D\varphi \end{pmatrix} \\
&= \det\left(r^2(D\varphi)^T D\varphi\right) = \det\left([D(r\varphi)]^T[D(r\varphi)]\right)
\end{aligned}$$

Thus $\det Dg(r,y) = \sqrt{\det\left(D(r\varphi)(y)^T D(r\varphi)(y)\right)} = |||r\varphi(y)|||$. Holy shit. (Keep in mind that the D in $D(r\varphi)$ is the derivative of the function $r\varphi$ in the y variable, so the r is viewed as a constant here.)

Going back to the original integral, we now have that

$$\int_{B_N(0,R)} f\,dx = \int_0^R \int_U f(r\varphi(y)) \cdot |||r\varphi(y)|||\,dy,$$

which, by the surface integral formula, is just

$$= \int_0^R \int_{\partial B_N(0,r)} f(x)\,d\mathcal{H}^{N-1}(x)$$

because $y \mapsto r\varphi(y)$ parametrizes $\partial B_N(0,r)$. □

9.5 Surface Measure of the N-Ball

Theorem 9.4

$$\mathcal{H}^{N-1}(\partial B_N(0,1)) = N\mathcal{L}^N(B_N(0,1))$$

Proof. Taking $f \equiv 1$ and $R = 1$ in the previous theorem, we get that

$$\begin{aligned}
\int_{B_N(0,1)} dx &= \int_0^1 \int_{\partial B_N(0,r)} d\mathcal{H}^{N-1}\, dr \\
\mathcal{L}^N(B_N(0,1)) &= \int_0^1 \mathcal{H}^{N-1}(\partial B_N(0,r))\, dr \\
\mathcal{L}^N(B_N(0,1)) &= \frac{1}{N}\mathcal{H}^{N-1}(\partial B_N(0,1)).
\end{aligned}$$

$\square$

10 More Surface Integrals

10.1 A Brief Remark on the Jacobian

- For a function $f : \mathbb{R}^M \to \mathbb{R}^N$, its Jacobian $Df(y_0)$ has dimensions $N \times M$.
- Suppose $f = (f_1, f_2, \cdots, f_N)$. We can view $Df(y_0)$ as the matrix

$$\begin{pmatrix} \nabla f_1(y_0)^T \\ \nabla f_2(y_0)^T \\ \vdots \\ \nabla f_N(y_0)^T \end{pmatrix}.$$

  That is, the rows are given by the gradients of the components.
- However, we can also view $Df(y_0)$ as the matrix

$$\begin{pmatrix} \frac{\partial f}{\partial y_1}(y_0) & \frac{\partial f}{\partial y_2}(y_0) & \cdots & \frac{\partial f}{\partial y_M}(y_0) \end{pmatrix}.$$

  That is, the columns are given by the directional derivatives of f.

10.2 Tangent Spaces and Normals

Let M be a differentiable manifold, and let $x_0 \in M$. What does it mean for a vector v to be "tangent" to M at x_0? Intuitively, we think of this as meaning that if we start at x_0 and move in the "v direction", then we "just touch" M. This is a bit hard to formalize.

Fortunately there is a different way to imagine this by flipping things a bit: If we *travel along* the manifold, and pass through x_0, then the direction v that we're currently moving in when we're at x_0 is a tangent vector.

Definition 10.1 (Tangent Vector, Tangent Space)

Let $M \subseteq \mathbb{R}^N$ be a differentiable k-dimensional manifold, and let $x_0 \in M$. Then a vector $t \in \mathbb{R}^N$ is a *tangent vector* at x_0 if there exists a differentiable function $h : (-\delta, \delta) \to M$ (our "path" on M) such that

- $h(0) = x_0$ (we pass through x_0), and
- $h'(0) = t$ (we're moving in direction t when we're at x_0).

The space of all tangent vectors is denoted $T_M(x_0)$.

The key property that we will prove in lecture:

- $T_M(x_0)$ is a k-dimensional vector space.
- Suppose that $\varphi : U \to M$ is a parameterization whose image contains x_0, and let $y_0 \in U$ be such that $\varphi(y_0) = x_0$. Then it turns out that a basis for $T_M(x_0)$ is given b the columns of $D\varphi(y_0)$. Recall that these columns are given exactly by $\frac{\partial \varphi}{\partial y_1}(y_0), \cdots, \frac{\partial \varphi}{\partial y_k}(y_0)$, where $y_1, \cdots, y_k$ form a basis for $\mathbb{R}^k$.

This leads to a natural way to define *normals* to a surface, which intuitively are vectors that are somehow "perpendicular to the surface".

Definition 10.2 (Normals, Normal Space)

The *normal space* at a point $x_0 \in M$ is given by

$$N_M(x_0) := T_M(x_0)^\perp.$$

That is, it is the orthogonal complement of $T_M(x_0)$, consisting of exactly those vectors $\nu \in \mathbb{R}^N$ for which $\nu \cdot t = 0$ for all $t \in T_M(x_0)$.

Note in particular that for a parameterization φ, we have that $\nu \in N_M(x_0)$ iff $D\varphi(y_0)^T \nu = 0 \in \mathbb{R}^k$.

Example 10.1: For $x_0 \in \partial B(0,1)$, compute the unit outward normal to $\partial B(0,1)$ at x_0.

Solution. Let $\varphi : U \to \partial B(0,1)$ be a parametrization of some subset of $\partial B(0,1)$ that contains x_0. Let $y_0 \in U$ be such that $\varphi(y_0) = x_0$. The columns of $D\varphi(y_0)$ form a basis for the tangent space at x_0, so a vector ν will be normal to the surface at x exactly when $D\varphi(y_0)^T \nu = 0$.

But in the last recitation, we showed that $D\varphi(y_0)^T \varphi(y_0) = 0$, so we may take $\nu = \varphi(y_0) = x_0$ to be a normal to $\partial B(0,1)$ at x_0. It turns out that this choice of ν has unit norm and points outward, so we're done. ■

In general, if $M \subseteq \mathbb{R}^N$ is an $N-1$-dimensional differentiable manifold, then the normal space at some $x_0 \in M$ is always dimension 1. For certain manifolds, called *orientable* manifolds, we will have that there exists a notion of a normal vector to point "outward", so that at every point on M there exists a unique unit outward normal.

10.3 A Totally Not Suspicious Surface Integral

Example 10.2: Let $R > 0$. Let $H \subseteq \mathbb{R}^3$ be the hemisphere $H := B_3(0,R) \cap \{z > 0\}$. Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $f(x,y,z) := (z^2, y, 2^x)$. Let $\nu : \partial H \to \mathbb{R}^3$ be the unique unit outward normal vector at a given point on ∂H. Evaluate the surface integral

$$\int_{\partial H} f(x,y,z) \cdot \nu(x,y,z)\, d\mathcal{H}^2.$$

Solution. We need to split the integral into two parts: One integral over the curved surface, and another over the flat circular bottom.

Let's get the easy one (the bottom side) out of the way, $H_0 := B_2(0,R) \times \{0\}$. A dumb chart for H_0 is given by $\varphi : B_2(0,R) \to \mathbb{R}^3$ with $\varphi(x,y) = (x,y,0)$. The Jacobian is, unsurprisingly, 1. And, at every point $(x,y,z) \in H_0$, the normal vector is given by $\nu(x,y,z) = (0,0,-1)$.

Thus, the integral over H_0, is given by

$$\int_{H_0} f(x,y,z) \cdot (0,0,-1)\, d\mathcal{H}^2 = \int_{B_2(0,R)} f(\varphi(x,y)) \cdot (0,0,-1)\, d(x,y)$$

$$= \int_{B_2(0,R)} f(x,y,0) \cdot (0,0,-1)\, d(x,y) = \int_{B_2(0,R)} (0^2, y, 2^x) \cdot (0,0,-1)\, d(x,y)$$

$$= \int_{B_2(0,R)} -2^x\, d(x,y).$$

Let's leave this for now and start tackling the hard part, which is the curved surface $\partial H \cap \{z > 0\} = \{(x,y,z) : x^2 + y^2 + z^2 = R^2, z > 0\} =: H^+$.

We observe that H^+ is the graph of the function $g(x,y) := \sqrt{R^2 - x^2 - y^2}$. Since

$$\nabla g(x,y) = \begin{pmatrix} \frac{-x}{\sqrt{R^2-x^2-y^2}} \\ \frac{-y}{\sqrt{R^2-x^2-y^2}} \end{pmatrix},$$

we see that

$$\sqrt{1 + \|\nabla g(x,y)\|^2} = \sqrt{1 + \frac{x^2}{R^2 - x^2 - y^2} + \frac{y^2}{R^2 - x^2 - y^2}} = \frac{R}{\sqrt{R^2 - x^2 - y^2}}.$$

Moreover the unit outward normal at $(x,y,z) \in H^+$ is given by $\nu(x,y,z) = \frac{(x,y,z)}{R}$. Thus

$$\int_{H^+} f(x,y,z) \cdot \nu(x,y,z)\, d\mathcal{H}^2 = \int_{B_2(0,R)} f(x,y,g(x,y)) \cdot \frac{(x,y,g(x,y))}{R} \cdot \sqrt{1 + \|\nabla g(x,y)\|^2}\, d(x,y)$$

$$\begin{aligned}
&= \int_{B_2(0,R)} (g(x,y)^2, y, 2^x) \cdot \frac{(x,y,g(x,y))}{R} \cdot \frac{R}{\sqrt{R^2-x^2-y^2}}\, d(x,y) \\
&= \int_{B_2(0,R)} \big(g(x,y)^2 x + y^2 + 2^x g(x,y)\big) \cdot \frac{1}{\sqrt{R^2-x^2-y^2}}\, d(x,y) \\
&= \int_{B_2(0,R)} \big(g(x,y)^2 x + y^2 + 2^x g(x,y)\big) \cdot \frac{1}{\sqrt{R^2-x^2-y^2}}\, d(x,y) \\
&= \int_{B_2(0,R)} \frac{(R^2-x^2-y^2)x + y^2 + 2^x\sqrt{R^2-x^2-y^2}}{\sqrt{R^2-x^2-y^2}}\, d(x,y) \\
&= \int_{B_2(0,R)} \frac{(R^2-x^2-y^2)x + y^2 + 2^x\sqrt{R^2-x^2-y^2}}{\sqrt{R^2-x^2-y^2}}\, d(x,y) \\
&= \int_{B_2(0,R)} \sqrt{R^2-x^2-y^2}\,x + \frac{y^2}{\sqrt{R^2-x^2-y^2}} + 2^x\, d(x,y).
\end{aligned}$$

Let's pause this computation. We'll quickly bring the two integrals we got together to form the integral over ∂H.

$$\int_{\partial H} f(x,y,z)\cdot\nu(x,y,z)\, d\mathcal{H}^2 = \int_{H_0} f(x,y,z)\cdot\nu(x,y,z)\, d\mathcal{H}^2 + \int_{H^+} f(x,y,z)\cdot\nu(x,y,z)\, d\mathcal{H}^2$$

$$= \int_{B_2(0,R)} -2^x\, d(x,y) + \int_{B_2(0,R)} \sqrt{R^2-x^2-y^2}\,x + \frac{y^2}{\sqrt{R^2-x^2-y^2}} + 2^x\, d(x,y)$$

$$= \int_{B_2(0,R)} \sqrt{R^2-x^2-y^2}\,x\, d(x,y) + \int_{B_2(0,R)} \frac{y^2}{\sqrt{R^2-x^2-y^2}}\, d(x,y)$$

For the first integral, we can quickly observe that the change of variables $(x,y)\mapsto(-x,y)$ just negates the integrand, so this symmetry implies that the first integral is equal to 0. It hence remains to evaluate the second integral. By polar coordinates, we have that

$$\int_{B_2(0,R)} \frac{y^2}{\sqrt{R^2-x^2-y^2}}\, d(x,y) = \int_0^R\int_0^{2\pi} \frac{r^3\sin^2\theta}{\sqrt{R^2-r^2}}\, d\theta\, dr = \int_0^R \frac{r^3}{\sqrt{R^2-r^2}}\, dr \cdot \int_0^{2\pi} \sin^2\theta\, d\theta.$$

For the r integral, we make the substitution $r=\sqrt{R^2-u^2}$, so that $dr = \frac{-u}{\sqrt{R^2-u^2}}\, du$. Then

$$\int_0^R \frac{r^3}{\sqrt{R^2-r^2}}\, dr = \int_0^R \frac{(R^2-u^2)^{3/2}}{u}\cdot\frac{u}{\sqrt{R^2-u^2}}\, du = \int_0^R R^2-u^2\, du = \frac{2}{3}R^3.$$

For the θ integral, using the double-angle formula or something gives you $\int_0^{2\pi}\sin^2\theta\, d\theta = \pi$. Therefore,

$$\int_{\partial H} f(x,y,z)\cdot\nu(x,y,z)\, d\mathcal{H}^2 = \boxed{\frac{2}{3}\pi R^3}.$$

■

...this just happens to be the volume of the hemisphere H. This is not a coincidence.

10.4 Yet Another Not Sus At All Surface Integral

Example 10.3: Let $R, h > 0$. Consider the cone $C := \{(x,y,z) : x^2 + y^2 \leq \frac{z^2 R^2}{h^2}, 0 \leq z \leq h\}$, so that C is a circular cone with base radius R and height h. Let M be the curved surface of C, and for each $x \in M$ let $\nu(x,y,z)$ denote the unit outward normal. Compute the surface integral

$$\int_M x\nu_1(x,y,z)\, d\mathcal{H}^2,$$

where ν_1 is the first component of $\nu = (\nu_1, \nu_2, \nu_3)$.

Solution. First let's get a chart for the surface and compute its Jacobian. One chart is given by $\varphi : B_2(0,R) \to \mathbb{R}^3$ with $\varphi(x,y) := (x, y, g(x,y))$, where $g(x,y) := \frac{h\sqrt{x^2+y^2}}{R}$. Then

$$|||D\varphi(x,y)||| = \sqrt{1 + \|\nabla g(x,y)\|^2} = \sqrt{1 + \frac{h^2}{R^2}} = \frac{\sqrt{R^2+h^2}}{R}.$$

Now let's compute the unit outward normal. We have that

$$D\varphi(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{hx}{R\sqrt{x^2+y^2}} & \frac{hy}{R\sqrt{x^2+y^2}} \end{pmatrix}.$$

We need to find some vector that's perpendicular to these two columns. One way to do this is via the cross product

$$\left(1, 0, \frac{hx}{R\sqrt{x^2+y^2}}\right) \times \left(0, 1, \frac{hy}{R\sqrt{x^2+y^2}}\right) = \begin{vmatrix} 1 & 0 & \frac{hx}{R\sqrt{x^2+y^2}} \\ 0 & 1 & \frac{hy}{R\sqrt{x^2+y^2}} \\ \hat{i} & \hat{j} & \hat{k} \end{vmatrix}$$

$$= \left(-\frac{hx}{R\sqrt{x^2+y^2}}, -\frac{hy}{R\sqrt{x^2+y^2}}, 1\right).$$

To make this point outward, we'll have to negate all these components. To make this have unit norm, we'll also have to divide by the norm it currently has, which is $\sqrt{1 + \frac{h^2}{R^2}} = \frac{\sqrt{R^2+h^2}}{R}$. Thus

$$\nu(x,y,z) = \frac{R}{\sqrt{R^2+h^2}} \left(\frac{hx}{R\sqrt{x^2+y^2}}, \frac{hy}{R\sqrt{x^2+y^2}}, -1\right).$$

In particular, $\nu_1(x,y,z) = \frac{hx}{\sqrt{R^2+h^2} \cdot \sqrt{x^2+y^2}}$.

We are now able to compute the very random surface integral.

$$\begin{aligned}
\int_M x\nu_1(x,y,z)\,d\mathcal{H}^2 &= \int_{B_2(0,R)} x\nu_1(x,y,g(x,y)) \cdot \frac{\sqrt{R^2+h^2}}{R}\,d(x,y) \\
&= \int_{B_2(0,R)} x\nu_1(x,y,g(x,y)) \cdot \frac{\sqrt{R^2+h^2}}{R}\,d(x,y) \\
&= \int_{B_2(0,R)} \frac{hx^2}{\sqrt{R^2+h^2}\sqrt{x^2+y^2}} \cdot \frac{\sqrt{R^2+h^2}}{R}\,d(x,y) \\
&= \frac{h}{R}\int_{B_2(0,R)} \frac{x^2}{\sqrt{x^2+y^2}\,d(x,y)} \\
&= \frac{h}{R}\int_0^R \int_0^{2\pi} \frac{r^2\cos^2\theta}{r} \cdot r\,d\theta\,dr \\
&= \boxed{\frac{1}{3}\pi R^2 h}
\end{aligned}$$

...this just happens to be the volume of the cone. This is not a coincidence.

■

11 The Divine Divergence Theorem

11.1 Stating the Theorem

Throughout analysis, we've often been focusing on *real-valued* functions, i.e. functions $\mathbb{R}^N \to \mathbb{R}$. Here we will focus more on *vector-valued functions* of the form $\mathbb{R}^N \to \mathbb{R}^N$. We also call these *vector fields*.

Definition 11.1 (Vector Field)

A *vector field* is a function $F : \mathbb{R}^N \to \mathbb{R}^N$.

...that's literally the entire definition.

We can think of a vector field as an assignment to each point in $\mathbb{R}^N$ a "direction" to move in. The "expandiness" induced by a vector field at a certain point is measured by the *divergence*.

Definition 11.2 (Divergence)

For a vector field $F : \mathbb{R}^N \to \mathbb{R}^N$, its *divergence* at a point $x \in \mathbb{R}^N$ is given by

$$\operatorname{div} F(x) := \sum_{i=1}^{N} \frac{\partial F_i}{\partial x_i}(x).$$

The notation $\nabla \cdot F$ is sometimes used instead of div, which is pretty sacrilegious but whatever. (Do you see where this notation comes from?)

Example 11.1: Let $F(x, y, z) := (3x + 2y + z, y^2, \log\log\log x)$. Then

$$\operatorname{div} F(x, y, z) = 3 + 2y + 0.$$

Example 11.2: Let $F(x, y, z) := (99x, 99y, 99z)$. Then $\operatorname{div} F = 297$, which makes sense because it's very expandy.

Example 11.3: If $F(x, y) := (999y, 999x)$ then $\operatorname{div} F = 0$. Intuitively this is consistent with the "expandiness" view of divergence because alot of the "movement" induced by F actually cancels out. For example, near $(0.1, 0.1)$ we have that F is pointing away from $(0, 0)$. On the other hand, at $(-0.1, 0.1)$, F is actually pointing *towards* $(0, 0)$.

Next, we need to define a notion of a boundary's "regularity". Intuitively, the boundary of an open set such as $B(0, 1)$ should be pretty smooth. But it's weird to talk about regularity of a *set*, considering that sets aren't functions. The workaround is to say that such boundaries are "locally functions".

Definition 11.3 (Regularity of Boundary)

Let $\Omega \subseteq \mathbb{R}^N$ be open and bounded. We say that $\partial\Omega$ is [some regularity class] if near every point on the boundary, we can zoom in and rotate our perspective so that the boundary looks like the graph of a function is [that regularity class].

To be more precise, for every $x_0 \in \partial\Omega$ there exists $r > 0$ so small that $T(B(x_0, r) \cap \Omega)$ is the supergraph of a [that regularity class] function for some rigid motion T.

The most typical choices for [some regularity class] are C^k and Lipschitz.

Example 11.4 (Regularity of $\partial B(0,1)$): At every $x_0 \in \partial B(0,1)$, the boundary near x_0 looks like the graph of $y \mapsto \sqrt{1 - \|y\|^2}$, which is smooth. Hence the boundary $\partial B(0,1)$ is of class C^∞.

Example 11.5 (Regularity of Cubic Boundaries): Let $Q = (0,1)^N \subseteq \mathbb{R}^N$ be a cube. ∂Q is *not* of class C^k for any $k \in \mathbb{N} \cup \{\infty\}$. This is because the graph that ∂Q "looks like" near the point $\vec{0} = (0, 0, \cdots, 0)$ will always have a sharp, non-differentiable bend.
This graph is, however, Lipschitz. Hence we may say that ∂Q is Lipschitz.

We are finally ready to state the Divergence Theorem.

Theorem 11.1 (Divergence Theorem)

Let $\Omega \subseteq \mathbb{R}^N$ be open and bounded such that $\partial\Omega$ is Lipschitz. Let $F : \overline{\Omega} \to \mathbb{R}^N$ be Lipschitz. Then

$$\int_\Omega \operatorname{div} F \, dx = \int_{\partial\Omega} F \cdot \nu \, d\mathcal{H}^{N-1},$$

where $\nu(x)$ is understood to be the *unit outward normal* at $x \in \partial\Omega$ to the manifold $\partial\Omega$.

This might sound like complete nonsense, but there is a way to think about the statement that makes it actually seem pretty intuitive.

Imagine lowering a cage into a pool of water, and suppose we start pumping *more* water into the cage (via a hose or something). The flow of water within the cage looks like a "source", with lots of water flowing "outwards" (i.e. "expandiness", or "divergence"). The amount of water added to the cage would just be the total "expandiness" of water. The water can't just keep accumulating in the cage though - a bunch of it has to escape the cage. The total amount of water that passes through the boundary of the cage is the amount that we put in. That's the divergence theorem.

11.2 Redoing the Previous Recitation

Example 11.6: Let $R > 0$. Let $H \subseteq \mathbb{R}^3$ be the hemisphere $H := B_3(0, R) \cap \{z > 0\}$. Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $f(x, y, z) := (z^2, y, 2^x)$. Let $\nu : \partial H \to \mathbb{R}^3$ be the unique unit outward normal vector at a given point on ∂H. Evaluate the surface integral

$$\int_{\partial H} f(x, y, z) \cdot \nu(x, y, z)\, d\mathcal{H}^2.$$

Solution. Note that $\operatorname{div} f(x, y, z) = 1$. Now apply the Divergence Theorem.

$$\int_{\partial H} f \cdot \nu\, d\mathcal{H}^2 = \int_H \operatorname{div} f\, dx = \int_H 1\, dx = \mathcal{L}^3(H) = \boxed{\frac{2}{3}\pi R^3}$$

■

Example 11.7: Let $R, h > 0$. Consider the cone $C := \{(x, y, z) : x^2 + y^2 \leq \frac{z^2 R^2}{h^2}, 0 \leq z \leq h\}$, so that C is a circular cone with base radius R and height h. Let M be the curved surface of C, and for each $x \in M$ let $\nu(x, y, z)$ denote the unit outward normal. Compute the surface integral

$$\int_M x\nu_1(x, y, z)\, d\mathcal{H}^2,$$

where ν_1 is the first component of $\nu = (\nu_1, \nu_2, \nu_3)$.

Solution. Apply the Divergence Theorem to the function $(x, y, z) \mapsto (x, 0, 0)$, whose divergence is exactly 1. Then

$$\int_M x\nu_1\, d\mathcal{H}^2 = \int_M (x, 0, 0) \cdot \nu\, d\mathcal{H}^2 = \int_C 1\, dx = \mathcal{L}^3(C) = \boxed{\frac{1}{3}\pi R^2 h}.$$

■

11.3 Redoing the Previous Previous Recitation

By the Divergence Theorem applied to the identity function $\vec{x} \mapsto \vec{x}$ (vector arrows for emphasis), we have that

$$\int_{B_N(0,1)} \operatorname{div} \vec{x}\, d\vec{x} = \int_{\partial B_N(0,1)} \vec{x} \cdot \nu(\vec{x})\, d\mathcal{H}^{N-1}.$$

But $\operatorname{div}\vec{x} = \operatorname{div}(x_1, x_2, \cdots, x_N) = \frac{\partial x_1}{\partial x_1} + \cdots + \frac{\partial x_N}{\partial x_N} = 1 + \cdots + 1 = N$. Also, $\nu(\vec{x}) = \vec{x}$, and $\vec{x} \cdot \vec{x} = \|\vec{x}\|^2 = 1$ for all $\vec{x} \in \partial B(0,1)$. Thus the equation turns into

$$\int_{B_N(0,1)} N \, d\vec{x} = \int_{\partial B_N(0,1)} 1 \, dx$$

or

$$N\mathcal{L}^N(B(0,1)) = \mathcal{H}^{N-1}(\partial B(0,1)).$$

...well that was easy.

11.4 Line Integrals

A *curve* is essentially some continuous path in $\mathbb{R}^N$. Unsurprinsgly you can integrate over these in the sense of integration with respect to $\mathcal{H}^1$. We call this a *line integral*.

Definition 11.4 (Line Integral of Real-Valued Functions)

For $f : \mathbb{R}^N \to \mathbb{R}$ and a curve $C \subseteq \mathbb{R}^N$, the integral $\int_C f \, d\mathcal{H}^1$ is called a *line integral*.

...were you expecting something more?

To evaluate a line integral, you find a parameterization $\varphi : [0, T] \to \mathbb{R}^N$ for the curve C, and then the integral is just $\int_0^T f(\varphi(t))\|\varphi'(t)\| \, dt$.

This is boring and we won't care about this for this recitation. What we're actually interested in is a different notion of line integral: That of *vector-valued functions* (i.e. a vector field $\mathbb{R}^N \to \mathbb{R}^N$) instead.

Here, it is important to assign a direction for the curve (are you going one way, or the other?) in order to get an *oriented curve*. A prototypical example is a circle: The orientation of a circle is either *counter-clockwise* or *clockwise*.

Definition 11.5 (Line Integral of Vector-Valued Functions)

For $F : \mathbb{R}^N \to \mathbb{R}^N$ a vector field and C an oriented curve, the *line integral* of F over C is given by

$$\int_C F := \int_0^T F(\varphi(t)) \cdot \varphi'(t) \, dt.$$

where $\varphi : [0, T] \to C$ is a parameterization of C respecting the chosen orientation for C.

- If you choose the opposite orientation, then line integral will be negated.

- The definition suggests that the line integral does not change if we choose a different φ (that still respects the chosen orientation). This is indeed true.
- In a sense, the line integral of a vector field measures how much the vector field "agrees" with your movement along the curve.

  Take, for instance, the vector field $F(x, y) := (-y, x)$. Let C be the unit circle *oriented counter-clockwise*. Then C is parameterized by $\varphi(t) := (\cos t, \sin t)$, and the line integral is given by

  $$\int_C F = \int_0^{2\pi} (-\sin t, \cos t) \cdot (-\sin t, \cos t)\, dt = 2\pi,$$

  suggesting that F *agrees alot* with the counter-clockwise movement along C. This makes perfect sense. In fact, this counter-clockwise movement literally follows the vector field F.

  What if we instead took the vector field $G(x, y) := (x, y)$? Then the vector field always runs perpendicularly to our movement, so this should suggest not much agreement. Indeed,

  $$\int_C G = \int_0^{2\pi} (\cos t, \sin t) \cdot (-\sin t, \cos t)\, dt = \int_0^{2\pi} -2\sin t \cos t\, dt = 0.$$

This discussion of "line integration measuring agreement" may suggest that we may write

$$\int_C F \overset{?}{=} \int_C F \cdot t\, d\mathcal{H}^1$$

where $t(x)$ is understood to be a unit tangent vector at x pointing in the same direction as the orientation of C. This is true, provided that C is nice enough. If C is an *absolutely continuous* curve, then it turns out that we can move along C at a constant speed. That is, we may find a parameterization $\varphi : [0, T] \to C$ such that $\|\varphi'(s)\| = 1$ for all s. Then the tangent vector $t(x)$ at a point $x \in C$ should be "the derivative of φ at x", i.e. $\varphi'(\varphi^{-1}(x))$. This is indeed of unit length by choice of φ. Now,

$$\int_C F \cdot \varphi'(\varphi^{-1})\, d\mathcal{H}^1 = \int_0^T F(\varphi(s))\varphi'(\varphi^{-1}(\varphi(s)))\|\varphi'(s)\|\, ds = \int_0^T F(\varphi(s))\varphi'(s)\, ds = \int_C F.$$

A brief remark on notation: You'll often see

$$\int_C F =: \int_C F_1\, dx + F_2\, dy.$$

I guess the logic behind this is that we can write

$$\int_C F = \int_0^T F_1(\varphi(t))\varphi_1'(t) + F_2(\varphi(t))\varphi_2'(t)\, dt,$$

and so essentially we're taking $\int_C M\,dx := \int_0^T M(\varphi(t))\varphi_1'(t)\,dt$ and $\int_C N\,dx := \int_0^T N(\varphi(t))\varphi_2'(t)\,dt$. You're supposed to read "$M\,dx$" as "as we move along the curve, multiply M(my location) by our change in x", and likewise for "$N\,dy$".

This kinda makes sense. We can view $M\,dx + N\,dy$ as the "dot product" $(M, n)\cdot(dx, dy)$. Intuitively (dx, dy) is the direction you're moving in, so we're essentially measuring the agreement between $(-y, x)$ and the direction we're moving, i.e. the line integral $\int_C (M, N)$.

(There ***is*** *a rigorous meaning behind writing $M\,dx + N\,dy$, but it is above my paygrade and beyond my understanding.)*

11.5 Green's Theorem

I think I will basically never have Green's Theorem memorized and I'll be doomed to forever keep rederiving it whenever I need it on an exam. And that's ok. This section reflects this sentiment by not stating Green's Theorem until it is proven.

To remember Green's Theorem, you essentially need to remember two things:

1. It's something about relating a line integral in $\mathbb{R}^2$ to an integral of the enclosed region.
2. It's a stupid consequence of the Divergence Theorem.

Alright, let $\Omega \subseteq \mathbb{R}^2$ be open and bounded. $\partial\Omega$ is a curve, so let's orient it counter-clockwise (and let's just assume the boundary is nice enough). Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a vector field. Supposedly Green's Theorem gives a way to simplify $\int_{\partial\Omega} F$.

To do it, we blindly use the Divergence Theorem on F. This gives

$$\int_\Omega \frac{\partial F_1}{\partial x}(x,y) + \frac{\partial F_2}{\partial y}(x,y)\,d(x,y) = \int_{\partial\Omega} F_1(x,y)\nu_1(x,y) + F_2(x,y)\nu_2(x,y)\,d\mathcal{H}^2. \qquad (*)$$

We somehow want to turn the right side into a line integral. How? Well, from the discussion in the previous section, we'd like to write it as $\int_{\partial\Omega} F \cdot t\,d\mathcal{H}^2$, where t is the *tangent* vector ("pointing counter-clockwise"). How can we get the tangent vector from the normal vector's components?

...we can just rotate it 90° can't we? Observe that if (ν_1, ν_2) is the (unit outward) normal vector, then $(-\nu_2, \nu_1)$ is the normal vector rotated 90° counter-clockwise, and hence must be the (unit) tangent vector oriented counter-clockwise.

This tells us exactly what we need to do now. In equation $(*)$, we just need to replace the F_2 with $-F_1$, and replace the F_1 with F_2. That is, let's instead *apply the Divergence*

Theorem to $(F_2, -F_1)$. This gives

$$\int_\Omega \frac{\partial F_2}{\partial x}(x,y) - \frac{\partial F_1}{\partial y}(x,y)\, d(x,y) = \int_{\partial\Omega} F_2(x,y)\nu_1(x,y) - F_1(x,y)\nu_2(x,y)\, d\mathcal{H}^2$$

$$= \int_{\partial\Omega} (F_1, F_2) \cdot (-\nu_2, \nu_1)\, d\mathcal{H}^2 = \int_{\partial\Omega} F \cdot t\, d\mathcal{H}^2 = \int_{\partial\Omega} F.$$

We have thus derived Green's Theorem.

Theorem 11.2 (Green's Theorem)

Let $\Omega \subseteq \mathbb{R}^2$ be open and bounded with Lipschitz boundary. Let $F : \overline{\Omega} \to \mathbb{R}^2$ be a Lipschitz vector field. Then

$$\int_{\partial\Omega} F = \int_\Omega \frac{\partial F_2}{\partial x}(x,y) - \frac{\partial F_1}{\partial y}(x,y)\, d(x,y),$$

where $\partial\Omega$ is taken to have counter-clockwise orientation.

Written in alternative notation:

$$\int_{\partial\Omega} M\, dx + N\, dy = \int_\Omega \frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\, dA$$

By choosing F_1, F_2 in a dumb way, Green's Theorem can be used to find area in an incredibly dumb way.

Corollary 11.1

Let $\Omega \subseteq \mathbb{R}^2$ be open and bounded with Lipschitz boundary. Then

$$\begin{aligned} \mathcal{L}^2(\Omega) &= \int_{\partial\Omega} (0, x) && \left(= \int_{\partial\Omega} x\, dy\right) \\ &= \int_{\partial\Omega} (-y, 0) && \left(= \int_{\partial\Omega} -y\, dx\right) \\ &= \int_{\partial\Omega} \left(\frac{-y}{2}, \frac{x}{2}\right) && \left(= \int_{\partial\Omega} \frac{-y}{2}\, dx + \frac{x}{2}\, dy\right) \end{aligned}$$

where all line integrals are taken counter-clockwise.

Example 11.8 (Area of a Cartioid): Find the area of the region U enclosed by the curve parameterized by

$$\begin{cases} x(\theta) = (1 - \cos\theta)\cos\theta, \\ y(\theta) = (1 - \cos\theta)\sin\theta, \end{cases} \qquad \theta \in [0, 2\pi].$$

Solution. Let's use the formula $\mathcal{L}^2(U) = \int_{\partial U}(0, x)$. The parameterization

$$\varphi(\theta) := ((1-\cos\theta)\cos\theta, (1-\cos\theta)\sin\theta)$$

has been given to us on a silver platter, so it remains to compute.

$$\begin{aligned}
\int_{\partial U}(0,x) &= \int_0^{2\pi} (0, (1-\cos\theta)\cos\theta)\cdot\varphi'(\theta)\,d\theta \\
&= \int_0^{2\pi} (1-\cos\theta)\cos\theta\varphi_2'(\theta)\,d\theta \\
&= \int_0^{2\pi} (1-\cos\theta)\cos\theta\Big((1-\cos\theta)\cos\theta + \sin\theta\sin\theta\Big)\,d\theta \\
&= \int_0^{2\pi} \cos\theta - 3\cos^3\theta + 2\cos^4\theta\,d\theta \\
&= \int_0^{2\pi} 2\cos^4\theta\,d\theta \\
&= \boxed{\frac{3\pi}{2}}
\end{aligned}$$

■

11.6 Examples from the Math GRE

> **Example 11.9:** What is the value of the flux of the vector field $F(x, y, z) = x^2\hat{\imath} - 2xy\hat{\jmath} + x^2y^2\hat{k}$ through the surface $z = \sqrt{4 - x^2 - y^2}$ oriented upwards?

Solution. The main difficulty is understanding the terrible notation that the world has decided to accept. The flux is just the surface integral dotted with some choice of normal vectors over the surface. $F(x, y, z) = x^2\hat{\imath} - 2xy\hat{\jmath} + x^2y^2\hat{k}$ literally just means $F(x, y, z) = (x^2, -2xy, x^2y^2)$. By oriented upwards, we mean that we choose the normal vectors that point upwards, i.e. outwards.

Let H be the obvious hemisphere, and let H_0, H^+ be the bottom disk and curved surface of H, respectively. The question is asking for $\int_{H^+} F\cdot\nu\,d\mathcal{H}^2$. By the Divergence Theorem,

$$\int_{H^+} F\cdot\nu\,d\mathcal{H}^2 + \int_{H_0} F\cdot\nu\,d\mathcal{H}^2 = \int_{\partial H} F\cdot\nu\,d\mathcal{H}^2 = \int_H \operatorname{div} F\,d(x,y,z) = 0,$$

so the answer is whatever $-\int_{H_0} F\cdot\nu\,d\mathcal{H}^2$ is. But over H_0 we have that $\nu = (0, 0, -1)$, hence

$$\int_{H_0} F\cdot\nu\,d\mathcal{H}^2 = \int_{H_0} -x^2y^2\,d\mathcal{H}^2 = \int_{B_2(0,2)} -x^2y^2\,d(x,y)$$

$$= \int_0^2 \int_0^{2\pi} -r^4 \cos^2\theta \sin^2\theta \, dr \, d\theta = \frac{-\pi}{20}.$$

The final answer is $\boxed{\dfrac{\pi}{20}}$. ■

Example 11.10: Let C be the ellipse with center $(0,0)$, major axis of length $2a$, and minor axis of length $2b$. What is the value of $\oint_C x\,dy - y\,dx$?

(It didn't say it on the test but surely we assume that this is oriented counter-clockwise.)

Solution. The main difficulty is understanding the terrible notation that the world has decided to accept. First of all, $\oint_C$ is just $\int_C$ if you're trying to look intelligent. Second, the $x\,dy - y\,dx$ is just representing the the line integral

$$\int_C (-y, x).$$

By Green's Theorem, this is just

$$\int_{\text{Ellipse}} 1 - (-1)\, d(x,y) = 2\mathcal{L}^2(\text{Ellipse}) = \boxed{2ab\pi}.$$

■

11.7 Integration by Parts

Let $f, g : \mathbb{R}^N \to \mathbb{R}$. Suppose we're staring at an integral of the form

$$\int_\Omega \frac{\partial f}{\partial x_i} g \, dx.$$

The goal here is to find a way to yeet the derivative over to the g, as in the integration by parts theorem in one dimension.

If you ponder it, it's quite natural to try applying the Divergence Theorem to the vector field given by

$$F(x) := f(x)g(x)e_i.$$

Then $\operatorname{div} F = \frac{\partial}{\partial x_i} f(x)g(x) = \frac{\partial f}{\partial x_i}(x)g(x) + f(x)\frac{\partial g}{\partial x_i}(x)$. It follows that

$$\int_\Omega \frac{\partial f}{\partial x_i}(x)g(x) + f(x)\frac{\partial g}{\partial x_i}(x)\, dx = \int_\Omega \operatorname{div} F(x)\, dx$$

$$= \int_{\partial\Omega} F(x) \cdot \nu(x)\, d\mathcal{H}^{N-1} = \int_{\partial\Omega} f(x)g(x)\nu(x)e_i \, d\mathcal{H}^{N-1} = \int_{\partial\Omega} f(x)g(x)\nu_i(x)\, d\mathcal{H}^{N-1}$$

We have hence derived integration by parts.

Theorem 11.3 (Integration by Parts)

Suppose $\Omega \subseteq \mathbb{R}^N$ is open and bounded with Lipschitz boundary, and suppose $f, g : \overline{\Omega} \to \mathbb{R}$ are Lipschitz. Then

$$\int_\Omega \frac{\partial f}{\partial x_i} g \, dx = \int_{\partial\Omega} f g \nu_i \, d\mathcal{H}^{N-1} - \int_\Omega f \frac{\partial g}{\partial x_i} \, dx.$$

Plugging in $g = 1$ gives the following corollary.

Corollary 11.2

$$\int_\Omega \frac{\partial f}{\partial x_i} \, dx = \int_{\partial\Omega} f \nu_i \, d\mathcal{H}^{N-1}$$

Plugging in a "test function" for g that vanishes on the boundary will give another corollary.

Corollary 11.3

- We have that
$$\int_\Omega \frac{\partial f}{\partial x_i} \varphi \, dx = - \int_\Omega f \frac{\partial \varphi}{\partial x_i} \, dx$$
where φ is a function that vanishes on the boundary, i.e. $\varphi(x) = 0$ for all $x \in \partial\Omega$.
- Taking this further: For $f : \mathbb{R}^N \to \mathbb{R}$ of class C^k and $\varphi \in C_0^k(\Omega)$, i.e. φ is class C^k and satisfies $\varphi(x) = 0$ for all $x \in \partial\Omega$, we have the identity
$$\int_\Omega \frac{\partial^k f}{\partial x_i^k} \varphi \, dx = (-1)^k \int_\Omega f \frac{\partial^k \varphi}{\partial x_i^k} \, dx.$$
- Taking this even further, we have that
$$\int_\Omega \partial^\alpha f \varphi \, dx = (-1)^{|\alpha|} \int_\Omega f \partial^\alpha \varphi \, dx.$$
Here, α denotes a multi-index.

Proof. Induction. ∎

If you *really really want to* you can also write something like this.

Corollary 11.4 (help what is a tensor)

For $f : \mathbb{R}^N \to \mathbb{R}$ we have

$$\int_\Omega \nabla f \, dx = \int_{\partial\Omega} f\nu \, d\mathcal{H}^{N-1}.$$

For $f : \mathbb{R}^N \to \mathbb{R}^d$ we have

$$\int_\Omega Df \, dx = \int_{\partial\Omega} f \otimes \nu \, d\mathcal{H}^{N-1}.$$

11.8 Divergence is Independent of Basis

It may seem strange that the divergence operator is so tied to the standard basis $x_1, \cdots, x_N$. How does accounting for "expandiness" in the $\{x_i\}_i$ directions give us the general "expandiness" in all the directions? This oddity would be explained if, by some miracle, choosing a different set of basis vectors would not change the divergence. This happens to be true.

Theorem 11.4 (Divergence is Coordinate-Free)

Let $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$ be bases for $\mathbb{R}^N$. Then for any $F : \mathbb{R}^N \to \mathbb{R}^N$ we have

$$\sum_{i=1}^N \frac{\partial F_i}{\partial x_i} = \sum_{i=1}^N \frac{\partial F_i}{\partial y_i}.$$

Proof. The proof, in short, is:

1. The divergence is the trace of the Jacobian matrix.
2. Trace is independent of choice of basis.

Let the $N \times N$ matrix DF be the Jacobian matrix with respect to the $\{x_i\}$ basis. Then the Jacobian matrix with respect to the $\{y_i\}$ basis is $A^{-1}(DF)A$, where A is the a matrix that sends the $\{x_i\}$ basis to the $\{y_i\}$ basis.

(If you're skeptical, try going back to the definition of differentiability + the Jacobian matrix in order to show that the differential of the function $x \mapsto A^{-1}F(Ax)$ is just $A^{-1}(DF)A$.)

But now

$$\sum_{i=1}^N \frac{\partial F_i}{\partial y_i} = \operatorname{tr}(A^{-1}(DF)A) = \operatorname{tr}((DF)AA^{-1}) = \operatorname{tr} DF = \sum_{i=1}^N \frac{\partial F_i}{\partial x_i}.$$

□

12 Special Vector Fields

12.1 Conservative vs. Irrotational

Definition 12.1 (Conservative)

1. A vector field $F : \Omega \to \mathbb{R}^N$ is *conservative* if it is a gradient. That is, there exists $f : \Omega \to \mathbb{R}$ such that
$$F = \nabla f.$$
2. An equivalent property is that
$$\int_C F = 0$$
for every (Lipschitz) closed curve C.
3. A (trivially...?) equivalent property is that for any $x, y \in \Omega$, and any two (Lipschitz) paths γ_1, γ_2 from x to y, we have that
$$\int_{\gamma_1} F = \int_{\gamma_2} F.$$

Example 12.1: The vector field $F(x, y, z) = (y^2z^3, 2xyz^3, 3xy^2z^2)$ is conservative because $F = \nabla f$ where $f(x, y, z) = xy^2z^3$.

Example 12.2: The force of gravity is conservative. For a point mass $x_0 \in \mathbb{R}^N$ with mass M, we may construct a vector field $F_g : \mathbb{R}^N \setminus \{x_0\} \to \mathbb{R}^N$ representing the gravitational pull on a mass m induced by x_0 via

$$F_g(x) := \frac{GMm}{\|x - x_0\|^2} \cdot -\frac{x - x_0}{\|x - x_0\|} = \frac{-GMm(x - x_0)}{\|x - x_0\|^3},$$

where G is the universal gravitational constant. To see that F_g is a conservative vector field, we must find $-U_g : \mathbb{R}^2 \setminus \{x_0\} \to \mathbb{R}$ for which $\nabla(-U_G) = F_g$. Indeed, we can take

$$-U_g(x) := \frac{GMm}{\|x - x_0\|}.$$

To convince yourself that $\nabla(-U_g) = F_g$, you can either just manually compute the patital derivatives or use some v e c t o r c a l c u l u s via the chain rule.

$$\nabla(-U_g) = \left(\frac{-GMm}{\|x - x_0\|^2}\right) \nabla(\|x - x_0\|) = \frac{-Gm}{\|x - x_0\|^2} \cdot \frac{x - x_0}{\|x - x_0\|}$$

Thankfully this is consistent with our understanding of physics: $U_g(x) = \frac{-GMm}{\|x - x_0\|}$ is the gravitational potential energy.

Two intuitions for conservative vector fields:

1. A conservative vector field F must be able to represent the “upward slope”-ness of some function $f : \Omega \to \mathbb{R}$. This is what it means for F to be the gradient of f, after all. As in the gravity example, you can imagine F_g to be a bunch of arrows pointing towards the mass x_0, and indeed these arrows are all pointing “upward” when placed on a “mountain” whose peak is at x_0. This mountain is the graph of $-U_g$.

   (We take $-U_g$ because in physics/nature, things like to go down in energy, rather than up.)

2. You can think of a conservative F as a force, where if I move along any loop, there must be moments in which I move *against* the force and moments where I move *with* the force. As with gravity, “what goes up must come down”.

   This thinking leads to an easy example of a non-conservative vector field: Take $F(x, y) := (-y, x)$. Then if I move around the unit circle counter-clockwise, I'm always moving with the force, which would not be possible if F were conservative.

Definition 12.2 (Irrotational)

A vector field $F : \Omega \to \mathbb{R}^N$ is *irrotational* or *curl-free* if

$$\frac{\partial F_i}{\partial x_j} = \frac{\partial F_j}{\partial x_i}$$

for all i, j.

To see why we call this "curl-free", let's define a notion of curl.

Definition 12.3 (Curl)

For a vector field F on $\mathbb{R}^2$, its *curl* is a scalar function given by

$$\operatorname{curl} F := \frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}.$$

For a vector field F on $\mathbb{R}^3$, its curl is a vector field given by

$$\operatorname{curl} F := \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}, \frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}, \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right).$$

For higher dimensions it's some kind of unholy garbage.

(Ominous Foreshadowing: Isn't $\operatorname{curl} F$ *for* $N = 2$ *the expression you see in Green's Theorem? Hmm...)*

Now it's clear: For $N = 2, 3$, it is evident that a vector field F is curl-free exactly when $\operatorname{curl} F = 0$. Just don't ask me what happens in higher dimensions.

In the interest of memorizing the $N = 3$ formula, we note that $\operatorname{curl} F$ may be "written" as the "determinant"

$$\operatorname{curl} F = \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_1 & F_2 & F_3 \\ \hat{\imath} & \hat{\jmath} & \hat{k} \end{vmatrix}$$

where $\hat{\imath} = (1, 0, 0)$, et cetera. The word "determinant" is in quotes because this isn't actually a determinant (algebraically speaking, the elements of a matrix should come from a common field!).

If you're having trouble remembering the order of the rows in this "determinant", we note that sometimes, due to this "determinant" form, the notation

$$\nabla \times F := \operatorname{curl} F$$

is used for the curl. Indeed, the "determinant" does kinda look like a cross product between "differentiating" and "the vector field F". So as long as you know how to take a cross product, you can memorize this easily, more or less.

The intuition for curl is that it measures "rotatiness" around a point. A good way to visualize $\operatorname{curl} F(x_0)$ is to place a fixed, physical spinning ball centered at x_0 and let F, as a force, act on this ball, where the ball is free to rotate but cannot move otherwise. If $\operatorname{curl} F(x_0) = 0$ then the ball is still. Otherwise it rotates in a way dictated by the value of $\operatorname{curl} F(x_0)$. See `https://mathinsight.org/curl_components` for pretty pictures and a better explanation.

An important property is as follows.

Theorem 12.1 (Conservative Implies Irrotational)

Every C^1 conservative vector field is irrotational.

Proof. Basically the C^1 property allows us to swap derivatives.

If F is conservative then $F = \nabla f$ for some f. Note then that $F_i = \frac{\partial f}{\partial x_i}$. Since F is C^1, f is C^2, so

$$\frac{\partial F_i}{\partial x_j} = \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial F_j}{\partial x_i}.$$

$\square$

So *being conservative is a stronger property than being irrotational*, provided that the vector field is nice enough.

(Remark: This property is often written as "$\operatorname{curl} \nabla F = 0$" or "$\nabla \times (\nabla F) = 0$". Do you see why?)

We also have the other direction (!) provided that the *domain* is nice enough!

Theorem 12.2 (Irrotational Implies Conservative)

A C^1 irrotational vector field $F : \Omega \to \mathbb{R}^N$ is conservative, **provided that Ω is *simply connected***.

"Simply connected" is a new term. Roughly speaking, an open set is *simply connected* if it is in one piece (i.e. connected) and it has no holes. For example, a donut is connected but not simply connected.

Here is a slightly more precise definition.

Definition 12.4 (Simply Connected)

A connected set U is simply connected if the following property holds: For any loop we draw inside U, we can continuously *shrink* the loop until it becomes a point.

(Note that consequently, the punctured disk $B_2(0,1) \setminus \{(0,0)\}$ is not simply connected, witnessed by drawing a loop around $(0,0)$ and trying to shrink it. However, the punctured sphere $B_3(0,1) \setminus \{(0,0,0)\}$ *is* simply connected.)

The following example shows that simply-connectedness is crucial.

Example 12.3: Consider the vector field $F : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}^2$ defined as

$$F(x,y) := \left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2} \right).$$

Intuitively, the vector field F is made by drawing a unit direction vector "counter-clockwise" at every point.
To verify that F is irrotational, we need only show that $\operatorname{curl} F = 0$. Indeed,

$$\operatorname{curl} F(x,y) = \frac{\partial F_1}{\partial y}(x,y) - \frac{\partial F_2}{\partial x}(x,y)$$

$$= \frac{-(x^2+y^2) - (-y)(2y)}{(x^2+y^2)^2} - \frac{(x^2+y^2) - x(2x)}{(x^2+y^2)^2} = 0.$$

However, F is not conservative! Intuitively this is because going around the origin can be done without every moving against the force F. If you want to do the math, we can. Let C be the unit circle oriented clockwise. Then we may compute

$$\int_C F = \int_0^{2\pi} F(\cos t, \sin t) \cdot (-\sin t, \cos t)\, dt$$

$$= \int_0^{2\pi} (-\sin t, \cos t) \cdot (-\sin t, \cos t)\, dt = 2\pi \neq 0.$$

An interesting observation is that $(-y, x)$ has curl, but $\left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$ does not. The rescaling is crucial to eliminate curl. The "rotating ball" may help you visualize why that is (though, note that the division by x^2+y^2 doesn't actually make the norm exactly 1...).

13 Wrap-Up

13.1 Some Things We Skipped

- Divergence does not depend on basis
- GRE problem on Divergence Theorem
- Finding the potential function

13.2 Finding the Potential Function

Example 13.1: Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined via

$$F(x, y) = (ye^x + \sin y, e^x + x \cos y).$$

Is F conservative? If so, find its gradient potential, i.e. a function $f : \mathbb{R}^2 \to \mathbb{R}$ for which $\nabla f = F$.

Solution. $\frac{\partial F_1}{\partial y}(x, y) = e^x + \cos y$ and $\frac{\partial F_2}{\partial x}(x, y) = e^x + \cos y$, so F is irrotational. Since $\mathbb{R}^2$ is simply connected and F is C^1, we may conclude that F is conservative. But this doesn't tell us what f is.

To figure that out, we suppose that $F = \nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$. Then we have the system of equations:

$$\frac{\partial f}{\partial x}(x, y) = ye^x + \sin y$$
$$\frac{\partial f}{\partial y}(x, y) = e^x + x \cos y$$

None of the following manipulations are necessarily rigorous. Remember, we just need to propose a correct f in order to win. The way we arrive at the answer doesn't really matter. To wit, let us start by applying FTC to the first equation to get

$$f(x, y) - f(0, y) = \int_0^x ye^x + \sin y \, dx = ye^x + x \sin y - y.$$

We just need to figure out what $f(0, y)$ is. To do this we differentiate with respect to y to get

$$\frac{\partial f}{\partial y}(x, y) - \frac{\partial f}{\partial y}(0, y) = e^x + x \cos y - 1.$$

But we know what $\frac{\partial f}{\partial y}(x, y)$ is! So

$$e^x + x \cos y - \frac{\partial f}{\partial y}(0, y) = e^x + x \cos y - 1$$

and hence $\frac{\partial f}{\partial y}(0,y) = 1$. By the FTC yet again, $f(0,y) = y + f(0,0)$. Therefore

$$f(x,y) = f(0,y) + ye^x + x\sin y - y = \boxed{ye^x + x\sin y + f(0,0)}.$$

$f(0,0)$ can be anything we want it to be, which makes sense because the potential function is unique only up to a constant. So we can e.g. just take $f(0,0)$ to be 0. ■

Example 13.2: Define a vector field $F : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}^2$ via

$$F(x,y) := \left(\frac{2xy}{x^4+y^2}, \frac{-x^2}{x^4+y^2}\right).$$

Is F conservative?

Solution. As a first attempt, let's check if F is irrotational. If it isn't then we can immediately say "no" because F is C^1, so conservative must imply irrotational.

...drat. $\frac{\partial F_1}{\partial y} = \frac{\partial F_2}{\partial x}$, so F is irrotational. This doesn't necessarily tell us that F is conservative, because the domain isn't simply connected!

Fine. We'll have to do this the hard way. We will find $f : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$ satisfying the following system of differential equations:

$$\frac{\partial f}{\partial x}(x,y) = \frac{2xy}{x^4+y^2}$$
$$\frac{\partial f}{\partial y}(x,y) = \frac{-x^2}{x^4+y^2}$$

It's probably easier to integrate the second equality here. I'll do the first one instead to highlight a possible error. We have

$$f(x,y) = f(0,y) + \int_0^x \frac{2ty}{t^4+y^2}\,dt = f(0,y) + \tan^{-1}\left(\frac{x^2}{y}\right).$$

(Boring u-substitutions omitted.) Now we need to find $f(0,y)$. We differentiate with respect to y to obtain

$$\frac{-x^2}{x^4+y^2} = \frac{\partial f}{\partial y}(x,y) = \frac{\partial f}{\partial y}(0,y) - \frac{x^2}{x^4+y^2},$$

so $\frac{\partial f}{\partial y}(0,y) = 0$ and hence $f(0,y)$ is constant in y.

...or is it? Actually, note that the domain of $y \mapsto f(0,y)$ is $(-\infty,0) \cup (0,\infty)$, so $f(0,y)$ isn't necessarily constant. We can only say that it is *piece-wise* constant. Choosing the correct constant values for $f(0,y)$ is not too bad, so I'll just spoil the fun and say that we can take

$$\boxed{f(x,y) = \begin{cases} \tan^{-1}\left(\frac{x^2}{y}\right) - \frac{\pi}{2}, & y > 0 \\ 0, & y = 0 \\ \tan^{-1}\left(\frac{x^2}{y}\right) - \frac{\pi}{2}, & y < 0 \end{cases}}.$$

If we instead started by integrating the second equality, we would have gotten something that looks more like

$$\boxed{f(x,y) = -\tan^{-1}\left(\frac{y}{x^2}\right)},$$

and it turns out that this is the same function (!).

Of course, now that we have what the potential *has* to be provided that it exists, you should verify that it actually works. ■

We do one last example.

Example 13.3: Consider the vector field

$$F(x,y,z) = (2z^4 - 2y - y^3, z - 2x - 3xy^2, 6 + y + 8xz^3).$$

Is F conservative? If so, find its gradient potential.

Solution. After some boring computation we see that indeed curl $F = 0$, so F is irrotational, and since F is C^1 and $\mathbb{R}^3$ is simply-connected, we have that F is conservative. Brilliant.

Now we want to actually find the gradient potential f. We need to solve a system of three differential equations:

$$\frac{\partial f}{\partial x}(x,y,z) = 2z^4 - 2y - y^3$$
$$\frac{\partial f}{\partial y}(x,y,z) = z - 2x - 3xy^2$$
$$\frac{\partial f}{\partial z}(x,y,z) = 6 + y + 8xz^3$$

Welp. Let's just start by integrating the z equation I guess.

$$f(x,y,z) = f(x,y,0) + 6z + yz + 2xz^4 \qquad (*)$$

On one hand, if we differentiate $(*)$ with respect to x, we get

$$2z^4 - 2y - y^3 = \frac{\partial f}{\partial x}(x,y,0) + 2x^4$$

so that $\frac{\partial f}{\partial x}(x,y,0) = -2y - y^3$. On the other hand, if we differentiate $(*)$ with respect to y, we get

$$z - 2x - 3xy^2 = \frac{\partial f}{\partial y}(x,y,0) + z$$

so that $\frac{\partial f}{\partial y}(x,y,0) = -2x - 3xy^2$.

Hey wait a minute, now the equations

$$\frac{\partial f}{\partial x}(x,y,0) = -2y - y^3$$

$$\frac{\partial f}{\partial y}(x, y, 0) = -2x - 3xy^2$$

form a system of equations for finding the gradient potential for the vector field $(-2y - y^3, -2x - 3xy^2)$! And we know how to do this. Let's integrate the y equation.

$$f(x, y, 0) = f(x, 0, 0) - 2xy - xy^3$$

Now differentiate with respect to x.

$$-2y - y^3 = \frac{\partial f}{\partial x}(x, 0, 0) - 2y - y^3$$

Oh wow, so $\frac{\partial f}{\partial x}(x, 0, 0) = 0$ and $f(x, 0, 0)$ is just a constant, so let's just take it to be 0. Then

$$f(x, y, 0) = -2xy - xy^3$$

and so

$$f(x, y, z) = f(x, y, 0) + 6z + yz + 2xz^4 = \boxed{-2xy - xy^3 + 6z + yz + 2xz^4}.$$

(This time I don't think we need to actually verify this works, because we *know* F is conservative so there *has* to be a gradient potential, and if there exists a gradient potential, it *must* be this one.) ■

Note: If you're reading this I just realized that this process can be streamlined slightly. But I have a headache so I don't feel like deleting half the page.

14 (Appendix) Stoke's Theorem

In a previous recitation, we derived **Green's Theorem**, which states that

$$\int_{\Omega} \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\, d(x,y) = \int_{\partial\Omega} F.$$

This can also be viewed as the identity

$$\int_{\Omega} \operatorname{curl} F\, d(x,y) = \int_{\partial\Omega} F.$$

Hence, we may view Green's Theorem has a statement that "the total swirliness inside the set is equal to the swirly along the outside".

Stoke's Theorem says that this is still true, even if the set isn't completely flat (i.e. $\subseteq \mathbb{R}^2$). That is, this "swirly equation" still holds if the set in question is only *locally* flat, i.e. a 2D manifold.

Making this statement rigorous requires us to answer an essential question however: What does it mean to integrate along the "boundary of a manifold"? Heck, what even is the "boundary" of a manifold?

14.1 Manifolds with Boundary

Recall the definition of a differentiable manifold.

Definition 14.1 (Manifold)

A set $M \subseteq \mathbb{R}^N$ is a k-dimensional manifold of class C^m if for every $x_0 \in M$ there exists a C^m homeomorphism $\varphi : V \to M \cap U$ where $V \subseteq \mathbb{R}^k$ is open and $U \subseteq \mathbb{R}^N$ is open and contains x_0, such that $D\varphi$ has rank k.

Essentially V is "where we are charting out a map of M". It is required that V is open, it can't have an edge. A manifold with *boundary*, however, relaxes this requirement on V, allowing it to have an edge.

Definition 14.2 (Manifold with Boundary)

A set $M \subseteq \mathbb{R}^N$ is a k-dimensional manifold of class C^m if for every $x_0 \in M$ there exists a C^m homeomorphism $\varphi : V \to M \cap U$ where $V \subseteq \overline{\mathbb{R}^k_+}$ is **relatively open** in $\overline{\mathbb{R}^k_+}$ and $U \subseteq \mathbb{R}^N$ is open and contains x_0, such that $D\varphi$ has rank k.

Here, $\overline{\mathbb{R}^k_+} = \{(\vec{x}, x) \in \mathbb{R}^k : x \geq 0\}$, where we view $\mathbb{R}^k$ as $(\mathbb{R}^{k-1} \times \mathbb{R})$.

Example 14.1: The hemisphere $\{(x,y,z) : x^2+y^2+z^2 = 1, z > 0\}$ is a manifold. But it does not have boundary.
The "closed" hemisphere $\{(x,y,z) : x^2+y^2+z^2 = 1, z \geq 0\}$ is a manifold with boundary. This is because if I stand on the "edge" of the hemisphere, i.e. at a point $(x,y,0)$ with $x^2+y^2 = 1$, then the part of the manifold near me looks like the upper part of a plane, i.e. $\overline{\mathbb{R}^2_+}$.

The definition leads to a natural notion of the boundary a manifold with boundary.

Definition 14.3 (Boundary)

Let M be a manifold with boundary, let $x_0 \in M$ and find an appropriate chart $\varphi : V \to U \cap M$ for which $x_0 \in U$. Let $y_0 \in V$ be such that $\varphi(y_0) = x_0$. If $y_0 = (\vec{y}, 0)$ (i.e. y_0 lies on the "edge" of the set $\overline{\mathbb{R}^k_+}$) then we say that x_0 is a *boundary point* of M. The set of M's boundary points is denoted by ∂M, and is called the *boundary* of M.

An exercise is to prove that the definition of a boundary point doesn't depend on which chart we use.

In Stoke's Theorem, we ultimately want to integrate over the boundary of a manifold M. A priori, this isn't very well-defined. We can be at ease with the following theorem, though!

Theorem 14.1 (Boundary of a Manifold is a Manifold)

Let $k \geq 2$. Let $M \subseteq \mathbb{R}^N$ be a k-dimensional differentiable manifold with boundary. Then ∂M is a $k-1$-dimensional differentiable manifold.
Moreover, if M is orientable then so is ∂M.

In particular, when $N = 3$, then ∂M is a 1-dimensional manifold, i.e. a path in $\mathbb{R}^3$. In fact it will be a looping path. When M is orientable, then ∂M will be orientable, so we can view ∂M as going in one of two directions.

Now we can state Stoke's.

Theorem 14.2 (Stoke's Theorem)

Let $M \subseteq \mathbb{R}^3$ be a compact C^1 manifold with boundary. Then

$$\int_M \operatorname{curl} F \cdot \nu \, d\mathcal{H}^2 = \int_{\partial M} F,$$

where

- $F : \Omega \to \mathbb{R}^3$ is a C^1 vector field defined on an open set Ω containing M,
- ∂M is an oriented curve, and
- the continuous unit outward normal ν is chosen such that, if we view $\nu(y)$ as pointing "up", then the orientation of ∂M is such that ∂M is "travelling counter-clockwise" around $\nu(y)$.

Ensuring that the directions we choose for ν and ∂M are compatible is crucial.

Example 14.2: Let $M = \{(x, y, z) : x^2 + y^2 + z^2 = 1, z \geq 0\}$ be a manifold with boundary. Consider the vector field $F(x, y, z) := (xy, yz, xz)$. What is

$$\int_M \operatorname{curl} F \cdot \nu \, d\mathcal{H}^2,$$

where ν is the normal pointing **up** or **outward**, i.e. $\nu(x, y, z) = (x, y, z)$?

Solution. We want to apply Stoke's. Because ν is pointing up, the direction we want to assign the boundary $\partial M = \{(x, y, 0) : x^2 + y^2 = 1\}$ is the counter-clockwise one, given by the parameterization $\varphi(t) := (\cos t, \sin t, 0)$. (If instead ν were pointing down, i.e. $\nu(x, y, z) = (-x, -y, -z)$, then we would have to choose the clockwise orientation.)

By Stoke's we just need to compute the line integral $\int_{\partial M} F$. This is given by

$$\int_0^{2\pi} F(\varphi(t)) \cdot \varphi'(t) \, dt = \int_0^{2\pi} (\cos t \sin t, 0, 0) \cdot (-\sin t, \cos t, 0) \, dt = 0.$$

■

Some other notes on Stoke's:

- Why do we take M to be compact? That's to make sure that there is a boundary that we're not just "ignoring". For example, the "open" hemisphere does not technically have a boundary. However, we can't plug it into Stoke's because it's not compact. By closing the hemisphere to include the circular perimeter of its base, we are forced to include the boundary. The compactness condition also ensures that the boundary will be a nice loop, and not something weird that shoots off to infinity.

- A fascinating observation is that the value of $\int_M \operatorname{curl} F \cdot \nu \, d\mathcal{H}^2$ does not depend on which manifold M we choose, as long as its boundary is the same. For example, as in the example above, we still would get $\int_M \operatorname{curl} F \cdot \nu \, d\mathcal{H}^2 = 0$ if we chose M to look really wacky, like a cactus or something, but still with that circular boundary at the bottom (and as long as ν still points "up").

- Consider $M = \partial B_3(0,1)$. Let F be any C^1 vector field. Then $\int_M \operatorname{curl} F \cdot \nu \, d\mathcal{H}^2 = 0$. This is because M is a (compact!) manifold with boundary, and specifically **its boundary is the empty set**. The same identity holds for any M that "encloses" a region.

- As suggested at the beginning of these notes, Stoke's Theorem generalizes Green's Theorem. A nice understanding check is to demonstrate this.