Math 320-3: Real Analysis
Northwestern University, Lecture Notes

Written by Santiago Cañez

These are notes which provide a basic summary of each lecture for Math 320-3, the third quarter of "Real Analysis", taught by the author at Northwestern University. The book used as a reference is the 4th edition of *An Introduction to Analysis* by Wade. Watch out for typos! Comments and suggestions are welcome.

Contents

Lecture 1: Limits and Continuity

Welcome to the final quarter of real analysis! This quarter, to use an SAT-style analogy, is to multivariable calculus what the first quarter was to single-variable calculus. That is, this quarter is all about making precise the various concepts you would see in a multivariable calculus course—such as multivariable derivatives and integrals, vector calculus and Stokes' Theorem—and pushing them further. After giving a broad introduction to this, we started talking about limits.

The Euclidean norm. First we clarify some notation we'll be using all quarter long. The book discusses this in Chapter 8, which is essentially a review of linear algebra.

For a point $\mathbf{a} \in \mathbb{R}^n$ written in components as $\mathbf{a} = (a_1, \ldots, a_n)$, the (Euclidean) *norm* of $\mathbf{a}$ is

$$\|\mathbf{a}\| = \sqrt{a_1^2 + \ldots + a_n^2},$$

and is nothing but the usual notion of length when we think of $\mathbf{a}$ as a vector. Thus for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, the norm of their difference:

$$\|\mathbf{a} - \mathbf{b}\| = \sqrt{(a_1 - b_1)^2 + \cdots + (a_n - b_n)^2}$$

is nothing but the Euclidean *distance* between $\mathbf{a}$ and $\mathbf{b}$, so in the language of metrics from last quarter we have

$$d(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|$$

and $\|\mathbf{a}\|$ is then the distance from $\mathbf{a}$ to the origin $\mathbf{0}$. Using what we know about this Euclidean distance, we know that sometimes we can instead phrase things in terms of the taxicab or box metrics instead, but $\|\cdot\|$ will always denote *Euclidean* length.

Limits. Suppose that $V \subseteq \mathbb{R}^n$ is open and that $\mathbf{a} \in \mathbb{R}^n$. For a function $f : V \backslash \{\mathbf{a}\} \to \mathbb{R}^m$, we say that the *limit* of f as $\mathbf{x}$ approaches $\mathbf{a}$ is $\mathbf{L} \in \mathbb{R}^m$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that

$$0 < \|\mathbf{x} - \mathbf{a}\| < \delta \text{ implies } \|f(\mathbf{x}) - \mathbf{L}\| < \epsilon.$$

Limits are unique when they exist, and we use the notation $\lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = \mathbf{L}$ when this is the case.

The book discusses such limits in Chapter 9, which we skipped last quarter in favor of the metric space material in Chapter 10. These notes should cover everything we'll need to know about limits, but it won't hurt to briefly glance over Chapter 9 on your own.

Remarks. A few remarks are in order. First, since the two norms used in the definitions are giving distances in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, it is clear that a similar definition works for metric spaces in general. Indeed, all we do is take the ordinary definition of a limit for single-variable functions from first quarter analysis and replace absolute values by metrics; the special case where we use the Euclidean metric on $\mathbb{R}^n$ and $\mathbb{R}^m$ results in the definition we have here. But, this quarter we won't be using general metric spaces, so the special case above will be enough.

Second, we should clarify why we are considering the domain of the function to be $V \backslash \{\mathbf{a}\}$ where V is open in $\mathbb{R}^n$. The fact that we exclude $\mathbf{a}$ from the domain just indicates that the definition works perfectly well for functions which are not defined at $\mathbf{a}$ itself, since when considering limits we never care about what is happening *at* $\mathbf{a}$, only what is happening *near* $\mathbf{a}$. (The $0 < \|\mathbf{x} - \mathbf{a}\|$ in the definition is what excludes $\mathbf{x} = \mathbf{a}$ as a possible value.)

More importantly, why are we assuming that V is open? The idea is that in order for the limit as $\mathbf{x} \to \mathbf{a}$ to exist, we should allow $\mathbf{x}$ to approach $\mathbf{a}$ from *any* possible direction, and in order for this

to be the case we have to guarantee that our function is defined along all such "possible directions" near $\mathbf{a}$. Saying that V is open guarantees that we can find a ball around $\mathbf{a}$ which is fully contained within V, so that it makes sense to take any "possible direction" towards $\mathbf{a}$ and remain within V. This will be a common theme throughout this quarter, in that whenever we have a definition which can be phrased in terms of limits, we will assume that the functions in question are defined on open subsets of $\mathbb{R}^n$.

Proposition. Here are two basic facts when working with multivariable limits. First, just as we saw for single-variable limits in Math 320-1, instead of using ϵ's and δ's we can characterize limits in terms of sequences instead:

$$\lim_{x\to\mathbf{a}} f(\mathbf{x}) = \mathbf{L} \text{ if and only if for any sequence } \mathbf{x}_n \to \mathbf{a} \text{ with all } \mathbf{x}_n \neq \mathbf{a}, \text{ we have } f(\mathbf{x}_n) \to \mathbf{L}.$$

One use of this is the following: if we can find two sequences $\mathbf{x}_n$ and $\mathbf{y}_n$ which both converge to $\mathbf{a}$ (and none of those terms are equal to $\mathbf{a}$) such that either $f(\mathbf{x}_n)$ and $f(\mathbf{y}_n)$ converge to different things or one of these image sequences does not converge, then $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})$ does not exist.

The second fact is the one which makes working with multivariable limits more manageable, since it says that we can just work with component-wise limits instead. To be precise, we can write $f : V\backslash\{\mathbf{a}\} \to \mathbb{R}^m$ in terms of its components as

$$f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$$

where each $f_i : V\backslash\{\mathbf{a}\} \to \mathbb{R}$ maps into $\mathbb{R}^1$, and we can write $\mathbf{L} \in \mathbb{R}^m$ in terms of components as $\mathbf{L} = (L_1, \ldots, L_m)$. Then:

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = \mathbf{L} \text{ if and only if for each } i = 1, \ldots, m, \lim_{\mathbf{x}\to\mathbf{a}} f_i(\mathbf{x}) = L_i,$$

so that a multivariable limit exists if and only if the component-wise limits exist, and the value of the multivariable limit is a point whose components are the individual component-wise limits. This is good, since working with inequalities in $\mathbb{R}$ is simpler than working with inequalities in $\mathbb{R}^m$.

The proof of the first fact is the same as the proof of the corresponding statement for single-variable limits, only replacing absolute values in $\mathbb{R}$ by norms in $\mathbb{R}^n$. The second fact comes from the sequential characterization of limits and the fact that, as we saw last quarter, a sequence in $\mathbb{R}^m$ converges if and only if its individual component sequences converge.

Example. Define $f : \mathbb{R}^2\backslash\{(0,0)\} \to \mathbb{R}^2$ by

$$f(x,y) = \left(\frac{x^4+y^4}{x^2+y^2}, \frac{\sqrt{|xy|}}{\sqrt[3]{x^2+y^2}}\right).$$

We show that the limit of f as $(x,y) \to (0,0)$ is $(0,0)$. To get a sense for why this is the correct value of the limit, recall the technique of converting to polar coordinates from multivariable calculus. Making the substitution $x = r\cos\theta$ and $y = r\sin\theta$, we get

$$\frac{x^4+y^4}{x^2+y^2} = r^2(\cos^4\theta + \sin^4\theta) \text{ and } \frac{\sqrt{|xy|}}{\sqrt[3]{x^2+y^2}} = r^{1/3}\sqrt{|\cos\theta\sin\theta|}.$$

The pieces involving sine and cosine are bounded, and so the factors of r leftover will force the limits as $r \to 0$ to be zero. Still, we will prove this precisely using the definition we gave above, or rather the second fact in the proposition above.

Before doing so, note how messy it would be verify the definition directly without the fact about component-wise limits. For a fixed $\epsilon > 0$, we would need to find $\delta > 0$ such that

$$0 < \sqrt{x^2+y^2} < \delta \text{ implies } \|f(x,y) - (0,0)\| = \sqrt{\left(\frac{x^4+y^4}{x^2+y^2}\right)^2 + \left(\frac{\sqrt{|xy|}}{\sqrt[3]{x^2+y^2}}\right)^2} < \epsilon.$$

The complicated nature of the term on the right suggests that this might not be very straightforward, and is why looking at the component-wise limits instead is simpler.

Thus, first we claim that

$$\lim_{(x,y)\to(0,0)} \frac{x^4+y^4}{x^2+y^2} = 0.$$

Indeed, for a fixed $\epsilon > 0$ let $\delta = \sqrt{\epsilon} > 0$. Since $x^4 + y^4 \leq x^4 + 2x^2y^2 + y^4 = (x^2+y^2)^2$, for (x,y) such that $0 < \sqrt{x^2+y^2} < \delta$ we have:

$$|f(x,y) - 0| = \left|\frac{x^4+y^4}{x^2+y^2}\right| \leq x^2 + y^2 < \delta^2 = \epsilon$$

as required. To justify that

$$\lim_{(x,y)\to(0,0)} \frac{\sqrt{|xy|}}{\sqrt[3]{x^2+y^2}} = 0,$$

for a fixed $\epsilon > 0$ we set $\delta = \epsilon^3 > 0$. Since

$$|xy| \leq \max\{x^2, y^2\} \leq x^2 + y^2,$$

for (x,y) such that $0 < \sqrt{x^2+y^2} < \delta$ we have:

$$\frac{\sqrt{|xy|}}{\sqrt[3]{x^2+y^2}} \leq \frac{(x^2+y^2)^{1/2}}{(x^2+y^2)^{1/3}} = (x^2+y^2)^{1/6} = (\sqrt{x^2+y^2})^{1/3} < \delta^{1/3} = \epsilon,$$

as required. Hence since the component-wise limits of f as $(x,y) \to (0,0)$ both exist and equal zero, we conclude that $\lim_{(x,y)\to(0,0)} f(x,y)$ exists and equals $(0,0)$.

Important. A multivariable limit exists if and only if its component-wise limits exist. Thus, in order to determine the value of such limits, most of the time it will be simpler to look at the component-wise limits instead.

Continuity. We have already seen what it means for a function $\mathbb{R}^n \to \mathbb{R}^m$ to be continuous, by taking any of the versions of continuity we had for functions between metric spaces last quarter and specializing them to the Euclidean metric. For instance, the ϵ-δ definition looks like: $f : \mathbb{R}^n \to \mathbb{R}^m$ is *continuous* at $\mathbf{a} \in \mathbb{R}^n$ if for any $\epsilon > 0$ there exists $\delta > 0$ such that

$$\|\mathbf{x} - \mathbf{a}\| < \delta \text{ implies } \|f(\mathbf{x}) - f(\mathbf{a})\| < \epsilon.$$

We also have the characterization in terms of sequences and the one in terms of pre-images of open sets, both of which will be useful going forward.

But now we can add one more phrased in terms of limits, which is really just a rephrasing of either the ϵ-δ definition of the sequential definition: $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\mathbf{a} \in \mathbb{R}^n$ if and

only if $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$. In other words, saying that a function is continuous at a point means that the limit as you approach that point is the value of the function at that point.

Back to the example. The function from the previous example was undefined at $(0,0)$, but using the value we found for the limit in that problem we can now conclude that the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f(x,y) = \begin{cases} \left(\frac{x^4+y^4}{x^2+y^2}, \frac{\sqrt{|xy|}}{\sqrt[3]{x^2+y^2}} \right) & (x,y) \neq (0,0) \\ (0,0) & (x,y) = (0,0) \end{cases}$$

is actually continuous on all of $\mathbb{R}^2$.

Important. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ (or defined on some smaller domain) is continuous at $\mathbf{a} \in \mathbb{R}^n$ if and only if $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$.

Lecture 2: Linear Transformations, Partial Derivatives

Today we continued talking a bit about continuous functions, in particular focusing on properties of linear transformations, and then began talking about partial derivatives. Partial derivatives are the first step towards formalizing the concept of differentiability in higher dimensions, but as we saw, they alone aren't enough to get the job done.

Warm-Up. Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ is a function and that for some $\mathbf{a} \in \mathbb{R}^n$, $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = Ł$ exists. We show that f is then bounded on some open set containing $\mathbf{a}$. To be precise, we show that there exists an open subset $V \subseteq \mathbb{R}^n$ which contains $\mathbf{a}$ and a constant M such that

$$\|f(\mathbf{x})\| \leq M \text{ for all } \mathbf{x} \in V.$$

(We'll take this inequality as what it means for a multivariable function to be bounded and is equivalent to saying that the image of f is contained in a ball of finite radius, which matches up with the definition of bounded we had for metric spaces last quarter.)

The ϵ-δ definition of a limit guarantees that for $\epsilon = 1$ there exists $\delta > 0$ such that

$$\text{if } 0 < \|\mathbf{x} - \mathbf{a}\| < \delta, \text{ then } \|f(\mathbf{x}) - Ł\| < 1.$$

The reverse triangle inequality

$$\|f(\mathbf{x})\| - \|Ł\| \leq \|f(\mathbf{x}) - Ł\|$$

then implies that $\|f(\mathbf{x})\| < 1 + \|Ł\|$, so f is bounded among points of $B_\delta(\mathbf{a})$ which are different from $\mathbf{a}$. To include $\mathbf{a}$ as well we can simply make the bound $1 + \|Ł\|$ larger if need be, say to be the maximum of $1 + \|Ł\|$ and $1 + \|f(\mathbf{a})\|$. Thus for $M = \max\{1 + \|Ł\|, 1 + \|f(\mathbf{a})\|\}$ and $V = B_\delta(\mathbf{a})$, we have

$$\|f(\mathbf{x})\| \leq M \text{ for all } \mathbf{x} \in V,$$

so f is bounded on the open set V containing $\mathbf{a}$ as required. The point of this is to show that a function cannot be unbounded near a point at which it has a limit.

Proposition. You might recall the following fact from a multivariable calculus course, which is essentially a rephrasing of the sequential characterization of limits: $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})$ exists and equals $\mathbf{L}$

if and only if the single-variable limit as you approach $\mathbf{a}$ along any possible path exists and equals $\mathbf{L}$. This gives a nice rephrasing, and is useful when trying to show that limits don't exist.

Linear transformations. We recall a concept from linear algebra, which will play an important role when discussing multivariable differentiability. A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be a *linear transformation* if it has the following properties:

- $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and
- $T(c\mathbf{x}) = cT(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

Thus, linear transformations preserve addition (the first property) and preserve scalar multiplication (the second property).

Apart from the definition, the key fact to know is that these are precisely the types of functions which can be defined via matrix multiplication: for any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ there exists an $m \times n$ matrix B such that

$$T(\mathbf{x}) = B\mathbf{x} \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

Here, when multiplying an element $\mathbf{x}$ of $\mathbb{R}^n$ by a matrix, we are expressing $\mathbf{x}$ as a column vector. Thus, linear transformations concretely look like:

$$T(x_1, \dots, x_n) = (a_{11}x_1 + \cdots + a_{1n}x_n, a_{21}x_1 + \cdots + a_{2n}x_n, \dots, a_{m1}x_1 + \cdots + a_{mn}x_n)$$

for scalars $a_{ij} \in \mathbb{R}$. Here, the a_{ij} are the entries of the matrix B and this expression is the result—written as a row—of the matrix product $B\mathbf{x}$, where we write $\mathbf{x} = (x_1, \dots, x_n)$ as a column. This expression makes it clear that linear transformations are always continuous since their component functions are continuous.

Norms of linear transformations. Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, we define its *norm* $\|T\|$ by:

$$\|T\| := \sup_{\|\mathbf{x}\|=1} \|T(\mathbf{x})\|.$$

That is, $\|T\|$ is the supremum of the vector norms $\|T(\mathbf{x})\|$ in $\mathbb{R}^m$ as $\mathbf{x}$ ranges through all vectors in $\mathbb{R}^n$ of norm 1. (To see why such a restriction on the norm of $\mathbf{x}$ is necessary, note that nonzero linear transformations are always unbounded since for $x \neq 0$,

$$\|T(c\mathbf{x})\| = \|cT(\mathbf{x})\| = |c| \, \|T(\mathbf{x})\|$$

gets arbitrarily large as $c \to \infty$. Thus for nonzero T, $\sup_{\mathbf{x} \in \mathbb{R}^n} \|T(\mathbf{x})\|$ is always infinite.)

To justify that the supremum used in the definition of $\|T\|$ is always finite, note that the set $\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1\}$ the supremum is being taken over is closed and bounded, so it is compact. Thus the composition

$$\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1\} \to \mathbb{R}^m \to \mathbb{R}$$

defined by $\mathbf{x} \mapsto T(\mathbf{x}) \mapsto \|T(\mathbf{x})\|$, which is continuous since it is the composition of continuous functions, has a maximum value by the Extreme Value Theorem, and this (finite) maximum value is then $\|T\|$.

For us, the most important property of this concept of the norm of a linear transformation is the following inequality, which will soon give us a way to bound expressions involving multivariable derivatives: for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, we have

$$\|T(\mathbf{x})\| \leq \|T\| \, \|\mathbf{x}\| \text{ for any } \mathbf{x} \in \mathbb{R}^n.$$

To see this, note first that when $\mathbf{x} = \mathbf{0}$, $T(\mathbf{x}) = \mathbf{0}$ and both sides of the inequality are 0 in this case. When $\mathbf{x} \neq \mathbf{0}$, $\frac{\mathbf{x}}{\|\mathbf{x}\|}$ has norm 1 and so

$$\left\| T\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \right\| \leq \|T\|$$

since the left side is among the quantities of which the right side is the supremum. Thus for $\mathbf{x} \neq 0$:

$$\|T(\mathbf{x})\| = \left\| T\left(\|\mathbf{x}\| \frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \right\| = \left\| \|\mathbf{x}\| T\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \right\| = \|\mathbf{x}\| \left\| T\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \right\| \leq \|\mathbf{x}\| \, \|T\|$$

as claimed. The exact value of $\|T\|$ for a given linear transformation will not be so important, only that it always finite and nonnegative.

Remark. The book goes through this material in Chapter 8, where it gives a slightly different definition of $\|T\|$ as:

$$\|T\| := \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|T(\mathbf{x})\|}{\|\mathbf{x}\|}.$$

This is equivalent to our definition since we can rewrite the expressions of which we are taking the supremum as

$$\frac{\|T(\mathbf{x})\|}{\|\mathbf{x}\|} = \frac{1}{\|\mathbf{x}\|} \|T(\mathbf{x})\| = \left\| \frac{1}{\|x\|} T(\mathbf{x}) \right\| = \left\| T\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \right\|$$

where $\frac{\mathbf{x}}{\|\mathbf{x}\|}$ at the end as norm 1. Thus the book's definition of $\|T\|$ can be reduced to one which only involves vectors of norm 1, which thus agrees with our definition.

The book then goes onto to prove that $\|T\|$ is always finite and that $\|T(\mathbf{x})\| \leq \|T\| \, \|\mathbf{x}\|$ using purely linear-algebraic means and without using the fact that T is continuous; the (in fact uniform) continuity of T is then derived from this using the bound:

$$\|T(\mathbf{x}) - T(\mathbf{y})\| = \|T(\mathbf{x} - \mathbf{y})\| \leq \|T\| \, \|\mathbf{x} - \mathbf{y}\| \, .$$

I think our approach is nicer and more succinct, especially since it builds off of the Extreme Value Theorem and avoids linear algebra. But feel free to check the book for full details about this alternate approach to defining $\|T\|$.

Important. For a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, $\|T\mathbf{x}\| \leq \|T\| \, \|\mathbf{x}\|$ for any $\mathbf{x} \in \mathbb{R}^n$.

Partial derivatives. Let $f : V \to \mathbb{R}$ be a function defined on an open subset V of $\mathbb{R}^n$, let $\mathbf{x} = (x_1, \ldots, x_n)$ and let $\mathbf{a} = (a_1, \ldots, a_n) \in V$. The *partial derivative* of f with respect to x_i at $\mathbf{a}$ is the derivative—if it exists—of the single-variable function

$$g(x_i) := f(a_1, \ldots, x_i, \ldots, a_n)$$

obtained by holding x_j fixed at a_j for $j \neq i$ and only allowing x_i to vary. To be clear, denoting this partial derivative by $\frac{\partial f}{\partial x_i}$ (or f_{x_j}) we have:

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{x_i \to a_i} \frac{f(a_1, \ldots, x_i, \ldots, a_n) - f(a_1, \ldots, a_i, \ldots, a_n)}{x_i - a_i}$$

if this limits exists. (The fraction on the right is simply $\frac{g(x_i) - g(a_i)}{x_i - a_i}$ where g is the single-variable function introduced above.)

Equivalently, by setting $h = x_i - a_i$, we can rewrite this limit as:

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_i, \ldots, a_n)}{h},$$

which is analogous to the expression

$$\lim_{h \to 0} \frac{g(a_i + h) - g(a_i)}{h} \text{ as opposed to } \lim_{x_i \to a_i} \frac{g(x_i) - g(a_i)}{x_i - a_i}$$

for single-variable derivatives. Even more succinctly, by introducing the vector $\mathbf{e}_i$ which has a 1 in the i-th coordinate and zeros elsewhere, we have:

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{e}_i) - f(\mathbf{a})}{h},$$

again whenever this limit exists.

For a function $f : V \to \mathbb{R}^m$ written in components as $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_n(\mathbf{x}))$, we say that the partial derivative of f with respect to x_i at $\mathbf{a}$ exists when the partial derivative of each component function with respect to x_i exists and we set

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) := \left(\frac{\partial f_1}{\partial x_i}(\mathbf{a}), \ldots, \frac{\partial f_n}{\partial x_i}(\mathbf{a}) \right).$$

Example. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x, y) = \begin{cases} \frac{x^2 y^2}{x^4 + y^4} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0). \end{cases}$$

We claim that both partial derivatives $f_x(0, 0)$ and $f_y(0, 0)$ exist and equal 0. Indeed, we have:

$$f_x(0, 0) = \frac{\partial f}{\partial x}(0, 0) = \lim_{h \to 0} \frac{f(h, 0) - f(0, 0)}{h} = \lim_{h \to 0} \frac{0 - 0}{h} = 0$$

and

$$f_y(0, 0) = \frac{\partial f}{\partial y}(0, 0) = \lim_{h \to 0} \frac{f(0, h) - f(0, 0)}{h} = \lim_{h \to 0} \frac{0 - 0}{h} = 0,$$

where in both computations we use the fact that $f(h, 0) = \frac{0}{h^4} = f(0, h)$ since we are considering h approaching 0 but never equal to 0 in such limits.

Partial derivatives and continuity. Here's the punchline. Ideally, we would like the fact that "differentiability implies continuity" to be true in the higher-dimensional setting as well, and the previous example shows that if we try to characterize "differentiability" solely in terms of partial derivatives this won't be true. Indeed, both partial derivatives of that function exist at the origin and yet that function is not continuous at the origin: taking the limit as we approach $(0, 0)$ along the x-axis gives:

$$\lim_{x \to 0} f(x, 0) = \lim_{x \to 0} \frac{0}{x^4} = 0$$

while taking the limit as we approach $(0, 0)$ along the line $y = x$ gives:

$$\lim_{x \to 0} f(x, x) = \lim_{x \to 0} \frac{x^4}{x^4 + x^4} = \frac{1}{2},$$

so $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist, let alone equal $f(0,0)$ as would be required for continuity at the origin. The issue is that differentiability in higher-dimensions is a trickier beat and requires more than the existence of partial derivatives alone, and will force us to think harder about what derivatives are *really* meant to measure. We'll get to this next week.

Important. Partial derivatives of a function $\mathbb{R}^n \to \mathbb{R}$ are defined as in a multivariable calculus course, by taking single-variable derivatives of the function in question as we hold all variables except for one constant. For functions mapping into $\mathbb{R}^m$ where $m > 1$, partial derivatives are defined component-wise. The existence of all partial derivatives at a point is not enough to guarantee continuity at that point.

Lecture 3: Second-Order Partial Derivatives

Today we spoke about second-order partial derivatives, focusing on an example where the mixed second-order derivatives are unequal. We then stated Clairaut's Theorem, which guarantees that such mixed second-order derivatives *are* in fact equal as long as they're continuous.

Warm-Up 1. We determine the partial differentiability of the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} \frac{-2x^3+3y^4}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

First, for any $(x,y) \neq (0,0)$, there exists an open set around (x,y) which excludes the origin and hence the function f on this open set agrees with the function

$$g(x,y) = \frac{-2x^3+3y^4}{x^2+y^2}.$$

Since the limits defining partial derivatives only care about what's happening close enough to the point being approached, this means that the partial differentiability of f at any $(x,y) \neq (0,0)$ is the same as that of f. Thus because both partial derivatives $g_x(x,y)$ and $g_y(x,y)$ exist at any $(x,y) \neq (0,0)$ since g is a quotient of partially-differentiable functions with nonzero denominator, we conclude that $f_x(x,y)$ and $f_y(x,y)$ exist at any $(x,y) \neq (0,0)$ as well. (The values of these partial derivatives are obtained simply by differentiating with respect to the one variable at a time using the quotient rule, as you would have done in a multivariable calculus course.)

The reasoning above doesn't work at $(0,0)$, in which case we fall back to the limit definition of partial derivatives. We have:

$$f_x(0,0) = \lim_{h\to 0} \frac{f(h,0)-f(0,0)}{h} = \lim_{h\to 0} \frac{-2h-0}{h} = -2$$

and

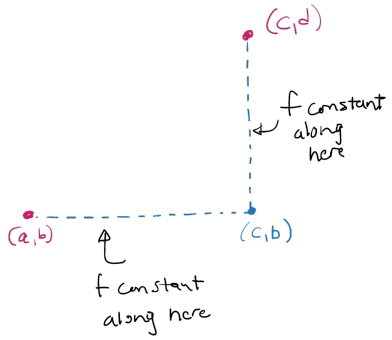
$$f_y(0,0) = \lim_{h\to 0} \frac{f(0,h)-f(0,0)}{h} = \lim_{h\to 0} \frac{3h^2-0}{h} = 0.$$

To be clear, for $f_x(0,0)$ we are taking a limit where only x varies and y is held constant at 0, and for $f_y(0,0)$ we hold x constant at 0 and take a limit where only y varies. We conclude that f_x and f_y exist on all of $\mathbb{R}^2$.

Warm-Up 2. Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ is such that all partial derivatives of f exist and are equal to zero throughout $\mathbb{R}^n$. We show that f must be constant. To keep the notation and geometric

picture simpler, we only do this for the case where $n = 2$ and $m = 1$, but the general case is very similar.

The issue is that f_x and f_y only gives us information about how f behaves in two specific directions, whereas to say that f is constant we have to consider how f behaves on all of $\mathbb{R}^2$. Concretely, suppose that $(a, b), (c, d) \in \mathbb{R}^2$; we want to show that $f(a, b) = f(c, d)$. To get some intuition we will use the following picture which assumes that (a, b) and (c, d) do not lie on the same horizontal nor straight line in $\mathbb{R}^2$, but our proof works even when this is the case:



Consider the point (c, b). Since $f_x(x, y) = 0$ for all $(x, y) \in \mathbb{R}^2$, f is constant along any horizontal line, and thus along the line connecting (a, b) and (c, b). Hence $f(a, b) = f(c, b)$. Now, since $f_y(x, y) = 0$ for all $(x, y) \in \mathbb{R}^2$, f is constant along any vertical line and thus along the vertical line connecting (c, b) and (c, d). Hence $f(c, b) = f(c, d)$ and together with the previous equal we get

$$f(a, b) = f(c, b) = f(c, d)$$

as required. We conclude that f is constant on $\mathbb{R}^2$, and mention again that a similar argument works in a more general $\mathbb{R}^n \to \mathbb{R}^m$ setting.

Remark. The key take away from the second Warm-Up is that we were able to use information about the behavior of f in specific directions to conclude something about the behavior of f overall. In particular, the fact that $f_x = 0$ everywhere implies that f is constant along horizontal lines is a consequence of the single-variable Mean Value Theorem applied to the x-coordinate, and the fact that $f_y = 0$ everywhere implies that f is constant along vertical lines comes from the single-variable Mean Value Theorem applied to the y-coordinate. We'll see a similar application of the Mean Value Theorem one coordinate at-a-time in the proof of Clairaut's Theorem.

Important. Sometimes, but not always, behavior of a function along specific directions can be pieced together to obtain information about that function everywhere.

Higher-order partial derivatives. Second-order partial derivatives are defined by partial differentiating first-order partial derivatives, and so on for third, fourth, and higher-order partial derivatives. In particular, for a function f of two variables, there are four total second-order partial derivatives we can take:

$$\frac{\partial^2 f}{\partial x^2} := \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \quad \frac{\partial^2 f}{\partial y \partial x} := \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right), \quad \frac{\partial^2 f}{\partial x \partial y} := \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right), \quad \frac{\partial^2 f}{\partial y^2} := \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right).$$

We say that a function $f : V \to \mathbb{R}^m$, where $V \subseteq \mathbb{R}^n$ is open, is C^p on V if all partial derivatives up to and including the p-th order ones exist and are continuous throughout V.

Example. This is a standard example showing that f_{xy} and f_{yx} are not necessary equal, contrary to what you might remember from a multivariable calculus course. The issue is that these so-called mixed second-order derivatives are only guaranteed to be equal when at least one is continuous, which is thus not the case in this example. This example is so commonly used that even after hours of searching I was unable to find a different one which illustrates this same concept. This example is in our book, but we'll flesh out some of the details which the book glosses over.

Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

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

We claim that this function is C^1 and that both second-order derivatives $f_{xy}(0,0)$ and $f_{yx}(0,0)$ exist at the origin and are *not* equal. First, as in the Warm-Up, the existence of $f_x(x,y)$ and $f_y(x,y)$ for $(x,y) \neq (0,0)$ follow from the fact that f agrees with the function

$$g(x,y) = xy\left(\frac{x^2-y^2}{x^2+y^2}\right)$$

near such (x,y). By the quotient rule we have:

$$f_x(x,y) = xy\left(\frac{(x^2+y^2)2x - (x^2-y^2)2x}{(x^2+y^2)^2}\right) + y\left(\frac{x^2-y^2}{x^2+y^2}\right) = xy\frac{4xy^2}{(x^2+y^2)^2} + y\left(\frac{x^2-y^2}{x^2+y^2}\right)$$

for $(x,y) \neq (0,0)$. To check the existence of $f_x(0,0)$ we compute:

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

so $f_x(0,0) = 0$. Thus we have:

$$f_x(x,y) = \begin{cases} xy\frac{4xy^2}{(x^2+y^2)^2} + y\left(\frac{x^2-y^2}{x^2+y^2}\right) & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

Now, f is continuous at $(x,y) \neq (0,0)$ again because f agrees with the continuous expression

$$xy\frac{4xy^2}{(x^2+y^2)^2} + y\left(\frac{x^2-y^2}{x^2+y^2}\right)$$

at and near such points. To check continuity at the origin we must show that

$$\lim_{(x,y)\to(0,0)} f_x(x,y) = 0 = f_x(0,0).$$

We use the inequality $2|xy| \leq x^2 + y^2$ which comes from rearranging the terms in

$$0 \leq (|x| - |y|)^2 = x^2 - 2|xy| + y^2.$$

This gives $4|xy|^2 \leq (x^2+y^2)^2$, so for $(x,y) \neq (0,0)$ we have:

$$|f_x(x,y)| = \left| xy\frac{4xy^2}{(x^2+y^2)^2} + y\left(\frac{x^2-y^2}{x^2+y^2}\right)\right|$$

$$
\begin{aligned}
&\leq \left| xy \frac{4xy^2}{(x^2+y^2)^2} \right| + \left| y \left(\frac{x^2-y^2}{x^2+y^2} \right) \right| \\
&= \frac{4|xy|^2|y|}{(x^2+y^2)^2} + |y| \left| \frac{x^2-y^2}{x^2+y^2} \right| \\
&\leq \frac{(x^2+y^2)^2|y|}{(x^2+y^2)^2} + |y| \left| \frac{x^2-y^2}{x^2+y^2} \right| \\
&= |y| + |y| \\
&= 2|y|,
\end{aligned}
$$

where we use the fact that $\frac{|x^2-y^2|}{x^2+y^2} \leq 1$ since the denominator is larger than the numerator. Since $2|y| \to 0$ as $(x,y) \to (0,0)$, $|f_x(x,y)| \leq 2|y|$ implies that $|f_x(x,y)| \to 0$ as $(x,y) \to (0,0)$ and hence that $f_x(x,y) \to 0$ as well. Thus f_x is continuous on all of $\mathbb{R}^2$ as claimed. Similar reasoning shows that f_y exists and is continuous on all of $\mathbb{R}^2$, but we'll leave this verification to the homework.

Next we claim that $f_{xy}(0,0)$ exists and equals -1. Note that for $y \neq 0$ we have

$$
f_x(0,y) = 0 + y \left(\frac{0-y^2}{0+y^2} \right) = -y,
$$

and that this expression also gives the correct value of $f_x(0,0) = 0$. Thus the single-variable function $f_x(0,y)$ is given by

$$
f_x(0,y) = -y
$$

for all y, and hence its single-variable derivative at $y=0$ is -1. But this single-variable derivative is precisely the definition of $f_{xy}(0,0)$, so $f_{xy}(0,0) = -1$ as claimed. A similar computation which is also left to the homework will show that $f_{yx}(0,0) = 1$, so $f_{xy}(0,0) \neq f_{yx}(0,0)$ as was to be shown.

Clairaut's Theorem. As the example above shows, f_{xy} and f_{yx} do not have to agree in general. However, if these second-order derivatives are continuous, then they must be equal. More generally, we have:

> Suppose that $f : V \to \mathbb{R}^m$ is C^2 on an open subset V of $\mathbb{R}^n$ and that $x_1, \ldots, x_n$ are variables on V. Then $f_{x_i x_j}(a,b) = f_{x_j x_i}(a,b)$ for any i and j and any $(a,b) \in V$.

Thus continuity of mixed second order derivatives guarantees their equality. The name "Clairaut's Theorem" for this result is very common although our book doesn't use it and instead only refers to this as Theorem 11.2.

The book actually proves a slightly stronger version which instead of assuming that f is C^2 only assumes it is C^1 and that *one* of the mixed second-order derivatives $f_{x_i x_j}$ exists and is continuous at (a,b)—-the existence of the other mixed derivative $f_{x_j x_i}(a,b)$ and the fact that it's equal to $f_{x_i x_j}(a,b)$ is then derived as a consequence. The ideas behind the proofs of this stronger version and our version are the same, so we'll only focus on our version. We'll save the proof for next time.

Finally, we'll state that this result about second-order derivatives implies similar ones about higher-order derivatives. For instance, if f is C^3, then applying Clairaut's Theorem to the C^2 function f_{x_i} gives the equality of $f_{x_i x_j x_k}$ and $f_{x_i x_k x_j}$, and so on.

Important. Mixed second (and higher) order derivatives do not always agree, except when they're continuous as is the case for C^2 (or C^p with $p > 2$) functions.

What Clairaut's Theorem is really about. We'll finish today by going off on a bit of a tangent, to allude to some deeper meaning behind Clairaut's Theorem. This is something we might come

back to later on when we talk about surfaces, but even if we do we won't go into it very deeply at all. Take a course on differential geometry to really learn what this is all about.

We denote the sphere in $\mathbb{R}^3$ by S^2. (In general, S^n denotes the set of points in $\mathbb{R}^{n+1}$ at distance 1 from the origin.) Then we can develop much of what we're doing this quarter for functions $f : S^2 \to \mathbb{R}$ defined on S^2. In particular, we can make sense of partial derivatives of such a function with respect to variables (s, t) which are variables "along" the sphere. (In more precise language, the sphere can be parametrized using parametric equations, and s and t are the parameters of these equations.) Then we can also compute second-order partial derivatives, and ask whether the analog of Clairaut's Theorem still holds. The fact is that this analog does *not* hold in this new setting:

$$f_{st} \text{ is not necessarily equal to } f_{ts}$$

for $f : S^2 \to \mathbb{R}$ *even if* both of these second-order derivatives are continuous!

The issue is that there sphere S^2 is a *curved* surface while $\mathbb{R}^2$ (or more generally $\mathbb{R}^n$) is what's called *flat*. The fact that Clairaut's Theorem holds on $\mathbb{R}^n$ but not on S^2 reflects the fact that $\mathbb{R}^n$ has zero curvature but that S^2 has nonzero (in fact positive) curvature. The difference

$$f_{xy} - f_{yx}$$

measures the extent to which a given space is curved: this difference is identically zero on $\mathbb{R}^2$ but nonzero on S^2. This is all we'll say about this for now, but in the end this gives some nice geometric meaning to Clairaut's Theorem. Again, take a differential geometry course to learn more.

Lecture 4: Clairaut's Theorem, Differentiability

Today we proved Clairaut's Theorem and then started talking about differentiability in the higher-dimensional setting. Here differentiability is defined in terms of matrices, and truly gets at the idea which derivatives in general are meant to capture.

Warm-Up. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} \frac{-2x^3+3y^4}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

from the Warm-Up last time. We now show that the second-order derivatives $f_{yy}(0,0)$ and $f_{yx}(0,0)$ exist and determine their values.

First, for $(x,y) \neq (0,0)$ we have

$$f_y(x,y) = \frac{(x^2+y^2)(12y^3) - 2y(-2x^3+3y^4)}{(x^2+y^2)^2} = \frac{12x^2y^3 + 6y^5 - 4x^3y}{(x^2+y^2)^2}.$$

Since $f(0,y) = 3y^2$ for all y, we get $f_y(0,0) = 0$ after differentiating with respect to y and setting $y = 0$. (Note one subtlety: the expression $f(0,y) = 3y^2$ was obtained by setting $x = 0$ in the fraction defining f for $(x,y) \neq (0,0)$ so that technically at this point we only have $f(0,y) = 3y^2$ for $y \neq 0$. But since this happens to also give the correct value for $f(0,0) = 0$, we indeed have $f(0,y) = 3y^2$ for all y as claimed.)

Thus the function f_y is defined by

$$f_y(x,y) = \begin{cases} \frac{12x^2y^3+6y^5-4x^3y}{(x^2+y^2)^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

Now, from this we see that $f_y(0,y) = 6y$ for all y (including $y = 0$), so that $f_{yy}(0,0) = 6$. (We can also use the fact that f_{yy} should be the ordinary second derivative of the single variable function $f(0,y) = 3y^2$.) The second-order derivative $f_{yx}(0,0)$ should be the ordinary derivative at $x = 0$ of the single-variable function obtained by holding y constant at 0 in $f_y(x,y)$, so since

$$f_y(x,0) = 0 \text{ for all } x \text{ including } x = 0,$$

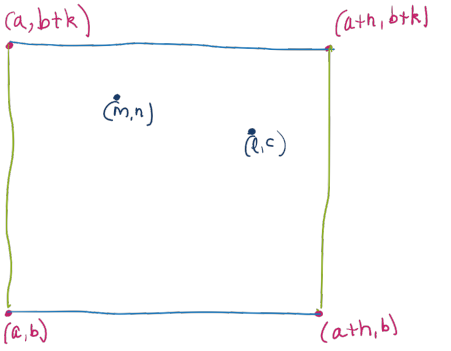
we get $f_{yx}(0,0) = 0$ too. Thus both second-order partial derivatives $f_{yy}(0,0)$ and $f_{yx}(0,0)$ exist and equal 6 and 0 respectively.

Back to Clairaut's Theorem. Recall the statement of Clairaut's Theorem: if $f : V \to \mathbb{R}^m$ is C^2 on an open subset V of $\mathbb{R}^n$, then $f_{x_ix_j}(a,b) = f_{x_jx_i}(a,b)$ at any $(a,b) \in V$ for any i and j. We now give a proof in the case where V is a subset of $\mathbb{R}^2$, so that f is a function of two variables x and y, and where $m = 1$. This is only done to keep the notation simpler, but the proof in the most general case is very similar.

Before giving the proof, let us talk about the ideas which go into it. Fix $(a,b) \in V$. For small values of h and k (small enough to guarantee that $(a+h, b+k), (a+h, b)$, and $(a, b+k)$ are all still in domain of f; this is where openness of V is used and is where the book gets the requirement that $|h|, |k| < r/\sqrt{2}$) we introduce the function

$$\Delta(h,k) = f(a+h, b+k) - f(a, b+k) - f(a+h, b) + f(a,b),$$

which measures the behavior of f at the corners of a rectangle:



Apply MVT in one direction, then the other

The proof amounts to computing

$$\lim_{(h,k)\to(0,0)} \frac{\Delta(h,k)}{hk}$$

in two ways: computing it one way gives the value $f_{yx}(a,b)$ and computing it the other way gives $f_{xy}(a,b)$, so since these two expressions equal the *same* limit, they must equal each other as Clairaut's Theorem claims.

This limit is computed in the first way by focusing on what's happening in the picture above "vertically" and then "horizontally", and in the second way by focusing on what's happening "horizontally" and then "vertically". To be precise, both computations come from the applying the single-variable Mean Value Theorem in one direction and then in the other, and so the proof is a reflection of the idea mentioned last time that sometimes we can check the behavior of a function in one direction at a time to determine its overall behavior.

Looking at the definition of $\Delta(h,k)$, note that the first and third terms have the same first input and only differ in their second input; applying the Mean Value Theorem in the y-coordinate to these gives

$$f(a+h,b+k)-f(a+h,b)=f_y(a+h,c)k$$

for some c between b and $b+k$. Similarly, applying the Mean Value Theorem in the y-coordinate to the second and fourth terms in the definition of $\Delta(h,k)$ gives

$$f(a,b+k)-f(a,b)=f_y(a,d)k$$

for some d between b and $b+k$, so that overall we get

$$\Delta(h,k)=f_y(a+h,c)k-f_(a,d)k=k[f_y(a+h,c)-f_y(a,d)].$$

At this point we would like to now apply the Mean Value Theorem in the x-coordinate to get an expression involving f_{yx}, but the problem is that the two terms above also differ in their y-coordinates; we can only apply the single-variable Mean Value Theorem to expressions where the "other" coordinate is the same. In the proof below we will see how to guarantee we can take c and d here to be the same—this is why we use the functions F and G defined below—but this is a point which the book glosses over without explaining fully. Apart from this though, the book's approach works fine, but hopefully we'll make it a little easier to follow.

Proof of Clairaut's Theorem. For h and k small enough we define

$$\Delta(h,k)=f(a+h,b+k)-f(a,b+k)-f(a+h,b)+f(a,b)$$

and compute $\lim_{(h,k)\to(0,0)}\frac{\Delta(h,k)}{hk}$ in two ways. First, introduce the single-variable function

$$F(y)=f(a+h,y)-f(a,y).$$

Then we have $\Delta(h,k)=F(b+k)-F(b)$. By the single-variable Mean Value Theorem, there exists c between b and $b+k$ such that

$$F(b+k)-F(b)=F_y(c)k,$$

which gives

$$\Delta(h,k)=k[f_y(a+h,c)-f_y(a,c)].$$

(The book uses the fact that any number c between b and $b+k$ can be written as $c=b+tk$ for some $t\in(0,1)$ in its proof.) Now, applying the single-variable Mean Value Theorem again in the x-coordinate gives the existence of ℓ between a and $a+h$ such that

$$f_y(a+h,c)-f_y(a,c)=f_{yx}(\ell,c)h.$$

(The book writes ℓ as $\ell=a+sh$ for some $s\in(0,1)$.) Thus we have that

$$\Delta(h,k)=khf_{yx}(\ell,c)\text{ so }\frac{\Delta(h,k)}{hk}=f_{yx}(\ell,c).$$

Since ℓ is between a and $a+h$ and c is between b and $b+k$, $(\ell,c)\to(a,b)$ as $(h,k)\to(0,0)$ so since f_{yx} is continuous at (a,b) we have:

$$\lim_{(h,k)\to(0,0)}\frac{\Delta(h,k)}{hk}=\lim_{(h,k)\to(0,0)}f_{yx}(\ell,c)=f_{yx}(a,b).$$

Now go back to the original definition of $\Delta(h,k)$ and introduce the single-variable function

$$G(x) = f(x, b+k) - f(x,b).$$

Then $\Delta(h,k) = G(a+h) - G(a)$. By the Mean Value Theorem there exists m between a and $a+h$ such that

$$G(a+h) - G(a) = G_x(m)h,$$

which gives

$$\Delta(h,k) = G(a+h) - G(a) = G_x(m)h = h[f_x(m, b+k) - f_x(m,b)].$$

By the Mean Value Theorem again there exists n between b and $b+k$ such that

$$f_x(m, b+k) - f_x(m,b) = f_{xy}(m,n)k,$$

so

$$\Delta(h,k) = h[f_x(m, b+k) - f_x(m,b)] = hkf_{xy}(m,n).$$

As $(h,k) \to (0,0)$, $(m,n) \to (a,b)$ since m is between a and $a+h$ and n between b and $b+k$, so the continuity of f_{xy} at (a,b) gives:

$$\lim_{(h,k)\to(0,0)} \frac{\Delta(h,k)}{hk} = \lim_{(h,k)\to(0,0)} f_{xy}(m,n) = f_{xy}(a,b).$$

Thus $f_{xy}(a,b) = f_{yx}(a,b)$ since these both equal the same limit $\lim_{(h,k)\to(0,0)} \Delta(h,k)/hk$. □

Motivating higher-dimensional derivatives. The idea of viewing derivatives as "slopes" for single-variable functions is nice visually, but doesn't capture the correct essence of what differentiability means in higher-dimensions. (It is also harder to visualize what "slope" might mean in these higher-dimensional settings.) Instead, we use another point of view—the"linear approximation" point of view—of single-variable derivatives to motivate the more general notion of differentiability.

If $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$, then values of f at points near a are pretty well approximated by the tangent line to the graph of f at the point a; to be clear, we have that

$$f(a+h) \approx f(a) + f'(a)h \text{ for small } h,$$

where the expression on the right is the usual tangent line approximation. Rewriting gives

$$f(a+h) - f(a) \approx f'(a)h$$

again for small h. Here's the point: if we view h as describing the "small" difference between the inputs a and $a+h$, then the expression on the left $f(a+h) - f(a)$ gives the resulting difference between the outputs at these inputs. Thus we have:

$$(\text{change in output}) \approx f'(a)(\text{change in input}),$$

and this approximation gets better and better as $h \to 0$. Thus, we can view the single-variable derivative $f'(a)$ at a as the object which tells us how to go from small changes in input to corresponding changes in output, or even better: $f'(a)$ transforms "infinitesimal" changes in input into "infinitesimal" changes in output. From this point of view, $f'(a)$ is not simply a number but is better thought of as the *transformation* obtained via multiplication by that number.

And now we claim that the same idea works for functions $f : \mathbb{R}^n \to \mathbb{R}^m$ as well. The derivative of such an f at some $\mathbf{a} \in \mathbb{R}^n$ should be something which transforms "small" changes in input into corresponding "changes" in output, or said another way transforms infinitesimal changes in input into infinitesimal changes in output. But changes in inputs in this case look like

$$(\mathbf{a} + \mathbf{h}) - \mathbf{a},$$

which is a difference of vectors and hence is itself a vector, and similarly changes in outputs look like

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}),$$

which is also a vector quantity. Thus the "derivative" of f at $\mathbf{a}$, whatever it is, should be something which transforms (infinitesimal) input vectors into (infinitesimal) output vectors. Throwing in the point of view that derivatives should also be "linear" objects in some sense gives the conclusion that the derivative of f at $\mathbf{a}$ should be a *linear transformation*, i.e. a matrix! This will lead us to the following definition, where indeed the *derivative* of $f : \mathbb{R}^n \to \mathbb{R}^m$ at $\mathbf{a} \in \mathbb{R}^n$ is not just a single number, but is rather an entire matrix of numbers.

Differentiability. Suppose that $f : V \to \mathbb{R}^m$ is defined on some open subset V of $\mathbb{R}^n$. We say that f is *differentiable* at $\mathbf{a} \in V$ if there exists an $m \times n$ matrix B such that

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

We will soon see that if there is a matrix with this property, there is only one and we can give an explicit description of what it has to be. We call this matrix the *(total) derivative of f at $\mathbf{a}$*, or the *Jacobian matrix* of f at $\mathbf{a}$. In a sense, this matrix geometrically captures the "slopes" of f in all possible "directions".

This limit precisely captures the intuition we outlined earlier. The first two terms in the numerator $f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})$ measure the change in outputs and the $B\mathbf{h}$ term measures the "approximate" change in output corresponding to the change in input $\mathbf{h}$; saying that this limit is 0 means that the approximated change in output gets closer and closer to the actual change in output as $\mathbf{h} \to \mathbf{0}$. Note that the numerator has limit 0 as long as f is continuous, so the $\|\mathbf{h}\|$ in the denominator is there to guarantee that the overall limit is zero not just because of the continuity of f but rather as a result of of how well f is approximated by the linear transformation given by B.

The single-variable case. Finally, we note that the definition of differentiability for a function $f : \mathbb{R}^n \to \mathbb{R}^m$ give above becomes the usual notion of differentiability in the case $n = m = 1$. So, suppose that $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ in the above sense. Then there exists a 1×1 matrix B such that

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

Denote the single entry in B by b, so that $B = \begin{pmatrix} b \end{pmatrix}$ as a matrix and thus $Bh = bh$ in the limit expression above. Then we can rewrite this expression as:

$$0 = \lim_{h \to 0} \frac{f(a + h) - f(a) - bh}{h} = \lim_{h \to 0} \left(\frac{f(a + h) - f(a)}{h} - b \right) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h} - b.$$

Thus we see that the remaining limit on the right exists and equals b:

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

which says that f is differentiable in the sense we saw in first-quarter analysis and that $f'(a) = b$. Hence, this new notion of differentiability really is a generalization of the previous version for single-variable functions, and the derivative in this new sense viewed as a matrix agrees with the ordinary derivative in the previous sense for single-variable functions.

Important. A function $f : V \to \mathbb{R}^m$ defined on an open subset V of $\mathbb{R}^n$ is differentiable at $\mathbf{a} \in V$ if there exists an $m \times n$ matrix B such that

$$\frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - B\mathbf{h}}{\|\mathbf{h}\|} \to 0 \text{ as } \mathbf{h} \to \mathbf{0}.$$

We call B the derivative (or Jacobian matrix) of f at $\mathbf{a}$ and this definition captures the idea that B transforms infinitesimal changes in input into infinitesimal changes in output. When $n = m = 1$, this agrees with the usual notion of a derivative for single-variable functions. We also use $Df(\mathbf{a})$ to denote the Jacobian matrix of f at $\mathbf{a}$.

Lecture 5: More on Differentiability

Today we continued talking about differentiability of multivariable functions, focusing on examples and general properties. The main result is an explicit description in terms of partial derivatives of the Jacobian matrices which are used in the definition of differentiability.

Warm-Up 1. We show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = x^2 + y^2$ is differentiable at $(0, 1)$ using $B = \begin{pmatrix} 0 & 2 \end{pmatrix}$ as the candidate for the Jacobian matrix. (We will see in a bit how to determine that this is the right matrix to use.) So, we must verify that for this matrix B we have:

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

where $\mathbf{a} = (0, 1)$ and in $B\mathbf{h}$ we think of $\mathbf{h} = (h, k)$ as a column vector. The numerator is:

$$f(h, k+1) - f(0, 1) - \begin{pmatrix} 0 & 2 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = h^2 + (k+1)^2 - 1 - 2k = h^2 + k^2.$$

Thus:

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - B\mathbf{h}}{\|\mathbf{h}\|} = \lim_{(h,k)\to(0,0)} \frac{h^2 + k^2}{\sqrt{h^2 + k^2}} = \lim_{(h,k)\to(0,0)} \sqrt{h^2 + k^2} = 0$$

as required. Thus f is differentiable at $(0, 1)$ and $Df(0, 1) = \begin{pmatrix} 0 & 2 \end{pmatrix}$.

Warm-Up 2. We show that the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = (x^2 + y^2, xy + y)$ is differentiable at $(0, 1)$ using $B = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 1 \end{smallmatrix}\right)$ as the candidate for the Jacobian matrix. Setting $\mathbf{a} = (0, 1)$ and $\mathbf{h} = (h, k)$, we compute:

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - B\mathbf{h} &= f(h, k+1) - f(0, 1) - \begin{pmatrix} 0 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} \\ &= (h^2 + (k+1)^2, h(k+1) + k + 1) - (1, 1) - (2k, h + k) \\ &= (h^2 + k^2, hk). \end{aligned}$$

Note that when computing the product $B\mathbf{h}$ we have the written the result as a row vector so that we can combine it with the $f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})$ part of the expression.

Now we claim that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - B\mathbf{h}}{\|\mathbf{h}\|} = \lim_{(h,k)\to(0,0)} \frac{(h^2+k^2, hk)}{\sqrt{h^2+k^2}} = 0.$$

For this we need to show that the limit of each component function

$$\frac{h^2+k^2}{\sqrt{h^2+k^2}} \text{ and } \frac{hk}{\sqrt{h^2+k^2}}$$

is zero. But the limit of the first component is precisely the one we considered in the first Warm-Up, where we showed that it was indeed zero. This is no accident: the first component x^2+y^2 of our function f in this case was the function we looked at previously, and our work here shows that a function is differentiable if and only if each component function is differentiable. So, we can use the result of the first Warm-Up to say that the first component of f here is differentiable, so we need only consider the second component. For this, we use $2|hk| \leq h^2+k^2$ to say

$$\left|\frac{hk}{\sqrt{h^2+k^2}}\right| \leq \frac{1}{2}\frac{h^2+k^2}{\sqrt{h^2+k^2}} = \frac{1}{2}\sqrt{h^2+k^2},$$

and since this final expression goes to 0 as $(h,k)\to(0,0)$, the squeeze theorem gives

$$\lim_{(h,k)\to(0,0)} \frac{hk}{\sqrt{h^2+k^2}} = 0$$

as desired. We conclude that f is differentiable at $(0,1)$ and that $Df(0,1) = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 1 \end{smallmatrix}\right)$.

Important. A function $f: V \to \mathbb{R}^m$ written in components as $f = (f_1, \ldots, f_m)$ is differentiable at $\mathbf{a} \in V \subseteq \mathbb{R}^n$ if and only if each component function $f_i : V \to \mathbb{R}$ is differentiable at $\mathbf{a} \in V$.

Differentiability implies continuity. And now we verify something we would hope is true: differentiability implies continuity. Recall that existence of all partial derivatives at a point alone is not enough to guarantee continuity at that point, but the stronger version of differentiability we've given will make this work. This will also be our first instance of using the norm of a linear transformation to bound expressions involving multivariable derivatives.

Suppose that $f: V \to \mathbb{R}^m$ is differentiable at $\mathbf{a} \in V \subseteq \mathbb{R}^n$. We must show that

$$\lim_{\mathbf{h}\to\mathbf{0}} f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}).$$

Since f is differentiable at $\mathbf{a}$ there exists an $m \times n$ matrix B such that

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

Then there exists $\delta > 0$ such that

$$\frac{\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - B\mathbf{h}\|}{\|\mathbf{h}\|} < 1 \text{ when } 0 < \|\mathbf{h}\| < \delta.$$

Thus for such $\mathbf{h}$ we have

$$\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - B\mathbf{h}\| \leq \|\mathbf{h}\|, \text{ which gives } \|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})\| \leq \|\mathbf{h}\| + \|B\mathbf{h}\|$$

after using the reverse triangle inequality $\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})\| - \|B\mathbf{h}\| \leq \|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - B\mathbf{h}\|$. Since $\|B\mathbf{h}\| \leq \|B\|\,\|\mathbf{h}\|$ where $\|B\|$ denotes the norm of the linear transformation induced by B, we get

$$\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})\| \leq (1 + \|B\|)\,\|\mathbf{h}\|\,,$$

which goes to 0 as $\mathbf{h} \to \mathbf{0}$. Thus $f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) \to 0$, so $f(\mathbf{a}+\mathbf{h}) \to f(\mathbf{a})$ as $\mathbf{h} \to \mathbf{0}$ as was to be shown. Hence f is continuous at $\mathbf{a}$.

Description of Jacobian matrices. The definition of differentiable at a point requires the existence of a certain matrix, but it turns out that we can determine which matrix this has to be. In particular, there can only be one matrix B satisfying the requirement that

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

and its entries have to consist of the various partial derivatives of f. Thus, when determining whether or not a function is differentiable, we first find this matrix—which requires that the partial derivatives of our function all exist—and then try to compute the limit above where B is this matrix.

To be precise, suppose that $f : V \to \mathbb{R}^m$ is differentiable at $\mathbf{a} \in V \subseteq \mathbb{R}^n$. We claim that then all partial derivatives of f at $\mathbf{a}$ exist, and moreover that $Df(\mathbf{a})$ is the matrix whose ij-th entry is the value of $\frac{\partial f_i}{\partial x_j}(\mathbf{a})$ where f_i is the i-th component of f. The proof is in the book, but we'll include it here anyway. Since f is differentiable at $\mathbf{a}$ we know that there is an $m \times n$ matrix B such that

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

Since this limit exists, we should get the same limit no matter the direction from which we approach $\mathbf{0}$. Approaching along points of the form $\mathbf{h} = h\mathbf{e}_i$ where $\mathbf{e}_i$ is the standard basis vector with 1 in the i-th entry and zeroes elsewhere—so we are approaching $\mathbf{0}$ along the x_i-axis—we have:

$$\lim_{h\to 0} \frac{f(\mathbf{a}+h\mathbf{e}_i) - f(\mathbf{a}) - B(h\mathbf{e}_i)}{|h|} = \mathbf{0}$$

where in the denominator we use the fact that $\|h\mathbf{e}_i\| = |h|$. For $h > 0$ we have $|h| = h$ so

$$\frac{f(\mathbf{a}+h\mathbf{e}_i) - f(\mathbf{a}) - B(h\mathbf{e}_i)}{|h|} = \frac{f(\mathbf{a}+h\mathbf{e}_i) - f(\mathbf{a})}{h} - \frac{hB\mathbf{e}_i}{h} = \frac{f(\mathbf{a}+h\mathbf{e}_i) - f(\mathbf{a})}{h} - B\mathbf{e}_i,$$

while for $h < 0$ we have $|h| = -h$ so

$$\frac{f(\mathbf{a}+h\mathbf{e}_i) - f(\mathbf{a}) - B(h\mathbf{e}_i)}{|h|} = \frac{f(\mathbf{a}+h\mathbf{e}_i) - f(\mathbf{a})}{-h} - \frac{hB\mathbf{e}_i}{-h} = -\frac{f(\mathbf{a}+h\mathbf{e}_i) - f(\mathbf{a})}{h} + B\mathbf{e}_i,.$$

Thus since

$$\lim_{h\to 0} \frac{f(\mathbf{a}+h\mathbf{e}_i) - f(\mathbf{a}) - B(h\mathbf{e}_i)}{|h|} = \mathbf{0}$$

we get that

$$\lim_{h\to 0} \frac{f(\mathbf{a}+h\mathbf{e}_i) - f(\mathbf{a})}{h} = B\mathbf{e}_i.$$

The left side is the definition of the partial derivative $\frac{\partial f}{\partial x_i}(\mathbf{a}) = (\frac{\partial f_1}{\partial x_i}(\mathbf{a}), \ldots, \frac{\partial f_m}{\partial x_i}(\mathbf{a}))$, so it exists, and the right side is precisely the i-th column of B. Thus the i-th column of B is $(\frac{\partial f_1}{\partial x_i}(\mathbf{a}), \ldots, \frac{\partial f_m}{\partial x_i}(\mathbf{a}))$ written as a column vector so

$$B = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}$$

as claimed.

Example. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} \frac{x^3+y^3}{\sqrt{x^2+y^2}} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

We claim that f is differentiable at $(0,0)$. Since the matrix needed in the definition of differentiability must be the one consisting of the partial derivatives of f at $(0,0)$, we must first determine the values of these partials. We have:

$$f_x(0,0) = \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} = \lim_{h\to 0} \frac{h^3}{h|h|} = 0$$

and

$$f_y(0,0) = \lim_{h\to 0} \frac{f(0,h) - f(0,0)}{h} = \lim_{h\to 0} \frac{h^3}{h|h|} = 0.$$

Thus $Df(\mathbf{0}) = \begin{pmatrix} 0 & 0 \end{pmatrix}$. Now we must show that

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

Setting $\mathbf{h} = (h,k)$, the numerator is

$$f(h,k) - f(0,0) - \begin{pmatrix} 0 & 0 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = \frac{h^3+k^3}{\sqrt{h^2+k^2}}.$$

Thus

$$\frac{f(\mathbf{0}+\mathbf{h}) - f(\mathbf{0}) - Df(\mathbf{0})\mathbf{h}}{\|\mathbf{h}\|} = \frac{h^3+k^3}{h^2+k^2},$$

and converting to polar coordinates $(h,k) = (r\cos\theta, r\sin\theta)$ gives

$$\frac{h^3+k^3}{h^2+k^2} = r(\cos^3\theta + \sin^3\theta)$$

which goes to 0 as $r \to 0$ since the $\cos^3\theta + \sin^3\theta$ part is bounded. Hence we conclude that f is differentiable at $(0,0)$ as claimed.

Important. To check if a function f is differentiable at $\mathbf{a}$, we first compute all partial derivatives of f at $\mathbf{a}$; if at least one of these does not exist then f is not differentiable at $\mathbf{a}$. If they all exist, we form the matrix $Df(\mathbf{a})$ having these as entries and then check whether or not the limit

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

is zero. If it is zero, then f is differentiable at $\mathbf{a}$ and hence also continuous at $\mathbf{a}$.

Lecture 6: Yet More on Derivatives

Yet another day talking about derivatives! Today we showed that having continuous partial derivatives implies being differentiable, which often times gives a quick way to show differentiability without having to compute a limit. Take note, however, that having non-continuous partial derivatives does NOT imply non-differentiability. We also mentioned properties of higher-dimensional differentiable functions related to sums and products, which are analogous to properties we saw for single-variable functions.

Warm-Up 1. We show that the function

$$f(x,y) = \begin{cases} \frac{-2x^3+3y^4}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

is not differentiable at $(0,0)$. We computed the partial derivatives $f_x(0,0)$ and $f_y(0,0)$ in a previous Warm-Up where we found that

$$f_x(0,0) = -2 \text{ and } f_y(0,0) = 0.$$

Thus the candidate for the Jacobian matrix is $Df(0,0) = \begin{pmatrix} -2 & 0 \end{pmatrix}$. For this matrix we have:

$$f(\mathbf{0}+\mathbf{h}) - f(\mathbf{0}) - Df(\mathbf{0})\mathbf{h} = \frac{-2h^3+3k^4}{h^2+k^2} - 0 + 2h = \frac{2hk^2+3k^4}{h^2+k^2},$$

so

$$\frac{f(\mathbf{0}+\mathbf{h}) - f(\mathbf{0}) - Df(\mathbf{0})\mathbf{h}}{\|\mathbf{h}\|} = \frac{2hk^2+3k^4}{(h^2+k^2)\sqrt{h^2+k^2}}.$$

However, taking the limit as we approach $\mathbf{0}$ along $h = k$ gives

$$\lim_{h\to 0} \frac{2h^3+3h^4}{2\sqrt{2}|h|h^2},$$

which does not exist since for $h \to 0^+$ this limit is $1/\sqrt{2}$ while for $h \to 0^-$ it is $-1/\sqrt{2}$. Thus this limit is not zero, so

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{0}+\mathbf{h}) - f(\mathbf{0}) - Df(\mathbf{0})\mathbf{h}}{\|\mathbf{h}\|}$$

is also not zero. (In fact it does not exist.) Hence f is not differentiable at $\mathbf{0}$.

Warm-Up 2. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ satisfies $|f(\mathbf{x})| \leq \|\mathbf{x}\|^2$ for all $\mathbf{x} \in \mathbb{R}^n$. We show that f is differentiable at $\mathbf{0}$. First, note that

$$|f(\mathbf{0})| \leq \|\mathbf{0}\|^2 = 0$$

implies $f(\mathbf{0}) = 0$. Now, we have

$$|f(t\mathbf{e}_i)| \leq \|t\mathbf{e}_i\|^2 = t^2 \text{ for all } i$$

where $\mathbf{e}_i$ is a standard basis vector, so

$$\left|\frac{f(\mathbf{0}+t\mathbf{e}_i) - f(\mathbf{0})}{t}\right| = \left|\frac{f(t\mathbf{e}_i)}{t}\right| \leq |t|.$$

The right side has limit 0 as $t \to 0$, so the squeeze theorem implies that

$$\frac{\partial f}{\partial x_i}(\mathbf{0}) = \lim_{t \to 0} \frac{f(\mathbf{0} + t\mathbf{e}_i) - f(\mathbf{0})}{t} = 0.$$

Hence the candidate for the Jacobian matrix of f at $\mathbf{0}$ is the zero matrix $Df(\mathbf{0}) = \mathbf{0}$.

Thus

$$f(\mathbf{0} + \mathbf{h}) - f(\mathbf{0}) - Df(\mathbf{0})\mathbf{h} = f(\mathbf{h}) - 0 - 0 = f(\mathbf{h}),$$

so

$$\frac{|f(\mathbf{0} + \mathbf{h}) - f(\mathbf{0}) - Df(\mathbf{0})\mathbf{h}|}{\|\mathbf{h}\|} = \frac{|f(\mathbf{h})|}{\|\mathbf{h}\|} \leq \frac{\|h\|^2}{\|\mathbf{h}\|} = \|\mathbf{h}\|.$$

This has limit 0 as $\mathbf{h} \to \mathbf{0}$, so the squeeze theorem again gives

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

and hence f is differentiable at $\mathbf{0}$ as claimed.

C^1 implies differentiable. The definition we have of differentiable is at times tedious to work with, due to the limit involved, but is usually all we have available. However, there is nice scenario in which we can avoid using this definition directly, namely the scenario when we're looking at a C^1 function.

To be precise, suppose that all partial derivatives of $f : V \to \mathbb{R}^m$ exist and are continuous at $\mathbf{a} \in V \subseteq \mathbb{R}^n$. Then the result is that f is differentiable at $\mathbf{a}$. However, note that the converse is not true: if f is differentiable at $\mathbf{a}$ it is NOT necessarily true that the partial derivatives of f are continuous at $\mathbf{a}$. In other words, just because a function has non-continuous partial derivatives at a point does not mean it is not differentiable at that point; in such cases we must resort to using the limit definition of differentiability. Indeed, the result we're stating here will really only be useful for functions whose partial derivatives are simple enough so that determining their continuity is relatively straightforward. For "most" examples we've seen this is not the case, so most of the time we'll have to use the limit definition anyway, or some of the other properties we'll soon mention.

This result is proved in the book, but the proof can be a little hard to follow. So here we'll give the proof only in the $n = 2, m = 1$ case when $V \subseteq \mathbb{R}^2$, which is enough to illustrate the general procedure. The main idea, as in the proof of Clairaut's Theorem, is to rewrite the expression we want to take the limit of in the definition of differentiability by applying the single-variable Mean Value Theorem one coordinate at a time.

Proof. Suppose that the partial derivatives of $f : V \to \mathbb{R}$ at $(a, b) \in V \subseteq \mathbb{R}^2$ both exist and are continuous. For h, k small enough so that $(a+h, b+k), (a, b+k)$, and $(a+h, k)$ also lie in V, write $f(a + h, b + k) - f(a, b)$ as

$$f(a + h, b + k) - f(a, b + k) + f(a, b + k) - f(a, b)$$

by subtracting and adding $f(a, b+k)$. The first two terms have the same second input, so applying the Mean Value Theorem in the x-direction gives

$$f(a + h, b + k) - f(a, b + k) = f_x(c, b + k)h$$

for some c between a and $a + h$. The second two terms in the previous expression have the same first input, so applying the Mean Value Theorem in the y-direction gives

$$f(a, b + k) - f(a, b) = f_y(a, d)k$$

for some d between b and $b+k$. All together we thus have

$$f(a+h,b+k)-f(a,b)=f_x(c,b+k)h+f_y(a,d)k=\begin{pmatrix}f_x(c,b+k) & f_y(a,d)\end{pmatrix}\begin{pmatrix}h\\k\end{pmatrix}$$

where at the end we have written our expression as a matrix product. This gives

$$f(a+h,b+k)-f(a,b)-Df(a,b)\begin{pmatrix}h\\k\end{pmatrix}=\begin{pmatrix}f_x(c,b+k)-f_x(a,b) & f_y(a,d)-f_y(a,b)\end{pmatrix}\begin{pmatrix}h\\k\end{pmatrix}$$

where we use $Df(a,b)=\begin{pmatrix}f_x(a,b) & f_y(a,b)\end{pmatrix}$. Thus setting $\mathbf{h}=(h,k)$, we have

$$\begin{aligned}\frac{\|f(\mathbf{a}+\mathbf{h})-f(\mathbf{a})-Df(\mathbf{a})\mathbf{h}\|}{\|\mathbf{h}\|}&=\frac{\left\|\begin{pmatrix}f_x(c,b+k)-f_x(a,b) & f_y(a,d)-f_y(a,b)\end{pmatrix}\begin{pmatrix}h\\k\end{pmatrix}\right\|}{\left\|\begin{pmatrix}h\\k\end{pmatrix}\right\|}\\&\le\frac{\left\|\begin{pmatrix}f_x(c,b+k)-f_x(a,b) & f_y(a,d)-f_y(a,b)\end{pmatrix}\right\|\left\|\begin{pmatrix}h\\k\end{pmatrix}\right\|}{\left\|\begin{pmatrix}h\\k\end{pmatrix}\right\|}\\&=\left\|\begin{pmatrix}f_x(c,b+k)-f_x(a,b) & f_y(a,d)-f_y(a,b)\end{pmatrix}\right\|\end{aligned}$$

where in the second step we use $\|B\mathbf{h}\|\le\|B\|\,\|\mathbf{h}\|$. As $(h,k)\to(0,0)$, $(c,b+k)\to(a,b)$ since c is between a and $a+h$ and $(a,d)\to(a,b)$ since d is between b and $b+h$. Thus since f_x and f_y are continuous at (a,b), we have

$$\begin{pmatrix}f_x(c,b+k)-f_x(a,b) & f_y(a,d)-f_y(a,b)\end{pmatrix}\to\begin{pmatrix}f_x(a,b)-f_x(a,b) & f_y(a,b)-f_y(a,b)\end{pmatrix}=\begin{pmatrix}0 & 0\end{pmatrix},$$

so $\left\|\begin{pmatrix}f_x(c,b+k)-f_x(a,b) & f_y(a,d)-f_y(a,b)\end{pmatrix}\right\|\to 0$ as $\mathbf{h}\to\mathbf{0}$. The inequality above thus implies that

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

by the squeeze theorem, so f is differentiable at $\mathbf{a}$ as claimed. □

Converse not true. An important word of warning: the converse of the above result is NOT true. That is, just because a function might have non-continuous partial derivatives at a point does NOT guarantee that it is not differentiable there. Indeed, the function

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

is differentiable at $(0,0)$ even though its partial derivatives are not continuous at $(0,0)$, as shown in Example 11.18 in the book. (Think of this function as an analog of the single-variable function $f(x)=x^2\sin\frac{1}{x}$ we saw back in first-quarter analysis, as an example of a differentiable function with non-continuous derivative.)

Example. Consider the function

$$f(x,y)=\begin{cases}\frac{x^3+y^3}{\sqrt{x^2+y^2}} & (x,y)\neq(0,0)\\0 & (x,y)=(0,0).\end{cases}$$

Last time we showed this was differentiable at the origin by explicitly working out the Jacobian matrix and verifying the limit definition, but now we note that we can also conclude differentiability by verifying that the partial derivatives of f are both continuous at $(0,0)$.

The partial derivative of f with respect to x is explicitly given by

$$f_x(x,y) = \begin{cases} \dfrac{3x^2\sqrt{x^2+y^2} - x\frac{x^3+y^3}{\sqrt{x^2+y^2}}}{x^2+y^2} = \dfrac{3x^2(x^2+y^2)-x(x^3+y^3)}{(x^2+y^2)\sqrt{x^2+y^2}} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

Converting to polar coordinates $x = r\cos\theta$ and $y = r\sin\theta$, we have

$$|f_x(x,y)| = \frac{3r^4\cos^2\theta - r^4\cos\theta(\cos^3\theta + \sin^3\theta)}{r^3} = r(\text{some bounded expression}),$$

which goes to 0 as $r \to 0$. Thus $f_x(x,y) \to 0 = f_x(0,0)$ as $(x,y) \to (0,0)$, so f_x is continuous at $(0,0)$ as claimed. The justification that f_y is continuous at $(0,0)$ involves the same computation only with the roles of x and y reserved. Since the partial derivatives of f are both continuous at $(0,0)$, f is differentiable at $(0,0)$.

Important. If a function has continuous partial derivatives at a point, then it is automatically differentiable at that point. HOWEVER, a function can still be differentiable at a point even if its partial derivatives are not continuous there.

More properties. Finally, we list some properties which allow us to construct new differentiable functions out of old ones, analogous to properties we saw in first-quarter analysis for single-variable functions. Suppose that $f, g : V \to \mathbb{R}^m$ are both differentiable at $\mathbf{a} \in V \subseteq \mathbb{R}^n$. Then:

- $f + g$ is differentiable at $\mathbf{a}$ and $D(f+g)(\mathbf{a}) = Df(\mathbf{a}) + Dg(\mathbf{a})$,
- cf is differentiable at $\mathbf{a}$ and $D(cf)(\mathbf{a}) = cDf(\mathbf{a})$ for $c \in \mathbb{R}$,
- $f \cdot g$ is differentiable at $\mathbf{a}$ and $D(f \cdot g)(\mathbf{a}) = f(\mathbf{a})Dg(\mathbf{a}) + g(\mathbf{a})Df(\mathbf{a})$.

The first two say that "the derivative of a sum is the sum of derivatives" and "constants can be pulled out of derivatives" respectively.

The third is a version of the product rule and requires some explanation. The function $f \cdot g : V \to \mathbb{R}$ in question is the *dot product* of f and g defined by

$$(f \cdot g)(\mathbf{x}) = f(\mathbf{x}) \cdot g(\mathbf{x}).$$

The Jacobian matrix of this dot product is obtained via the "product rule"-like expression given in the third property, where the right side of that expression consists of matrix products: $f(\mathbf{a})$ is an element of $\mathbb{R}^m$ and so is a $1 \times m$ matrix, and $Dg(\mathbf{a})$ is an $m \times n$ matrix so the $f(\mathbf{a})Dg(\mathbf{a})$ is the ordinary matrix product of these and results in a $1 \times n$ matrix; similarly $g(\mathbf{a})Df(\mathbf{a})$ is also a matrix product which gives a $1 \times n$ matrix, so the entire right hand side of that expression is a row vector in $\mathbb{R}^n$, just as the Jacobian matrix of the map $f \cdot g : V \subseteq \mathbb{R}^n \to \mathbb{R}$ should be.

There is also a version of the "quotient rule" for Jacobian matrices of functions $\mathbb{R}^n \to \mathbb{R}$, and a version of the product rule for the *cross product* of functions $f, g : \mathbb{R}^n \to \mathbb{R}^m$ defined by $(f \times g)(\mathbf{x}) = f(\mathbf{x}) \times g(\mathbf{x})$. These are given in some of the exercises in the book, but we won't really use them much, if at all. The main point for us is that these all together give yet more ways of justifying that various functions are differentiable.

Important. Sums, scalar multiples, and (appropriately defined) products of differentiable functions are differentiable, and the Jacobian matrices of sums, scalar multiples, and products obey similar differentiation rules as do single-variable derivatives.

Lecture 7: The Chain Rule

Today we spoke about the chain rule for higher-dimensional derivatives phrased in terms of Jacobian matrices. This is the most general version of the chain rule there is and subsumes all versions you might have seen in a multivariable calculus course.

Warm-Up. Define $f : \mathbb{R}^4 \to \mathbb{R}^4$ by

$$f(x, y, z, w) = (x^2 + yz, xy + yw, xz + wz, zy + w^2).$$

We claim that f is differentiable everywhere. Indeed, one way to justify this is to note that all the partial derivatives of f are polynomial expressions, and so are continuous everywhere. Or, we can say that each component of f is made up by taking sums and products of differentiable functions and so are each differentiable. The point is that we can justify that f is differentiable without any hard work using other tools we've built up.

So that's it for the Warm-Up, but see below for why we looked at this example; in particular, this function isn't just some random function I came up with, but has some actual meaning.

Derivatives of matrix expressions. All examples we've seen of differentiable functions in higher dimensions were thought up of in order to illustrate how to use a certain definition or property, but aren't truly representative of all types of functions you might seen in actual applications. So, here we see how to apply the material we've been developing to a more interesting and "relevant" example, motivated by the fact that higher-dimensional functions expressed in terms of matrices tend to turn up quite a bit in concrete applications. This is a bit of a tangent and isn't something we'll come back to, but hopefully it illustrates some nice ideas.

Let $M_2(\mathbb{R})$ denote the "space" of 2×2 matrices and let $f : M_2(\mathbb{R}) \to M_2(\mathbb{R})$ denote the *squaring* function defined by

$$f(X) = X^2 \text{ for } X \in M_2(\mathbb{R}).$$

Here X^2 denotes the usual matrix product XX. We want to make sense of what it means for f to be differentiable and what the derivative of f should be. (The idea is that we want to differentiate the matrix expression X^2 with respect to the matrix X.) Going by what we know about the analogous function $g(x) = x^2$ on $\mathbb{R}$ we might guess that the derivative of f should be $f'(X) = 2X$, which is not quite right but understanding what this derivative actually is really serves to illustrate what differentiability means in higher dimensions.

First we note that we can identity $M_2(\mathbb{R})$ with $\mathbb{R}^4$ by associating to a 2×2 matrix the vector in $\mathbb{R}^4$ whose coordinates are the entries in that matrix:

$$\begin{pmatrix} x & y \\ z & w \end{pmatrix} \mapsto (x, y, z, w).$$

So, under this identification, the squaring map we're looking at should really be thought us as a function $f : \mathbb{R}^4 \to \mathbb{R}^4$. Concretely, since

$$\begin{pmatrix} x & y \\ z & w \end{pmatrix}^2 = \begin{pmatrix} x & y \\ z & w \end{pmatrix} \begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} x^2 + yz & xy + yw \\ xz + wz & yz + w^2 \end{pmatrix},$$

the squaring function $f : \mathbb{R}^4 \to \mathbb{R}^4$ is given by:

$$f(x, y, z, w) = (x^2 + yz, xy + yw, xz + wz, yz + w^2),$$

which lo-and-behold is the function we looked at in the Warm-Up! Now we see that the point of that Warm-Up was to show that this squaring function was differentiable everywhere.

Now, the Jacobian of this squaring function at $\mathbf{x} = (x, y, z, w)$ is given by:

$$Df(\mathbf{x}) = \begin{pmatrix} 2x & z & y & 0 \\ y & x+w & 0 & y \\ x & 0 & x+w & z \\ 0 & z & y & 2w \end{pmatrix}.$$

It is not at all clear that we can interpret this "derivative" as $2X$ as we guessed above might be the case. But remember that $Df(\mathbf{x})$ is meant to represent a linear transformation, in this case a linear transformation from $\mathbb{R}^4 \to \mathbb{R}^4$. When acting on a vector $\mathbf{h} = (h, k, \ell, m)$, this linear transformation gives:

$$Df(\mathbf{x})\mathbf{h} = \begin{pmatrix} 2x & z & y & 0 \\ y & x+w & 0 & y \\ x & 0 & x+w & z \\ 0 & z & y & 2w \end{pmatrix} \begin{pmatrix} h \\ k \\ \ell \\ m \end{pmatrix} = \begin{pmatrix} 2xh + zk + y\ell \\ hy + xk + wk + ym \\ xh + x\ell + w\ell + zm \\ xk + y\ell + 2wm \end{pmatrix}.$$

Phrasing everything in terms of matrices again, this says that the Jacobian $Df(X)$ of the squaring function is the linear transformation $M_2(\mathbb{R}) \to M_2(\mathbb{R})$ defined by

$$H = \begin{pmatrix} h & k \\ \ell & m \end{pmatrix} \mapsto \begin{pmatrix} 2xh + zk + y\ell & hy + xk + wk + ym \\ xh + x\ell + w\ell + zm & xk + y\ell + 2wm \end{pmatrix}.$$

Perhaps we can interpret our guess that $f'(X) = 2X$ as saying that this resulting linear transformation should be given by the matrix $2X$ in the sense that

$$H \mapsto 2XH.$$

However, a quick computation:

$$2XH = 2\begin{pmatrix} x & y \\ z & w \end{pmatrix}\begin{pmatrix} h & k \\ \ell & m \end{pmatrix} = \begin{pmatrix} 2xh + 2y\ell & 2xk + 2ym \\ 2zh + 2w\ell & 2zk + 2wm \end{pmatrix}$$

shows that this is not the case.

The crucial observation is that the matrix derived above when computing the linear transformation $M_2(\mathbb{R}) \to M_2(\mathbb{R})$ induced by $Df(X)$ is precisely:

$$\begin{pmatrix} 2xh + zk + y\ell & hy + xk + wk + ym \\ xh + x\ell + w\ell + zm & xk + y\ell + 2wm \end{pmatrix} = \begin{pmatrix} x & y \\ z & w \end{pmatrix}\begin{pmatrix} h & k \\ \ell & m \end{pmatrix} + \begin{pmatrix} h & k \\ \ell & m \end{pmatrix}\begin{pmatrix} x & y \\ z & w \end{pmatrix},$$

so that $Df(X)$ is the linear transformation $M_2(\mathbb{R}) \to M_2(\mathbb{R})$ defined by

$$H \mapsto XH + HX.$$

Thus, the conclusion is that the "derivative" $f'(X)$ of the squaring map $X \mapsto X^2$ at $X \in M_2(\mathbb{R})$ is the linear transformation which sends a 2×2 matrix H to $XH + HX$! (Note that, in a sense, this derivative is "almost" like $2X = X + X$ in that there are two X terms which show up and are being added, only after multiplying by H on opposite sides. So, our guess wasn't totally off.) With this in mind, we can now verify that $f(X) = X^2$ is differentiable at any X directly using the definition of differentiability, only phrased in terms of matrices:

$$\lim_{H \to 0} \frac{f(X+H) - f(X) - Df(X)H}{\|H\|} = \lim_{H \to 0} \frac{(X+H)^2 - X^2 - (XH + HX)}{\|H\|} = \lim_{H \to 0} \frac{H^2}{\|H\|} = 0.$$

Note that $(X+H)^2 = X^2 + XH + HX + H^2$, so since $XH \neq HX$ in general we indeed need to have $XH + HX$ in the Jacobian expression in order to have the numerator above simplify to H^2.

Finally, we note that a similar computation shows that the squaring map $X \mapsto X^2$ for $n \times n$ matrices in general is differentiable everywhere and that the derivative at any X is the linear tranformation $H \mapsto XH + HX$. In particular, for 1×1 matrices the squaring map is the usual $f(x) = x^2$ and the derivative acts as

$$h \mapsto xh + hx = 2xh$$

Thus the 1×1 Jacobian matrix $Df(x)$ representing this linear transformation is simply $Df(x) = \begin{pmatrix} 2x \end{pmatrix}$, which agrees with the well known fact that $f'(x) = 2x$ is true in this case. Thus the above results really do generalize what we know about $f(x) = x^2$. Using similar ideas you can then define differentiability and derivatives of more general functions expressed in terms of matrices. Huzzah!

Second Derivatives. As one more aside, we note that we can also make sense of what *the* second derivative (as opposed to the second-order partial derivatives) of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ should mean. To keep notation cleaner, we'll only look at the $f : \mathbb{R}^2 \to \mathbb{R}$ case.

As we've seen, the derivative of f at (x, y) is now interpreted as the 1×2 Jacobian matrix

$$Df(\mathbf{x}) = \begin{pmatrix} f_x(\mathbf{x}) & f_y(\mathbf{x}) \end{pmatrix}.$$

We view this now as a function $Df : \mathbb{R}^2 \to \mathbb{R}^2$ which assigns to any $\mathbf{x} \in \mathbb{R}^2$ the vector in $\mathbb{R}^2$ given by this Jacobian matrix:

$$Df : \mathbf{x} \mapsto Df(\mathbf{x}).$$

To say that f is *twice-differentiable* should mean that this "first derivative" map $Df : \mathbb{R}^2 \to \mathbb{R}^2$ is itself differentiable, in which case the *second derivative* of f should be given by the Jacobian matrix of Df, which we think of as the "derivative of the derivative of f":

$$D^2 f(\mathbf{x}) := D(Df)(\mathbf{x}) = \begin{pmatrix} f_{xx}(\mathbf{x}) & f_{xy}(\mathbf{x}) \\ f_{yx}(\mathbf{x}) & f_{yy}(\mathbf{x}) \end{pmatrix},$$

where these entries are found by taking the partial derivatives of the components of the function $Df : \mathbb{R}^2 \to \mathbb{R}^2$ defined above. This resulting matrix is more commonly known as the *Hessian* of f at $\mathbf{x}$ and is denoted by $Hf(\mathbf{x})$; it literally does play the role of the "second derivative" of f in this higher-dimensional setting. And so on, you can keep going and define higher orders of differentiability in analogous ways.

Chain Rule. The single-variable chain rule says that derivatives of compositions are given by products of derivatives, and now we see that the same is true in the higher-dimensional setting. To be clear, suppose that $g : \mathbb{R}^n \to \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}^p$ are functions with g differentiable at $\mathbf{a} \in \mathbb{R}^n$ and f differentiable at $g(\mathbf{a}) \in \mathbb{R}^m$. (Of course, these functions could be defined on smaller open domains—the only requirement is that the the image of g is contained in the domain of f so that the composition $f \circ g$ makes sense.) The claim is that $f \circ g$ is then differentiable at $\mathbf{a}$ as well and the Jacobian matrix of the composition is given by:

$$D(f \circ g)(\mathbf{a}) = Df(g(\mathbf{a}))Dg(\mathbf{a}),$$

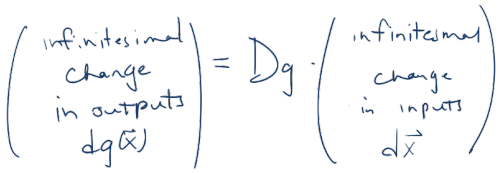
where the right sides denote the ordinary product of the $p \times m$ matrix $Df(g(\mathbf{a}))$ with the $m \times n$ matrix $Dg(\mathbf{a})$. Thus the derivative of $f \circ g$ is indeed the product of the derivatives of f and of g.

Note that in the case $n = m = p = 1$, all of the matrices involved are 1×1 matrices with ordinary derivatives as their entries, and the expression $D(f \circ g)(\mathbf{a}) = Df(g(\mathbf{a}))Dg(\mathbf{a})$ becomes

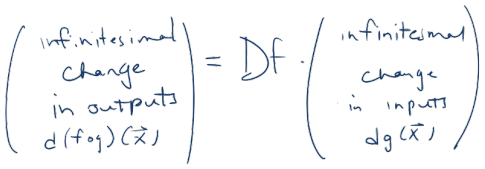
$$(f \circ g)'(a) = f'(g(a))g'(a),$$

which is the usual single-variable chain rule.

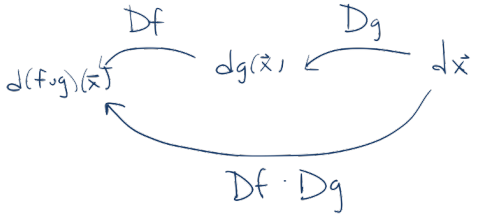
Intuition behind the chain rule. The chain rule is actually pretty intuitive from the point of view that derivatives are meant to measure how a small change in inputs into a function transforms into a small change in outputs. Intuitively, we have:



But the outputs of g can then be fed in as inputs of f so:



and putting it all together gives



Thus the product $Df(g(\mathbf{a}))Dg(\mathbf{a})$ tells us how to transform an infinitesimal change in input into $f \circ g$ into an infinitesimal change in output of $f \circ g$, but this is precisely what the Jacobian matrix of $f \circ g$ at $\mathbf{a}$ should do as well so it makes sense that we should have $D(f \circ g)(\mathbf{a}) = Df(g(\mathbf{a}))Dg(\mathbf{a})$ as the chain rule claims.

Deriving other chain rules. Note also that this statement of the chain rule encodes within it all other versions you would have seen in a multivariable calculus course. To be specific, suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ is a function of two variables $f(x, y)$ and that each of $x = x(s, t)$ and $y = y(s, t)$ themselves depend on some other variables s and t. In a multivariable calculus course you would have seen expressions such as:

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}.$$

We can view the fact that x and y depend on s and t as defining a map $g : \mathbb{R}^2 \, to \mathbb{R}^2$:

$$g(s,t) = (x(s,t), y(s,t)).$$

Then according to the chain rule, the composition $f \circ g$ has Jacobian matrix given by:

$$D(f \circ g) = Df \cdot Dg = \begin{pmatrix} f_x & f_y \end{pmatrix} \begin{pmatrix} x_s & x_t \\ y_s & y_t \end{pmatrix} = \begin{pmatrix} f_x x_s + f_y y_s & f_x x_t + f_y y_t \end{pmatrix}.$$

The second entry of this resulting 1×2 Jacobian should be f_t, so we get

$$f_t = f_x x_t + f_y y_t, \text{ or } \frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}$$

as the well-known formula above claims. In a similar way, all other types of expressions you get when differentiating with respect to the variables which themselves depend on other variables (which themselves possibly depend on other variables, and so on) can be derived from this one chain rule we've given in terms of Jacobian matrices.

Important. For differentiable functions f and g for which the composition $f \circ g$ makes sense, the chain rule says that $f \circ g$ is differentiable and that its Jacobian matrices are given by

$$D(f \circ g)(\mathbf{a}) = Df(g(\mathbf{a}))Dg(\mathbf{a}).$$

Thus, the chain rule is essentially nothing but a statement about matrix multiplication!

Proof of Chain Rule. The proof of the chain rule is in the book and is not too difficult to follow once you understand what the book is trying to do. So here we only give the basic idea and leave the full details to the book. The key idea is to rewrite the numerator in the limit

$$\lim_{\mathbf{h} \to 0} \frac{f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - Df(g(\mathbf{a}))Dg(\mathbf{a})\mathbf{h}}{\|\mathbf{h}\|}$$

which says that $f \circ g$ is differentiable at $\mathbf{a}$ with Jacobian matrix $Df(g(\mathbf{a}))Dg(\mathbf{a})$ in a way which makes it possible to see that this limit is indeed zero.

We introduce the "error" functions

$$\epsilon(\mathbf{h}) = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a}) - Dg(\mathbf{a})\mathbf{h}$$

and

$$\delta(\mathbf{k}) = f(g(\mathbf{a})+\mathbf{k}) - f(g(\mathbf{a})) - Df(g(\mathbf{a}))\mathbf{k}$$

for g at $\mathbf{a}$ and f at $g(\mathbf{a})$ respectively which measure how good/bad the linear approximations to f and g given by their "first derivatives" are. With this notation, note that saying g is differentiable at $\mathbf{a}$ and f is differentiable at $g(\mathbf{a})$ translates to

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\epsilon(\mathbf{h})}{\|\mathbf{h}\|} = \mathbf{0} \text{ and } \lim_{\mathbf{k} \to \mathbf{0}} \frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} = \mathbf{0}$$

respectively. (In other books you might see the definition of differentiability phrased in precisely this way.) After some manipulations, the book shows that the numerator we want to rewrite can be expressed in terms of these error functions as:

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - Df(g(\mathbf{a}))Dg(\mathbf{a})\mathbf{h} = Df(g(\mathbf{a}))\epsilon(\mathbf{h}) + \delta(\mathbf{k}).$$

The proof is finished by showing that

$$\frac{Df(g(\mathbf{a}))\epsilon(\mathbf{h})}{\|\mathbf{h}\|} \text{ and } \frac{\delta(\mathbf{k})}{\|\mathbf{h}\|}$$

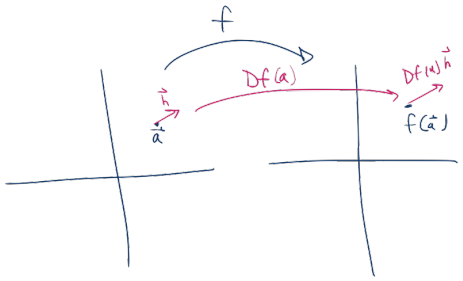
both go to $\mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$. Again, you can check the book for all necessary details.

Lecture 8: Mean Value Theorem

Today we spoke about the Mean Value Theorem in the higher-dimensional setting. For certain types of functions—namely scalar-valued ons—the situation is the same as what we saw for single-variable functions in first-quarter analysis, but for other functions—those which are vector-valued—the statement of the Mean Value Theorem requires some modification.

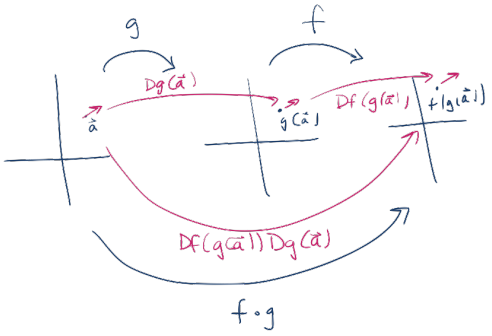
Geometric meaning of the chain rule. Before moving on, we give one more bit of intuition behind the chain rule, this time from a more geometric point of view. The interpretation of Jacobian matrices we've seen in terms of measuring what a function does to infinitesimal changes in input does not seem to be as nice as the interpretation in the single-variable case in terms of slopes, but if we interpret "slope" in a better way we do get a nice geometric interpretation of Jacobians. The key is that we should no longer think of a "slope" as simply being a number, but rather as a vector! (We won't make this more precise, but we are considering what are called "tangent vectors".)

So, for a function $f : \mathbb{R}^n \to \mathbb{R}^m$ we think of having an "infinitesimal" vector at a given point $\mathbf{a}$. The behavior induced by the function f should transform this infinitesimal vector at $\mathbf{a}$ into an infinitesimal vector at $f(\mathbf{a})$:



and the point is that this "infinitesimal" transformation is described precisely by the Jacobian matrix $Df(\mathbf{a})$! That is, the matrix $Df(\mathbf{a})$ tells us how to transform infinitesimal vectors at $\mathbf{a}$ into infinitesimal vectors at $f(\mathbf{a})$, which is the geometric analog of the view that $Df(\mathbf{a})$ transforms small changes in input into small changes in output.

Now, given functions $g : \mathbb{R}^n \to \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}^p$, starting with an infinitesimal vector at $\mathbf{a} \in \mathbb{R}^n$, after applying $Dg(\mathbf{a})$ we get an infinitesimal vector at $f(\mathbf{a})$, which then gets transformed by $Df(g(\mathbf{a}))$ into an infinitesimal vector at $f(g(\mathbf{a}))$:



But the same infinitesimal vector should be obtained by applying the Jacobian of the composition $f \circ g$, so we get that $D(f \circ g)(\mathbf{a})$ should equal $Df(g(\mathbf{a}))Dg(\mathbf{a})$ as the chain rules states.

The point of view that Jacobian matrices tells us how to transform "infinitesimal" vectors (or more precisely *tangent vectors*) would be further developed in a differential geometry course. Being a geometer myself, I'm probably biased but I do believe that this point of view provides the best intuition behind the meaning of higher-dimensional derivatives and differentiability.

Warm-Up. Suppose that $F : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at (a, b) and that $F_y(a, b) \neq 0$. Suppose further that $f : I \to \mathbb{R}$ is a differentiable function on some open interval $I \subseteq \mathbb{R}$ and that $F(x, f(x)) = 0$ for all $x \in I$. We claim that then the derivative of f is given by the expression

$$f'(a) = -\frac{F_x(a, b)}{F_y(a, b)}.$$

Before proving this, here is the point of this setup and claim. Suppose that F is something like $F(x, y) = x^2y + y^3 - xy + x - 1$. Then the equation $F(x, y) = 0$ defines some curve in the xy-plane:

$$x^2y + y^3 - xy + x = 1.$$

The idea is that we should think of this equation as *implicitly* defining $y = f(x)$ as a function of x, so that this curve looks like the graph of this function f. Even if we don't know what f explicitly is, the result of this Warm-Up says that we can nonetheless compute the derivatives of f in terms of the partial derivatives of F, giving us a way to find the slope of the curve at some given point as long as F_y is nonzero at that point. This is a first example of what's called the *Implicit Function Theorem*, which we will talk about in more generality later. Indeed, the result of this Warm-Up is the justification behind the method of "implicit differentiation" you might have seen in a previous calculus course.

Consider the function $g : I \to \mathbb{R}^2$ defined by $g(x) = (x, f(x))$ and look at the composition $F \circ g$. On the one hand, this composition is constant since $F(x, f(x)) = 0$ for all $x \in I$, so its derivative is 0. On the other hand, the chain rule gives:

$$D(F \circ g)(a) = Df(g(\mathbf{a}))Dg(\mathbf{a}) = \begin{pmatrix} F_x(a, b) & F_y(a, b) \end{pmatrix} \begin{pmatrix} 1 \\ f'(a) \end{pmatrix} = F_x(a, b) + F_y(a, b)f'(a).$$

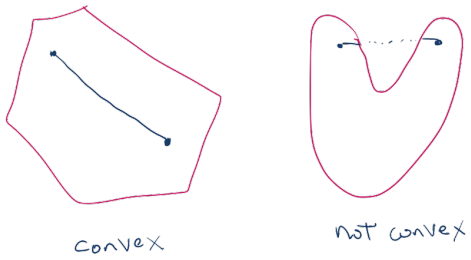
Thus we must have $0 = F_x(a, b) + F_y(a, b)f'(a)$, and solving for $f'(a)$—which we can do since $F_y(a, b)$ is nonzero—gives the desired conclusion.

Convexity. Recall that in the statement of the single-variable Mean Value Theorem, at the end we get the existence of c between some values x and a satisfying some equality. To get a higher-dimensional analog of this we first need to generalize what "between" means in $\mathbb{R}^n$ for $n > 1$. For $\mathbf{x}, \mathbf{a} \in \mathbb{R}^n$ we define $L(\mathbf{x}; \mathbf{a})$ to be the line segment in $\mathbb{R}^n$ between $\mathbf{x}$ and $\mathbf{a}$, so $L(\mathbf{x}; \mathbf{a})$ consists of vectors of the form $\mathbf{a} + t(\mathbf{x} - \mathbf{a})$ for $0 \leq t \leq 1$:

$$L(\mathbf{x}; \mathbf{a}) = \{\mathbf{a} + t(\mathbf{x} - \mathbf{a}) \mid 0 \leq t \leq 1\}.$$

We think of a vector $\mathbf{c} \in L(\mathbf{x}; \mathbf{a})$ as one which is "between" $\mathbf{x}$ and $\mathbf{a}$.

We say that a set $V \subseteq \mathbb{R}^n$ is *convex* if for any $\mathbf{x}, \mathbf{a} \in V$, the entire line segment $L(\mathbf{x}; \mathbf{a})$ lies in V. In the case of $\mathbb{R}^2$, we have:



So, a set is convex if any point "between" two elements of that set is itself in that set. In the case of $\mathbb{R}$, a subset is convex if and only if it is an interval.

First version of Mean Value Theorem. For a first version of the Mean Value Theorem, we consider functions which map into $\mathbb{R}$. Then the statement is the same as it is for single-variable functions, only we replace the single-variable derivative by the Jacobian derivative.

To be precise, suppose that $f : V \to \mathbb{R}$ is differentiable where $V \subseteq \mathbb{R}^n$ is open and convex. Then for any $\mathbf{x}, \mathbf{a} \in V$ there exists $\mathbf{c} \in L(\mathbf{x}; \mathbf{a})$ such that

$$f(\mathbf{x}) - f(\mathbf{a}) = Df(\mathbf{c})(\mathbf{x} - \mathbf{a}).$$

The right side is an ordinary matrix product: since $Df(\mathbf{c})$ is $1 \times n$ And $\mathbf{x} - \mathbf{a}$ (thinking of it as a column vector) is $n \times 1$, the product $Df(\mathbf{c})(\mathbf{x} - \mathbf{a})$ is 1×1, just as the difference $f(\mathbf{x}) - f(\mathbf{a})$ of numbers in $\mathbb{R}$ should be. Note that the book writes this right side instead as a dot product $\nabla f(\mathbf{c}) \cdot (\mathbf{x} - \mathbf{a})$, where $\nabla f(\mathbf{c}) = Df(\mathbf{c})$ is the (row) gradient of f and we think of $\mathbf{x} - \mathbf{a}$ as a row vector. We'll use the $Df(\mathbf{c})(\mathbf{x} - \mathbf{a})$ notation to remain consistent with other notations we've seen involving Jacobians.

Proof. The proof of this Mean Value Theorem works by finding a way to convert $f : \mathbb{R}^n \to \mathbb{R}$ into a single-variable function and then using the single-variable Mean Value Theorem. Here are the details, which are also in the book. The idea is that, since at the end we want to get $\mathbf{c}$ being on the line segment $L(\mathbf{x}; \mathbf{a})$, we restrict our attention to inputs which come only from this line segment, in which case we can indeed interpret f after restriction as a single-variable function.

Define $g : [0, 1] \to V$ by $g(t) = \mathbf{a} + t(\mathbf{x} - \mathbf{a})$, so that g parameterizes the line segment in question, and consider the composition $f \circ g : [0, 1] \to \mathbb{R}$. Since f and g are differentiable, so is this composition and so the single-variable chain rule gives the existence of $t \in (0, 1)$ such that

$$f(g(1)) - f(g(0)) = (f \circ g)'(c)(1 - 0) = (f \circ g)'(0).$$

The left side is $f(\mathbf{x}) - f(\mathbf{a})$ since $g(1) = \mathbf{x}$ and $g(0) = \mathbf{a}$. By the chain rule, the right side is:

$$(f \circ g)'(0) = Df(g(t))Dg(t) = Df(g(t)))(\mathbf{x} - \mathbf{a})$$

where we use the fact that $g'(t) = \mathbf{x} - \mathbf{a}$ for all t. Thus for $\mathbf{c} = g(t) = \mathbf{a} + t(\mathbf{x} - \mathbf{a}) \in L(\mathbf{x}; \mathbf{a})$, we have

$$f(\mathbf{x}) - f(\mathbf{a}) = Df(\mathbf{c})(\mathbf{x} - \mathbf{a})$$

as desired. Note that the convexity of V guarantees that the image of g, which is the line segment $L(\mathbf{x}; \mathbf{a})$, remains within the domain of f, which is needed in order to have the composition $f \circ g$ make sense. □

Incorrect guess for a more general version. The above statement of the Mean Value Theorem for a function $f : \mathbb{R}^n \to \mathbb{R}$ looks to be pretty much the same as the statement

$$f(x) - f(a) = f'(c)(x - a)$$

for a function $f : \mathbb{R} \to \mathbb{R}$, only we use the more general Jacobian derivative. This together with the fact that the equation

$$f(\mathbf{x}) - f(\mathbf{a}) = Df(\mathbf{c})(\mathbf{x} - \mathbf{a})$$

also makes sense when $f : \mathbb{R}^n \to \mathbb{R}^m$ might lead us to believe that the same result should be true even in this more general setting. Indeed, now $Df(\mathbf{c})$ is an $m \times n$ matrix and so its product with the $n \times 1$ matrix $\mathbf{x} - \mathbf{a}$ will give an $m \times 1$ matrix, which is the type of thing which $f(\mathbf{x}) - f(\mathbf{a})$ should be.

However, this more general version is NOT true, as the following example shows. Take the function $f : \mathbb{R} \to \mathbb{R}^2$ defined by

$$f(x) = (\cos x, \sin x).$$

If the above generalization of the Mean Value Theorem were true, there should exist $c \in (0, 2\pi)$ such that

$$f(2\pi) - f(0) = Df(c)(2\pi - 0).$$

But $f(2\pi) = (1, 0) = f(0)$ so this means that $Df(c) = \begin{pmatrix} -\sin c \\ \cos c \end{pmatrix}$ should be the zero matrix, which is not possible since there is no value of c which makes $\sin c$ and $\cos c$ simultaneously zero.

Why the incorrect guess doesn't work. The issue is that, although we can apply the Mean Value Theorem to each *component* of $f : \mathbb{R}^n \to \mathbb{R}^m$, the point $\mathbf{c}$ we get in the statement might vary from one component to the next. To be clear, if $f = (f_1, \dots, f_m)$ are the components of a differentiable $f : \mathbb{R}^n \to \mathbb{R}^m$, for $\mathbf{x}, \mathbf{a} \in \mathbb{R}^n$ the Mean Value Theorem applied to $f_i : \mathbb{R}^n \to \mathbb{R}$ gives the existence of $\mathbf{c}_i \in L(\mathbf{x}; \mathbf{a})$ such that

$$f_i(\mathbf{x}) - f_i(\mathbf{a}) = Df_i(\mathbf{c}_i)(\mathbf{x} - \mathbf{a}).$$

Note that this gives m different points $\mathbf{c}_1, \dots, \mathbf{c}_m$ on this line segment, one for each component of f. However, to get

$$f(\mathbf{x}) - f(\mathbf{a}) = Df(\mathbf{c})(\mathbf{x} - \mathbf{a})$$

we would need the *same* $\mathbf{c} = \mathbf{c}_1 = \mathbf{c}_2 = \dots = \mathbf{c}_m$ to satisfy the component equations above, and there is no way to guarantee that this will happen. In the concrete example we looked at above, we can find c_1 such that

$$1 - 1 = (-\sin c_1)(2\pi - 0)$$

by applying the Mean Value Theorem to the first component $f_1(x) = \cos x$ and we can find c_2 such that

$$1 - 1 = (\cos c_2)(2\pi - 0)$$

by applying it to the second component $f_2(x) = \sin x$, but c_1 and c_2 are different.

The upshot is that such a direct generalization of the Mean Value Theorem for functions $\mathbb{R} \to \mathbb{R}$ or $\mathbb{R}^n \to \mathbb{R}$ does not work for functions $\mathbb{R}^n \to \mathbb{R}^m$ when $m > 1$. Instead, we'll see next time that the best we can do in this higher-dimensional setting is to replace the equality being asked for by an *inequality* instead, which for many purposes is good enough.

Important. For a differentiable function $f : V \to \mathbb{R}$ with $V \subseteq \mathbb{R}^n$ open and convex, for any $\mathbf{x}, \mathbf{a} \in V$ there exists $\mathbf{c} \in L(\mathbf{x}; \mathbf{a})$ such that $f(\mathbf{x}) - f(\mathbf{a}) = Df(\mathbf{c})(\mathbf{x} - \mathbf{a})$. When $n = 1$ and $V = (a, b)$, this is the ordinary single-variable Mean Value Theorem. There is no direct generalization of this to functions $f : V \to \mathbb{R}^m$ when $m > 1$ if we require a similar equality to hold.

Lecture 9: More on Mean Value, Taylor's Theorem

Today we finished talking about the Mean Value Theorem, giving the correct generalization for functions which map into higher-dimensional spaces. We also spoke about Taylor's Theorem, focusing on the second-order statement.

Warm-Up. Suppose that $f : V \to \mathbb{R}^m$ is differentiable on the open, convex set $V \subseteq \mathbb{R}^m$ and that there exists $\mathbf{a} \in V$ such that $Df(\mathbf{x}) = Df(\mathbf{a})$ for all $\mathbf{x} \in V$. (So, the derivative of f is "constant" in the sense that the Jacobian matrix at any point is the constant matrix given by $Df(\mathbf{a})$.) We show that f then has the form $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ for some $m \times n$ matrix A and some $\mathbf{b} \in \mathbb{R}^m$. This is a higher-dimensional analogue of the claim that a single-variable function with constant derivative must look like $f(x) = ax + b$, where the coefficient a is now replaced by a constant matrix and b is replaced by a constant vector.

Let $f = (f_1, \dots, f_m)$ denote the components of f and fix $\mathbf{x} \in V$. For each $1 \leq i \leq m$, applying the Mean Value Theorem to $f_i : V \to \mathbb{R}$ gives the existence of $\mathbf{c}_i \in L(\mathbf{x}; \mathbf{a})$ such that

$$f_i(\mathbf{x}) - f_i(\mathbf{a}) = Df_i(\mathbf{c}_i)(\mathbf{x} - \mathbf{a}).$$

But now $Df(\mathbf{c}_i) = Df(\mathbf{a})$ for any i, so we get

$$f_i(\mathbf{x}) - f_i(\mathbf{a}) = Df_i(\mathbf{a})(\mathbf{x} - \mathbf{a}) \text{ for each } i = 1, \dots, m.$$

The $1 \times n$ row vector $Df_i(\mathbf{a})$ gives the i-th row of the Jacobian matrix $Df(\mathbf{a})$ of f, so the right sides of the above expressions as i varies give the rows of the matrix product $Df(\mathbf{a})(\mathbf{x} - \mathbf{a})$. The left sides $f_i(\mathbf{x}) - f_i(\mathbf{a})$ give the rows of the difference

$$f(\mathbf{x}) - f(\mathbf{a}) = \begin{pmatrix} f_1(\mathbf{x}) - f_1(\mathbf{a}) \\ \vdots \\ f_m(\mathbf{x}) - f_m(\mathbf{a}) \end{pmatrix},$$

so all-in-all the equations above for the different f_i fit together to give the equality

$$f(\mathbf{x}) - f(\mathbf{a}) = Df(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

Thus $f(\mathbf{x}) = Df(\mathbf{a})\mathbf{x} + (f(\mathbf{a}) - Df(\mathbf{a})\mathbf{a})$, so setting $A = Df(\mathbf{a})$ and $\mathbf{b} = f(\mathbf{a}) - Df(\mathbf{a})\mathbf{a})$ gives $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ as the required form of f.

Second version of Mean Value Theorem. As we went through last time, a direct generalization of the Mean Value Theorem phrased as an equality doesn't work when f maps into $\mathbb{R}^m$ with $m > 1$, the issue being that the $\mathbf{c}_i$'s we get when applying the Mean Value Theorem component-wise might be different. Note that in the Warm-Up above we were able to get around this since the function there satisfied $Df(\mathbf{x}) = Df(\mathbf{a})$ for all $\mathbf{x}$, so that the different $Df_i(\mathbf{c}_i)$'s could be replaced using Jacobians evaluated at the same $\mathbf{a}$ throughout.

Instead, for a function mapping into a higher-dimensional $\mathbb{R}^m$ we have the following result, which we still view as a type of Mean Value Theorem. (This is sometimes called the *Mean Value Inequality*.) Suppose that $f : V \to \mathbb{R}^m$ is continuously differentiable (i.e. C^1) on the open subset $V \subseteq \mathbb{R}^n$. (Note the new requirement that the partial derivatives of f be continuous on V.) Then for any *compact* and convex $K \subseteq V$ there exists $M > 0$ such that

$$\|f(\mathbf{x}) - f(\mathbf{a})\| \leq M \|\mathbf{x} - \mathbf{a}\| \text{ for all } \mathbf{x}, \mathbf{a} \in K.$$

So, in the most general higher-dimensional setting the Mean Value Theorem only gives an inequality instead of an equality, but for many purposes we'll see this is good enough. Also, the constant M can be explicitly described as a certain supremum, which we'll elaborate on below; this explicit description is just as important as the fact that such a bound exists, and is where the requirements that f have continuous partial derivatives and that K is compact come in.

Idea behind the proof. The Mean Value Inequality is Corollary 11.34 in the book, and the proof is given there. Here we only give some insights into the proof and leave the full details to the book. To start with, the bound M is explicitly given by

$$M := \sup_{\mathbf{x} \in K} \|Df(\mathbf{x})\| .$$

In order to see that this quantity is defined, note that the continuity of the partial derivatives of f imply that the function $\mathbf{x} \mapsto Df(\mathbf{x})$ mapping $\mathbf{x}$ to the Jacobian matrix of f at $\mathbf{x}$ is continuous when we interpret the $m \times n$ matrix $Df(\mathbf{x})$ as being a vector in $\mathbb{R}^{mn}$. The matrix norm map $Df(\mathbf{x}) \mapsto \|Df(\mathbf{x})\|$ is also continuous, so the composition

$$K \to \mathbb{R} \text{ defined by } \mathbf{x} \mapsto \|Df(\mathbf{x})\|$$

is continuous as well. Since K is compact, the Extreme Value Theorem implies that this composition has a maximum value, which is the supremum M defined above. (So, this supremum is actually a maximum, meaning that $M = \|Df(\mathbf{c})\|$ for some $\mathbf{c} \in K$.)

The book's proof then works by turning $f : V \to \mathbb{R}^m$ into a function which maps into $\mathbb{R}$ alone by composing with the function $g : \mathbb{R}^m \to \mathbb{R}$ defined by $g(\mathbf{y}) = (f(\mathbf{x}) - f(\mathbf{a})) \cdot \mathbf{y}$, and then applying the equality-version of the Mean Value Theorem to this composition $g \circ f : V \to \mathbb{R}$ and using the chain rule to compute $D(g \circ f)$. (Actually, the book uses what it calls Theorem 11.32, which we didn't talk about but makes precise the idea mentioned above of "turning" a function mapping into $\mathbb{R}^m$ into one which maps into $\mathbb{R}$.) Various manipulations and inequalities then give the required statement, and again you can check the book for the details.

To give an alternate idea, we start with the point at which we got stuck before when trying to generalize the Mean Value Theorem as an equality directly, namely the fact that after applying the Mean Value Theorem component-wise we get

$$f_i(\mathbf{x}) - f_i(\mathbf{a}) = Df_i(\mathbf{c}_i)(\mathbf{x} - \mathbf{a})$$

for various $\mathbf{c}_1, \dots, \mathbf{c}_m$. These equalities for $i = 1, \dots, m$ then give the equality

$$f(\mathbf{x}) - f(\mathbf{a}) = \begin{pmatrix} Df_1(\mathbf{c}_1) \\ \vdots \\ Df_m(\mathbf{c}_m) \end{pmatrix} (\mathbf{x} - \mathbf{a})$$

where the $Df_i(\mathbf{c}_i)$ are row vectors. (Note again that we cannot express this resulting matrix as the Jacobian matrix at a single point unless the $\mathbf{c}_i$'s were all the same.) Taking norms and using $\|A\mathbf{y}\| \leq \|A\| \, \|\mathbf{y}\|$ gives:

$$\|f(\mathbf{x}) - f(\mathbf{a})\| \leq \left\| \begin{pmatrix} Df_1(\mathbf{c}_1) \\ \vdots \\ Df_m(\mathbf{c}_m) \end{pmatrix} \right\| \|\mathbf{x} - \mathbf{a}\|$$

Thus if we can bound the norm of the matrix in question we would be done.

The trouble is that, because the rows are evaluated at different $\mathbf{c}_i$'s it is not at all clear that this entire norm is indeed bounded by the M defined above since the that M depended on Jacbobian matrices in which all rows $Df_i(\mathbf{x})$ are evaluated at the same $\mathbf{x}$. Still, this suggests that is it not beyond the realm of reason to believe that the norm we end up with here should be bounded, in particular since, for the same reasons as before, the compactness of K and continuity of the partials of f imply that the norm of each row here attains a supremum value. Turning this into an argument that the given M works as a bound requires some linear algebra which we'll skip, but provides an alternate approach than the book's proof. □

Important. For $f : V \to \mathbb{R}^m$ continuously differentiable with $V \subseteq \mathbb{R}^n$, for any convex and compact $K \subseteq V$ we have

$$\|f(\mathbf{x}) - f(\mathbf{a})\| \leq M \|\mathbf{x} - \mathbf{a}\| \text{ for all } \mathbf{x}, \mathbf{a} \in K$$

where $M = \sup_{\mathbf{x} \in K} \|Df(\mathbf{x})\|$. Moreover, there exists $\mathbf{c} \in K$ such that $M = \|Df(\mathbf{c})\|$.

Example. An easy consequence of the Mean Value Theorem (inequality version) is that any C^1 function on a compact and convex set is uniformly continuous. This is something we already know to be true since continuous functions on compact sets are always uniformly continuous, but the point is that we can give a proof in this special case without making use of this more general fact.

So, suppose that $f : V \to \mathbb{R}^m$ is C^1 and that $K \subseteq V$ is compact and convex. By the Mean Value Theorem there exists $M > 0$ such that

$$\|f(\mathbf{x}) - f(\mathbf{a})\| \leq M \|\mathbf{x} - \mathbf{a}\| \text{ for all } \mathbf{x}, \mathbf{a} \in K.$$

Let $\epsilon > 0$. Then for $\delta = \frac{\epsilon}{M} > 0$ we have that if $\|\mathbf{x} - \mathbf{y}\| < \delta$ for some $\mathbf{x}, \mathbf{y} \in K$, then

$$\|f(\mathbf{x}) - f(\mathbf{y})\| \leq M \|\mathbf{x} - \mathbf{y}\| < M\delta = \epsilon,$$

showing that f is uniformly continuous on K as claimed.

Taylor's Theorem. We finish with a version of Taylor's Theorem in the multivariable setting. Recall that the single-variable Taylor's Theorem from first-quarter analysis is a generalization of the single-variable Mean Value Theorem: this latter theorem gives the existence of c such that

$$f(x) = f(a) + f'(c)(x - a),$$

and (the second order) Taylor's Theorem gives the existence of c satisfying

$$f(x) = f(a) + f'(a)(x - a) + \frac{1}{2} f''(c)(x - a)^2.$$

Of course, Taylor's Theorem gives higher-order statements as well, but the second-order statement is the only one we'll give explicitly in the multivariable setting.

Now we move to the multivariable setting. If $f : V \to \mathbb{R}$ is differentiable where $V \subseteq \mathbb{R}^n$ is open and convex, the Mean Value Theorem gives $\mathbf{c} \in L(\mathbf{x}; \mathbf{a})$ such that

$$f(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{c})(\mathbf{x} - \mathbf{a}).$$

In the case $n = 2$, with $\mathbf{x} = (x, y), \mathbf{a} = (a, b)$, and $\mathbf{c} = (c, d)$ this equation explicitly looks like

$$f(x, y) = f(a, b) + f_x(c, d)(x - a) + f_y(c, d)(y - b).$$

If now f has second-order partial derivatives on V, (the second-order) Taylor's Theorem gives the existence of $\mathbf{c} \in L(\mathbf{x}; \mathbf{a})$ such that

$$f(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T Hf(\mathbf{c})(\mathbf{x} - \mathbf{a}),$$

where Hf denotes the Hessian (i.e. second derivative) of f. Note that, as in the single-variable case, it is only the "second derivative" term $Hf(\mathbf{c})$ which is evaluated at the "between" point $\mathbf{c}$.

To make the notation clear, in the final term we are thinking of $\mathbf{x} - \mathbf{a}$ as a column vector, so the transpose $(\mathbf{x} - \mathbf{a})^T$ denotes the corresponding row vector. Then $(\mathbf{x} - \mathbf{a})^T Hf(\mathbf{c})(\mathbf{x} - \mathbf{a})$ denotes the usual matrix product of a $1 \times n$ vector, an $n \times n$ matrix, and an $n \times 1$ vector, which gives a 1×1 matrix in the end. In the $n = 2$ case, this product explicitly gives:

$$\begin{aligned}(\mathbf{x} - \mathbf{a})^T Hf(\mathbf{c})(\mathbf{x} - \mathbf{a}) &= \begin{pmatrix} x - a & y - b \end{pmatrix} \begin{pmatrix} f_{xx}(c, d) & f_{xy}(c, d) \\ f_{yx}(c, d) & f_{yy}(c, d) \end{pmatrix} \begin{pmatrix} x - a \\ y - b \end{pmatrix} \\ &= f_{xx}(c, d)(x - a)^2 + f_{xy}(c, d)(x - a)(y - b) \\ &\quad + f_{yx}(c, d)(y - b)(x - a) + f_{yy}(c, d)(y - b)^2,\end{aligned}$$

which is analogous to the $f''(c)(x - a)^2$ term in the single-variable version: we have one term for each possible second-order partial derivative, and each is multiplied by an $(x - a)$ and/or $(y - b)$ depending on which two variables the second-order partial derivative is taken with respect to. (The book calls this expression the "second-order total differential" of f and denotes it by $D^{(2)}f$.)

The fact that we can more succinctly encode all of these second-order terms using the single Hessian expression is the reason why we'll only consider the second-order statement of Taylor's Theorem—such a nice expression in terms of matrices is not available once we move to third-order and beyond.

Important. If $f : V \to \mathbb{R}$, where $V \subseteq \mathbb{R}^n$ is open and convex, has second-order partial derivatives on V, then for any $\mathbf{x}, \mathbf{a} \in V$ there exists $\mathbf{c} \in L(\mathbf{x}; \mathbf{a})$ such that

$$f(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T Hf(\mathbf{c})(\mathbf{x} - \mathbf{a}).$$

Setting $\mathbf{x} = \mathbf{a} + \mathbf{h}$ for some $\mathbf{h}$, we can also write this as

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2}\mathbf{h}^T Hf(\mathbf{c})\mathbf{h}.$$

This gives a way to estimate the "error" which arises when approximating $f(\mathbf{x})$ using the linear approximation $f(\mathbf{x}) \approx f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x} - \mathbf{a})$.

Lecture 10: Inverse Function Theorem

Today we spoke about the Inverse Function Theorem, which is one of our final big two theorems concerning differentiable functions. This result is truly a cornerstone of modern analysis as it leads to numerous other facts and techniques; in particular, it is the crucial point behind the Implicit Function Theorem, which is our other big remaining theorem. These two theorems give us ways to show that equations having solutions using only Jacobians.

Warm-Up 1. Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is C^1 and that there exists $0 < C < 1$ such that

$$\|Df(\mathbf{x})\| \leq C \text{ for all } \mathbf{x} \in B_r(\mathbf{a})$$

for some ball $B_r(\mathbf{a})$. (In class we only did this for a ball centered at $\mathbf{0}$ and for $C = \frac{1}{2}$, but this general version is no more difficult.) We claim that f then has a unique fixed point on the closed ball $\overline{B_r(\mathbf{a})}$, meaning that there exists a unique $\mathbf{x} \in \overline{B_r(\mathbf{a})}$ such that $f(\mathbf{x}) = \mathbf{x}$. The idea is that the given condition will imply that f is a contraction on this closed ball, and so the Contraction Mapping Principle we saw at the end of last quarter will apply.

First we show that the given bound also applies when $x \in \partial B_r(\mathbf{a})$, which will then tells us that it applies for all $x \in \overline{B_r(\mathbf{a})}$. Since f is continuously differentiable, the Jacobian map $\mathbf{x} \mapsto Df(\mathbf{x})$ is continuous, so $Df(\mathbf{x}) \to Df(\mathbf{a})$ as $\mathbf{x} \to \mathbf{a}$. The operation of taking the norm of a matrix is also continuous, so $\|Df(\mathbf{x})\| \to \|Df(\mathbf{a})\|$ as $\mathbf{x} \to \mathbf{a}$. Thus if $(\mathbf{x}_n)$ is a sequence of points in $B_r(\mathbf{a})$ which converges to $\mathbf{x} \in \partial B_r(\mathbf{a})$, taking limits in

$$\|Df(\mathbf{x}_n)\| \leq C$$

gives $\|Df(\mathbf{x})\| \leq C$ as claimed. Hence $\|Df(\mathbf{x})\| \leq C$ for all $\mathbf{x} \in \overline{B_r(\mathbf{a})}$, so the supremum of such norms is also less than or equal to C.

Now, since $\overline{B_r(\mathbf{a})}$ is compact (it is closed and bounded) and convex, the Mean Value Theorem implies that

$$\|f(\mathbf{x}) - f(\mathbf{y})\| \leq C\,\|\mathbf{x} - \mathbf{y}\| \text{ for all } \mathbf{x}, \mathbf{y} \in \overline{B_r(\mathbf{a})}.$$

Since $0 < C < 1$, this says that f is a contraction on $\overline{B_r(\mathbf{a})}$. Since $\overline{B_r(\mathbf{a})}$ is closed in the complete space $\mathbb{R}^n$, it is complete itself and so the Contraction Mapping Principle tells us that f has a unique fixed point in $\overline{B_r(\mathbf{a})}$ as claimed.

Warm-Up 2. We show that for h and k small enough, the values of

$$\cos\left(\frac{\pi}{4} + h\right)\sin\left(\frac{\pi}{4} + k\right) \text{ and } \frac{1}{2} - \frac{1}{2}h + \frac{1}{2}k$$

agree to 3 decimal places. Thus the expression on the right will provide a fairly accurate approximation to the expression on the left for such h, k.

Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = \cos x \sin y$. For a given $\mathbf{h} = (h, k)$ and $\mathbf{a} = (\frac{\pi}{4}, \frac{\pi}{4})$, by the second-order statement of Taylor's Theorem there exists $\mathbf{c} = (c, d) \in L(\mathbf{a}; \mathbf{a} + \mathbf{h})$ such that

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2}\mathbf{h}^T Hf(\mathbf{c})\mathbf{h}.$$

The term on the left is $\cos(\frac{\pi}{4} + h)\sin(\frac{\pi}{4} + k)$. We compute:

$$Df(\mathbf{x}) = \begin{pmatrix} -\sin x \sin y & \cos x \cos y \end{pmatrix} \text{ and } Df(\mathbf{a}) = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

so

$$f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} = \frac{1}{2} - \frac{1}{2}h + \frac{1}{2}k.$$

Thus

$$\cos\left(\frac{\pi}{4} + h\right)\sin\left(\frac{\pi}{4} + k\right) - \left(\frac{1}{2} - \frac{1}{2}h + \frac{1}{2}k\right) = f(\mathbf{a} + \mathbf{h}) - [f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h}] = \frac{1}{2}\mathbf{h}^T Hf(\mathbf{c})\mathbf{h},$$

and hence

$$\left|\cos\left(\frac{\pi}{4} + h\right)\sin\left(\frac{\pi}{4} + k\right) - \left(\frac{1}{2} - \frac{1}{2}h + \frac{1}{k}\right)\right| \leq \frac{1}{2}\left\|\mathbf{h}^T\right\|\,\|Hf(\mathbf{c})\|\,\|\mathbf{h}\|.$$

Now, we compute:

$$Hf(\mathbf{c}) = \begin{pmatrix} -\cos c \sin d & -\sin c \cos d \\ -\sin c \cos d & -\cos c \sin d \end{pmatrix}.$$

Note that each of the entries in this matrix has absolute value smaller or equal to than 1. We claim that $\|Hf(\mathbf{c})\| \leq 2\sqrt{2}$. Indeed, denoting this matrix by $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right)$ to keep notation simpler, we have for $\mathbf{x} = (x, y)$ of norm 1:

$$\begin{aligned}
\|Hf(\mathbf{c})\mathbf{x}\| &= \left\| \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right\| \\
&= \left\| \begin{pmatrix} px + qy \\ rx + sy \end{pmatrix} \right\| \\
&= \sqrt{|px + qy|^2 + |rx + sy|^2} \\
&\leq \sqrt{(|p||x| + |q||y|)^2 + (|r||x| + |s||y|)^2} \\
&\leq \sqrt{(1+1)^2 + (1+1)^2} \\
&= 2\sqrt{2}
\end{aligned}$$

where we use the fact that $|x|, |y| \leq 1$ since $\mathbf{x}$ has norm 1. Thus the supremum of the values $\|Hf(\mathbf{c})\mathbf{x}\|$ as $\mathbf{x}$ varies over vectors of norm 1 is also bounded by $2\sqrt{2}$, and this supremum is the definition of $\|Hf(\mathbf{c})\|$.

Hence we have:

$$\left| \cos\left(\frac{\pi}{4} + h\right) \sin\left(\frac{\pi}{4} + k\right) - \left(\frac{1}{2} - \frac{1}{2}h + \frac{1}{2}k\right) \right| \leq \frac{1}{2} \left\|\mathbf{h}^T\right\| \|Hf(\mathbf{c})\| \|\mathbf{h}\| \leq \sqrt{2} \|h\|^2,$$

and so as long as

$$\|\mathbf{h}\| \leq \frac{1}{\sqrt{10000\sqrt{2}}}$$

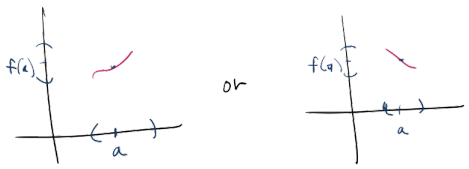
we have

$$\left| \cos\left(\frac{\pi}{4} + h\right) \sin\left(\frac{\pi}{4} + k\right) - \left(\frac{1}{2} - \frac{1}{2}h + \frac{1}{2}k\right) \right| < \frac{1}{10000},$$

which implies that $\cos(\frac{\pi}{4} + h)\sin(\frac{\pi}{4} + k)$ and $\frac{1}{2} - \frac{1}{2}h + \frac{1}{2}k$ agree to 3 decimal places as required.

Inverse Function Theorem. The Inverse Function Theorem is a result which derives local information about a function near a point from infinitesimal information about the function *at* that point. In the broadest sense, it gives a simple condition involving derivatives which guarantees that certain equations have solutions.

Consider first the single-variable version: if $f : (c, d) \to \mathbb{R}$ is differentiable and $f'(c) \neq 0$ for some $a \in (c, d)$, then there exist intervals around a and $f(a)$ on which f is invertible and where f^{-1} is differentiable and $(f^{-1})'(b) = \frac{1}{f'(a)}$, where $b = f(a)$. The key parts to this are that f^{-1} exists near a and $f(a)$ and that it is differentiable—the expression for the derivative of f^{-1} (which says that "the derivative of the inverse is the inverse of the derivative") comes from the chain rule applied to $f^{-1}(f(x)) = x$. This makes some intuitive sense geometrically: if $f'(a) \neq 0$, it is either positive or negative, and so *roughly* f should look something like:



near a. Thus, near a it does appear that f is invertible (if f sends x to y, the inverse f^{-1} sends y to x), and at least in this picture the inverse also appears to be differentiable. Of course, it takes effort to make all this intuition precise.

Here is then the multivariable analog:

> Suppose that $f : V \to \mathbb{R}^n$ is C^1 on an open $V \subseteq \mathbb{R}^n$ and that $Df(\mathbf{a})$ is invertible at some $\mathbf{a} \in V$. (This invertibility condition is equivalent to $\det Df(\mathbf{a}) \neq 0$, which is how the book states it.) Then there exists an open set $W \subseteq V$ containing a such that $f(W)$ is open in $\mathbb{R}^n$, $f : W \to f(W)$ is invertible, and $f^{-1} : f(W) \to W$ is C^1 with $D(f^{-1})(\mathbf{b}) = [Df(\mathbf{a})]^{-1}$ where $\mathbf{b} = f(\mathbf{a})$.

The key parts again are that f^{-1} exists near $\mathbf{a}$ and $f(\mathbf{a})$ and that it is continuously differentiable. The fact that the Jacobian matrix of the inverse of f is the inverse of Jacobian matrix of f follows from the chain rule as in the single-variable case. Often times we won't make the open set W explicit and will simply say that "f is locally invertible at $\mathbf{a}$" or "f is invertible near $\mathbf{a}$".

Intuition and proof. Thus, the Inverse Function Theorem says that if f is "infinitesimally invertible" at a point it is actually invertible near that point with a differentiable inverse. To give some idea as to why we might expect this to be true, recall that near $\mathbf{a}$ the function f is supposed to be well-approximated by $Df(\mathbf{a})$ in the sense that:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x} - \mathbf{a}) \text{ for } \mathbf{x} \text{ near } \mathbf{a}.$$

If $Df(\mathbf{a})$ is invertible, the function on the right is invertible (it is a sum of an invertible linear transformation with a constant vector) and so this approximation suggests that f should be invertible near $\mathbf{a}$ as well. Moreover, the function on the right has an inverse which looks like

$$\mathbf{c} + [Df(\mathbf{a})]^{-1}\mathbf{y}$$

where $\mathbf{y}$ is the variable and $\mathbf{c}$ is a constant vector, and such an inverse is itself differentiable with Jacobian matrix $[Df(\mathbf{a})]^{-1}$. We might expect that this inverse approximates f^{-1} well, so we would guess that f^{-1} is also differentiable with Jacobian matrix $[Df(\mathbf{a})]^{-1}$.

The hard work comes in actually proving all of these claims, which is possibly the most difficult proof we've come across so far among all quarters of analysis. Indeed, in the book this proof is broken up into pieces and takes multiple pages to go through. Going through all the details isn't so important for us, so here we will only give an outline and leave the precise proof to the book. (However, we will give a proof of the first step of the outline which is different than the book's proof.) Here is the outline:

- First, the condition that $Df(\mathbf{a})$ is invertible is used to show that f is invertible near $\mathbf{a}$. This is essentially the content of Lemma 11.40 in the book, which is roughly a big linear algebraic computation. I don't think that this sheds much light on what is actually going on, so we will give an alternate proof of this claim below. In particular, showing that f is invertible essentially requires that we show we can solve $\mathbf{y} = f(\mathbf{x})$ for $\mathbf{y}$ in terms of $\mathbf{x}$, and the book's proof gives no indication as to how this might be done. (This is the sense in which I claimed before that the Inverse Function Theorem is a result about showing that certain equations have solutions.)

- Second, once we know that f^{-1} exists near $\mathbf{a}$, the invertible of $Df(\mathbf{a})$ is used again to show that f^{-1} is actually continuous near $\mathbf{a}$. This is essentially the content of Lemma 11.39 in the book. (Note that the book does this part before the first part.)

- Finally, it is shown that f^{-1} is continuously differentiable and that $D(f^{-1}) = (Df)^{-1}$, which is done in the proof of Theorem 11.41 in the book.

Again, this is a tough argument to go through completely, and I wouldn't suggest you do so until you have some free time, say over the summer when you're feeling sad that Math 320 is over. But, as promised, we give an alternate proof of the first part, which helps to make it clear why $\mathbf{y} = f(\mathbf{x})$ should be solvable for $\mathbf{y}$ in terms $\mathbf{x}$, as needed to define the inverse function f^{-1}. The proof we give depends on properties of contractions and the result of the first Warm-Up from today. The point is that we can rephrase the existence of a solution to $\mathbf{y} = f(\mathbf{x})$ as a fixed point problem, and then the Warm-Up applies.

Proof of first step of the Inverse Function Theorem. Fix $\mathbf{y} \in \mathbb{R}^n$ near $f(\mathbf{a})$, with how "near" this must be still to be determined. Define the function $g : V \to \mathbb{R}^n$ by setting

$$g(\mathbf{x}) = \mathbf{x} + [Df(\mathbf{a})]^{-1}(\mathbf{y} - f(\mathbf{x})).$$

Note that $g(\mathbf{x}) = \mathbf{x}$ if and only if $[Df(\mathbf{a})]^{-1}(\mathbf{y} - f(\mathbf{x})) = \mathbf{0}$, which since $[Df(\mathbf{a})]^{-1}$ is invertible is true if and only if $\mathbf{y} - f(\mathbf{x})$. Thus $\mathbf{x}$ satisfies

$$g(\mathbf{x}) = \mathbf{x} \text{ if and only if } \mathbf{y} = f(\mathbf{x}),$$

which gives the connection between solving $\mathbf{y} = f(\mathbf{x})$ for $\mathbf{x}$ and finding a fixed point $\mathbf{x}$ of g.

The function g is differentiable on V since it is a sum of differentiable functions, and its Jacobian matrix is given by:

$$Dg(\mathbf{x}) = I + [Df(\mathbf{a})]^{-1}](0 - Df(\mathbf{x})) = I - [Df(\mathbf{a})]^{-1}Df(\mathbf{x}),$$

where I denotes the $n \times n$ identity matrix (which is the Jacobian matrix of the identity function $\mathbf{x} \mapsto \mathbf{x}$), $[Df(\mathbf{a})]^{-1}$ remains as is since it behaves like a constant with respect to $\mathbf{x}$ and the Jacobian of the $\mathbf{y}$ term is 0 since it does not depend on $\mathbf{x}$. Writing the identity matrix as $I = [Df(\mathbf{a})]^{-1}Df(\mathbf{a})$, we can rewrite the above as:

$$Dg(\mathbf{x}) = [Df(\mathbf{a})[^{-1}[Df(\mathbf{a}) - Df(\mathbf{x})], \text{ so } \|Dg(\mathbf{x})\| \leq \left\|[Df(\mathbf{a})]^{-1}\right\| \|Df(\mathbf{a}) - Df(\mathbf{x})\|.$$

Since f is C^1, the map $\mathbf{x} \mapsto Df(\mathbf{x})$ is continuous so $Df(\mathbf{x}) \to Df(\mathbf{a})$ as $\mathbf{x} \to \mathbf{a}$. Thus for $\mathbf{x}$ close enough to $\mathbf{a}$ we can make $\|Df(\mathbf{a}) - Df(\mathbf{x})\|$ however small we like, so in particular we can make it small enough so that

$$\|Dg(\mathbf{x})\| \leq \left\|[Df(\mathbf{a})]^{-1}\right\| \|Df(\mathbf{a}) - Df(\mathbf{x})\| \leq \frac{1}{2} \text{ for } \mathbf{x} \text{ close enough to } \mathbf{a}.$$

This "$\mathbf{x}$ close enough to $\mathbf{a}$" business is where the open set $W \subseteq V$ in the statement of the Inverse Function Theorem comes from, and hence the $\mathbf{y}$'s we are considering should be "near" $f(\mathbf{a})$ in the sense that they lie in the image $f(W)$.

Thus g is a contraction on this W, and so the Warm-Up (or rather the Warm-Up applied to some smaller closed ball $\overline{B_r(\mathbf{a})}$ contained in W) tells us that g has a unique fixed point $\mathbf{x}$ in W, which as mentioned previously is then a unique $\mathbf{x}$ satisfying $f(\mathbf{x}) = \mathbf{y}$. The existence of such an $\mathbf{x}$ says that f is onto near $\mathbf{a}$, and the uniqueness of $\mathbf{x}$ says that f is one-to-one, so f is invertible on W with inverse given by $f^{-1}(\mathbf{y}) = \mathbf{x}$. □

Important. If f is C^1 and $Df(\mathbf{a})$ is invertible, then f is invertible near $\mathbf{a}$ and has a C^1 inverse. We should view this as a statement that equations of the form $f(\mathbf{x}) = \mathbf{y}$ can be solved for $\mathbf{x}$ in terms of $\mathbf{y}$ in a "continuously differentiable" way.

What's the point? The Inverse Function Theorem seems like a nice enough result in that it relates "infinitesimal invertibility" to actual invertibility, but nonetheless it's fair to ask why we care about this. Although the Inverse Function Theorem has many applications we won't talk about, the point for us is that one of the main applications is to derive the so-called *Implicit Function Theorem*, which is truly remarkable. We'll spend the next few lectures talking about this Implicit Function Theorem and its applications, and I hope you'll be convinced at the end that it and hence the Inverse Function Theorem really are amazing in that they tell us that things exist without having to know what those things actually are.

Lecture 11: Implicit Function Theorem

Today we started talking about the Implicit Function Theorem, which is our final topic in the "differentiability" portion of the course. As mentioned last time, this theorem is a consequence of the Inverse Function Theorem and gives a way to show that equations have differentiable solutions without having to explicitly derive them. We will continue talking about this next time, where we'll see some interesting applications.

Warm-Up. Suppose that A is an $n \times n$ invertible matrix and that $f : \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$f(\mathbf{x}) = A\mathbf{x} + g(\mathbf{x})$$

where $g : \mathbb{R}^n \to \mathbb{R}^n$ is a C^1 function such that $\|g(\mathbf{x})\| \leq M \|\mathbf{x}\|^2$ for some $M > 0$ and all $\mathbf{x} \in \mathbb{R}^n$. We show that f is locally invertible near $\mathbf{0} \in \mathbb{R}^n$, meaning that there exists an open set containing $\mathbf{0}$ such that f is invertible on this open set. The intuition is that the condition on g says roughly that g is "negligible" as $\mathbf{x} \to \mathbf{0}$, so that the $A\mathbf{x}$ term (with A invertible) should dominate the behavior of f.

First note that $\|g(\mathbf{0})\| \leq M \|\mathbf{0}\|^2 = 0$ implies that $g(\mathbf{0}) = \mathbf{0}$. Moreover,

$$\left\| \frac{g(\mathbf{h})}{\|\mathbf{h}\|} \right\| \leq \frac{M \|\mathbf{h}\|^2}{\|\mathbf{h}\|} = M \|\mathbf{h}\| \to \mathbf{0} \text{ as } \mathbf{h} \to \mathbf{0},$$

so

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{g(\mathbf{h}) - g(\mathbf{0}) - 0\mathbf{h}}{\|\mathbf{h}\|} = 0$$

where 0 denotes the zero matrix and we use the fact that $g(\mathbf{0}) = \mathbf{0}$. This shows that $Dg(\mathbf{0}) = 0$ since the zero matrix satisfies the required property of $Dg(\mathbf{0})$ in the definition of differentiability at $\mathbf{0}$ for g. Since $A\mathbf{x}$ and $g(\mathbf{x})$ are C^1 at $\mathbf{0}$, f is as well and

$$Df(\mathbf{0}) = A + Dg(\mathbf{0}) = A.$$

Since this is invertible the Inverse Function Theorem implies that f is locally invertible near $\mathbf{0}$ as claimed.

Implicit functions. Say we are given an equation of the form

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

where F is a function of two variables. We want to understand conditions under which such an equation *implicitly* defines y as a function x. For instance, we know that

$$x^2 + y^2 - 1 = 0$$

is the equation of the unit circle in $\mathbb{R}^2$ and that we can solve for x or y to get

$$x = \pm\sqrt{1 - y^2} \text{ or } y = \pm\sqrt{1 - x^2}$$

respectively. In this case, we can solve for either x or y *explicitly* in terms of the other variable, at least "locally", meaning that such an explicit solution characterizes one variable in terms of the other over a portion of the entire unit circle: $x = -\sqrt{1 - y^2}$ gives the left half of the circle while the positive square root gives the right half, and $y = -\sqrt{1 - x^2}$ gives the bottom half while the positive square root the top half. Moreover, once we have these explicit functions we can compute derivatives such as $\frac{dy}{dx}$ (giving slopes along the unit circle) or $\frac{dx}{dy}$.

However, if $F(x, y) = 0$ was given by something more complicated, say:

$$x^3y^2 - xy + y - x^2y^4 - 4 = 0,$$

solving for x or y explicitly in terms of the other becomes impossible. Nonetheless, the *Implicit Function Theorem* guarantees that, under some mild Jacobian condition, it will be possible to implicitly solve for x or y in terms of the other, so that we can locally think of the curve $F(x, y) = 0$ as the graph of a single-variable function.

Higher dimensional example. Consider now the system of (non-linear) equations

$$\begin{aligned} xu^2 + yv^2 + xy &= 11 \\ xv^2 + yu^2 - xy &= -1 \end{aligned}$$

for $(u, v, x, y) \in \mathbb{R}^4$, with one solution given by $(1, 1, 2, 3)$. We can ask if there are any other solutions, and if so, what type of geometric object these equations characterize in $\mathbb{R}^4$.

We claim that it is possible to (implicitly) solve these equations for u and v in terms of x and y. Define $F : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ (we're writing the domain $\mathbb{R}^4$ as $\mathbb{R}^2 \times \mathbb{R}^2$ in order to separate the variables u, v we want to solve for from the variables x, y they will be expressed in terms of) by

$$F(u, v, x, y) = (xu^2 + yv^2 + xy - 11, xv^2 + yu^2 - xy + 1),$$

so that the given system of equations is the same as $F(u, v, x, y) = \mathbf{0}$. Denoting the components of F by $F = (F_1, F_2)$, we use $DF_{(u,v)}$ to denote the *partial Jacobian matrix* obtained by differentiating F with respect to only u and v:

$$DF_{(u,v)} = \begin{pmatrix} \frac{\partial F_1}{\partial u} & \frac{\partial F_1}{\partial v} \\ \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v} \end{pmatrix} = \begin{pmatrix} 2xu & 2yv \\ 2yu & 2xv \end{pmatrix}.$$

(Note that the book does NOT use this notation.) The *partial Jacobian determinant* is then

$$\frac{\partial(F_1, F_2)}{\partial(u, v)} := \det DF_{(u,v)} = 4x^2uv - 4y^2uv.$$

Evaluating at the point $(1, 1, 2, 3)$ which we know is a solution of the original system of equations, we get:

$$\frac{\partial(F_1, F_2)}{\partial(u, v)}(1, 1, 2, 3) = 16 - 36 = -20,$$

so since this is nonzero the partial Jacobian matrix

$$DF_{(u,v)}(1,1,2,3) = \begin{pmatrix} 4 & 6 \\ 6 & 4 \end{pmatrix}$$

is invertible. The Implicit Function Theorem (which we'll state in a bit) thus guarantees that it is possible to solve for $u = u(x,y)$ and $v = v(x,y)$ in terms of x and y (at least implicilty) locally near $(1,1,2,3)$, giving many points $(u(x,y), v(x,y), x, y)$ which satisfy the equations $F(u,v,x,y) = \mathbf{0}$.

Thus, even without knowing what $u(x,y)$ and $v(x,y)$ explicitly are, we know that the given equations have infinitely many solutions near $(1,1,2,3)$ and that near this point the equations define a 2-dimensional *surface* in $\mathbb{R}^4$, where we know this is 2-dimensional since the resulting "parametric equations" $u(x,y)$ and $v(x,y)$ in the end depend on two parameters. In this case, explicitly solving for u and v in terms of x and y is not possible, but the point is that we can answer the questions we want without knowing these explicit solutions. Moreover, as we'll see, the Implicit Function Theorem also gives a way to *explicitly* compute the Jacobian matrix obtained by differentiating $u(x,y)$ and $v(x,y)$ with respect to x and y, so that in the end, even though we don't know $u(x,y)$ and $v(x,y)$, we do know their partial derivatives and the fact that they satisfy the system of equations $F(u,v,x,y) = \mathbf{0}$.

Implicit Function Theorem. Here then is the statement. Suppose that $F : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ is C^1 and that $F(\mathbf{x}_0, \mathbf{t}_0) = \mathbf{0}$, where $\mathbf{x}_0 \in \mathbb{R}^n$ and $\mathbf{t}_0 \in \mathbb{R}^k$. (The $\mathbf{x}$'s in $\mathbb{R}^n$ are the variables we want to solve for and the $\mathbf{t}$'s in $\mathbb{R}^k$ are the variables the $\mathbf{x}$'s will be expressed in terms of.) If the partial Jacobian matrix $DF_{\mathbf{x}}(\mathbf{x}_0, \mathbf{t}_0)$ is invertible, then there exists an open set $W \subseteq \mathbb{R}^k$ containing $\mathbf{t}_0$ and a unique C^1 function $g : W \to \mathbb{R}^n$ such that $F(g(\mathbf{t}), \mathbf{t}) = \mathbf{0}$. Moreover, the Jacobian matrix of g is given by

$$Dg(\mathbf{t}) = -[DF_{\mathbf{x}}(g(\mathbf{t}), \mathbf{t}))]^{-1} DF_{\mathbf{t}}(g(\mathbf{t}), \mathbf{t}).$$

Some remarks are in order. Thinking of $F(\mathbf{x}, \mathbf{t}) = \mathbf{0}$ as a system of n non-linear equations in terms of $n + k$ variables, the goal is to solve for the n variables in $\mathbf{x}$ in terms of the k variables in $\mathbf{t}$. Note that we, in theory, we can solve for as many variables as there are equations. The partial Jacobian matrix $DF_{\mathbf{x}}$ is obtained by differentiating with respect to the variables we want to solve for, and the fact that there are as many of these as there are equations (i.e. components of F) guarantees that $DF_{\mathbf{x}}$ is a square matrix so that it makes sense to talk about it being invertible. The function $g : W \to \mathbb{R}^n$ obtained has as its components the expressions for $\mathbf{x} = g(\mathbf{t})$ in terms of $\mathbf{t}$ we are after, which hold only "locally" on W near $\mathbf{t}_0$. The result that $F(g(\mathbf{t}), \mathbf{t}) = \mathbf{0}$ says that indeed the points $(\mathbf{x} = g(\mathbf{t}), \mathbf{t})$ satisfying the equations given by the components of F, so we get one such simultaneous solutions for each $t \in W$. Said another way, this says that locally near $(\mathbf{x}_0, \mathbf{t}_0)$, the object defined by $F(\mathbf{x}, \mathbf{t}) = \mathbf{0}$ is given by the graph of g.

So, the Implicit Function Theorem essentially says that if we have one solution of a system of non-linear equations and an invertibility condition on a certain partial Jacobian, we can solve for some variables in terms of the others. In the higher-dimensional example given above, the function g is the one defined by

$$g(x,y) = (u(x,y), v(x,y))$$

where $u(x,y)$ and $v(x,y)$ are the implicit expressions for u and v in terms of x and y. We will look at the idea behind the proof (which depends on the Inverse Function Theorem) next time. Note that the statement of this theorem in the book does not include the part about the explicit expression for the Jacobian of g.

Important. Using an invertibility criteria, the Implicit Function Theorem gives a way to show that systems of equations have solutions, by showing that locally these equations characterize the graph

of a function. This helps to understand the object defined by those equations, by "parameterizing" solutions in terms of only some of the variables involved.

Back to two-variable case. Let us see what the Implicit Function Theorem looks like in the case where $F : \mathbb{R}^2 \to \mathbb{R}$, so that $F(x, y) = 0$ describes a curve in $\mathbb{R}^2$. Say $F(a, b) = 0$. The partial Jacobian matrix

$$DF_y(a, b) = \begin{pmatrix} F_y(a, b) \end{pmatrix}$$

is just a 1×1 matrix, so the condition required is that $F_y(a, b) \neq 0$. Then the Implicit Function Theorem gives a function $y = f(x)$ which expresses y implicitly as a function of x, so that $F(x, y) = 0$ is (locally) the graph of $f : I \to \mathbb{R}$ where $I \subseteq \mathbb{R}$ is an open interval. In particular, for the point (a, b) satisfying $F(x, y) = 0$ we have $b = f(a)$. The expression

$$Dg(\mathbf{t}) = -[DF_{\mathbf{x}}(g(\mathbf{t}), \mathbf{t}))]^{-1} DF_{\mathbf{t}}(g(\mathbf{t}), \mathbf{t})$$

in this case becomes

$$f'(a) = -\frac{F_x(a, b)}{F_y(a, b)},$$

allowing us to compute the slope of the graph of f (and hence of the curve $F(x, y) = 0$) at $x = a$ in terms of the partial derivatives of F. (Note that this expression for f' was derived back in the Warm-Up from April 15th using the chain rule.)

Lecture 12: More on Implicit Functions

Today we spoke more about the Implicit Function Theorem, giving an outline of the proof and some possible applications. We also pointed out a linear algebraic analog to keep in mind which helps to get at the heart of the Implicit Function Theorem.

Warm-Up. Suppose that $T : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ is a linear transformation, which we write as

$$T(\mathbf{x}, \mathbf{t}) = A\mathbf{x} + B\mathbf{t}$$

where $(\mathbf{x}, \mathbf{t}) \in \mathbb{R}^n \times \mathbb{R}^k$, A is an $n \times n$ matrix, and B is an $n \times k$ matrix. (All together A and B form an $n \times (n+k)$ matrix $\begin{pmatrix} A & B \end{pmatrix}$, which is the matrix of $T : \mathbb{R}^{n+k} \to \mathbb{R}^n$. With this notation, the A part is multiplied only by the first n entries $\mathbf{x}$ of a vector in $\mathbb{R}^{n+k}$ and the B part is multiplied only by the final k entries $\mathbf{t}$.) Supposing that A is invertible and that $T(\mathbf{x}, \mathbf{t}) = \mathbf{0}$, we show that we can solve for $\mathbf{x}$ in terms of $\mathbf{t}$.

The partial Jacobian matrix $DT_{\mathbf{x}}$ is simply A since only A involves the variables given in $\mathbf{x}$. Since this is invertible and linear transformations are C^1, the Implicit Function Theorem implies that we can indeed solve for $\mathbf{x}$ in terms of $\mathbf{t}$ in the system of equations $T(\mathbf{x}, \mathbf{t}) = \mathbf{0}$, so we are done.

Of course, using what we know about matrices, we can actually solve for $\mathbf{x}$ in terms of $\mathbf{t}$ explicitly in this case. After all, $T(\mathbf{x}, \mathbf{t}) = 0$ is

$$A\mathbf{x} + B\mathbf{t} = 0, \text{ or } A\mathbf{x} = -B\mathbf{t},$$

and since A is invertible we have

$$\mathbf{x} = -A^{-1}B\mathbf{t}.$$

So actually, we didn't need the Implicit Function Theorem to solve this, only linear algebra. The point is not to view this result as a consequence of the Implicit Function Theorem as we did above, but rather to view it as *motivation* for the Implicit Function Theorem. The equation

$$A\mathbf{x} + B\mathbf{t} = \mathbf{0}$$

gives a system of n *linear* equations in $n+k$ unknowns $(\mathbf{x}, \mathbf{t})$, and the condition that A is invertible says that the corresponding *augmented matrix* $\begin{pmatrix} A & B \end{pmatrix}$ has rank n. Solving these equations using Gaussian elimination (i.e. using row operations to reduce to echelon form) will in the end express the first n variables $\mathbf{x}$ in terms of the final k "free" variables $\mathbf{t}$, an expression which is explicitly given by $\mathbf{x} = -A^{-1}B\mathbf{t}$.

Thus, we should view the Implicit Function Theorem as a non-linear analog of this linear algebraic fact! The requirement that $DF_{\mathbf{x}}$ be invertible is a type of "rank" or "linear independence" condition, and the implicit function g expresses the "dependent" variables $\mathbf{x}$ in terms of the "independent" or "free" variables $\mathbf{t}$. The claimed expression for the Jacobian matrix of g:

$$Dg(\mathbf{t}) = -[DF_{\mathbf{x}}(g(\mathbf{t}), \mathbf{t}))]^{-1} DF_{\mathbf{t}}(g(\mathbf{t}), \mathbf{t})$$

mimics the $\mathbf{x} = -A^{-1}B\mathbf{t}$ equation obtained in the linear-algebraic case.

Important. The Implicit Function Theorem is a non-linear analog of the fact that systems of linear equations with more unknowns than equations and full rank always have infinitely many solutions, which can moreover be expressed solely in terms of "free" variables.

Idea behind the proof of Implicit Function Theorem. We will here give an idea behind the proof of the Implicit Function Theorem, only going as far as explaining where the implicit function g and its Jacobian come from. Check the book for full details and for the rest of the proof.

First, the idea that the Implicit Function Theorem should be a consequence of the Inverse Function Theorem comes from the point of view that both the idea that, under suitable conditions on Jacobians, certain equations have solutions: $F(\mathbf{x}, \mathbf{t}) = \mathbf{0}$ in the case of the Implicit Function Theorem and $\mathbf{y} = f(\mathbf{x})$ in the case of the Inverse Function Theorem since finding f^{-1} amounts to solving $\mathbf{y} = f(\mathbf{x})$ for $\mathbf{x}$ in terms of $\mathbf{y}$. However, the Inverse Function Theorem only applies to functions between spaces of the same dimension, while here in the Implicit Function Theorem we have $F : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$, so we need a way of "extending" this function F to one between spaces of the same dimension.

Define $\widetilde{F} : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n \times \mathbb{R}^k$ by

$$\widetilde{F}(\mathbf{x}, \mathbf{t}) = (F(\mathbf{x}, \mathbf{t}), \mathbf{t}),$$

so the "extra dimensions" required come from keeping track of $\mathbf{t}$. The Jacobian matrix of $\widetilde{F}$ then looks like:

$$D\widetilde{F} = \begin{pmatrix} DF_{\mathbf{x}} & DF_{\mathbf{t}} \\ 0 & I \end{pmatrix},$$

where $DF_{\mathbf{x}}$ and $DF_{\mathbf{t}}$ are partial Jacobian $n \times n$ and $n \times k$ matrices respectively and come from differentiating the $F(\mathbf{x}, \mathbf{t})$ component of $\widetilde{F}$, 0 in the lower left denotes the $k \times n$ zero matrix and comes from differentiating the $\mathbf{t}$ component of $\widetilde{F}$ with respect to $\mathbf{x}$, and I in the lower right denotes the $k \times k$ identity matrix and comes from differentiating the $\mathbf{t}$ component with respect to $\mathbf{t}$. At $(\mathbf{x}_0, \mathbf{t}_0)$ we have thus have:

$$D\widetilde{F}(\mathbf{x}_0, \mathbf{t}_0) = \begin{pmatrix} DF_{\mathbf{x}}(\mathbf{x}_0, \mathbf{t}_0) & DF_{\mathbf{t}}(\mathbf{x}_0, \mathbf{t}_0) \\ 0 & I \end{pmatrix}.$$

Since the two main diagonal "blocks" $DF_{\mathbf{x}}(\mathbf{x}_0, \mathbf{t}_0)$ and I are invertible (recall the assumptions of the Implicit Function Theorem), this entire Jacobian matrix is invertible as well. Thus by the Inverse Function Theorem, $\widetilde{F}$ is locally invertible near $(\mathbf{x}_0, \mathbf{t}_0)$. The portion of the inverse of $\widetilde{F}$:

$$(\widetilde{F})^{-1} : (F(\mathbf{x}, \mathbf{t}), \mathbf{t}) \mapsto (\mathbf{x}, \mathbf{t})$$

which sends the second component $\mathbf{t}$ of the input to the first component $\mathbf{x}$ of the output then gives the implicit function $g(\mathbf{t}) = \mathbf{x}$ which is asked for in the Implicit Function Theorem. The fact that g is C^1 comes from the fact that the inverse of $\widetilde{F}$ is C^1, which is part of the statement of the Inverse Function Theorem.

The claimed expression for Dg as

$$Dg(\mathbf{t}) = -[DF_{\mathbf{x}}(g(\mathbf{t}), \mathbf{t}))]^{-1} DF_{\mathbf{t}}(g(\mathbf{t}), \mathbf{t})$$

can be justified in multiple ways. One way is via the chain rule, which is done in the solution to Problem 7 from the collection of practice problems for the first midterm. Another way is as follows. First consider a 2×2 matrix of the form

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix},$$

which is a simplified version of the form for $D\widetilde{F}$ we have above. This has inverse

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}^{-1} = \frac{1}{a} \begin{pmatrix} 1 & -b \\ 0 & a \end{pmatrix} = \begin{pmatrix} a^{-1} & -a^{-1}b \\ 0 & 1 \end{pmatrix}.$$

It turns out that the same type of expression works for the matrix we have consisting of four "blocks", so that:

$$(D\widetilde{F})^{-1} = \begin{pmatrix} [DF_{\mathbf{x}}]^{-1} & -[DF_{\mathbf{x}}]^{-1} DF_{\mathbf{t}} \\ 0 & I \end{pmatrix}.$$

Since g is the portion of $\widetilde{F}$ which sends the second component of $(F(\mathbf{x}, \mathbf{t}), \mathbf{t})$ to the first component of $(\mathbf{x}, \mathbf{t})$, Dg should come from the upper-right part of this matrix, giving the desired expression for Dg. (This approach requires knowing some linear algebraic properties of block matrices. Instead, the chain rule method described in Problem 7 of the practice problems is much more rigorous.)

Approximating implicit functions. Let us comment on one aspect of the implicit function theorem which might seem troubling: the fact that the function g is usually only implicitly and not explicitly defined. Although explicitly knowing Dg and the fact that g satisfies $F(g(\mathbf{x}), \mathbf{x}) = \mathbf{0}$ is often times good enough, other times having some better information about g itself is necessary. While we can almost never explicitly find g, it turns out that there are fairly good ways of *approximating* g.

The key comes from the contraction/fixed-point approach to the Inverse Function Theorem we outlined a few lectures ago. In the end, the function g we're after comes from inverting some function $\widetilde{F}$, and finding $\widetilde{F}^{-1}$ (i.e. solving $\mathbf{w} = \widetilde{F}(\mathbf{z})$ for $\mathbf{z}$ in terms of $\mathbf{w}$) amounts to finding a fixed point of a function of the form

$$h(\mathbf{z}) = \mathbf{z} + [D\widetilde{F}(\mathbf{a})]^{-1}(\mathbf{w} - \widetilde{F}(\mathbf{z})).$$

As we saw when looking at contractions and fixed points at the end of last quarter, this fixed point is obtained as the limit of a sequence

$$\mathbf{z},\ h(\mathbf{z}),\ h(h(\mathbf{z})),\ \dots$$

obtained by repeated application of h. Thus, we can approximate the implicit function g via an iterated procedure where we take some $\mathbf{z}$ and apply h over and over again. This description is a

bit vague, but rest assured that there do exist honest workable algorithms and techniques which turn this idea into a concrete way of approximating implicit functions.

Application 1. We finish by outlining some applications of the Implicit Function Theorem. To start with, the Implicit Function Theorem lies behind the fact that any randomly drawn sufficiently "smooth" surface in $\mathbb{R}^3$ can be described using equations. This is clear for something like a sphere which can be described using

$$x^2 + y^2 + z^2 = R^2,$$

and for other familiar surfaces, but is not so clear for a surface which does not come from such a recognizable shape. Nonetheless, the Implicit Function Theorem implies that such random-looking surfaces can indeed be described via equations, and in particular locally as the graphs of smooth functions. This gives rise to the idea that we can describe such surfaces *parametrically*, at least implicitly.

The same is true for sufficiently smooth geometric objects in higher dimensions, and such applications are at the core of why computations in *differential geometry* work the way they do. Without the Implicit Function Theorem, we would really have no way of approaching higher-dimensional geometric structures in a systematic way.

Application 2. For an economic application, suppose we have some products we're producing at various quantities $\mathbf{x} = (x_1, \dots, x_n)$ we have control over. But there are also variables we have no control over, say based on prices the market determines or the amount of labor available, and so on—denote these variables by $\mathbf{t} = (t_1, \dots, t_k)$. Given some relation between these different variables

$$F(\mathbf{x}, \mathbf{t}) = \mathbf{0}$$

we look to be able to determine how to adjust our production in response to changes in the uncontrolled various; in other words, we'd like to be able to solve for $\mathbf{x}$ in terms of $\mathbf{t}$. Under some suitable "non-degeneracy" conditions the Implicit Function Theorem says that we can in fact do so. Moreover, even though we may have an explicit expression for $\mathbf{x}$ in terms of $\mathbf{t}$, we can indeed get an explicit expression for the rate of change of $\mathbf{x}$ with respect to $\mathbf{t}$ via the Jacobian Dg of the implicit function, so that we can predict how we should adjust production in order to maximize profit given a change in the market.

Application 3. The method of Lagrange multipliers is no doubt a basic technique you saw in a multivariable calculus course. The statement is that at a point $\mathbf{a}$ at which a function $f : \mathbb{R}^n \to \mathbb{R}$ achieves a maximum or minimum among points satisfying some constraint equations:

$$g_1(\mathbf{x}) = 0, \ \dots, \ g_k(\mathbf{x}) = 0,$$

there exists scalars $\lambda_1, \dots, \lambda_k$ such that

$$\nabla f(\mathbf{a}) = \lambda_1 \nabla g_1(\mathbf{a}) + \cdots + \lambda_k \nabla g_k(\mathbf{a}).$$

The point is that if we want to optimize f subject to the given constraints, we should first try to solve the equation

$$\nabla f(\mathbf{a}) = \lambda_1 \nabla g_1(\mathbf{a}) + \cdots + \lambda_k \nabla g_k(\mathbf{a})$$

in order to find the candidate points $\mathbf{a}$ at which maximums or minimums can occur.

The derivation of the equation above which must be satisfies at such a maximum or minimum depends on the Implicit Function Theorem. The idea is that saying such scalars λ_i exist at a

critical point $\mathbf{a}$ amounts to saying that we can find (i.e. solve for) the λ_i's in terms $\mathbf{a}$. Applying the Implicit Function Theorem to the expression

$$F(\mathbf{a}, \lambda) = \nabla f(\mathbf{a}) - \lambda_1 \nabla g_1(\mathbf{a}) - \cdots - \lambda_k \nabla g_k(\mathbf{a})$$

in the end gives $\lambda = g(\mathbf{a})$, which says that such λ_i's exist. Check the optional section at the end of Chapter 11 in the book for a complete derivation.

Application 4. Finally, we give an application in the infinite-dimensional setting, which is quite powerful in modern mathematics. (We didn't talk about this example in class.) The Implicit Function Theorem we've given applies to functions $F : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ between finite dimensional spaces, but it turns out that there are version of this theorem for functions between infinite-dimensional (even of uncountable dimension) spaces as well. We won't state this version, but will say that the book's proof of the Inverse Function Theorem does not generalize to this infinite dimensional setting since it involves solving systems of finitely many linear equations in terms of finitely many unknowns; instead, to get the infinite-dimensional version of the Inverse and hence Implicit Function Theorems you need to use the contraction/fixed-point approach, which *does* work in this setting since it applies to general metric spaces.

A *partial differential equation* (abbreviated PDE) is an equation involving an unknown function $f : \mathbb{R}^n \to \mathbb{R}^m$ and its partial derivatives. For instance, the equation

$$u_{xx} + u_{yy} + u_{zz} = u_t$$

for a function $u(x, y, z, t)$ is called the *heat equation* and models those functions which describe the dissipation of heat in a 3-dimensional region. To solve this PDE means to find all such functions. In rare instances this is possible to do explicitly, but the behavior of such equations in general is complicated enough that explicit solutions are impossible to come by. Nonetheless, versions of the Implicit Function Theorem in infinite-dimensions guarantee that certain PDEs do have solutions and gives a way to *parametrize* the space of solutions. (The infinite-dimensional space being considered here is not solely made up of the variables $\mathbf{x}$, but also consists of the functions being considered themselves; this is analogous to looking at metric spaces like $C[a, b]$ from last quarter where the "points" in such spaces are actually functions.) Moreover, the fact that we can approximate these implicit solutions using iterations as described earlier gives methods for approximating solutions of PDEs, which in many applications is good enough. These ideas are crucial to understanding the various types of PDEs which pop-up across diverse disciplines: the Schrödinger equation in physics, the Black-Scholes equation in economics and finances, the wave equation in engineering, etc. The (infinite-dimensional) Implicit Function Theorem *is* the only reason we know how to work with such equations in general.

Important. The Implicit Function Theorem has some truly amazing applications, most of the interesting of which go far beyond the scope of this course.

Lecture 13: Jordan Measurability

Today we moved on to the "integration" portion of the course, starting with characterizing regions in $\mathbb{R}^n$ which have a well-defined "volume". Such regions will form the domains of integration for multivariable integrals.

Moral of first four weeks. Before moving on, let us emphasize the key take-away from the first four weeks of the course:

locally, multivariable calculus is linear algebra.

The idea is that basically every concept we've seen when dealing with higher-dimensional differentiability stems from some corresponding concept in linear algebra—the linear algebra describes what happens "infinitesimally" at a point and the calculus describes how that translates into some "local" property near that point.

Indeed, here is a table of analogies to keep in mind, where the concept in the first column is meant to be the nonlinear analog of the linear concept in the second column:

Calculus (nonlinear)	Linear Algebra
differentiable function	matrix or linear transformation
chain rule	matrix multiplication gives composition of linear transformations
Mean Value Equality	linearity of A: $A\mathbf{x} - A\mathbf{y} = A(\mathbf{x} - \mathbf{y})$
Mean Value Inequality	Cauchy-Schwarz: $\|A\mathbf{x} - A\mathbf{y}\| \leq \|A\| \, \|\mathbf{x} - \mathbf{y}\|$
Inverse Function Theorem	invertible matrices define invertible linear transformations
Implicit Function Theorem	in a system of n equations in $n+k$ unknowns of rank n, can solve for n variables in terms of other k variables

In particular, the last line corresponding to the Implicit Function Theorem was the point of the Warm-Up from last time. In a nutshell, whenever you have some nice linear algebraic fact available, there is likely to be some corresponding non-linear analog in calculus.

Regions of integration. Before looking at multivariable integration, we recall the single-variable integral. We consider functions $f : [a, b] \to \mathbb{R}$ satisfying the property that the "upper" and "lower" sums can be made arbitrarily close, as expressed by the inequalities

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

we saw back in the first quarter. A key point here is that we only defined integration over closed intervals $[a, b]$. More generally, we can extend the definition of the integral to finite unions of disjoint closed intervals, where we define, say, the integral of f over $[a, b] \cup [c, d]$ to be the sum of the integrals of f over $[a, b]$ and over $[c, d]$:

$$\int_{[a,b]\cup[c,d]} f(x)\,dx := \int_{[a,b]} f(x)\,dx + \int_{[c,d]} f(x)\,dx,$$

and similarly for finite unions of more than two closed intervals.

However, this is as far as we can go with the Riemann integral of the first quarter. (On the last day of the first quarter we briefly looked at what's called the *Lebesgue integral*, which allows us to extend single-variable integration further, but that is not what we are talking about here.) What all of these regions of integration, obtained as finite unions of disjoint closed intervals, have in common is that they all have a well-defined "length"—indeed, the length of $[a, b] \cup [c, d]$, when this union is disjoint, is just the length of $[a, b]$ plus the length of $[c, d]$.

Thus, the upshot is that the single-variable Riemann integral is only defined over subsets of $\mathbb{R}$ with a well-defined length. By analogy, we might expect that a similar thing should be true for integration in higher-dimensional spaces, only replacing "length" by "volume". (In the case of $\mathbb{R}^2$, "volume" just means "area".) This is correct, but it requires that we give a precise notion of what it means for a subset of $\mathbb{R}^n$ to have a volume, a complication which wasn't readily apparent in the case of single-variable integration.

Jordan outer sums. How can we precisely measure the volume of a subset of $\mathbb{R}^n$? First, the set should definitely be bounded, which implies that we can encase it in a large enough rectangular box:

$$R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n],$$

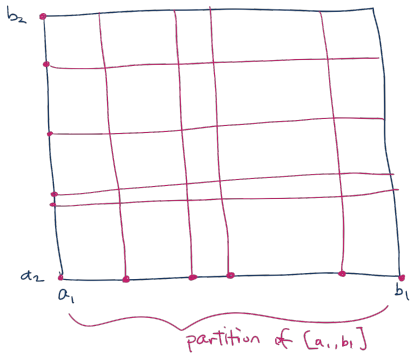
which denotes the region in $\mathbb{R}^n$ consisting of points $(x_1, \ldots, x_n)$ where the i-th coordinate x_i is in the closed interval $[a_i, b_i]$. In the case of $\mathbb{R}^2$ this gives an ordinary rectangle. (Most pictures we draw from now on will take place in $\mathbb{R}^2$, since things are simpler to visualize there.)

But we know geometrically that the "volume", whatever that means, of such a box should just be obtained by multiplying the lengths of the various edges, so we define the volume $|R|$ of R to be:

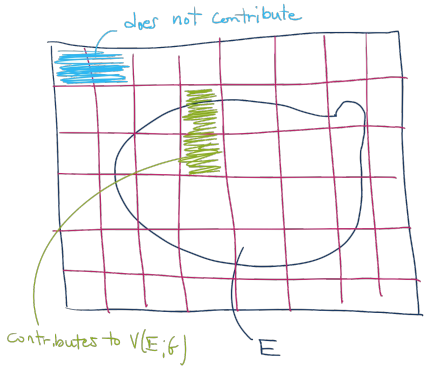
$$|R| = (b_1 - a_1) \cdots (b_n - a_n).$$

(Once we give a precise notion of "volume" we will show that this is indeed the correct volume of a rectangular box; for now, we think of $|R|$ as simply being some quantity which we expect to be "volume" based on geometric intuition.) The idea now is that, if $E \subseteq R$ is a region which we want to define the volume of, we can use rectangular boxes which get smaller and smaller to better and better approximate the "volume" of S; if we get a well-defined number by doing so, then that number should be defined to be the volume of S. Here are the first definitions.

A *grid* G on R is a partition of R obtained by partitioning (in the sense of how "partition" was used in the first quarter) each of the edges $[a_i, b_i]$ making up R. Each grid cuts R up into smaller rectangular boxes:



Clearly, adding together the volumes $|R_i|$ of all the R_i will just give the volume $|R|$ of R, so to bring the region E into play we should only take those smaller rectangular boxes which intersect $\overline{E}$:

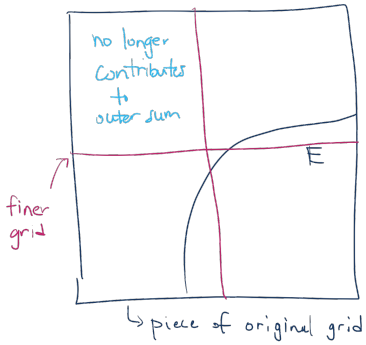


We define the *outer sum* $V(E;G)$ to be the sum of the volumes of all such rectangular boxes:

$$V(E;G) = \sum_{R_i \cap \overline{E} \neq \emptyset} |R_i|.$$

The idea is that such an outer sum *overestimates* the volume of E.

Now, as the grid G gets finer (meaning that we split the small rectangular boxes making up G into even smaller pieces), the outer sum also gets smaller since when splitting up some R_i we might be left with pieces which no longer intersect $\overline{E}$, meaning that they will no longer contribute to the new outer sum:



We expect that as the grid gets finer and finer, the outer sums provide a better and better approximation to the volume E, and that "in the limit" they should approach said volume. Thus we guess that we should define the volume of E as the infimum of all such outer sums:

$$\text{Vol}(E) := \inf\{V(E;G) \mid G \text{ is a grid on } R\},$$

where R is a rectangular box containing E.

Clarifying example. The above procedure *almost* works, only that there are sets E for which the resulting "volume" does not behave how we expect volume to behave. In particular, one key property we would expect is that if A and B are disjoint, then the volume of $A \cup B$ should be obtained by adding together the volumes of A and B separately:

$$\text{Vol}(A \cup B) = \text{Vol}\, A + \text{Vol}\, B.$$

Indeed, it makes intuitive sense in $\mathbb{R}^3$ at least that if we take some solid with a well-defined volume and break it up into two pieces, the sum of the volumes of those pieces should give the volume of the original solid.

However, here is an examples where this equality cannot hold. Take $E \subseteq [0,1] \times [0,1]$ to be the subset of the unit square in $\mathbb{R}^2$ consisting of points which rational coordinates:

$$E = \{(x,y) \in [0,1] \times [0,1] \mid x, y \in \mathbb{Q}\}.$$

Given any grid G on $[0,1] \times [0,1]$, any smaller rectangle R_i contains a point of E since $\mathbb{Q}^2$ is dense in $\mathbb{R}^2$, so every smaller rectangle of the grid will contribute to $V(E;G)$, giving:

$$V(E;G) = \sum_{R_i \cap \overline{E} \neq \emptyset} |R_i| = \sum_{R_i} |R_i| = \text{Vol}([0,1] \times [0,1]) = 1.$$

(Volume in this case just means area.) But note that we get the same result using the complement E^c of E: since the irrationals are dense in $\mathbb{R}$, any R_i will also contain a point of E^c so every such rectangle contributes to the outer sum $V(E^c; G)$, giving:

$$V(E^c; G) = \sum_{R_i} |R_i| = 1$$

as well. Thus the infimum of such outer sums will be 1 for both E and E^c, so even though $[0,1] \times [0,1] = E \cup E^c$, we have:

$$\text{Vol}(E) + \text{Vol}(E^c) = 2 \neq 1 = \text{Vol}\, E \cup E^c$$

if $\text{Vol}(E)$ and $\text{Vol}(E^c)$ are defined as in our attempt above.

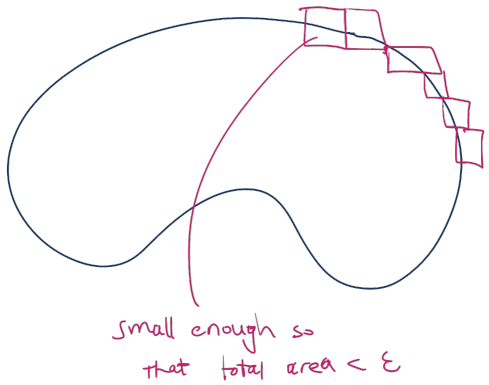
Regions of zero volume. The above example shows that there are regions for which our attempted definition of volume gives something which does not necessarily behave how we expect volumes to behave. The issue in that example is that both E and E^c are dense in $[0,1] \times [0,1]$, which is what ends up giving a value of 1 for all outer sums. Last quarter we saw that this property is equivalent to the claim that $\partial E = [0,1] \times [0,1]$, and so the point is that in some sense the boundary of E is "too large" in order to allow for E to have a well-defined volume.

The upshot is that volume should only be defined for regions which have as "small" a boundary as possible, which we will now interpret as meaning that the boundary should have volume zero. We can make this precise as follows:

> We say that a bounded set $S \subseteq \mathbb{R}^n$ contained in a rectangular box R has *Jordan measure zero* (or *zero volume*) if for any $\epsilon > 0$ there exists a grid G on R such that $V(S; G) < \epsilon$.

The intuition is that we can cover S by small enough rectangles which have total area less than any small quantity like—this would imply that if $\text{Vol}\, S$ were defined, it would have to satisfy $\text{Vol}\, S < \epsilon$ for all $\epsilon > 0$, which means that $\text{Vol}\, S$ would have to be zero.

For example, a continuous curve drawn in $\mathbb{R}^2$ will have Jordan measure zero, since intuitively we can cover it with finitely many small rectangles of arbitrarily small total area:



Note that we are not talking here about the area of the region *enclosed* by the curve, but rather about the area of the curve *itself*, which should be zero since a curve only has "length" but no "width". Similarly, a 2-dimensional surface in $\mathbb{R}^3$ should have Jordan measure zero in $\mathbb{R}^3$, where we use small rectangular boxes to cover the surface. Note that a square such as $[0,1] \times [0,1]$ is does

not have Jordan measure zero when viewed as a subset of $\mathbb{R}^2$ but that it *does* have Jordan measure zero when viewed as a subset of $\mathbb{R}^3$. The point is that this notion of "measure zero" is relative to the "ambient space" the region in question is sitting inside of.

Jordan measurability. Now armed with the notion of zero volume, we can characterize the types of regions which have well-defined volumes, with the intuition being that these are the regions whose boundaries are as small as possible:

> We say that a bounded set $E \subseteq \mathbb{R}^n$ is *Jordan measurable* (or is a *Jordan region*) if its boundary ∂E has Jordan measure zero. In this case, we define the *volume* (or *Jordan measure* of E to be the infimum of all outer sums obtained by varying through all possible grids covering a rectangular box R containing E:
>
> $$\text{Vol}(E) = \inf\{V(E;G) \mid G \text{ is a grid on } R\}.$$

These will be the regions over which the notion of *Riemann integration* will make sense in the multivariable setting. After having defined integration in this setting, we will see a better reason as to why Jordan measurable sets are indeed the right types of regions to consider.

Important. Volumes in $\mathbb{R}^n$ are computed using grid approximations, where we define

$$V(E;G) = \sum_{R_i \cap \overline{E} \neq \emptyset} |R_i| \quad \text{and} \quad \text{Vol}(E) = \inf\{V(E;G) \mid G \text{ is a grid on } R\}$$

for a region E contained in a rectangular box R. This only makes sense for regions whose boundary has volume (or measure) zero, meaning that $V(\partial E; G)$ can be made arbitrary small by choosing appropriate grids.

A comment on the term "measure". The book does not use the phrases *Jordan measure zero*, *Jordan measurable*, nor *Jordan measure*, instead using "zero volume", "Jordan region", and "volume" respectively. But, I think emphasizing the term "measure" is important since it emphasizes the connection between what we're doing and the more general subject of *measure theory*, which provides the most general theory of integration available. The word "measure" in this sense is simply meant to be a general notion of "volume".

To say a bit more, note that in our approach we started with outer sums which overestimate the volume we're interested in. Similarly, we can define "inner sums" by underestimating this volume, using rectangles which lie fully inside E as opposed to ones which can extend beyond. In this case, as grids get finer and finer, inner sums get larger and larger, so it is the supremum of the inner sums which should approach the volume we want. For a region to have a well-defined volume we should expect that the outer sum and inner sum approach give the same value, meaning that

$$\inf\{\text{outer sums}\} = \sup\{\text{inner sums}\}.$$

The infimum on the left is more generally called the *outer Jordan measure* of E and the supremum on the right is the *inner Jordan measure*, and a set should be Jordan measurable precisely when these two numbers are the same. (For the clarifying example above, the inner Jordan measure turns out to be zero since the interior of the set in question is empty.) It turns out that the outer and inner Jordan measure agree if and only if ∂E has Jordan measure zero, which thus reproduces the definition we gave of Jordan measurability. The idea is that that volume of ∂E should be sandwiched between the inner and outer Jordan measures since when subtracting inner sums from

outer sums you're left only with rectangular boxes which cover ∂E. Thus the inner and outer Jordan measures agree if and only if the difference between the inner and outer sums can be made arbitrarily small, which says that the volume of ∂E can be made arbitrarily small.

We won't look at these notions of inner sums and measures in class, but the book has some optional material you can look at if interested. In all of our definitions we only consider grids which consist of finitely many rectangular boxes, but in more generality we can consider collections of countable infinitely many rectangular boxes (at least those for which the infinite sum $\sum_{R_i} |R_i|$ converges) covering a given region—the resulting "measure" obtained is what's called the *Lebesgue measure* of a region in $\mathbb{R}^n$, so what we're doing is essentially a simplified version of Lebesgue measure theory, namely the simplification where we only allow finitely many rectangular boxes. Again, the book has some optional sections which elaborate on this if interested.

Lecture 14: Riemann Integrability

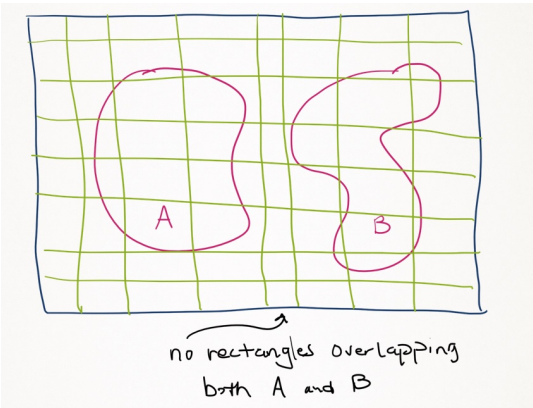
Today we started talking about integrability of multivariable functions, using an approach analogous to what we saw for single-variable functions back in the first quarter. For the most part we get the same properties and facts we had in the single-variable case, although there are some new things which happen in the multivariable setting which we'll elaborate on next time.

Warm-Up 1. One of the main reasons we restrict the types of regions we want to consider to those whose boundaries have volume zero is the desire for

$$\operatorname{Vol}(A \cup B) = \operatorname{Vol} A + \operatorname{Vol} B$$

to be true when A and B are disjoint. We show that this equality holds in the simpler case where A and B are *closed* Jordan measurable subsets of $\mathbb{R}^2$, although it is in fact true for any disjoint Jordan measurable sets. (In fact, it's true as long as $A \cap B$ has Jordan measure zero!)

We use the following picture as a guide:



(The point of requiring that A and B are closed—and hence compact—and disjoint is to guarantee that there is some positive distance between them.) The idea is that for fine enough grids, the small rectangles making up the grid can be separated into only those which cover A and only those which cover B, as in the picture above. First, we need to know that if A and B are Jordan measurable, $A \cup B$ is as well. This is done in the book, but for completeness here is the argument. Since A and B are Jordan measurable, ∂A and ∂B have Jordan measure zero, so for a fixed $\epsilon > 0$ there exist grids G_1 and G_2 such that

$$V(\partial A; G_1) < \frac{\epsilon}{2} \quad \text{and} \quad V(\partial B; G_2) < \frac{\epsilon}{2}.$$

Since $\partial(A \cup B) \subseteq \partial A \cup \partial B$, for a grid G finer than both G_1 and G_2 we have:

$$V(\partial(A \cup B); G) \leq V(\partial A; G) + V(\partial B; G) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

where the first inequality comes from separating out the rectangles which cover $\partial(A \cup B)$ into the union of those which cover ∂A with those which cover ∂B. (If G_1 and G_2 are not fine enough, some rectangles might cover both ∂A and ∂B, which is why we can only guarantee a non-strict inequality in this step.) Thus $\partial(A \cup B)$ has Jordan measure zero, so $A \cup B$ is Jordan measurable as claimed.

Now, take a fine enough grid G where all small rectangles have diagonal length less than the distance between A and B, which guarantees that any small rectangle can only intersect A or B but not both simultaneously. Thus we can separate the rectangles R_i in this grid which intersect $A \cup B$ into those which intersect A and those which intersect B with no overlap between these, so:

$$V(A \cup B; G) = \sum_{R_i \cap (A \cup B) \neq \emptyset} |R_i| = \sum_{R_i \cap A \neq \emptyset} |R_i| + \sum_{R_i \cap B \neq \emptyset} |R_i| = V(A; G) + V(B; G).$$

Hence taking infimums of both sides of this equality once G is fine enough gives

$$\inf\{V(A \cup B; G)\} = \inf\{V(A; G)\} + \inf\{V(B; G)\},$$

so $\operatorname{Vol}(A \cup B) = \operatorname{Vol} A + \operatorname{Vol} B$ as claimed.

(In the more general setting where A and B are not necessarily closed nor disjoint but $A \cap B$ still has Jordan measure zero, we would have to separate the rectangles into three categories: those which intersect $\overline{A}$ but not $\overline{B}$, those which intersect $\overline{B}$ but not $\overline{A}$, and those which intersect $\overline{A \cap B}$. In this case, the contribution to $V(A \cup B; G)$ coming from the third type of rectangle can be made arbitrarily small using the fact that $A \cap B$ has Jordan measure zero, so that when taking infimums this contribution can be ignored. There are details to fill in, but that's the basic idea for this general case.)

Important. If $A, B \subseteq \mathbb{R}^n$ are Jordan measurable, then $A \cup B$ is Jordan measurable. If in addition $A \cap B$ has Jordan measure zero, then $\operatorname{Vol}(A \cup B) = \operatorname{Vol} A + \operatorname{Vol} B$.

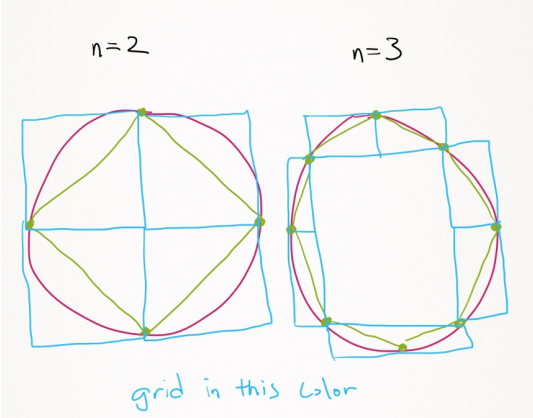
Banach-Tarski Paradox. As an aside, we give an example which truly illustrates the importance of the of the property given in the first Warm-Up, which seems like it should be an "obvious" fact. Here is the result, which is known as the *Banach-Tarski Paradox*: it is possible to take a solid unit ball, break it up into a finite number of pieces, and rearrange those pieces to end up with *two* solid unit balls! To be clear, this is a true statement and the "paradox" in the name only refers to the paradoxical observation that it seems as if we've doubled the total volume we started out with through this process. Indeed, if this could be done in the "real world" this would say that you could turn a solid bar of gold into two solid bars of gold, each of the same size as the original one.

The point is that the "pieces" which we break the sphere up into originally will not be Jordan measurable and so do not have a well-defined volume, thus since at some point in the process you work with sets for which "volume" does not make sense there is no reason to expect that the total volume you end up with at the end should be the same as the one you started with. This results shows why care has to be taken when making claims such as the one in the Warm-Up. (Here's a joke for you all: What's the anagram of Banach-Tarski? Banach-Tarski Banach-Tarsiki.)

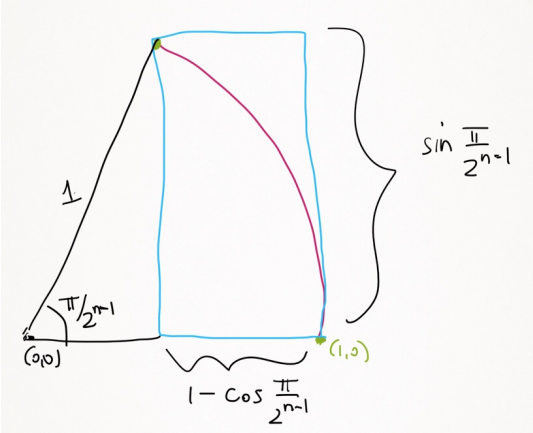
Warm-Up 2. We show that the unit disk $\overline{B_1(\mathbf{0})}$ in $\mathbb{R}^2$ is Jordan measurable by showing that the unit circle, which is the boundary of $\overline{B_1(\mathbf{0})}$, has Jordan measure zero. Here we do this by finding

explicit grids which do the job, but later we will see a quicker approach, which generalizes to show that any continuous curve has Jordan measure zero.

Given some $\epsilon > 0$ we want to cover the unit circle by a finite number of small rectangles whose total summed area is smaller than ϵ. For each $n \geq 2$, take the point $(1, 0)$ together with the other $2^n - 1$ points which all together give the vertices of a convex regular 2^n-gon inscribed in the circle:



Let G_n be the grid on $[0, 1] \times [0, 1]$ determined by all these points, meaning start with the small rectangles which have "adjacent" vertices as corners and then fill in the rest of the grid by translating these. Focus on the small rectangle which has its lower right corner at $(1, 0)$:



The central angle of the 2^n-gon is $\frac{2\pi}{2^n} = \frac{\pi}{2^{n-1}}$, so this rectangle has height $\sin(\pi/2^{n-1})$ and base length $1 - \cos(\pi/2^{n-1})$, and hence has area $\sin(\pi/2^{n-1})[1 - \cos(\pi/2^{n-1}]$. All together there are 2^n rectangles in our grid which touch the unit circle (namely those with a corner at one of the vertices of the 2^n-gon we used), and by symmetry all have the same area. Thus for this grid we get:

$$V(\partial B_1(\mathbf{0}); G_n) = \sum_{R_i \cap \partial B_1(\mathbf{0}) \neq \emptyset} |R_i| = 2^n \sin \frac{\pi}{2^{n-1}} \left(1 - \cos \frac{\pi}{2^{n-1}}\right).$$

It can be shown, say using L'Hopital's rule, that this converges to 0 as $n \to \infty$, implying that for any $\epsilon > 0$ there exists n such that $V(\partial B_1(\mathbf{0}); G_n) < \epsilon$, which shows that $\partial B_1(\mathbf{0})$ has Jordan measure zero as desired.

Again, after we've spoken about various ways of interpreting integrability, we will come back and show that the unit circle has volume zero without having to explicitly construct any grids.

Upper and lower sums. Given a Jordan region $E \subseteq \mathbb{R}^n$ and a bounded function $f : E \to \mathbb{R}^n$, we construct the *Riemann integral* of f over E using upper and lower sums in similar manner as we did for single-variable functions. The only difference is that while in the single-variable case we only integrated over intervals, now we are integrating over a Jordan region which may not be a rectangular box even though the upper and lower sums we want should depend on grids defined over rectangular boxes. Thus, we need some way of extending f, which is only a priori defined over E, to a function defined on an entire rectangular box containing E.

We proceed as follows. Pick a rectangular box R containing E and extend f to a function on all of R by defining it to be zero outside of E. The point of doing so is to ensure that the integral we end up with will only depend on how f behaves over E, since outside of E it is zero and zero functions contribute nothing to integrals. With this extended function in mind, given a grid G on R we define *upper* and *lower sums* just as we did in the single-variable case, only now summing over the small rectangular boxes arising from G which intersect E:

$$U(f; G) = \sum_{R_i \cap E \neq \emptyset} \left(\sup_{\mathbf{x} \in R_i} f(\mathbf{x}) \right) |R_i| \quad \text{and} \quad L(f; G) = \sum_{R_i \cap E \neq \emptyset} \left(\inf_{\mathbf{x} \in R_i} f(\mathbf{x}) \right) |R_i|.$$

Again to be clear, even if the given $f : E \to \mathbb{R}$ can actually already be viewed as a function defined on a larger domain, for the purposes of these definitions we nonetheless define the extension of f outside of E to be zero; for instance, if f was the constant function 1, which is clearly defined on all of $\mathbb{R}^n$, when constructing the integral of f over E we "change" the values of f outside of E to be zero instead of 1 to ensure that the integral only depends on E.

Integrability. Analogous to the single-variable case, as grids get finer upper sums can only get smaller and lower sums can only get larger. Thus we define the *upper* and *lower integrals* of f over E as:

$$(U) \int_E f(\mathbf{x})\, d\mathbf{x} = \inf\{(U(f; G) \mid G \text{ is a grid on } R\} \quad \text{and}$$
$$(L) \int_E f(\mathbf{x})\, d\mathbf{x} = \sup\{L(f; G) \mid G \text{ is a grid on } R\}.$$

We say that f is *(Riemann) integrable* over E if the upper and lower integrals agree, and define this common value to be the *integral* of f over E, which we denote by either

$$\int_E f(\mathbf{x})\, d\mathbf{x} \quad \text{or} \quad \int_E f\, dV.$$

As expected from a multivariable calculus course, in the two-variable case this gives the signed volume of the region in $\mathbb{R}^3$ between the graph of f and the Jordan region E in the xy-plane: the upper sums overestimate this volume while the lower sums underestimate it.

Alternate characterization. As in the single-variable case, we have the following alternate characterization: a bounded function $f : E \to \mathbb{R}$ on a Jordan measurable region $E \subseteq \mathbb{R}^n$ inside of a rectangular box R is integrable if and only if for any $\epsilon > 0$ there exists a grid G on R such that

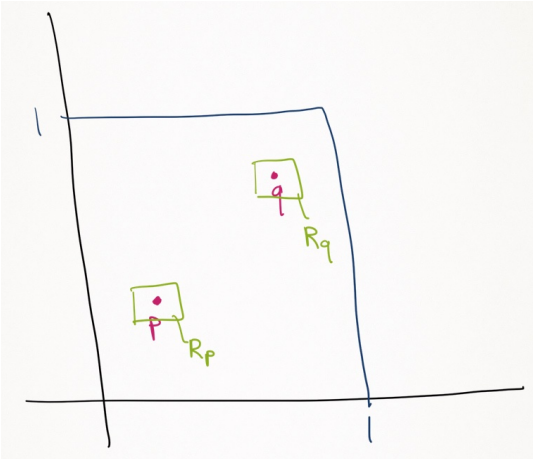
$$U(f; G) - L(f; G) < \epsilon.$$

The proof of this is exactly the same as in the single-variable case.

Important. The definition of the Riemann integral in the multivariable setting works in the same way as in the single-variable case, only that before constructing upper and lower sums we define

the function in question to be zero outside of the region of integration to ensure that this outside region does not contribute to the integral.

Example. Define $f : [0,1] \times [0,1] \to \mathbb{R}$ to be a function which has the value 1 at two points p and q of the unit square and the value 0 elsewhere:



We claim that this is integrable with integral zero. The value of zero for the integral comes from the fact that the lower sums corresponding to any grid is always 0 since the infimum of f over any small rectangle will always be zero. Thus (using dA instead of dV as is customary in the two-variable case):

$$(L)\int_{[0,1]\times[0,1]} f\,dA = \inf\{0\} = 0,$$

and so if f is integrable this will necessarily be the value of $\int_{[0,1]\times[0,1]} f\,dA$.

To show that f is integrable we must show that given any $\epsilon > 0$ we can find a grid G such that

$$U(f;G) - L(f;G) = U(f;G) < \epsilon.$$

But in any grid, the only small rectangles R_p and R_q which will contribute something nonzero to the upper sum are those which contain the points p and q at which f has the value 1 since f is zero everywhere else; hence the upper sum will look like:

$$U(f,G) = \left(\sup_{\mathbf{x}\in R_p} f(\mathbf{x})\right)|R_p| + \left(\sup_{\mathbf{x}\in R_q} f(\mathbf{x})\right)|R_q| = |R_p| + |R_q|.$$

Thus we can use a similar idea we used in the single-variable case: make R_p and R_q small enough to balance out the supremum value of 1 in order to make the entire sum smaller than ϵ.

Thus, for $\epsilon > 0$, pick a rectangle $R_p \subseteq [0,1] \times [0,1]$ around p whose area is less than $\frac{\epsilon}{2}$ and a rectangle $R_q \subseteq [0,1] \times [0,1]$ around q whose area is less than $\frac{\epsilon}{2}$. If necessary, make these rectangles smaller to ensure that they do not intersect each other. Letting G be a grid which contains these two as subrectangles, we get:

$$U(f;G) - L(f;G) = |R_p| + |R_q| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

showing that f is integrable on $[0,1] \times [0,1]$ as claimed.

Lecture 15: More on Integrability

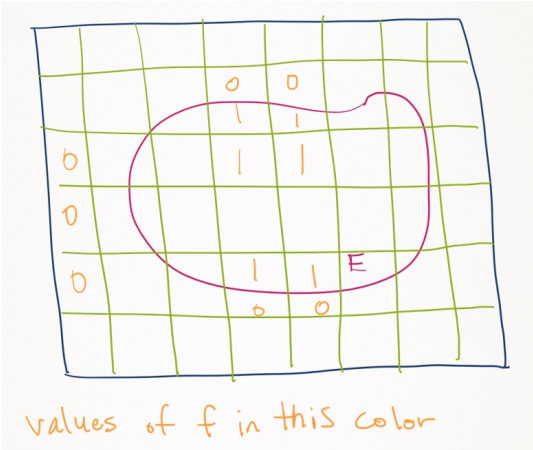
Today we continued talking about integrability in higher-dimensions, elaborating on a key new aspect which doesn't have a direct single-variable analog. We also gave yet another characterization of integrability which makes precise the geometric intuition that integrals should give volumes of regions between graphs and domains of integration; this also works in the single-variable case, we just didn't have enough machinery available in the first quarter to state the claim.

Warm-Up. Let $E \subseteq \mathbb{R}^n$ be a Jordan region. We show that the constant function $\mathbf{1}$ is integrable over E. On the one hand, this will be a consequence of the soon-to-be-mentioned fact that continuous functions on compact domains are always integrable, but we'll work it out here directly in order to demonstrated the subtleties involved. Based on what we know from a multivariable calculus course, we would expect that

$$\int_E \mathbf{1}\, dV = \operatorname{Vol} E,$$

which is indeed true; we won't prove this in detail here but it's worked out in the book.

We'll assume $n = 2$ for the sake of being able to draw nice pictures, but the general case is the same. Given a grid G on a rectangle R containing E, we want to work out what $U(f; G) - L(f; G)$ looks like. Recall that in order to define these upper and lower sums, we extend the given function to be zero outside of E; denote this extended function by f, so that $f = 1$ on E and $f = 0$ outside of E:



Note that we can separate the subrectangles R_i making up G into three categories: those which are fully contained in the interior of E, those which are fully contained in the interior of E^c, and those which intersect ∂E. Over the first type f always the value 1, so $\sup f - \inf f = 0$ in this case. Hence such rectangles contribute nothing to $U(f; G) - L(f; G)$. Over the second type, f always has the value 0 so again $\sup f - \inf f = 0$, meaning that these rectangles also do not contribute to upper sum minus lower sum.

Thus the only nonzero contributions to upper sum minus lower sum can come from the rectangles which intersect the boundary of E:

$$U(f; G) - L(f; G) = \sum_{R_i \cap \partial E \neq \emptyset} (\sup f - \inf f)|R_i|.$$

Now, any such R_i contains something in E, so $\sup f = 1$ on R_i, and something not in E, so $\inf f = 0$ on R_i. Thus we get

$$U(f;G) - L(f;G) = \sum_{R_i \cap \partial E \neq \emptyset} |R_i| = V(\partial E; G).$$

Since E is a Jordan region, ∂E has volume zero, so for any $\epsilon > 0$ there exists G such that this final expression is less than ϵ, and hence for this grid we have $U(f;G) - L(f;G) < \epsilon$, showing that $\mathbf{1}$ is integrable over E.

The fact that the value of the integral is actually $\operatorname{Vol} E$ can be derived from the fact that

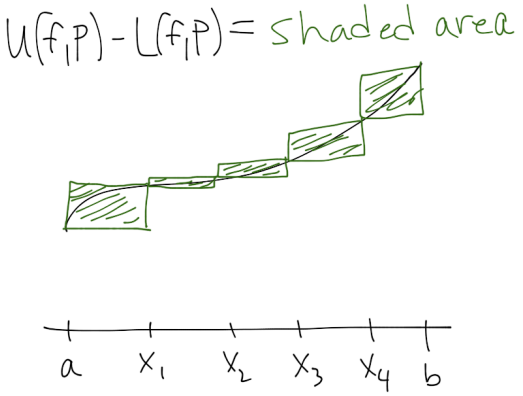
$$L(f;G) \leq V(E;G) \leq U(f;G)$$

for any grid G on R, so that the infimum of the terms in the middle (i.e. the volume of E) will be the common value of the supremum of the terms on the left and the infimum of the terms on the right. It would be good practice to understand why these final inequalities hold. In particular, the fact that $L(f;G) \leq V(E;G)$ depends on the fact that we extended the constant function $\mathbf{1}$ to be zero outside of E: if we had simply kept the value outside of E as 1 we would get that $L(f;G) = U(f;G)$ and so the lower sums would no longer *underestimate* the value of the integral as we expect them to do.

Integrability in terms of Jordan measurability. When f is integrable over E, given $\epsilon > 0$ we know that we can find a grid G such that

$$U(f;G) - L(f;G) < \epsilon.$$

Note what this difference looks like in the single-variable case, which we mentioned in passing back in the first quarter. For $f : [a,b] \to \mathbb{R}$ integrable (actually continuous in the picture) we get:



But, encasing the graph of f inside of a rectangle and viewing these small rectangles as the portions of a grid G which intersect the graph of f, we can now interpret this upper sum minus lower sum as an outer sum for the graph of f in $\mathbb{R}^2$:

$$U(f,P) - L(f,P) = V(\operatorname{graph} f; G).$$

Thus, the fact that we can make upper sum minus lower sum arbitrarily small says that we can find grids which make the outer sum for graph f arbitrarily small, which says that graph f has Jordan measure zero in $\mathbb{R}^2$! The same idea works for the graph of any integrable function over a Jordan region in $\mathbb{R}^n$, so we get that:

> If a bounded function $f : E \to \mathbb{R}$ is integrable over a Jordan region $E \subseteq \mathbb{R}^n$, then the graph of f has Jordan measure zero in $\mathbb{R}^{n+1}$.

This hints at a deep connection between integrability and Jordan measurability, and indeed we can now give the precise statement, which also applies in the single-variable case. For a nonnegative bounded function $f : E \to \mathbb{R}$ on a Jordan region $E \subseteq \mathbb{R}^n$, define the *undergraph* of f to be the region in $\mathbb{R}^{n+1}$ lying between the graph of f and the region E in the hyperplane $x_{n+1} = 0$ where $(x_1, \ldots, x_{n+1})$ are coordinates on $\mathbb{R}^{n+1}$. (So, when $n = 1$ or 2, the undergraph is literally the region under the graph of f and above E.) Then we have:

> f is integrable over E if and only if the undergraph of f is Jordan measurable, in which case the integral of f over E equals the volume of the undergraph.

Thus, we see that the basic intuition we've used all along (going back to the first quarter) when defining integration is the correct one: to say that $\int_E f\,dV$ exists should mean the region under the graph of f has a well-defined volume, and the value of the integral should be equal to this volume. Hence, Riemann integrability and Jordan measurability are indeed closely linked to one another.

The proof of this final equivalence is left as an optional problem on the homework, but the idea is as follows. The boundary of the undergraph should consist of three pieces: the graph of f on "top", the Jordan region E on the "bottom ", and the "sides" obtained by sliding the boundary of E "up" until you hit the graph of f. Each of these pieces should have Jordan measure zero in order for the undergraph to be Jordan measurable, and the fact that each piece does so relates to something showing up in the definition of integrability: that the "bottom" E has Jordan measure zero in $\mathbb{R}^{n+1}$ is related to the fact that E is Jordan measurable in $\mathbb{R}^n$, that the "sides" have Jordan measure zero is related to the fact that ∂E has Jordan measure zero in $\mathbb{R}^n$, and that the "top" (i.e. the graph of f) has Jordan measure zero relates to the fact that we can make upper sums minus lower sums arbitrarily small.

Important. The graph of an integrable function has Jordan measure zero, and integrability is equivalent to the undergraph being Jordan measurable: the Jordan measure of the undergraph is equal to the integral of f.

Properties analogous to single-variable case. Multivariable integrals have the same types of properties (with similar proofs) that single-variable integrals do. Here are a few key ones:

- continuous functions on compact domains are integrable (the proof, as in the single-variable case, uses the fact that continuous functions on compact domains are uniformly continuous),
- sums and scalar multiples of integrable functions are integrable, integrals of sums split up into sums of integrals, and constants can be pulled outside of integrals
- regions of integration can be split up when each piece is Jordan measurable, and the integrals split up accordingly
- integrals preserve inequalities: if $f \leq g$ and each is integrable, then $\int f \leq \int g$.

Check the book for full details, but again the proofs are very similar to the ones for the analogous single-variable claims.

Projectable regions. We didn't talk about projectable regions in class, but it's worth mentioning. You can check the book for more details. A bounded subset $E \subseteq \mathbb{R}^n$ is *projectable* essentially if its

boundary can be described using graphs of continuous functions, or in other words if we can view E as the region enclosed by different graphs of continuous functions. (Check the book for the precise definition.) For instance, the unit disk in $\mathbb{R}^2$ is projectable since the top portion of its boundary is the graph of $y = \sqrt{1-x^2}$ and the bottom portion of its boundary is the graph of $y = -\sqrt{1-x^2}$.

The basic fact is that projectable regions are always Jordan measurable. Indeed, since the boundary of such a region consists of graphs of continuous functions, it is enough to show that each such graph has Jordan measure zero. But continuous functions are always integrable, so the characterization of integrability in terms of Jordan measurability implies that such a graph indeed has Jordan measure zero, and we are done. This gives a much quicker proof of the fact that the unit circle has Jordan measure zero as opposed to the proof we gave last time in terms of explicit grids: we can find grids which make the outer sums for the top half and bottom half each arbitrarily small since these halves are each graphs of continuous functions, so putting these grids together gives grids which make the outer sum for the entire circle arbitrarily small.

More generally, all standard types of geometric objects you're used to seeing in $\mathbb{R}^2$ and $\mathbb{R}^3$ (spheres, cones, ellipses, paraboloids, hyperboloids, etc.) are projectable regions and hence are Jordan measurable. To argue along these lines, note that if we are given some nice enough continuous (or better yet C^1) curve in $\mathbb{R}^2$ or surface in $\mathbb{R}^3$, something like the Implicit Function Theorem (!!!) will imply that we can view this object locally as being made up of graphs of continuous functions, and so the same reasoning as above will imply that this is Jordan measurable.

Regions of volume zero don't matter. Although there are many similarities between multivariable and single-variable integrals, here is one which does not have a direct analog in the single-variable case, at least using the Riemann integral. (This DOES have a direct analog when using the single-variable *Lebesgue integral*, but this is not a topic we'll explore further.) The difference is that in the multivariable case we can integrate over more general types of regions than we can in the single-variable case.

Here is the claim: if $E \subseteq \mathbb{R}^n$ has Jordan measure zero, then any bounded function $f : E \to \mathbb{R}$ is integrable and $\int_E f\,dV = 0$. Thus, regions of volume zero can never contribute anything to values of integrals. The proof depends on Theorem 12.20 in the book, which essentially says that for any Jordan region E, upper and lower sums (and hence upper and lower integrals) can be approximated to whatever degree of accuracy we want using only subrectangles in a grid which are contained fully in the *interior* of E. The idea is that in any grid we can separate the rectangles intersecting E into those contained in the interior and those which intersect the boundary, and the fact that boundary has volume zero can be used to make these contributions negligible, so that only the contributions from those rectangles in the interior actually matter. If E has empty interior, then this implies that the upper and lower sums themselves can be made arbitrarily small (since they can be approximated by a bunch of zeroes), so that $\int_E f\,dV = 0$. In particular, it can be shown that if E has volume zero, then it must have empty interior and we have our result.

Since regions of volume zero never contribute to integrals, it follows that if we take an integrable function and change its values only over a set of volume zero, the resulting function is still integrable and has the same integral as the original one. The idea is as follows. If $f : E \to \mathbb{R}$ is integrable and $S \subseteq E$ has volume zero, then we can split up the integral of f over E as:

$$\int_E f\,dV = \int_S f\,dV + \int_{E \setminus S} f\,dV,$$

assuming that $E \setminus S$ is also Jordan measurable, which it will be. The first integral on the right is zero, and so

$$\int_E f\,dV = \int_{E \setminus S} f\,dV,$$

meaning that the entire value of $\int_E f\,dV$ comes solely from the region outside S. If $g : E \to \mathbb{R}$ has the same values on $E \setminus S$ as does f, then $\int_{E\setminus S} f\,dV = \int_{E\setminus S} g\,dV$ and the claim follows.

This is all clarified in detail in the book, and is an important fact to remember. We saw something analogous in the first quarter, where a homework problem showed that changing the value of an integrable function at a *finite* number of points does not affect the value of the integral, but now we have the most general statement available.

Important. Integrating over a region of volume zero (or with empty interior) always gives the value zero, so that regions of volume zero never actually contribute to integrals. Thus, altering an integrable function on a set of volume zero does not affect the existence of nor value of the integral.

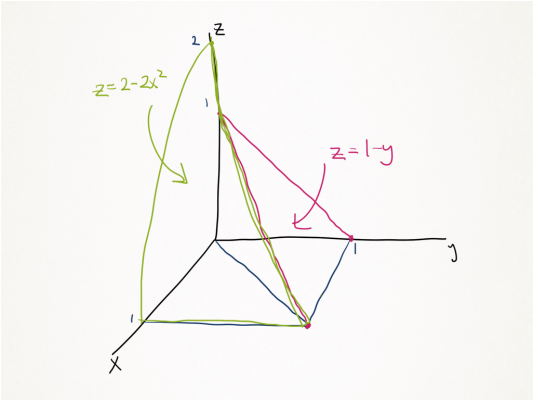
Lecture 16: Fubini's Theorem

Today we spoke about Fubini's Theorem, which gives a way to compute multivariable integrals using iterated integrals. No doubt you did this plenty of times in a previous multivariable calculus courses, but here we delve into the reasons as to why this actually works since, at first glance, the definitions of multivariable integrals and iterated integrals are quite different.

Warm-Up. Define $f : [0,1] \times [0,1] \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} 1-y & y \geq x \\ 2-2x^2 & y < x. \end{cases}$$

We show that f is integrable. To be clear, the graph of f is a plane over the upper-left half of the square and looks like a downward curved parabolic surface over the bottom-right half:



Over the upper-left half of the square, $f(x,y) = 1-y$ is continuous and so is integrable. Over the bottom-right half, $f(x,y)$ equals the continuous function $g(x,y) = 2-2x^2$ expect along the diagonal. Since the diagonal has Jordan measure zero, what happens along the diagonal does no affect integrability, so since f equals an integrable function expect on a set of volume zero, it too is integrable. Thus f is integrable over the upper-left half of the square and over the bottom-right half, so it is integrable over the entire square. To be clear, this is a property we mentioned last time and which is proved in the book: if f is integrable over a Jordan region A and also over a Jordan region B such that $A \cap B$ has Jordan measure zero, then f is integrable over $A \cup B$.

Iterated integrals. For $f : [a, b] \times [c, d] \to \mathbb{R}$, the *iterated integrals* of f are the expressions:

$$\int_c^d \left(\int_a^b f(x, y)\, dx \right) dy \quad \text{and} \quad \int_a^b \left(\int_c^d f(x, y)\, dy \right) dx.$$

(For simplicity, we will only give the definitions and results we'll look at in the two-variable case, even though the same applies to functions of more variables.) Computing these is what you no doubt spent plenty of time doing in a multivariable calculus course, where you first compute the "inner" integral and then the "outer" integral. We are interesting in knowing when these computations give the value of the multivariable integral

$$\iint_{[a,b]\times[c,d]} f(x, y)\, d(x, y)$$

we've defined in terms of two-dimensional upper and lower sums. (We use two integrals symbols in the notation here simply to emphasize that we are integrating over a 2-dimensional region.)

Here are some basic observations. First, in order for the iterated integrals to exist we need to know that that inner integrals exist. In the case of:

$$\int_a^b f(x, y)\, dx,$$

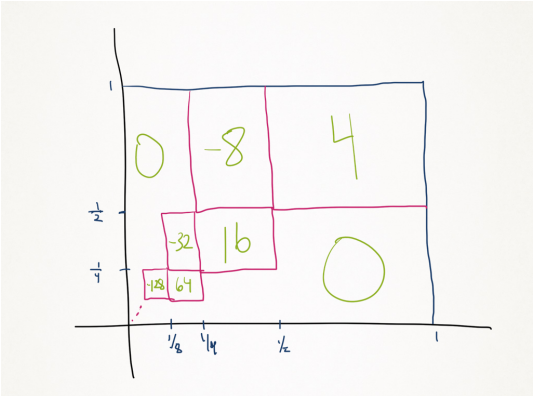
this requires knowing that for any fixed $y \in [c, d]$, the function $x \mapsto f(x, y)$ is integrable over the interval $[a, b]$; it is common to denote this function by $f(\cdot, y)$, where y is fixed and the $\cdot$ indicates the variable we are allowed to vary. Similarly, in order for the inner integral

$$\int_c^d f(x, y)\, dy$$

in the second iterated integral to exist we need to know that for any $x \in [a, b]$, the single-variable function $f(x, \cdot)$ is integrable over $[c, d]$. In addition, in order for the double integral $\iint_{[a,b]\times[c,d]} f(x, y)\, d(x, y)$ to exist we need to know that f is integrable (in the two-variable sense) over $[a, b] \times [c, d]$.

It turns out that when all these hypotheses are met all integrals in question are indeed equal, which is the content of *Fubini's Theorem*. Before giving a proof, we look at some examples which illustrate what can happen when some of the requirements are not met.

Example: iterated integrals need not be equal. (This example is in the book; here we're just making the idea a little more explicit.) Define the function $f : [0, 1] \times [0, 1] \to \mathbb{R}$ according to the following picture:

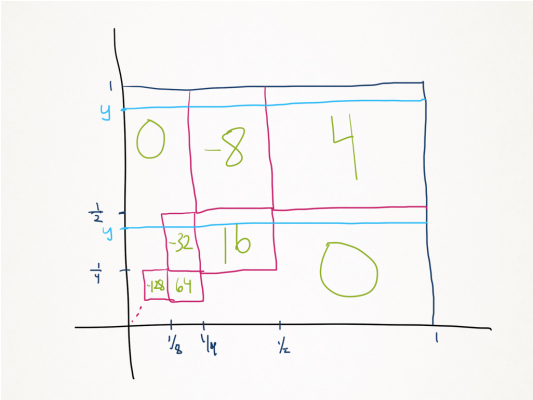


So, all nonzero values of f are positive or negative powers of 2: start with 2^2 in the "upper-right" quarter of the unit square, then -2^3 for the left-adjacent "eighth"; then move down to the next smaller square "along the diagonal" and give f the value 2^4, then -2^5 for the next left-adjacent rectangle of half the size; then take the vale 2^6 for the next diagonal square, and -2^7 for the left-adjacent half, and so on, and everywhere else f has the value zero. We claim both iterated integrals of this function exist but are not equal.

Consider

$$\int_0^1 f(x,y)\,dx$$

for a fixed $y \in [0,1]$. Along the horizontal line at this fixed y the function f has only two possible nonzero values: either a positive power of 2 or the negative next larger power of 2. (So in particular $f(\cdot, y)$ is integrable for any $y \in [0,1]$.) For instance, in the picture:



at the first fixed y value f only has the value 4 or -8. But the interval along which it has the value -8 has length $\frac{1}{4}$ and the interval along which it has the value 4 has length $\frac{1}{2}$, so at this fixed y we get:

$$\int_0^1 f(x,y)\,dx = -8\left(\frac{1}{4}\right) + 4\left(\frac{1}{2}\right) = 0.$$

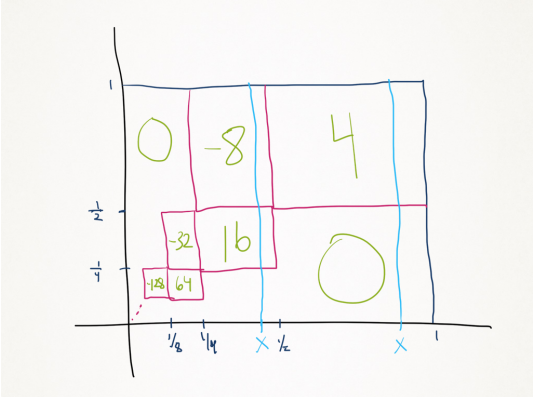
At the next fixed y in the picture, f only has the values -32 and 16, but the value -32 occurs along an interval of length $\frac{1}{8}$ and the value 16 along an interval of length $\frac{1}{4}$, so at this y we get

$$\int_0^1 f(x,y)\,dx = -32\left(\frac{1}{8}\right) + 16\left(\frac{1}{4}\right) = 0$$

as well. The same thing happens at any fixed y, so in the end $\int_0^1 f(x,y)\,dx = 0$ for any $y \in [0,1]$. Thus:

$$\int_0^1 \int_0^1 f(x,y)\,dx\,dy = \int_0^1 0\,dy = 0.$$

Now we consider the other iterated integral. For the inner integral we now fix $x \in [0,1]$ and look at what happens to f along the vertical line at this fixed x:



For $0 \le x \le \frac{1}{2}$ we get the same behavior as before: f has two possible values—a positive power of 2 and a negative power—along intervals whose lengths make the total integral equal 0. For instance, at the first x in the picture, f has the value 16 along a vertical interval of length $\frac{1}{4}$ and the value -8 along a vertical interval of length $\frac{1}{2}$, so at this x we have:

$$\int_0^1 f(x,y)\,dy = 16\left(\frac{1}{4}\right) - 8\left(\frac{1}{2}\right) = 0.$$

Thus $\int_0^1 f(x,y)\,dy = 0$ for $0 \le x \le \frac{1}{2}$, so:

$$\int_0^1 \int_0^1 f(x,y)\,dy\,dx = \int_0^{1/2} \underbrace{\int_0^1 f(x,y)\,dy}_{0}\,dx + \int_{1/2}^1 \int_0^1 f(x,y)\,dy\,dx = \int_{1/2}^1 \int_0^1 f(x,y)\,dy\,dx.$$

For $\frac{1}{2} \le x \le 1$, f only has one nonzero value: 4 along a vertical interval of length $\frac{1}{2}$. Thus

$$\int_0^1 f(x,y)\,dy = 4\left(\frac{1}{2}\right) = 2 \text{ for } \frac{1}{2} \le x \le 1,$$

and hence

$$\int_{1/2}^1 \int_0^1 f(x,y)\,dy\,dx = \int_{1/2}^1 2\,dx = 1,$$

so

$$\int_0^1 \int_0^1 f(x,y)\,dy\,dx = 1 \neq \int_0^1 \int_0^1 f(x,y)\,dx\,dy = 0.$$

Therefore the iterated integrals of this function both exist but are not equal. (Fubini's Theorem will not apply here because f is not integrable over the unit square in the two-variable sense.)

Example: equality of iterated integrals does not imply integrability. (This example is also in the book.) Define $f : [0,1] \times [0,1] \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} 1 & (x,y) = \left(\frac{p}{2^n}, \frac{q}{2^n}\right) \in \mathbb{Q}^2 \\ 0 & \text{otherwise.} \end{cases}$$

So, f gives the value 1 on points whose coordinates are both rational with denominators equal to the same power of 2 and f gives the value 0 otherwise. For a fixed $y \in [0,1]$, there are only finitely

many x such that $f(x,y)=1$: indeed, if y is not of the form $\frac{q}{2^n}$ there are no such x since $f(x,y)=0$ for all x in this case, whereas if $y=\frac{q}{2^n}$ only

$$x = \frac{1}{2^n}, \frac{2}{2^n}, \dots, \frac{2^n}{2^n}$$

will satisfy $f(x,y)=1$. Thus for any $y \in [0,1]$, $f(\cdot, y)$ differs from the constant zero function only on a finite set of volume zero, so $f(\cdot, y)$ is integrable on $[0,1]$ with integral 0. The same reasoning applies to show that for any $x \in [0,1]$, $f(x,\cdot)$ is integrable on $[0,1]$ with integral zero. Hence:

$$\int_0^1 \underbrace{\int_0^1 f(x,y)\,dx}_{0}\,dy = 0 \quad \text{and} \quad \int_0^1 \underbrace{\int_0^1 f(x,y)\,dy}_{0}\,dx = 0,$$

so the iterated integrals of f exist and are equal.

However, we claim that f is not integrable over $[0,1]\times[0,1]$. Indeed, the set of points

$$E = \left\{ \left(\frac{p}{2^n}, \frac{q}{2^n} \right) \;\middle|\; p, q \in \mathbb{Z} \text{ and } n \geq 0 \right\}$$

is dense in $\mathbb{R}^2$ as a consequence of the fact that the set of rationals with denominator a power of 2 is dense in $\mathbb{R}$, and the complement of E is also dense in $\mathbb{R}^2$ since it contains points with irrational coordinates. Thus given any grid R on $[0,1]\times[0,1]$, any small rectangle will always contain a point of E and a point of E^c, so $\sup f = 1$ and $\inf f = 0$ on any small rectangle. Thus

$$U(f,G) = 1 \text{ and } L(f,G) = 0$$

for any grid, so the upper integral of f is 1 and the lower integral is 0, showing that f is not integrable over the unit square. Thus, existence and equality of iterated integrals does not imply integrability of the function in question.

Fubini's Theorem. The above examples show that iterated integrals do not always behave in expected ways, but if all integrals in question—the double integral and both iterated integrals—do exist, then they will all have the same value, which is the statement of Fubini's Theorem:

> Suppose that $f : [a,b] \to [c,d] \to \mathbb{R}$ is integrable, that for each $y \in [c,d]$ the single-variable function $f(\cdot, y)$ is integrable on $[a,b]$, and that for each $x \in [a,b]$ the single-variable function $f(x,\cdot)$ is integrable on $[c,d]$. Then
>
> $$\iint_{[a,b]\times[c,d]} f(x,y)\,d(x,y) = \int_c^d \int_a^b f(x,y)\,dx\,dy = \int_a^b \int_c^d f(x,y)\,dy\,dx.$$

More generally, even if one of the iterated integrals does not exist, we will still have equality among the remaining iterated integral and the double integral. A similar statement holds in higher dimensions, and for other Jordan regions apart from rectangles as well.

We give a different and easier to follow proof than the book's. The book's proof depends on Lemma 12.30, which is a nice but somewhat complicated result on its own; however, if all we are interested in is a proof of Fubini's Theorem, avoiding the use of this lemma gives a simper argument. The key point is that the double integral $\iint_{[a,b]\times[c,d]} f(x,y)\,d(x,y)$ is defined in terms of two-dimensional upper and lower sums, whereas the iterated integrals in question do not explicitly involve these two-dimensional sums; yet nonetheless, we can come up with inequalities which relate the required upper and lower sums to the iterated integrals we want.

Proof. The partitions of $[a, b]$ and $[c, d]$ respectively:

$$a = x_0 < x_1 < \cdots < x_n = b \quad \text{and} \quad c = y_0 < y_1 < \cdots < y_n = b$$

and let G be the corresponding grid. Let m_{ij} and M_{ij} respectively denote the infimum and supremum of f over the small rectangle $R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$, and set $\Delta x_i = x_i - x_{i-1}$ and $\Delta y_j = y_j - y_{j-1}$. For a fixed $y \in [y_{j-1}, y_j]$, we have:

$$m_{ij} \leq f(x, y) \leq M_{ij} \text{ for } x \in [x_{i-1}, x_i],$$

so

$$m_{ij}\Delta x_i = \int_{x_{i-1}}^{x_i} m_{ij}\, dx \leq \int_{x_{i-1}}^{x_i} f(x, y)\, dx \leq \int_{x_{i-1}}^{x_i} M_{ij}\, dx = M_{ij}\Delta x_i.$$

Note that the middle integral exists since we are assuming that the single-variable function $f(\cdot, y)$ is integrable for any y. Summing these terms up over all subintervals $[x_{i-1}, x_i]$ gives:

$$\sum_i m_{ij}\Delta x_i \leq \int_a^b f(x, y)\, dx \leq \sum_i M_{ij}\Delta x_i$$

where for the middle term we use the fact that the subintervals in question cover $[a, b]$ so that:

$$\int_a^b f(x, y)\, dx = \int_{x_0}^{x_1} f(x, y)\, dx + \int_{x_1}^{x_2} f(x, y)\, dx + \cdots + \int_{x_{n-1}}^{x_n} f(x, y)\, dx.$$

Taking integrals throughout with respect to y over the interval $[y_{j-1}, y_j]$ preserves the inequalities, giving:

$$\int_{y_{j-1}}^{y_j} \left(\sum_i m_{ij}\Delta x_i \right) dy \leq \int_{y_{j-1}}^{y_j} \int_a^b f(x, y)\, dx\, dy \leq \int_{y_{j-1}}^{y_j} \left(\sum_i M_{ij}\Delta x_i \right) dy,$$

which is the same as

$$\sum_i m_{ij}\Delta x_i \Delta y_j \leq \int_{y_{j-1}}^{y_j} \int_a^b f(x, y)\, dx\, dy \leq \sum_i M_{ij}\Delta x_i \Delta y_j$$

since $\Delta y_j = \int_{y_{j-1}}^{y_j} dy$. Summing up over all the subintervals $[y_{j-1}, y_j]$ gives:

$$\sum_{i,j} m_{ij}\Delta x_i \Delta y_j \leq \int_c^d \int_a^b f(x, y)\, dx\, dy \leq \sum_{i,j} M_{ij}\Delta x_i \Delta y_j$$

where we use

$$\int_c^d \text{blah} = \int_{y_0}^{y_1} \text{blah} + \int_{y_1}^{y_2} \text{blah} + \cdots + \int_{y_{n-1}}^{y_n} \text{blah}.$$

The left hand side in the resulting inequalities is precisely the lower sum $L(f, G)$ and the right hand side is the upper sum $U(f, G)$, so we get that

$$L(f, G) \leq \int_c^d \int_a^b f(x, y)\, dx\, dy \leq U(f, G)$$

for any grid G, saying that the value of the iterated integral is sandwiched between all lower and upper sums. Taking the supremums of the lower sums and infimums of the upper sums thus gives:

$$\sup\{L(f,G)\} \leq \int_c^d \int_a^b f(x,y)\,dx\,dy \leq \inf\{U(f,G)\}.$$

Since f is integrable over $[a,b]\times[c,d]$, this supremum and infimum are equal, and thus must equal the iterated integral in the middle so

$$\iint_{[a,b]\times[c,d]} f(x,y)\,d(x,y) = \int_c^d \int_a^b f(x,y)\,dx\,dy$$

as claimed.

A similar argument switching the roles of x and y shows that

$$\iint_{[a,b]\times[c,d]} f(x,y)\,d(x,y) = \int_a^b \int_c^d f(x,y)\,dy\,dx$$

as the rest of Fubini's Theorem claims. To be clear, we start with fixing $x \in [x_{i-1}, x_i]$ and use

$$m_{ij} \leq f(x,y) \leq M_{ij} \text{ for } y \in [y_{j-1}, y_j],$$

to get

$$m_{ij}\Delta y_j = \int_{y_{j-1}}^{y_j} m_{ij}\,dy \leq \int_{y_{j-1}}^{y_j} f(x,y)\,dy \leq \int_{y_{j-1}}^{y_j} M_{ij}\,dy = M_{ij}\Delta y_j.$$

Then we take sums over all subintervals $[y_{j-1}, y_j]$, which gives:

$$\sum_j m_{ij}\Delta y_j \leq \int_c^d f(x,y)\,dy \leq \sum_j M_{ij}\Delta y_j,$$

and finally we take integrals with respect to x over $[x_{i-1}, x_i]$ and sum up over such intervals to get

$$L(f,G) \leq \int_a^b \int_c^d f(x,y)\,dy\,dx \leq U(f,G).$$

The same conclusions as before then hold. □

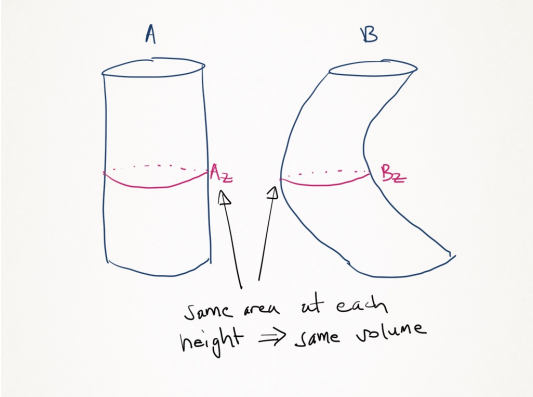
Remark. If the double integral and only one of the iterated integrals in question exists, the proof of Fubini's Theorem still gives equality between these two integrals. It is possible to find examples where $\iint_R f(x,y)\,dx\,dy$ and one of the iterated integrals exists, but the other iterated integral does not. The book has such an example based on what I declared to be my "favorite function" of all time in the first quarter, check there for details.

Important. If f is integrable over a Jordan region $D \subseteq \mathbb{R}^n$, then its integral can be computed using any iterated integrals which exist as well. However, it is possible that these iterated integrals exist (with the same or different values) even if f is not integrable over D.

Lecture 17: Change of Variables

Today we spoke about the change of variable formula for multivariable integrals, which will allow us to write integrals given in terms of one set of coordinates in terms of another set. Together with Fubini's Theorem, this gives the most general method for computing explicit integrals, and we'll see later on that it also justifies us the formulas we use for so-called line and surface integrals taken over curves and surfaces respectively.

Warm-Up. We justify *Cavieleri's Principle*: given two solids A and B of the same height, if for any z the two-dimensional regions A_z and B_z obtained by intersecting A and B respectively with the horizontal plane at height z have the same area, then A and B have the same volume. The picture to have in mind is the following, where on the left we have a cylinder and the right a cylinder with a "bulge":



The volumes in question are given by the integrals:

$$\operatorname{Vol} A = \iiint_A \mathbf{1}\, dV \quad \text{and} \quad \operatorname{Vol} B = \iiint_B \mathbf{1}\, dV.$$

Since the constant function $\mathbf{1}$ is continuous, Fubini's Theorem applies to say that each of these integrals can be computed using iterated integrals of the form:

$$\operatorname{Vol} A = \int_0^h \left(\iint_{A_z} \mathbf{1}\, dx\, dy \right) dz \quad \text{and} \quad \operatorname{Vol} B = \int_0^h \left(\iint_{B_z} \mathbf{1}\, dx\, dy \right) dz$$

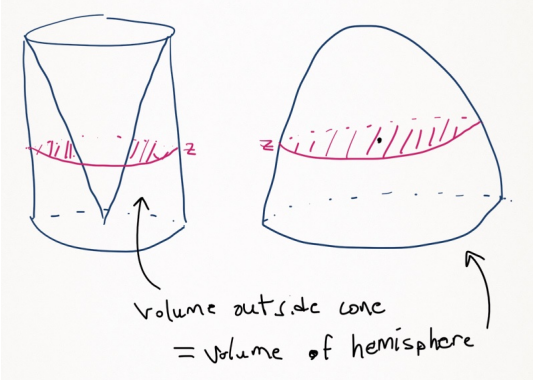
where h is the common height of the two solids. To be clear about the notation, at a fixed z the inner integrals in terms of x and y are taken over the piece of the solid which occurs at that fixed height, which are precisely what A_z and B_z denote. But integrating $\mathbf{1}$ over these two-dimensional regions gives their area, so we can write the integrals above as:

$$\operatorname{Vol} A = \int_0^h (\text{area of } A_z)\, dz \quad \text{and} \quad \operatorname{Vol} B = \int_0^h (\text{area of } B_z)\, dz.$$

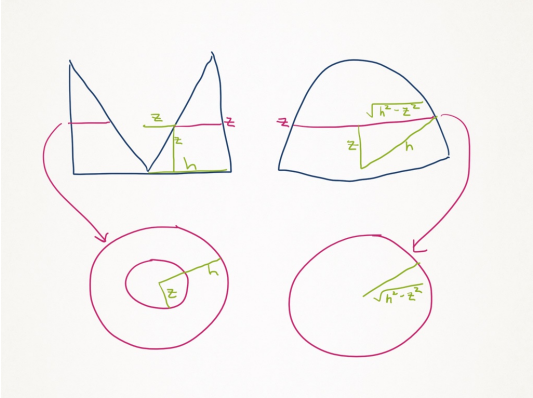
We are assuming that for any z, A_z and B_z have the same area, and thus the integrals above are the same, showing that $\operatorname{Vol} A = \operatorname{Vol} B$ as claimed.

Fun applications of Cavieleris' Principle. For this class, the important part of the above Warm-Up was the use of Fubini's Theorem. But just for fun, here are some well-known uses of Cavieleri's Principle.

First, take a cone of height and radius h sitting within a cylinder of height and radius h, and take the upper-half of a sphere of radius h:



The claim is that the volume of this half-sphere is equal to the volume of the region within the cylinder but outside the cone. Indeed, fix a height z and look at the intersections of these two solids with the horizontal plane at height z:



For the solid consisting of the region inside the cylinder but outside the cone, this intersection is the two-dimensional region inside a circle of radius h and outside a circle of radius z, so it has area

$$(\text{area of larger disk}) - (\text{area of smaller disk}) = \pi h^2 - \pi z^2 = \pi(h^2 - z^2).$$

For the solid within the top half of the sphere, this intersection is a disk of radius $\sqrt{h^2 - z^2}$ by the Pythagorean Theorem, so it has area

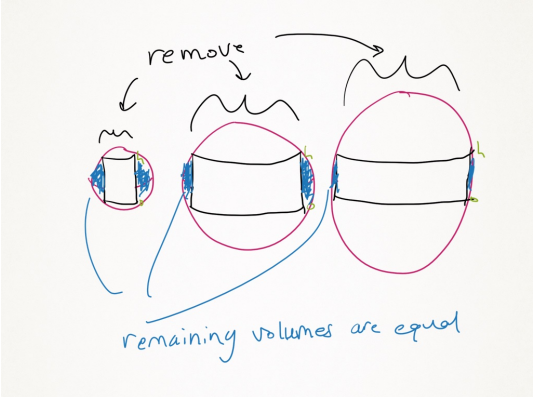
$$\pi\sqrt{h^2 - z^2}^2 = \pi(h^2 - z^2).$$

Thus since these two intersections have the same area at any height z, Cavieleri's Principle implies that the two solids in question have the same volume. If we know that the volume of the cone in question is $\frac{1}{3}\pi h^3$, then the volume of the region outside the cone is

$$(\text{volume of cylinder}) - (\text{volume of cone}) = \pi h^3 - \frac{1}{3}\pi h^3 = \frac{2}{3}\pi h^3,$$

which is thus the volume of the upper half of the sphere. Hence the volume of the entire sphere is twice this amount, which gives the well-known formula for the volume enclosed by a sphere of radius h, provided we know the formula for the volume of a cone. Or conversely, if we happen to know the formula for the volume of a sphere, this will give a way to derive the formula for the volume of a cone.

Second, take spheres of different radii and cut out from each a cylindrical piece through the middle so that the remaining solids have the same height:



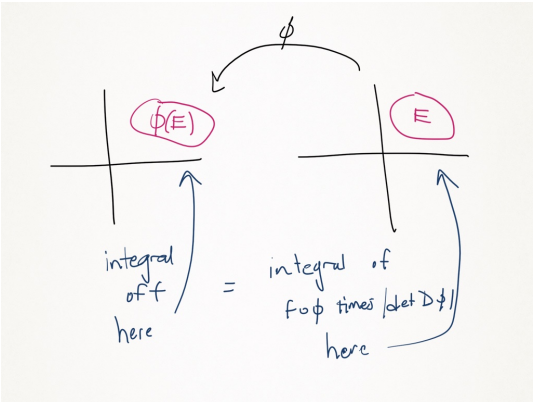
The claim is that these remaining solids have the same volume, so that this volume only depends on the height h used and not on the size of the sphere we started with. I'll leave the details to you, but the point is to show that the intersections of these left-over solids with any horizontal plane have the same area, so that Cavieleri's Principle applies again. This is known as the *Napkin Ring Problem* since the resulting solids which remain tend to look like the types of rings used to hold napkins in place in formal settings.

Change of Variables. The change of variables formula tells us how to express an integral written in terms of one set of coordinates as an integral written in terms of another set of coordinates. This "change of coordinates" is encoded by a function $\phi : V \to \mathbb{R}^n$ defined on some open subset V of $\mathbb{R}^n$. We'll denote the coordinates in the domain by $\mathbf{u} \in V$ and the coordinates in the codomain by $\mathbf{x} \in \mathbb{R}^n$, so that we are making a change of variables of the form $\mathbf{u} = \phi(\mathbf{x})$. In order for things to work out, we assume that ϕ is one-to-one, C^1, and that $D\phi$ is invertible. (We'll mention later where these assumptions come into play.)

Here is the statement. Suppose that $E \subseteq V$ is a Jordan region and that f is integrable over $\phi(E)$. (For this integrability statement to make sense we have to know that $\phi(E)$ is also a Jordan region, which is actually a consequence of the conditions imposed on ϕ as we'll soon clarify.) Then $f \circ \phi$ is integrable over E and:

$$\int_{\phi(E)} f(\mathbf{x})\, d\mathbf{x} = \int_E f(\phi(\mathbf{u}))\, |\det D\phi(\mathbf{u})|\, d\mathbf{u}.$$

Thus, visually:



The Jacobian determinant term plays the role of an "expansion factor" which tells us how volumes on one side relate to volumes on the other side.

Some of the assumptions can be relaxed a bit: we only need ϕ to be one-to-one on E away from a set of Jordan measure zero and similarly we only need $D\phi$ to be invertible on E away from a set of Jordan measure zero. This makes sense, since what happens on these sets of Jordan measure zero can't possibly affect the integrals in question. The requirement that ϕ be one-to-one guarantees that $\phi(E)$ is only "traced out once", which we'll clarify in the following example.

Example. Let us run through the standard conversion you would have seen in a multivariable calculus course from rectangular to polar coordinates in double integrals, in order to make sure that the various assumptions in the change of variables formula indeed hold in this case. The function $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ in this case is defined by $\phi(r, \theta) = (r\cos\theta, r\sin\theta)$. This is C^1 and has Jacobian determinant given by:

$$\det D\phi(r, \phi) = \det \begin{pmatrix} \frac{\partial \phi_1}{\partial r} & \frac{\partial \phi_1}{\partial \theta} \\ \frac{\partial \phi_2}{\partial r} & \frac{\partial \phi_2}{\partial \theta} \end{pmatrix} = \det \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r.$$

Thus this Jacobian matrix is invertible as long as $r \neq 0$, so this fails to be invertible only at the origin which is okay since $\{(0, 0)\}$ has Jordan measure zero.

Take $E = [0, 1] \times [0, 2\pi]$, which is a Jordan region. Then $\phi(E) = \overline{B_1(0, 0)}$, the closed unit disk of radius 1. Note that this is also a Jordan region, which as mentioned before is actually a consequence of the assumptions we have on ϕ. Note that ϕ is not one-to-one on all of E since

$$\phi(r, 0) = \phi(r, 2\pi) \text{ for all } r,$$

but that it only fails to be one-to-one along the bottom and top edges of the rectangle E, which have Jordan measure zero; as mentioned before, this is good enough for the change of variables formula to be applicable. Thus for some integrable function $f : \phi(E) \to \mathbb{R}$, the changes of variables formula gives:

$$\int_{\text{closed unit disk}} f(x, y)\, d(x, y) = \int_{[0,1]\times[0,2\pi]} f(r\cos\theta, r\sin\theta)\, r\, d(r, \theta),$$

just as you would expect.

Suppose that instead we took $E = [0, 1] \times [0, 4\pi]$, which still has image $\phi(E)$ equal to the closed unit disk. For the constant function $f = 1$, if the changes of variables formula we're applicable we would get:

$$\int_{\text{closed unit disk}} d(x, y) = \int_{[0,1]\times[0,4\pi]} r\, d(r, \theta) = \int_0^{4\pi} \int_0^1 r\, dr\, d\theta = 2\pi,$$

which is nonsense because we know that the left hand side should equal the area of the unit disk, which is π. (Note the use of Fubini's Theorem when computing the integral in polar coordinates as an iterated integral.) The problem is now that ϕ fails to be one-to-one throughout E since

$$\phi(r, \theta) = \phi(r, \theta + 2\pi) \text{ for all } r \text{ and } 0 \leq \theta \leq 2\pi,$$

and so the region on which ϕ is not one-to-one no longer has Jordan measure zero. Thus the change of variables formula is not applicable in this case. Geometrically, the problem is that allowing θ to go all the way up to 4π gives two copies of the unit disk superimposed on one another, which means that the unit disk is "traced out twice".

Outline of proof of change of variables formula. The proof of the change of variables formula is quite involved, requiring multiple steps. The book divides these into various lemmas, and you can see that it takes multiple pages to work it all out. Here we only give an outline, emphasizing where the various assumptions we make come into play. Here are the basic steps required:

- Step 0: show that if ϕ is C^1, one-to-one, and has invertible Jacobian matrix, then it sends Jordan regions to Jordan regions. I'm calling this Step 0 because it is actually done in Section 12.1 in the book as Theorem 12.10, so way before the change of variables section. This guarantees that the regions of integration in both integrals showing up in the change of variables formula are indeed Jordan regions. As mentioned previously, the one-to-one and invertibility requirements can be relaxed a bit.

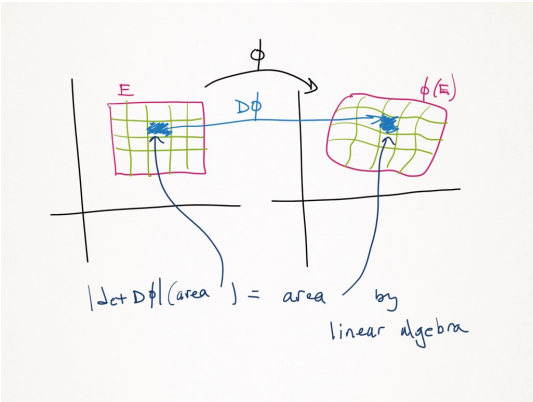
- Step 1: show that the change of variables formula holds in the special case where E is a rectangle and f is the constant function 1. This is Lemma 12.44 in the book, which we'll give some geometric intuition for in a bit.

- Step 2: show that the change of variables formula holds for various pieces of the Jordan region E and an arbitrary integrable function f. This is Lemma 12.43 in the book, and all steps so far are summarized as Lemma 12.45. This is also the step where it is shown that $f \circ \phi$ is automatically integrable as a consequence of our assumptions. The "various" pieces alluded to above come from the open sets on which ϕ is locally invertible as a consequence of the Inverse Function Theorem.

- Step 3: use compactness to show that what happens over the "local" pieces of E gives the require formula over all of E. This is Theorem 12.46 in the book. From the previous steps you end up covering $\overline{E}$ with various open rectangles, and compactness of $\overline{E}$ allows us to work with only finitely many of these.

As mentioned, the details are are quite involved, but since Jordan regions can in a sense be approximated by rectangles (an idea which is used in Step 2), the truly key part is Step 1. This is also the step which explains where the Jacobian determinant factor in the formula comes from, which has a linear algebraic origin. (Don't forget that locally, calculus is just linear algebra after all!) We finish with the geometric intuition behind Step 1.

Geometric intuition behind change of variables. Suppose that E is a rectangle and that we are given some grid on it. Then ϕ transforms this into a possibly "curved grid" on $\phi(E)$:



Since ϕ is differentiable, the Jacobian matrix $D\phi$ provides a good approximation to the behavior of f, and $D\phi$ roughly transforms a small rectangle in the grid on E into a small "curved parallelogram" in the "grid" on $\phi(E)$. According to the geometric interpretation of determinants from linear algebra, the volume of the this parallelogram is related to the area of the rectangle from which it came by:

$$\text{area of } \phi(R) = |\det D\phi(p)|(\text{area of } R).$$

Thus given an outer sum $\sum_i |R_i|$ for E, we get a corresponding "outer sum" for $\phi(E)$ of the form:

$$\sum_i |\phi(R_i)| = \sum_i |\det D\phi(p_i)||R_i|.$$

Note that this is not quite an outer sum since we don't have an honest grid on $\phi(E)$, but rather a curved grid. Also note that since ϕ is C^1, so that $D\phi$ is continuous, we have some control over how changing the grid on E will alter the grid on $\phi(E)$.

Now, since ϕ is C^1 with invertible Jacobian, the Inverse Function Theorem implies that ϕ^{-1} is also C^1. The Jacobian matrix $D\phi^{-1}$ transforms the curved grid on $\phi(E)$ into the original grid on E. Since $D\phi^{-1}$ is continuous, given an honest grid on $\phi(E)$ which approximates the curved grid, the curved grid on E resulting from this honest grid after applying $D\phi^{-1}$ will approximate the original grid on E fairly well. (A mouthful!) Using these ideas we can relate outer, lower, and upper sums on one side to those on the other side, and after some magic we get Step 1.

Important. For a one-to-one, C^1 change of variables $\mathbf{x} = \phi(\mathbf{u})$ with invertible Jacobian matrix, we have

$$\int_{\phi(E)} f(\mathbf{x})\, d\mathbf{x} = \int_E f(\phi(\mathbf{u}))|\det D\phi(\mathbf{u})|\, du$$

whenever everything involved in this formula makes sense, meaning that E should be a Jordan region and f should be integrable on $\phi(E)$. A key takeaway is that the Jacobian determinant $\det D\phi$ tells us how volumes are transformed under a change of variables.

Lecture 18: Curves

Today we started working towards our final topic: the theorems of vector calculus. As a first step we defined and looked at various properties of curves, most of which should be familiar from a previous multivariable calculus course. A key takeaway is that most things we do with curves are done with respect to specific parametric equations, but in the end the results we develop are independent of the choice of parametric equations.

Warm-Up 1. We compute the value of the integral

$$\iint_E \cos(3x^2 + y^2)\, d(x, y)$$

where E is the elliptical region defined by $x^2 + y^2/3 \leq 1$. Note that this integral exists since the integrand is continuous.

We use the change of variables $x = r\cos\theta, y = \sqrt{3}r\sin\theta$, which we can describe using the function $\phi : [0, 1] \times [0, 2\pi] \to \mathbb{R}^2$ defined by

$$\phi(r, \theta) = (r\cos\theta, \sqrt{3}r\sin\theta).$$

This is C^1, one-to-one away from a set of volume zero, and has Jacobian determinant

$$\det D\phi = \det \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sqrt{3}\sin\theta & \sqrt{3}r\cos\theta \end{pmatrix} = \sqrt{3}r,$$

which is nonzero away from a set of volume zero as well. Since $E = \phi([0, 1] \times [0, 2\pi])$ and $3x^2 + y^2 = 3r^2$, the change of variables formula applies to give:

$$\iint_E \cos(3x^2 + y^2)\, d(x, y) = \iint_{[0,1]\times[0,2\pi]} \cos(3r^2)|\sqrt{3}r|\, d(r, \theta).$$

By Fubini's Theorem, we have:

$$\iint_{[0,1]\times[0,2\pi]} \cos(3r^2)|\sqrt{3}r|\,d(r,\theta) = \int_0^{2\pi}\int_0^1 \sqrt{3}r\cos(3r^2)\,dr\,d\theta = \int_0^{2\pi} \frac{\sqrt{3}}{6}\sin 3\,d\theta = \frac{\pi\sin 3}{\sqrt{3}},$$

and thus $\iint_E \cos(3x^2+y^2)\,d(x,y) = \pi\sin 3/\sqrt{3}$ as well.

Warm-Up 2. Suppose that $f:\mathbb{R}^n\to\mathbb{R}^n$ is C^1, one-to-one, and has invertible Jacobian matrix at every point. We show that for each $\mathbf{x}_0\in\mathbb{R}^n$,

$$\lim_{r\to 0^+}\frac{\text{Vol}(f(B_r(\mathbf{x}_0)))}{\text{Vol}(B_r(\mathbf{x}_0))} = |\det Df(\mathbf{x}_0)|.$$

(This is almost the same as Exercise 12.4.6 on the homework, only there we do not assume that f is one-to-one. I'll leave it to you to think about how to get around this subtlety.)

First, we can rewrite the numerator of the fraction of which we are taking the limit as:

$$\text{Vol}(f(B_r(\mathbf{x}_0))) = \int_{f(B_r(\mathbf{x}_0))} d\mathbf{x} = \int_{B_r(\mathbf{x}_0)} |\det Df(\mathbf{u})|\,d\mathbf{u}$$

where we use $\mathbf{x}=f(\mathbf{u})$ as a change of variables, which we can do given the assumptions on f. Thus the fraction of which we are taking the limit is

$$\frac{\text{Vol}(f(B_r(\mathbf{x}_0)))}{\text{Vol}(B_r(\mathbf{x}_0))} = \frac{1}{\text{Vol}(B_r(\mathbf{x}_0))}\int_{B_r(\mathbf{x}_0)} |\det Df(\mathbf{u})|\,d\mathbf{u}.$$

Since f is C^1, the map $\mathbf{u}\mapsto Df(\mathbf{u})$ is continuous, and hence so is $\mathbf{u}\mapsto|\det Df(\mathbf{u})|$ since the operation of taking the determinant of a matrix is a continuous one given that determinants can be expressed as polynomials in the entries of a matrix. Thus the integrand in the resulting integral is continuous at $\mathbf{x}_0$, so Exercise 12.2.3 from the previous homework gives

$$\lim_{r\to 0^+}\frac{\text{Vol}(f(B_r(\mathbf{x}_0)))}{\text{Vol}(B_r(\mathbf{x}_0))} = \lim_{r\to 0^+}\frac{1}{\text{Vol}(B_r(\mathbf{x}_0))}\int_{B_r(\mathbf{x}_0)} |\det Df(\mathbf{u})|\,d\mathbf{u} = |\det Df(\mathbf{x}_0)|$$

as claimed.

Calculus is linear algebra, redux. The previous result should be viewed as another instance of the fact that, locally, multivariable calculus is just linear algebra. Indeed, the linear algebraic fact we are generalizing here is that for an invertible matrix A:

$$\frac{\text{Vol}(A(\Omega))}{\text{Vol}(\Omega)} = |\det A|$$

for any Jordan region Ω with nonzero volume. The result above says that is true "in the limit" for a more general C^1 function with invertible Jacobian. Rewriting this equality as

$$\text{Vol}(A(\Omega)) = |\det A|\,\text{Vol}(\Omega)$$

suggests that the entire change of variables formula itself should be viewed as the non-linear analog of this linear algebraic fact, which we alluded to when outlining the proof of the change of variables formula.

Curves. A "curve" in $\mathbb{R}^n$ should be what we expect it to be, namely some sort of 1-dimensional object. To give a precise definition, we say that a C^p *curve* in $\mathbb{R}^n$ is the image C of a C^p function $\phi : I \to \mathbb{R}^n$ where I is an interval and ϕ is one-to-one in the interior I^o. To be clear, the curve C being described is the one in $\mathbb{R}^n$ with *parametric equations* given by

$$\phi(t) = (x_1(t), \ldots, x_n(t)), \ t \in I$$

where $(x_1, \ldots, x_n)$ are the component of ϕ. We call $\phi : I \to \mathbb{R}^n$ a *parametrization* of C. For the most part, C^1 curves—i.e. curves which can be described using continuously differentiable parametric equations—are what we will be interested in.

We say that a curve with parametrization $\phi : I \to \mathbb{R}^n$ is an *arc* is I is a closed interval $[a, b]$ (so arcs have a start point and an end point), is *closed* if it is an arc and $\phi(a) = \phi(b)$, and is *simple* if it does not intersect itself apart from possibly the common start and end point in the case of a closed curve. Simple, closed curves in $\mathbb{R}^2$ are the ones which divide the plane into two pieces: a piece "interior" to the curve and a piece "exterior" to it. Look up the famous *Jordan Curve Theorem* to learn more about this.

Smooth Curves. A curve $C \subseteq \mathbb{R}^n$ is said to be *smooth* at a point $\mathbf{x}_0$ if there exists a C^1 parametrization $\phi : I \to \mathbb{R}^n$ of C such that $\phi'(t_0) \neq \mathbf{0}$ where t_0 is the point in I which gives $\phi(t_0) = \mathbf{x}_0$. A curve is smooth if it is smooth at each of its points.

As we will show in a bit, the point is that smooth curves are the ones which have well-defined tangent lines. To clarify one possible subtlety in the definition, to say that a curves is smooth at a point means that $\phi'(t_0) \neq \mathbf{0}$ for *some* parametrization, but it is not true that every parametrization of a smooth curve has this property. For instance, the unit circle has parametric equations

$$\phi(t) = (\cos t, \sin t) \text{ for } t \in [0, 2\pi],$$

and since $\phi'(t) = (-\sin t, \cos t)$ is never $\mathbf{0}$, the unit circle is smooth everywhere. However, we can also take

$$\psi(t) = (\cos t^3, \sin t^3) \text{ for } t \in [0, \sqrt[3]{2\pi}]$$

as parametric equations for the unit circle, but in this case

$$\psi'(t) = (-3t^2 \sin t^3, 3t^2 \cos t^3)$$

is $\mathbf{0}$ at $t = 0$, so this would be a *non-smooth* parametrization of the smooth unit circle. So, we can rephrase the definition of smoothness as saying that a smooth curve is one which has a smooth parametrization, even though it may have non-smooth parametrizations as well.

Why we care about smooth curves. To justify the definition of smoothness given above, we show that smooth curves in $\mathbb{R}^2$ have well-defined tangent lines. (Later we will also see that smooth curves are the ones which have well-defined "directions".) We assume that the only types of curves we can precisely define tangent lines for are those which are graphs of single-variable functions (indeed, the whole point of the definition of differentiability for a single-variable function is to capture the idea that the graph should have a well-defined tangent line), so the point is to show that a smooth curve in $\mathbb{R}^2$ can, at least locally, be described by such graphs. This will an application of the Implicit Function Theorem.

So, suppose that $C \subseteq \mathbb{R}^2$ is a smooth C^1 curve with smooth C^1 parametrization $\phi : I \to \mathbb{R}^2$. Pick a point $\mathbf{x}_0 \in C$ and $t_0 \in I$ such that $\phi(t_0) = \mathbf{x}_0$. Then we have parametric equations

$$x = \phi_1(t) \qquad y = \phi_2(t)$$

where (ϕ_1, ϕ_2) are the components of ϕ. Since ϕ is a smooth parametrization,

$$\phi'(t_0) = (\phi_1'(t_0), \phi_2'(t_0)) \neq (0, 0),$$

so at least one component is nonzero—say that $\phi_1'(t_0) \neq 0$. By the Implicit Function Theorem, we can then solve for y in terms of x locally near $\mathbf{x}_0$ in the given parametric equations, or to work it out a little more explicitly, by the Inverse Function Theorem we can solve for t in terms of x in

$$x = \phi_1(t)$$

locally near $\mathbf{x}_0$ to get $t = \phi^{-1}(x)$ where ϕ_1^{-1} is C^1, and then plugging into the second parametric equation gives

$$y = (\phi_2 \circ \phi_1^{-1})(x)$$

which expresses y as a C^1 function of x, at least near $\mathbf{x}_0$. Thus near $\mathbf{x}_0$, the curve C is the graph of the function $f = \phi_2 \circ \phi_1^{-1}$, so the curve has a well-defined tangent line at $\mathbf{x}_0 = (x_0, y_0)$ given by

$$y = f(x_0) + f'(x_0)(x - x_0).$$

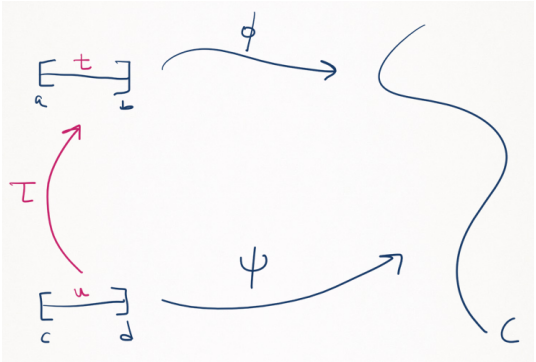
Important. A smooth curve is one which has a smooth parametrization. Geometrically, this is precisely the condition needed to guarantee that the curve has well-defined tangent lines and tangent vectors.

Arclength. Now we can define the notion of the *arclength* of a smooth C^1 arc. Suppose that $C \subseteq \mathbb{R}^n$ is a smooth C^1 arc with smooth C^1 parametrization $\phi : [a, b] \to \mathbb{R}^n$. The arclength of C is by the definition the value of

$$\int_a^b \|\phi'(t)\| \, dt$$

where $\|\cdot\|$ denotes the usual Euclidean norm. (This is likely a definition you saw in a previous course.) The intuition is that $\phi'(t)$ gives a vector tangent to the curve at a given point, so $\|\phi'(t)\|$ gives the length of a little infinitesimal piece of C and this integral is then adding up these infinitesimal lengths to get the total length.

Arclength is independent of parametrization. In order for this to be a good definition, we have to know that the number obtained solely depends on the curve C itself and not on the parametrization used. So, suppose that $\psi : [c, d] \to \mathbb{R}^n$ is another smooth parametrization of C. We take it as a given that for any two such parametrizations $\psi : [c, d] \to \mathbb{R}^n$ and $\phi : [a, b] \to \mathbb{R}^n$ can be related by a "change of variables" function $\tau : [c, d] \to [a, b]$ such that $\psi = \phi \circ \tau$, which describes how to move from the parameter u in ψ to the parameter $t = \tau(u)$ in ϕ:



That such a τ exists is given as Remark 13.7 in the book, which you should check on your own.

With this τ at hand, we have first using a change of variables:

$$\int_{[a,b]} |\phi'(t)|\,dt = \int_{\tau([c,d])} |\phi'(t)|\,dt = \int_{[c,d]} |\phi'(\tau(u))||\tau'(u)|\,du.$$

Since $\psi = \phi \circ \tau$, the chain rules gives $\psi'(u) = \phi'(\tau(u))\tau'(u)$, so this final integral is

$$\int_{[c,d]} |\phi'(\tau(u))||\tau'(u)|\,du = \int_{[c,d]} |\psi'(u)|\,du,$$

so

$$\int_a^b |\phi'(t)|\,dt = \int_c^d |\psi'(u)|\,du$$

as required in order to say that ϕ and ψ give the same arclength.

Line integrals. Given a smooth C^1 arc C in $\mathbb{R}^n$ with parametrization $\phi : [a,b] \to \mathbb{R}^n$ and a continuous function $f : C \to \mathbb{R}$, we define the *line integral* of f over C to be:

$$\int_C f\,ds = \int_a^b f(\phi(t))\,\left\|\phi'(t)\right\|\,dt.$$

This integral essentially "adds up" the values of f along the curve C, and the fact that ϕ is C^1 and that f is continuous guarantees that this integral exists. In a previous course you might have seen this referred to as a *scalar* line integral, to distinguish it from so-called *vector* line integrals which arise when integrating *vector fields* along curves—we'll look at this type of line integral later on.

Even though the definition given depends on a parametrization of C, an argument similar to that for arclength shows that the line integral of f over C is independent of the parametrization used. In fact, we can also try to define the integral $\int_C f\,ds$ via an upper/lower sum approach which makes no reference to parametrizations at all; we'll outline how to do this later on and argue that this approach gives the same value for $\int_C f\,ds$ as to how we've defined it here.

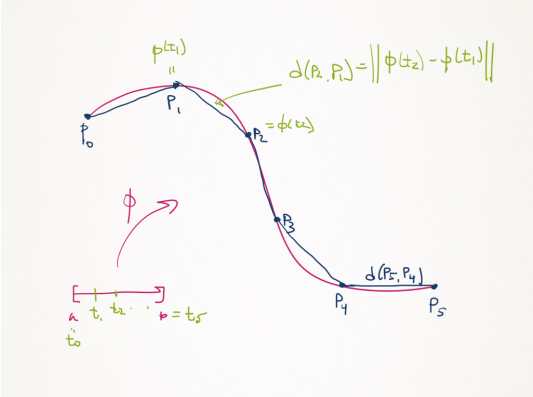
Important. The line integral of a continuous function $f : C \to \mathbb{R}$ over a smooth C^1 arc C is independent of parametrization. In the case where f is the constant function 1, this line integral gives the arclength of C.

Lecture 19: Surfaces

Today we spoke about surfaces, where as we did last time with curves, the point is to give precise definitions and justifications for concepts you would have seen in a previous multivariable calculus course. Note that everything we now do is an analog of something we did for curves.

Another approach to arclength. Before moving on to surfaces, we give another reason as to why the definition we gave for arclength last time makes sense geometrically, by relating it to another possible definition. This is all covered in the optional material at the end of Section 13.1 in the book.

The idea is that to define the length of a curve we can also argue using "line segment approximations". Say we are given some curve, and choose a finite number of points along it:



The distances $d(p_i, p_{i-1})$ (where d denotes the Euclidean distance) between successive points give some sort of approximation to the length of the curve between those points, and the total sum:

$$\sum_i d(p_i, p_{i-1})$$

then underestimates the actual length of the curve. Taking more and more points results in better and better approximations, so we can try to define the length as the supremum of such sums:

$$\text{arclength of } C = \sup \left\{ \sum_i d(p_i, p_{i-1}) \right\}.$$

Curves for which this quantity is finite are called *rectifiable*, and the claim is that for rectifiable smooth C^1 curves, this definition of arclength agrees with the previous one.

The idea behind the proof is as follows. Choose a parametrization $\phi : [a, b] \to \mathbb{R}^n$ of C and a partition of $[a, b]$. Then the points $\phi(x_0), \phi(x_1), \ldots, \phi(x_n)$ give successive points on the curve, and the distance between then are given by the expressions

$$\|\phi(x_i) - \phi(x_{i-1})\|.$$

Thus the sum over the entire partition is

$$\sum_i \|\phi(x_i) - \phi(x_{i-1})\|.$$

Now, using some version of the Mean Value Theorem, we can approximate these terms by

$$\|\phi(x_i) - \phi(x_{i-1})\| \approx \left\|\phi'(x_i)\right\| \Delta x_i,$$

so

$$\sum_i \|\phi(x_i) - \phi(x_{i-1})\| \approx \sum_i \left\|\phi'(x_i)\right\| \Delta x_i.$$

But the right-hand side can now be viewed as a Riemann sum for the integral $\int_a^b \|\phi'(t)\|\, dt$, so taking supremums of both sides gives the result. This is only meant to be a rough idea, but you can check the book for full details.

One final thing to note: this approach to defining arclength only depends on the Euclidean distance d, and so can in fact be generalized to arbitrary metric spaces. The result is a definition of arclength which makes sense in any metric space!

Warm-Up. Suppose that $f : I \to \mathbb{R}$ is a C^1 function on some interval $I \subseteq \mathbb{R}$ such that

$$|f(\theta)|^2 + |f'(\theta)|^2 \neq 0 \text{ for all } \theta \in I.$$

We show that the curve defined by the polar equation $r = f(\theta)$ is a smooth C^1 curve.

The curve we are looking at is the one consisting of all points in $\mathbb{R}^2$ whose polar coordinates satisfy $r = f(\theta)$. Thus our curve is given parametrically by:

$$x = r\cos\theta = f(\theta)\cos\theta \qquad y = r\sin\theta = f(\theta)\sin\theta.$$

Since f is C^1, the function $\phi : I \to \mathbb{R}^2$ given by $\phi(\theta) = (f(\theta)\cos\theta, f(\theta)\sin\theta)$ is C^1 as well so we do have a C^1 curve. To check smoothness we compute:

$$x'(\theta) = f'(\theta)\cos\theta - f(\theta)\sin\theta \qquad y'(\theta) = f'(\theta)\sin\theta + f(\theta)\cos\theta.$$

Then

$$\begin{aligned}\big\|(x'(\theta), y'(\theta))\big\| &= \sqrt{(f'(\theta)\cos\theta - f(\theta)\sin\theta)^2 + (f'(\theta)\sin\theta + f(\theta)\cos\theta)^2} \\ &= \sqrt{f'(\theta)^2 + f(\theta)^2} \neq 0\end{aligned}$$

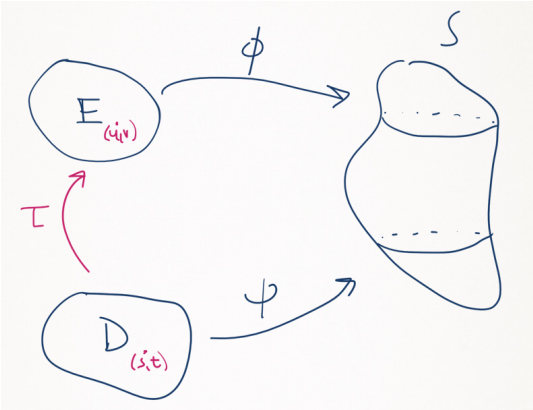
after simplification. Thus $\phi'(\theta) = (x'(\theta), y'(\theta))$ is nonzero everywhere, so C is smooth as claimed.

Surfaces. Intuitively, a surface should be a "2-dimensional" object in $\mathbb{R}^3$. To make this precise, we say that a C^p *surface* in $\mathbb{R}^3$ is the image S of a C^p function $\phi : E \to \mathbb{R}^3$ defined on a closed Jordan region E in $\mathbb{R}^2$ which is one-to-one on E^o. As with curves, we call $\phi : E \to \mathbb{R}^3$ a *parametrization* of S and its components (ϕ_1, ϕ_2, ϕ_3) give us parametric equations

$$x = \phi_1(u, v),\ y = \phi_2(u, v),\ z = \phi_3(u, v) \text{ with } (u, v) \in E$$

for S. The fact that there are two parameters in these equations is what makes S two-dimensional.

We note that, as with curves, a given surface can be expressed using different sets of parametric equations. In general, if $\phi : E \to \mathbb{R}^3$ and $\psi : D \to \mathbb{R}^3$ are two parametrizations of a surface S, there is a "change of variables" function $\tau : D \to E$ such that $\psi = \phi \circ \tau$, which tells us how to move from parameters (s, t) for ψ to parameters $(u, v) = \tau(s, t)$ for ϕ:



The proof of this fact is similar to the proof of the corresponding fact for curves given in the book.

Example. Suppose that S is the portion of the cone $z = \sqrt{x^2 + y^2}$ for $0 \leq z \leq h$, where h is some fixed height. One set of parametric equations for S is

$$\phi(u, v) = (u, v, \sqrt{u^2 + v^2}) \text{ with } (u, v) \in \overline{B_h(0, 0)}.$$

Note however that this parametrization is not C^1 since the z-component is not differentiable at $(u, v) = (0, 0)$, which corresponds to the point $(0, 0, 0)$ on the cone.

Instead, the parametrization given by

$$\psi(r, \theta) = (r\cos\theta, r\sin\theta, r) \text{ with } (r, \theta) \in [0, h] \times [0, 2\pi]$$

is C^1. (In fact, this parametrization is C^∞, showing that the cone is a C^∞ surface.)

Smooth surfaces and normal vectors. Suppose that $\phi : E \to \mathbb{R}^3$ is a C^p (C^1 will usually be enough) parametrization of a surface S. Denote the parameters by $(u, v) \in E$. Holding v fixed at a point and varying u results in a parametrization $\phi(\cdot, v)$ of a curve on S, and thus differentiating with respect to u gives a vector tangent to the surface, which we denote by ϕ_u. Similarly, holding u fixed gives a parametrization $\phi(u, \cdot)$ of another curve on S with tangent vector ϕ_v obtained by differentiating with respect to v.

We say that S is *smooth* at a point $(x_0, y_0, z_0) = \phi(u_0, v_0)$ if the cross product

$$(\phi_u \times \phi_v)(u_0, v_0)$$

is nonzero at that point, in which we case we call $(\phi_u \times \phi_v)(u_0, v_0)$ a *normal vector* to S at $\phi(u_0, v_0)$. (Recall that this cross product is perpendicular to both tangent vectors ϕ_u and ϕ_v, which intuitively suggests that it should indeed be perpendicular to the surface S.) We say that S is smooth if it is smooth everywhere. This smoothness condition is what guarantees that well-defined tangent planes exist, as we will see in the Warm-Up next time. For now we mention that, as we saw with curves, a smooth surface may have non-smooth parametrizations—all that matters is that a smooth parametrization exists.

Concretely, if $\phi = (\phi_1, \phi_2, \phi_3)$ are the components of ϕ, then:

$$\phi_u \times \phi_v = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial\phi_1}{\partial u} & \frac{\partial\phi_2}{\partial u} & \frac{\partial\phi_3}{\partial u} \\ \frac{\partial\phi_1}{\partial v} & \frac{\partial\phi_2}{\partial v} & \frac{\partial\phi_3}{\partial v} \end{vmatrix} = \left(\frac{\partial\phi_2}{\partial u}\frac{\partial\phi_3}{\partial v} - \frac{\partial\phi_2}{\partial v}\frac{\partial\phi_3}{\partial u}, -\frac{\partial\phi_1}{\partial u}\frac{\partial\phi_3}{\partial v} + \frac{\partial\phi_1}{\partial v}\frac{\partial\phi_3}{\partial u}, \frac{\partial\phi_1}{\partial u}\frac{\partial\phi_2}{\partial v} - \frac{\partial\phi_1}{\partial v}\frac{\partial\phi_2}{\partial u} \right).$$

But note that the first component here can be viewed as the Jacobian determiant of the matrix:

$$D(\phi_2, \phi_3) := \begin{pmatrix} \frac{\partial\phi_2}{\partial u} & \frac{\partial\phi_3}{\partial u} \\ \frac{\partial\phi_2}{\partial v} & \frac{\partial\phi_3}{\partial v} \end{pmatrix},$$

and similarly the other components can also be viewed as Jacobian determinants. (Technically, the matrix above is the transpose of the Jacobian matrix $D(\phi_2, \phi_3)$ of the function defined by the components (ϕ_2, ϕ_3), but since a matrix and its transpose have the determinant this will not affect our formulas.) That is, we have:

$$\phi_u \times \phi_v = (\det D(\phi_2, \phi_3), -\det D(\phi_1, \phi_3), \det D(\phi_1, \phi_2))$$

where $D(\phi_i, \phi_j)$ denotes the Jacobian matrix of the function defined by the components ϕ_i and ϕ_j. To save space, we will also denote this normal vector by N_ϕ.

Important. A surface is smooth if it has nonzero normal vectors at every point. (Often times, being smooth except on a set of volume zero will be enough.) Geometrically, this condition guarantees the existence of a well-defined tangent plane.

Back to cone example. Consider the C^1 parametrization $\psi(r,\theta) = (r\cos\theta, r\sin\theta, r)$ of the cone $z = \sqrt{x^2+y^2}$ we saw earlier. We have:

$$\psi_r = (\cos\theta, \sin\theta, 1) \quad \text{and} \quad \psi_\theta = (-r\sin\theta, r\cos\theta, 0),$$

so normal vectors are given by

$$\psi_r \times \psi_\theta = (-r\sin\theta, -r\cos\theta, r).$$

This is nonzero as long as $r \neq 0$, so we see that the cone is smooth everywhere except at $(0,0,0)$, which is the "tip" of the cone. Note that it makes sense geometrically that the cone should not be smooth a this point, since the cone does not have a well-defined tangent plane at this point.

How normal vectors change under a change of variables. Suppose that $\phi : E \to \mathbb{R}^3$ and $\psi : D \to \mathbb{R}^3$ are two parametrizations of a smooth C^1 surface S, with parameters $(u,v) \in E$ and $(s,t) \in D$. We can directly relate the normal vectors obtained by ϕ to those obtained by ψ as follows.

Recall that given these parametrizations there exists a C^1 function $\tau : D \to E$ such that $\psi = \phi \circ \tau$. Also recall the expression derived previously for normal vectors determined by ψ:

$$\psi_s \times \psi_t = (\det D(\psi_2, \psi_3), -\det D(\psi_1, \psi_3), \det D(\psi_1, \psi_2)) .$$

Since $\psi = \phi \circ \tau$, we also have

$$(\psi_i, \psi_j) = (\phi_i, \phi_j) \circ \tau, \text{ so } \det D(\psi_i, \psi_j) = (\det D(\phi_i, \phi_j))(\det D\tau)$$

by the chain rule. Thus

$$\psi_s \times \psi_t = (\det D\tau)\,(\det D(\phi_2, \phi_3), -\det D(\phi_1, \phi_3), \det D(\phi_1, \phi_2)) ,$$

so we get that

$$N_\psi(s,t) = (\det D\tau(s,t)) N_\phi(\tau(s,t)).$$

Hence normal vectors determined by ψ are obtained by multiplying those which are determined by ϕ by $\det D\tau$, so we should think of $\det D\tau$ as a type of "expansion factor" telling us how normal vectors are affected under a change of variables. This is analogous to the formula

$$\psi'(u) = \tau'(u)\phi'(\tau(u))$$

relating tangent vectors determined by two parametrizations of a curve, where τ (a single-variable change of coordinates in this case) plays a similar "expansion factor" role.

Surface area. Suppose that S is a smooth C^1 surface. Given a smooth C^1 parametrization $\phi : E \to \mathbb{R}^3$ with parameters $(u,v) \in E$, we define the *surface area* of S to be:

$$\iint_E \|N_\phi(u,v)\| \, d(u,v).$$

Intuitively, $\|\phi_u \times \phi_v\|$ gives the area of the little infinitesimal portion of S swept out by the tangent vectors ϕ_u and ϕ_v, so to obtain the total surface area we add up all of these infinitesimal areas.

As with arclength, this definition is independent of parametrization, as we now show. Let $\psi : D \to \mathbb{R}^3$ be another smooth C^1 parametrization with $\tau : D \to E$ satisfying $\psi = \phi \circ \tau$, so that $(u, v) = \tau(s, t)$. Then:

$$\begin{aligned}\iint_E \|N_\phi(u,v)\|\, d(u,v) &= \iint_{\tau(D)} \|N_\phi(u,v)\|\, d(u,v) \\ &= \iint_D \|N_\phi(\tau(s,t))\|\, |\det D\tau(s,t)|\, d(s,t) \\ &= \iint_D \|N_\psi(s,t)\|\, d(s,t)\end{aligned}$$

where in the second line we've used a change of variables and in the final line the relation between N_ψ and N_ϕ derived above. Thus the surface area as computed using ψ is the same as that computed using ϕ, so the surface area is independent of parametrization.

Surface integrals. Given a smooth C^1 surface S with parameterization $\phi : E \to \mathbb{R}^3$ and a continuous function $f : S \to \mathbb{R}^3$, we define the *surface integral* of f over S to be:

$$\iint_S f\, dS := \iint_E f(\phi(u,v))\, \|\phi_u \times \phi_v\|\, d(u,v),$$

which we interpret as adding up all the values of f as we vary throughout S. Note that the book uses $d\sigma$ instead of dS in the notation for surface integrals, but I like dS since it emphasizes better that we are integrating over a surface.

Using an argument similar to the one for surface area, it can be shown that this definition is also independent of parametrization. As such, it makes sense to ask whether we can define surface integrals without using parametrizations at all. We'll come back to this next time.

Important. Surface integrals arise when integrating functions over surfaces and are independent of the parametrization used. In the case where f is the constant function 1, the surface integral gives the surface area.

Lecture 20: Orientations

Today we spoke about orientations of both curves and surfaces. The point is that an orientation will give us a to turn vector-valued functions (i.e. vector fields) into scalar-valued functions which are then suitable for integration. Although curves are always orientable, we'll see a well-known example of a surface which is not orientable.

Warm-Up. Suppose that S is a C^1 surface which is smooth at $(x_0, y_0, z_0) \in S$. We show that S then has a well-defined tangent plane at (x_0, y_0, z_0). To be clear, we are assuming that the only type of surface for which tangent planes are well-defined are graphs of differentiable functions $f : V \to \mathbb{R}$ with $V \subseteq \mathbb{R}^2$(indeed, you can take the definition of differentiable in this setting as what it means for the graph to have a well-defined tangent plane), so the claim is that locally near (x_0, y_0, z_0) we can express S as the graph of such a function. This calls for the Implicit/Inverse Function Theorem.

Since S is smooth at (x_0, y_0, z_0) there exists a parametrization $\phi : E \to \mathbb{R}^3$ with $\phi(u_0, v_0) = (x_0, y_0, z_0)$ such that $N_\phi(u_0, v_0) \neq \mathbf{0}$. Recall from last time the expression

$$N_\phi(u_0, v_0) = (\det D(\phi_2, \phi_3)(u_0, v_0), -\det D(\phi_1, \phi_3)(u_0, v_0), \det D(\phi_1, \phi_2)(u_0, v_0))$$

for the normal vector N_ϕ, where $\phi = (\phi_1, \phi_2, \phi_3)$ are the components of ϕ. Since this is nonzero, at least one component is nonzero—we will assume that it is the third component which is nonzero. (This will result in expressing z as a function of x and y, so if instead one of the other components were nonzero we would end up expressing that corresponding variable as a function of the other two.) Our parametric equations looks like

$$x = \phi_1(u, v), \;\; y = \phi_2(u, v), \;\; z = \phi_3(u, v) \quad (u, v) \in E,$$

so since $\det D(\phi_1, \phi_2)(u_0, v_0) \neq 0$ and (ϕ_1, ϕ_2) is C^1, the Inverse Function Theorem implies that we can locally express u and v as a C^1 function of x and y:

$$(u, v) = \psi(x, y) \text{ for some } C^1 \text{ function } \psi$$

near (x_0, y_0, z_0). This gives

$$z = \phi_3(u, v) = (\phi_3 \circ \psi)(x, y)$$

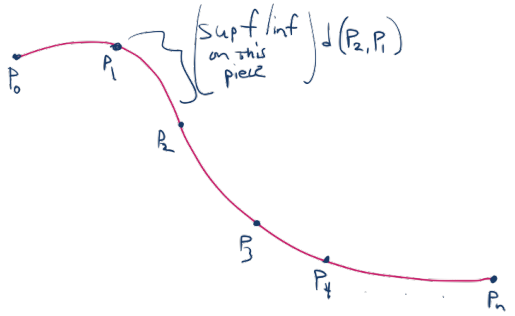
near (x_0, y_0, z_0), which expresses S near this point as the graph of the C^1 function $f := \phi_3 \circ \psi$. Thus S has a well-defined tangent plane at (x_0, y_0, z_0) which is explicitly given by

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

where f is the C^1 function defined above.

Line and surface integrals without parametrizations. Before moving on to orientations, we outline an attempt to define line and surface integrals without resorting to using parametric equations, mimicking the definitions we gave for integrals in terms of Riemann sums.

Suppose we are given some smooth C^1 arc $C \subseteq \mathbb{R}^n$ and a continuous function $f : C \to \mathbb{R}^n$. To define the integral of f over C we can proceed by "partitioning" C by choosing successive points $p_0, p_1, \ldots, p_n$ along C:



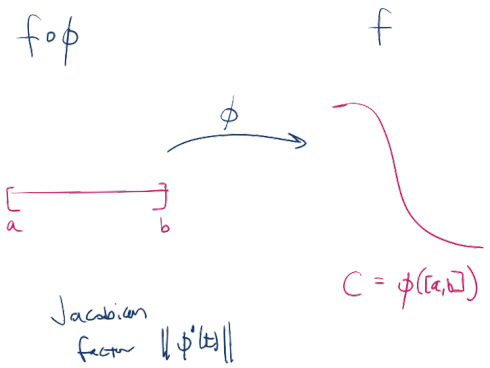
taking the infimum or supremum of f along the part of C between successive partition points, and then forming "lower" and "upper" sums

$$\sum_i (\inf f) d(p_i, p_{i-1}) \qquad \text{and} \qquad \sum_I (\sup f) d(p_i, p_{i-1})$$

where d denotes the ordinary Euclidean distance. (Note that we are essentially mimicking the alternate approach to defining arclength we outlined previously for "rectifiable" curves.) We would then define $\int_C f \, ds$ as the common value of the supremum of the lower sums or the infimum of the

upper sums. It turns out that this definition works perfectly well for rectifiable curves and gives the value we would expect.

However, note now that, in this setting, we would expect a "change of variables" formula to hold just as it does for other types of integrals we've seen, where a change of variables function $\phi : [a, b] \to \mathbb{R}^n$:



is precisely a parametrization of C! Thus, the change of variable formula in this setting would give

$$\int_C f\,ds = \int_{\phi([a,b])} f\,ds = \int_{[a,b]} f(\phi(t))\,\|\phi'(t)\|\,dt,$$

where $\|\phi'(t)\|$ is the Jacobian expansion factor, which is precisely the definition we gave for $\int_C f\,ds$ using a parametrization. Here's the point: even if we defined line integrals using a Riemann sum approach, the corresponding change of variables formula would imply that this definition to the one in terms of a parametrization, so rather than go through the trouble of defining such Riemann sums we simply take the parametrization approach as our definition, since in the end it would give the correct answer anyway.

The same is true for surface integrals. We can define surface integrals independently of parametrization by using certain upper and lower sums by picking "grids" on our surface, but in the end the change of variables formula will say that the resulting integral is equal to the one given in terms of a parametrization, so we simply take the latter approach as our definition of a surface integral, thereby avoiding having to develop some extra theory. This is morally why constructions involving curves and surfaces always come down to some computations in terms of parametrizations: even if we could perform these constructions without using parametric equations, in the end we would get the same types of objects using parametric equations anyway.

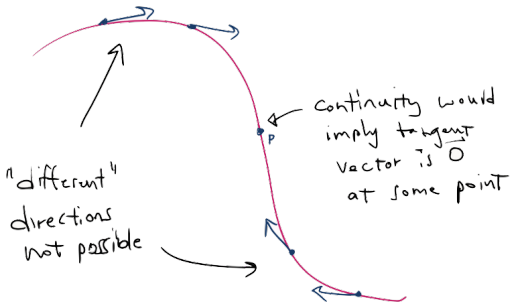
Orientations of curves. Given a smooth C^1 curve C, an *orientation* on C is simply a choice of continuously-varying unit tangent vectors along C. (We need the smoothness and C^1 assumptions to guarantee that our curves have well-defined tangent vectors, coming from the fact that they have well-defined tangent lines.) Visually this just amounts to choosing a "direction" in which to follow C. Concretely, if $\phi : I \to \mathbb{R}^n$ is a parametrization of C, then the unit tangent vectors

$$\frac{\phi'(t)}{\|\phi'(t)\|}$$

give us one possible orientation of C and the negative of these gives us the other possible orientation.

To be clear, the C^1 assumption says that the assignment $t \mapsto \phi'(t)$ of a tangent vector to each point on the curve is a continuous one, which is what we mean by saying that the tangent vectors

vary "continuously" along the curve. Note that because of this, for a connected curve we cannot have a scenario such as:



since some version of the Intermediate Value Theorem would imply that in this setting there must be some $p = \phi(t_0) \in C$ at which $\phi'(t_0) = \mathbf{0}$, which is ruled out by the smoothness assumption on C. This is what guarantees that we are indeed choosing a single direction of movement along C when we are picking an orientation.

When do parametrizations give the same orientation? We can easily determine when two parametrizations ϕ, ψ of a curve C give the same orientation. Recall that we have $\psi = \phi \circ \tau$ for some function $t = \tau(u)$. Then

$$\psi'(u) = \phi'(\tau(u))\tau'(u),$$

which implies that the unit vectors obtained from ψ and ϕ are the same precisely when $\tau'(u) > 0$. Thus two parametrizations give the same orientation when the "change of parameters" function τ has positive derivative everywhere.

For instance, consider the unit circle C in $\mathbb{R}^2$. This has possible parametric equations

$$\phi(t) = (\cos t, \sin t) \text{ for } 2\pi \leq t \leq 4\pi,$$

and also

$$\psi(u) = (\cos u^3, \sin u^3) \text{ for } \sqrt[3]{2\pi} \leq u \leq \sqrt[3]{4\pi}.$$

In this case, $\tau(u) = u^3$ is the function satisfying $\psi = \phi \circ \tau$, and since $\tau'(u) = 3u^2$ is positive for all $\sqrt[3]{2\pi} \leq u \leq \sqrt[3]{4\pi}$, we have that ϕ and ψ determine the same orientation on C, which makes sense since both sets of parametric equations give the "counterclockwise" direction on the unit circle.

Orientations of surfaces. Given a smooth C^1 surface S, an *orientation* on S is a choice of continuously-varying unit normal vectors across S. (Again, the smoothness and C^1 assumptions guarantee that S has well-defined tangent planes, so the the notion of a "normal vector" makes sense through S.) For a parametrization $\phi : E \to \mathbb{R}^3$ of S, the possible unit normal vectors are given by

$$\frac{\phi_u \times \phi_v}{\|\phi_u \times \phi_v\|} \quad \text{or} \quad \frac{\phi_v \times \phi_u}{\|\phi_u \times \phi_v\|},$$

which are negatives of one another.

The C^1 assumption says that the map $(u, v) \mapsto \phi_u \times \phi_v$ is continuous, which is what we mean by "continuously-varying" normal vectors across S. The smoothness assumption guarantees that for a connected surface we can't have one possible orientation along a portion of our surface but then opposite orientation along another portion: for this to be true the Intermediate Value Theorem would imply that at some point the normal vector would have to be zero, which is not allowed by

smoothness. However, there is a new subtlety with surfaces which we didn't see for curves: not all surfaces have well-defined orientations. We'll come back to this in a it.

Recalling that the normal vectors of two parametrizations ϕ and ψ are related by

$$N_\psi = (\det D\tau) N_\phi$$

where τ is the change of parameters function satisfying $\psi = \phi \circ \tau$, we see that two parametrizations determine the same orientation when $\det D\tau > 0$ throughout S.

Example. Parametric equations for the unit sphere are given by:

$$\mathbf{X}(\phi, \theta) = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi) \text{ for } (\phi, \theta) \in [0, \pi] \times [0, 2\pi].$$

This gives the normal vectors

$$\mathbf{X}_\phi \times \mathbf{X}_\theta = \sin\phi(\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi) = (\sin\phi)\mathbf{X}(\phi, \theta).$$

Since $\sin\phi > 0$ for $\phi \in (0, \pi)$, this gives normal vectors which point in the same direction as the vector $\mathbf{X}(\phi, \theta)$ extending from the origin to a point on the sphere, so this gives the "outward" orientation on the sphere. The negatives of these normal vectors give the inward orientation. Note that the possibility that certain normal vectors point outward and others inward is ruled out by the Intermediate Value Theorem and the fact that $\mathbf{X}$ is a C^1 parametrization, as mentioned earlier.

Non-orientable surfaces. Consider the surface S with C^1 parametrization

$$\phi(u, v) = \left(\left(1 + v\sin\frac{u}{2}\right)\cos u, \left(1 + v\sin\frac{u}{2}\right)\sin u, v\cos\frac{u}{2}\right) \text{ for } (u, v) \in [0, 2\pi] \times \left[-\frac{1}{2}, \frac{1}{2}\right].$$

A lengthy computation shows that

$$\begin{aligned}\phi_u \times \phi_v = \sin\frac{u}{2}\left(\cos u + 2v\left(\cos^3\frac{u}{2} - \cos\frac{u}{2}\right)\right)\mathbf{i} + \frac{1}{2}\left(4\cos\frac{u}{2} - 4\cos^3\frac{u}{2} + v(1 + \cos u - \cos^2 u)\right)\mathbf{j} \\ - \cos\frac{u}{2}\left(1 + v\cos\frac{u}{2}\right)\mathbf{k},\end{aligned}$$

and from this it is possible (although tedious) to show that S is smooth everywhere.

Now, note from the normal vector derived above that:

$$N_\phi(0, 0) = (0, 0, -1) \qquad \text{and} \qquad N_\phi(2\pi, 0) = (0, 0, 1).$$

However, going back to the parametric equations, we see that

$$\phi(0, 0) = (1, 0, 0) = \phi(2\pi, 0),$$

so the values $(u, v) = (0, 0)$ and $(u, v) = (2\pi, 0)$ of our parameters both determine the same point on S. This is bad: using $(u, v) = (0, 0)$ gave a normal vector of $(0, 0, -1)$ at this point while using $(u, v) = (2\pi, 0)$ gave a normal vector of $(0, 0, 1)$, which points in the opposite direction. The conclusion is that the given parametrization does not give a unique well-defined normal vector at $(1, 0, 0)$. We will show as a Warm-Up next time that no choice of parametrization of this surface S gives a well-defined normal vector at $(1, 0, 0)$, so the problem here is really a fault of the surface and not any specific choice of parametrization.

The lack of being able to define a unique normal vector at each point of S says that it is not possible to give S an orientation, so we say that S is *non-orientable*. The problem is that we

cannot make a consistent choice of "continuously-varying" normal vectors across S. We will avoid such surfaces in this class since the types of integrals we will soon consider depend on having an orientation. This surface in particular is known as the *Möbius strip*, and is what you get if you take a strip of paper, twist one end, and then glue the ends together—there's a picture in the book. Two other famous non-orientable surfaces are the *Klein bottle* and the *real projective plane*, although these really aren't "surfaces" according to our definition since they cannot be embedded in $\mathbb{R}^3$—rather, they would be examples of non-orientable surfaces in $\mathbb{R}^4$.

Important. An orientation of a curve is a continuous choice of unit tangent vectors along the curve, and an orientation of a surface is a continuous choice of unit normal vectors across the surface. All curves have orientations, but not all surfaces do.

Lecture 21: Vector Line/Surface Integrals

Today we started talking about vector line/surface integrals, which the book calls *oriented* line/surface integrals. These are the integrals which arise when wanting to integrate a *vector field* over a curve or surface, and are the types of integrals which the "Big Theorems of Vector Calculus" deal with. We reviewed some properties and even computed a couple of explicit examples.

Warm-Up. We justify a claim we made last time about the Möbius strip S, that the lack of having a well-defined normal vector at every point throughout the surface really is a characteristic of the surface itself and not a fault of the specific parametrization we previously gave. To be precise, the claim is that if $\psi : D \to \mathbb{R}^3$ is any parametrization of S, then ψ does not assign to the point $(1, 0, 0) \in S$ a well-defined unique normal vector.

Let $\phi : D \to \mathbb{R}^3$ be the parametrization we gave for S last time. The key thing to recall is that for these equations we have

$$N_\phi(0, 0) = -\mathbf{k} \qquad \text{but} \qquad N_\phi(2\pi, 0) = \mathbf{k}$$

even though $\phi(0, 0)$ and $\phi(2\pi, 0)$ both give the same point $(1, 0, 0)$ on S. If $\tau : D \to E$ is the change of parameters function satisfying $\psi = \phi \circ \tau$, then

$$N_\psi(s, t) = (\det D\tau(s, t))N_\phi(\tau(s, t)).$$

In particular, take (s_0, t_0) to be one set of values which give $(1, 0, 0)$ and (s_1, t_1) to be another, so that $\tau(s_0, t_0) = (0, 0)$ and $\tau(s_1, t_1) = (2\pi, 0)$. Then

$$N_\psi(s_0, t_0) = (\det D\tau(s_0, t_0))N_\phi(0, 0) = -(\det D\tau(s_0, t_0))\mathbf{k}$$

and

$$N_\psi(s_1, t_1) = (\det D\tau(s_1, t_1))N_\phi(2\pi, 0) = (\det D\tau(s_1, t_1))\mathbf{k}.$$

But $\det D\tau$ is either always positive or always negative since otherwise the Intermediate Value Theorem would imply that it was zero somewhere, contradicting the smoothness of these parametrizations. Thus in either case, the two vectors above will never point in the same direction since one is going to be a positive multiple of $\mathbf{k}$ and the other a negative multiple of $\mathbf{k}$. Since (s_0, t_0) and (s_1, t_1) both give the point $(1, 0, 0)$ on S, this shows that ψ does not assign a unique normal vector to this point as claimed.

Vector line integrals. Suppose that C is a smooth C^1 oriented arc in $\mathbb{R}^n$ and that $\mathbf{F} : C \to \mathbb{R}^n$ is a continuous function. (Such an $\mathbf{F}$ is said to be a continuous *vector field* along C, which we

normally visualize as a field of little vectors varying along C.) We define the *vector line integral* (or *oriented line integral*) of $\mathbf{F}$ over C as:

$$\int_C \mathbf{F} \cdot \mathbf{T}\, ds$$

where $\mathbf{T}$ denotes the unit tangent vector field along C determined by the given orientation. To be clear, the orientation determines $\mathbf{T}$, which is then used to turn the $\mathbb{R}^n$-valued function $\mathbf{F}$ into an $\mathbb{R}$-valued function $\mathbf{F} \cdot \mathbf{T}$, which is then integrated over the curve using a scalar line integral. This procedure is why we care about orientations! Given a parametrization $\phi : [a, b] \to \mathbb{R}^n$ of C, we get:

$$\int_C \mathbf{F} \cdot \mathbf{T}\, ds = \int_a^b \mathbf{F}(\phi(t)) \cdot \frac{\phi'(t)}{\|\phi'(t)\|} \left\|\phi'(t)\right\| \, dt,$$

which gives the familiar formula

$$\int_C \mathbf{F} \cdot \mathbf{T}\, ds = \int_a^b \mathbf{F}(\phi(t)) \cdot \phi'(t)\, dt$$

you would have seen in a multivariable calculus course. (Of course, the value is actually independent of the parametrization.)

Geometrically, this integral measures the extent to which you move "with" or "against" the flow of $\mathbf{F}$ as you move along the curve. To be concrete, the dot product $\mathbf{F} \cdot \mathbf{T}$ is positive when $\mathbf{F}$ and $\mathbf{T}$ point in the same "general" direction—i.e. when the angle between them is less than $90°$–and is negative when $\mathbf{F}$ and $\mathbf{T}$ point in "opposite" directions—i.e. when the angle between them is greater than $90°$— and the line integral then "adds up" all of these individual dot product contributions.

Example 1. Just in case it's been a while since you've looked at these types of integrals, we'll do an explicit computation. Let C be the piece of the parabola $y = x^2$ oriented from $(\pi/2, \pi^2/4)$ to $(5\pi/4, 25\pi^2/16)$ and let $\mathbf{F} : C \to \mathbb{R}^2$ be

$$\mathbf{F}(x, y) = \left(-\frac{y \sin x}{x^2}, \frac{\cos x}{2x}\right).$$

We use $\phi : [\pi/2, 5\pi/4] \to \mathbb{R}^2$ defined by $\phi(t) = (t, t^2)$ as a parametrization of C. Then:

$$\begin{aligned}
\int_C (\mathbf{F} \cdot \mathbf{T})\, ds &= \int_{\pi/2}^{5\pi/4} \mathbf{F}(\phi(t)) \cdot \phi'(t)\, dt \\
&= \int_{\pi/2}^{5\pi/4} \left(-\frac{t^2 \sin t}{t^2}, \frac{\cos t}{2t}\right) \cdot (1, 2t)\, dt \\
&= \int_{\pi/2}^{5\pi/4} (-\sin t + \cos t)\, dt \\
&= -\sqrt{2} - 1.
\end{aligned}$$

Vector surface integrals. Suppose S is a smooth C^1 oriented surface in $\mathbb{R}^3$ and that $\mathbf{F} : S \to \mathbb{R}^3$ is a continuous vector field on S. We define the *vector surface integral* (or *oriented surface integral*) of $\mathbf{F}$ over S as

$$\iint_S \mathbf{F} \cdot \mathbf{n}\, dS$$

where $\mathbf{n}$ denotes the unit normal vector field across S determined by the given orientation. Analogously to the line integral case, the point is that the orientation determines $\mathbf{n}$, which is then used to turn the vector-valued function $\mathbf{F}$ into a scalar-valued function $\mathbf{F} \cdot \mathbf{n}$, which is then integrated over S using a scalar surface integral. Given a parametrization $\phi : E \to \mathbb{R}^3$ of S such that $\phi_u \times \phi_v$ gives the correct direction for normal vectors, we get

$$\iint_S \mathbf{F} \cdot \mathbf{n} \, dS = \iint_E \mathbf{F}(\phi(u,v)) \cdot \frac{(\phi_u \times \phi_v)}{\|\phi_u \times \phi_v\|} \, \|\phi_u \times \phi_v\| \, d(u,v),$$

which gives the familiar formula

$$\iint_S \mathbf{F} \cdot \mathbf{n} \, dS = \iint_E \mathbf{F}(\phi(u,v)) \cdot (\phi_u \times \phi_v) \, d(u,v)$$

you would have seen in a multivariable calculus course. As with all of these types of integrals, the value is independent of parametrization.

Geometrically, this integral measures the extent to which $\mathbf{F}$ "flows" across S either "with" or "against" the orientation. Indeed, the dot product $\mathbf{F} \cdot \mathbf{n}$ is positive when $\mathbf{F}$ points in the same general direction as $\mathbf{n}$ and is negative when it points "opposite" the direction of $\mathbf{n}$, and the surface integral then adds up these individual quantities.

Example 2. We compute the vector surface integral of $\mathbf{F} = (-4, 0, -x)$ over the portion S of the plane $x + z = 5$ which is enclosed by the cylinder $x^2 + y^2 = 9$, oriented with upward-pointing normal vectors. Parametric equations for S are given by:

$$\phi(r,\theta) = (r\cos\theta, r\sin\theta, 5 - r\cos\theta) \text{ for } (r,\theta) \in [0,3] \times [0,2\pi].$$

This gives

$$\phi_r \times \phi_\theta = (\cos\theta, \sin\theta, -\cos\theta) \times (-r\sin\theta, r\cos\theta, r\sin\theta) = (r, 0, r),$$

which gives the correct orientation since these normal vectors point upwards due to the positive third-component. Thus

$$\begin{aligned}
\iint_S \mathbf{F} \cdot \mathbf{n} \, dS &= \iint_{[0,3]\times[0,2\pi]} \mathbf{F}(\phi(r,\theta)) \cdot (\phi_r \times \phi_\theta) \, d(r,\theta) \\
&= \int_0^{2\pi} \int_0^3 (-4, 0, -r\cos\theta) \cdot (r, 0, r) \, dr \, d\theta \\
&= \int_0^{2\pi} \int_0^3 (-4r - r^2\cos\theta) \, dr \, d\theta \\
&= -36\pi.
\end{aligned}$$

Important. Vector line and surface integrals are both defined by using orientations to turn vector-valued vector fields into scalar-valued functions (via taking dot products with tangent or normal vectors), and then integrating the resulting functions. In terms of parametric equations, we get the formulas we would have seen previously in a multivariable calculus course.

Fundamental Theorem of Line Integrals. We say that a continuous vector field $\mathbf{F} : E \to \mathbb{R}^n$ defined on some region $E \subseteq \mathbb{R}^n$ is *conservative* on E if there exists a C^1 function $f : E \to \mathbb{R}$ such that $\mathbf{F} = \nabla f$. We call such a function f a *potential* function for $\mathbf{F}$.

The value obtained when integrating a conservative field over a curve is given by the following analog of the Fundamental Theorem of Calculus for Line Integrals: if $f : E \to \mathbb{R}$ is C^1 and $C \subseteq E$ is a smooth oriented C^1 arc in E, then

$$\int_C \nabla f \cdot \mathbf{T}\, ds = f(\text{end point of } C) - f(\text{initial point of } C).$$

Indeed, suppose that $\phi : [a, b] \to \mathbb{R}^n$ is a parametrization of C. then

$$\int_C \nabla f \cdot \mathbf{T}\, ds = \int_a^b \nabla f(\phi(t)) \cdot \phi'(t)\, dt.$$

By the chain rule, the integrand in this integral is the derivative of the single-variable function obtained as the composition $f \circ \phi$:

$$(f \circ \phi)'(t) = \nabla f(\phi(t)) \cdot \phi'(t),$$

So by the single-variable Fundamental Theorem of Calculus we have:

$$\int_a^b \nabla f(\phi(t)) \cdot \phi'(t)\, dt = \int_a^b (f \circ \phi)'(t)\, dt = f(\phi(b)) - f(\phi(a)),$$

which gives

$$\int_C \nabla f \cdot \mathbf{T}\, ds = f(\text{end point of } C) - f(\text{initial point of } C)$$

as claimed.

In particular, two immediate consequences are the following facts:

- If C_1, C_2 are two smooth C^1 arcs with the same initial point and the same end point, then $\int_{C_1} \nabla f \cdot \mathbf{T}\, ds = \int_{C_2} \nabla f \cdot \mathbf{T}\, ds$. This property says that line integrals of conservative vector fields are *path-independent* in the sense that the value does only depends on the endpoints of path but not on the specific path chosen to go between those points.
- If C is a smooth closed curve, then $\int_C \nabla f \cdot \mathbf{T}\, ds = 0$.

We will see next time that a field which has either of these properties must in fact be conservative.

Important. Line integrals of conservative fields can be calculated by evaluating a potential function at the end and start points of a curve, and subtracting. This implies that line integrals of conservative fields are path-independent and that the line integral of a conservative field over a closed curve is always zero.

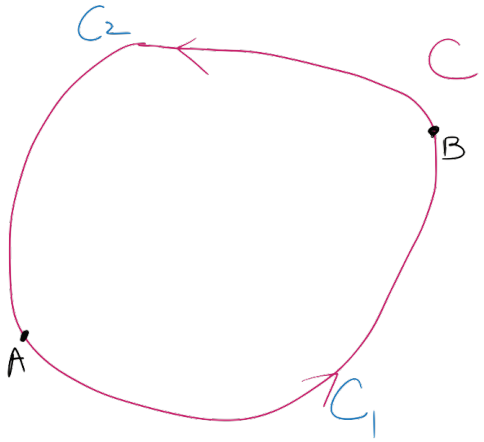
Lecture 22: Green's Theorem

Today we stated and proved *Green's Theorem*, which relates vector line integrals to double integrals. You no doubt saw applications of this in a previous course, and for us the point is understand why it is true and, in particular, the role which the single-variable Fundamental Theorem of Calculus plays in its proof.

Warm-Up. Suppose that $\mathbf{F} : E \to \mathbb{R}^n$ is a continuous vector field on a region $E \subseteq \mathbb{R}^n$. We show that line integrals of $\mathbf{F}$ are path-independent in E if and only if $\mathbf{F}$ has the property that its line

integral over any closed oriented curve in E is zero. Last time we saw that conservative fields have both of these properties, so now we show that these two properties are equivalent to one another for *any* field. (We'll see in a bit that either of these properties in fact implies being conservative.)

Suppose that line integrals of $\mathbf{F}$ in E are path-independent and let C be a closed oriented curve in E. Pick two points A and B on C and denote by C_1 the portion of C which goes from A to B along the given orientation, and C_2 the portion which goes from B back to A along the given orientation:



We use the notation $C_1 + C_2$ for the curve obtained by following C_1 and then C_2, so $C_1 + C_2 = C$. Now, denote by $-C_2$ the curve C_2 traversed with the opposite orientation, so going from A to B. Then C_1 and $-C_2$ both start at A and end at B, so by the path-independence assumption we have

$$\int_{C_1} \mathbf{F} \cdot \mathbf{T}\, ds = \int_{-C_2} \mathbf{F} \cdot \mathbf{T}\, ds.$$

Changing the orientation of a curve changes the sign of the line integral, so

$$\int_{-C_2} \mathbf{F} \cdot \mathbf{T}\, ds = -\int_{C_2} \mathbf{F} \cdot \mathbf{T}\, ds, \text{ and hence } \int_{C_1} \mathbf{F} \cdot \mathbf{T}\, ds + \int_{C_2} \mathbf{F} \cdot \mathbf{T}\, ds = 0.$$

Thus

$$\int_{C} \mathbf{F} \cdot \mathbf{T}\, ds = \int_{C_1+C_2} \mathbf{F} \cdot\, T\, ds = \int_{C_1} \mathbf{F} \cdot \mathbf{T}\, ds + \int_{C_2} \mathbf{F} \cdot\, ds = 0,$$

showing that the line integral of $\mathbf{F}$ over a closed curve is zero as claimed.

Conversely suppose that the line integral of $\mathbf{F}$ over any closed curve in E is zero. Let C_1 and C_2 be two oriented curves in E which start at the same point A and end at the same point B. Then the curve $C_1 + (-C_2)$ obtained by following C_1 from A to B and then C_2 in the reverse direction from B back to A is closed, so our assumption gives

$$\int_{C_1+(-C_2)} \mathbf{F} \cdot \mathbf{T}\, ds = 0.$$

But

$$\int_{C_1+(-C_2)} \mathbf{F} \cdot \mathbf{T}\, ds = \int_{C_1} \mathbf{F} \cdot \mathbf{T}\, ds + \int_{-C_2} \mathbf{F} \cdot \mathbf{T}\, ds = \int_{C_1} \mathbf{F} \cdot \mathbf{T}\, ds - \int_{C_2} \mathbf{F} \cdot \mathbf{T}\, ds,$$

so $\int_{C_1} \mathbf{F} \cdot \mathbf{T}\, ds = \int_{C_2} \mathbf{F} \cdot \mathbf{T}\, ds$, showing that line integrals of $\mathbf{F}$ in E are path-independent.

Path-independence implies conservative. We now show that conservative fields are the only ones with the properties given in the Warm-Up, at least for fields on $\mathbb{R}^2$. To be precise, we show that if $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ is a continuous vector field such that its line integrals in $\mathbb{R}^2$ are path-independent, then $\mathbf{F}$ must be conservative. The same holds for $\mathbb{R}^n$ instead of $\mathbb{R}^2$ using the same proof, only with more components. This result also holds for domains other than $\mathbb{R}^n$, although the proof we give below has to be modified. (We'll point out where the modification comes in.)

Denote the components of $\mathbf{F}$ by $\mathbf{F} = (P, Q)$, which are each continuous. Fix $(x_0, y_0) \in \mathbb{R}^2$ and define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x, y) = \int_{C(x,y)} \mathbf{F} \cdot \mathbf{T}\, ds$$

where $C(x, y)$ is any smooth C^1 path from (x_0, y_0) to (x, y). The path-independence property of $\mathbf{F}$ assures that f is well-defined in that the specific curve C used to connect (x_0, y_0) to (x, y) is irrelevant. The motivation for this definition comes from the Fundamental Theorem of Calculus: if $g : [a, b] \to \mathbb{R}$ is continuous, then the function $G(x) = \int_a^x g(t)\, dt$ obtained by integrating g is differentiable with derivative g.

We claim that f is differentiable (in fact C^1) and has gradient $\nabla f = \mathbf{F}$. First we compute the partial derivative of f with respect to x. For this we choose as a path connecting (x_0, y_0) to (x, y) the one consisting of the vertical line segment L_1 from (x_0, y_0) to (x_0, y), and then the horizontal line segment L_2 from (x_0, y) to (x, y). We have

$$f(x, y) = \int_{L_1} \mathbf{F} \cdot \mathbf{T}\, ds + \int_{L_2} \mathbf{F} \cdot \mathbf{T}\, ds.$$

Parametrizing L_1 with $\phi_1(t) = (x_0, t), y_0 \leq t \leq y$ gives

$$\int_{L_1} (P, Q) \cdot \mathbf{T}\, ds = \int_{y_0}^{y} (P(x_0, t), Q(x_0, t)) \cdot (0, 1)\, dt = \int_{y_0}^{y} Q(x_0, t)\, dt$$

and parametrizing L_2 with $\phi_2(t) = (t, y), x_0 \leq t \leq x$ gives

$$\int_{L_2} (P, Q) \cdot \mathbf{T}\, ds = \int_{x_0}^{x} (P(t, y), Q(t, y)) \cdot (1, 0)\, dt = \int_{x_0}^{x} P(t, y)\, dt,$$

so

$$f(x, y) = \int_{y_0}^{y} Q(x_0, t)\, dt + \int_{x_0}^{x} P(t, y)\, dt.$$

Differentiating with respect to x gives

$$f_x(x, y) = 0 + P(x, y) = P(x, y),$$

where the first term is zero since the first integral is independent of x and the second term is $P(x, y)$ by the Fundamental Theorem of Calculus. (Note the assumption that P is continuous is used here to guarantee that the derivative of the second integral is indeed $P(x, y)$.)

Now, to compute $f_y(x, y)$ we choose a path consisting of the line segment C_1 from (x_0, y_0) to (x, y_0) and the line segment C_2 from (x, y_0) to (x, y). Then after parametrizing these segments we can derive that

$$f(x, y) = \int_{C_1} \mathbf{F} \cdot \mathbf{T}\, ds + \int_{C_2} \mathbf{F} \cdot \mathbf{T}\, ds = \int_{x_0}^{x} P(t, y_0)\, dt + \int_{y_0}^{y} Q(x, t)\, dt.$$

Differentiating with respect to y gives 0 for the first integral and $Q(x, y)$ for the second, so

$$f_y(x, y) = Q(x, y), \text{ and hence } \nabla f = (f_x, f_y) = (P, Q) = \mathbf{F},$$

showing that $\mathbf{F}$ is indeed conservative on $\mathbb{R}^2$.

Important. For a vector field $\mathbf{F} : D \to \mathbb{R}^n$ which is C^1 on an open, connected region $D \subseteq \mathbb{R}^n$, the properties that $\mathbf{F}$ is conservative on D, that $\mathbf{F}$ has path-independent line integrals in D, and that the line integral of $\mathbf{F}$ over any closed curve in D is zero are all equivalent to one another.

Remarks. As mentioned before, the same is true for vector fields on $\mathbb{R}^n$ with the path-independence property: we just repeat the same argument as above when differentiating with respect to other variables, using appropriately chosen paths consisting of various line segments.

Also, the argument we gave works for other domains, as long as the line segments used are always guaranteed to be in those domains. This is true, for instance, for convex domains. More generally, a similar argument works for even more general domains—open connected ones in particular–but modifications are required since the line segments used may no longer be entirely contained in such domains; the way around this is to take paths consisting of multiple short horizontal and vertical segments, where we move from (x_0, y_0) a bit to the right, then a bit up, then a bit to the right, then a bit up, and so on until we reach (x, y). It can be shown that there is a way to do this while always remaining within the given domain. (Such a path is called a *polygonal* path, and it was a problem earlier in the book—which was never assigned—to show that such paths always exist in open connected regions.)

One last thing to note. Let us go back to the expression derived in the proof above when wanting to compute $f_x(x, y)$:

$$f(x, y) = \int_{y_0}^{y} Q(x_0, t)\, dt + \int_{x_0}^{x} P(t, y)\, dt.$$

When computing $f_y(x, y)$ we opted to use another path, but there is no reason why we couldn't try to differentiate this same expression with respect to y instead. The first term is differentiable with respect to y by the Fundamental Theorem of Calculus, and to guarantee that the second term is differentiable with respect to y we assume that P is C^1. Then the technique of "differentiation under the integral sign" (given back at the beginning of Chapter 11 in the book) says that the operations of differentiation and integration can be exchanged:

$$\frac{\partial}{\partial y} \int_{x_0}^{x} P(t, y)\, dt = \int_{x_0}^{x} \frac{\partial P}{\partial y}(t, y)\, dt,$$

so we get

$$f_y(x, y) = Q(x_0, y) + \int_{x_0}^{x} P_y(t, y)\, dt.$$

Since we know $\nabla f = \mathbf{F}$, it must be that this expression equals $Q(x, y)$:

$$Q(x, y) = Q(x_0, y) + \int_{x_0}^{x} P_y(t, y)\, dt.$$

This is indeed true(!), and is a reflection of the fact that $P_y = Q_x$ for a conservative C^1 vector field: $P_y = f_{xy} = f_{yx} = Q_x$ by Clairaut's Theorem, where f being C^2 is equivalent to $\mathbf{F} = \nabla f$ being C^1. Similarly, differentiating the expression

$$f(x, y) = \int_{x_0}^{x} P(t, y_0)\, dt + \int_{y_0}^{y} Q(x, t)\, dt$$

obtained when wanting to compute $f_y(x, y)$ with respect to x gives:

$$P(x, y) = P(x, y_0) + \int_{y_0}^{y} Q_x(x, t)\, dt,$$

which is also a true equality and reflects $P_y = Q_x$ as well.

Green's Theorem. Green's Theorem relates line integrals to double integrals, and is a special case of Stokes' Theorem, which we'll talk about next time. The statement is as follows. Suppose that $D \subseteq \mathbb{R}^2$ is a two-dimensional region whose boundary ∂D is a piecewise, smooth, simple C^1 curve with positive orientation, and that $F = (P, Q) : D \to \mathbb{R}^2$ is a C^1 vector field on D. Then

$$\int_{\partial D} (P, Q) \cdot \mathbf{T}\, ds = \iint_D (Q_x - P_y)\, dA.$$

The "positive" orientation on ∂D is the one where, if you were to walk along ∂D in that direction, the region D would be on your left side. (This a bit of a "hand-wavy" definition, but is good enough for most purposes. Giving a more precise definition of the positive orientation on the boundary would require having a more precise definition of "orientation", which is better left to a full-blown course in differential geometry.)

The proof of Green's Theorem is in the book. There are two key ideas: first, prove Green's Theorem in the special case where D is particular "nice", namely the case where pieces of its boundary can be described as graphs of single-variable functions, and second glue such "nice" regions together to get a more general D. The second step involves using the Implicit Function Theorem to say that smoothness of ∂D implies you can indeed describe portions of ∂D as single-variable graphs. The first step boils down to the observation that the types of expressions

$$P(x, g_1(x)) - P(x, g_2(x)) \quad \text{and} \quad Q(f_2(y), y) - Q(f_1(y), y)$$

you end up with (after using parametric equations) can be written—due the continuity of P and Q—as integrals:

$$P(x, g_1(x)) - P(x, g_2(x)) = \int_{g_1(x)}^{g_2(x)} -P_y(x, t)\, dt \text{ and } Q(f_2(y), y) - Q(f_1(y), y) = \int_{f_1(x)}^{f_2(x)} Q_x(t, y)\, dt$$

as a consequence of the Fundamental Theorem of Calculus. This introduction of an additional integral is what turns the line integral on one side of Green's Theorem into the double integral on the other side, and is where the $Q_x - P_y$ integrand comes from. Check the book for full details. I point this out now since the same idea is involved in the proof of Gauss's Theorem, suggesting that Green's Theorem and Gauss's Theorem are the "same" type of result, which is an idea we'll come back to on the final day of class.

Important. Green's Theorem converts line integrals into double integrals, and its proof boils down to an application of the single-variable Fundamental Theorem of Calculus.

Lecture 23: Stokes' Theorem

Today we spoke about Stokes' Theorem, the next "Big Theorem of Vector Calculus". Stokes' Theorem relates line integrals to surface integrals, and is a generalization of Green's Theorem to curved surfaces; its proof (at least the one we'll outline) uses the two-dimensional Green's Theorem at a key step.

Warm-Up 1. Suppose that $\mathbf{F} = (P, Q) : U \to \mathbb{R}^2$ is a C^1 vector field on an open set $U \subseteq \mathbb{R}^2$. We show that for any $p \in U$:

$$(Q_x - P_y)(p) = \lim_{r \to 0^+} \frac{1}{\pi r^2} \int_{\partial B_r(p)} \mathbf{F} \cdot \mathbf{T}\, ds$$

where $\partial B_r(p)$ is oriented counterclockwise for any $r > 0$. Note that U being open guarantees that for small enough $r > 0$, the circle $\partial B_r(p)$ is contained in U, so that the line integral in question makes sense since $\mathbf{F}$ is defined along points of $\partial B_r(p)$.

The justifies the geometric interpretation of the quantity $Q_x - P_y$ you might have seen in a previous course: it measures the "circulation" of $\mathbf{F}$ around any given point, where counterclockwise circulations are counted as positive and clockwise circulations are negative. Indeed, the line integral

$$\int_{\partial B_r(p)} \mathbf{F} \cdot \mathbf{T}\, ds$$

measures the circulation of $\mathbf{F}$ around the circle of radius r around p, and so the limit—where we take this circle getting smaller and smaller and closing in on p—is interpreted as the "circulation" around p itself. This is a special case of the geometric interpretation of the *curl* of a vector field, which we'll come to after stating Stokes' Theorem.

By Green's Theorem, we have

$$\frac{1}{\pi r^2} \int_{\partial B_r(p)} \mathbf{F} \cdot \mathbf{T}\, ds = \frac{1}{\text{Vol}(B_r(p))} \iint_{B_r(p)} (Q_x(x, y) - P_y(x, y))\, d(x, y).$$

Since $\mathbf{F}$ is C^1, $Q_x - P_y$ is continuous so a problem from a previous homework shows that

$$\lim_{r \to 0^+} \frac{1}{\text{Vol}(B_r(p))} \iint_{B_r(p)} (Q_x(x, y) - P_y(x, y))\, d(x, y) = (Q_x - P_y)(p),$$

which then gives the desired equality

$$(Q_x - P_y)(p) = \lim_{r \to 0^+} \frac{1}{\text{Vol}(B_r(p))} \iint_{B_r(p)} (Q_x(x, y) - P_y(x, y))\, d(x, y) = \lim_{r \to 0^+} \frac{1}{\pi r^2} \int_{\partial B_r(p)} \mathbf{F} \cdot \mathbf{T}\, ds.$$

Warm-Up 2. Let $\mathbf{F} = (P, Q)$ be the vector field

$$\mathbf{F}(x, y) = \left(-\frac{y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right).$$

We show that $\mathbf{F}$ is conservative on the set $\mathbb{R}^2 \setminus \{\text{nonnegative } x\text{-axis}\}$ obtained by removing the origin and the positive x-axis from $\mathbb{R}^2$.

Let C be any simple smooth closed curve in $\mathbb{R}^2 \setminus \{\text{nonnegative } x\text{-axis}\}$ and let D be the region it encloses, so that $C = \partial D$. Then $\mathbf{F}$ is C^1 on D and so Green's Theorem gives

$$\int_C \mathbf{F} \cdot \mathbf{T}\, ds = \pm \iint_D (Q_x - P_y)\, dA$$

where the $\pm$ sign depends on the orientation of C. For this field we have

$$Q_x = \frac{y^2 - x^2}{x^2 + y^2} = P_y,$$

so $Q_x - P_y = 0$ and hence $\iint_D (Q_x - P_y)\, dA = 0$. Thus the line integral of $\mathbf{F}$ over any simple smooth closed curve in $\mathbb{R}^2 \setminus \{\text{nonnegative } x\text{-axis}\}$ is zero, and so by an equivalence we established last time, this implies that $\mathbf{F}$ is conservative on $\mathbb{R}^2 \setminus \{\text{nonnegative } x\text{-axis}\}$.

Observation. Now, knowing that the field above $\mathbf{F}$ is conservative on $\mathbb{R}^2 \setminus \{\text{nonnegative } x\text{-axis}\}$, we can try to find a potential function for $\mathbf{F}$ over this region, which is a C^2 function $f : \mathbb{R}^2 \setminus \{\text{nonnegative } x\text{-axis}\} \to \mathbb{R}$ satisfying $\nabla f = \mathbf{F}$. This is not something you'd be expected to able to do on the final, and I'm just mentioning this in order to illustrate an important property of this specific vector field.

To start with, a direct computation involving derivatives of the arctangent function will show that:

$$\nabla \left(\tan^{-1} \frac{y}{x} \right) = \mathbf{F}.$$

However, $f(x, y) = \tan^{-1} \left(\frac{y}{x} \right)$ is not defined for $x = 0$, so this does not give us a potential function over all of $\mathbb{R}^2 \setminus \{\text{nonnegative } x\text{-axis}\}$ yet. For now, we view this as defining our sought-after potential only in the first quadrant. The idea is now to extend this function to the second, third, and fourth quadrants in a way so that we have $\nabla f = \mathbf{F}$ at each step along the way.

Another direction computation shows that the following is also true:

$$\nabla \left(-\tan^{-1} \frac{x}{y} \right) = \mathbf{F}.$$

This is good since this new potential function is defined for $x = 0$, although it is no longer defined for $y = 0$. The point is that we will only try to use this expression to define the sought-after potential over the second quadrant and on the positive y-axis, on which the potential we used above over the first quadrant was not defined. However, we have to do this in a way which guarantees the potential we're defining (by piecing together local potentials) over the first and second quadrants is in fact C^1 even along the positive y-axis. For (x, y) in the first quadrant, $\frac{y}{x} > 0$ so as $x \to 0$ we have $\frac{y}{x} \to +\infty$, and thus $\tan^{-1}(\frac{y}{x}) \to \frac{\pi}{2}$. The function $-\tan^{-1}(\frac{x}{y})$ has value 0 when $x = 0$, so the function

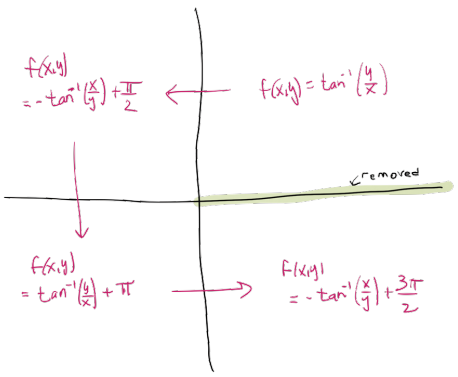
$$-\tan^{-1} \left(\frac{x}{y} \right) + \frac{\pi}{2}$$

has value $\frac{\pi}{2}$ on the positive y-axis. Thus the function defined by

$$\tan^{-1} \left(\frac{y}{x} \right) \text{ on the first quadrant and } -\tan^{-1} \left(\frac{x}{y} \right) + \frac{\pi}{2} \text{ on the second quadrant}$$

will indeed be C^1 even on the positive y-axis. Since adding a constant to a function does not alter its gradient, this piece over the second quadrant still has gradient equal to $\mathbf{F}$, so now we have a C^1 function on the upper-half plane which has gradient $\mathbf{F}$.

We continue in this way, using the two arctangent expressions above to define the sought-after potential over the remaining quadrants, where we add the necessary constants to ensure that we still get C^1 functions along way as we "jump" from one quadrant to another:



For (x, y) in the second quadrant, $\frac{x}{y} < 0$ so as $y \to 0$ (i.e. as we approach the x-axis), we have $\frac{x}{y} \to -\infty$ and thus $-\tan^{-1}\left(\frac{x}{y}\right) + \frac{\pi}{2} \to -(-\frac{\pi}{2}) + \frac{\pi}{2} = \pi$. Thus defining the potential function over the third quadrant and negative x-axis to be

$$\tan^{-1}\left(\frac{y}{x}\right) + \pi$$

maintains the C^1 condition and gradient equals $\mathbf{F}$ condition. For (x, y) in the third quadrant, $\frac{y}{x} > 0$ to as $x \to 0$ we get $\tan^{-1}(\frac{y}{x}) + \pi \to \frac{3\pi}{2}$, so we define our potential to be

$$-\tan^{-1}\left(\frac{x}{y}\right) + \frac{3\pi}{2}$$

over the fourth quadrant and on the negative y-axis. The conclusion is that the function $f : \mathbb{R}^2 \setminus \{\text{nonnegative } x\text{-axis}\} \to \mathbb{R}$ defined by:

$$f(x,y) = \begin{cases} \tan^{-1}\left(\frac{y}{x}\right) & x > 0, y > 0 \\ -\tan^{-1}\left(\frac{x}{y}\right) + \frac{\pi}{2} & x \leq 0, y > 0 \\ \tan^{-1}\left(\frac{y}{x}\right) + \pi & x < 0, y \leq 0 \\ -\tan^{-1}\left(\frac{x}{y}\right) + \frac{3\pi}{2} & x \geq 0, y < 0 \end{cases}$$

is C^1 and satisfies $\nabla f = \mathbf{F}$, so f is a potential function for the conservative field $\mathbf{F}$ over the region $\mathbb{R}^2 \setminus \{\text{nonnegative } x\text{-axis}\}$.

Now, note what happens if we take the same expression above, only we now consider it as a function defined on $\mathbb{R}^2 \setminus \{(0,0\}$, so including the positive x-axis. This seems plausible since the portion $\tan^{-1}\left(\frac{y}{x}\right)$ defining f over the first quadrant is in fact defined when $y = 0$, so we could change the first case in the definition above to hold for $x > 0, y \geq 0$, thereby giving a well-defined function on $\mathbb{R}^2 \setminus \{(0,0)\}$. However, if we take the limit of the portion of f defined in the third quadrant as (x, y) approaches the positive x-axis from below we would get a value of 2π, which means that in order to preserve continuous the portion defined over the first quadrant would have to be

$$\tan^{-1}\left(\frac{y}{x}\right) + 2\pi$$

instead of simply $\tan^{-1}(\frac{y}{x})$. Since we're off by an additional 2π, this says that there is no way to extend the definition of f above to the positive x-axis so that it remains continuous, let alone C^1,

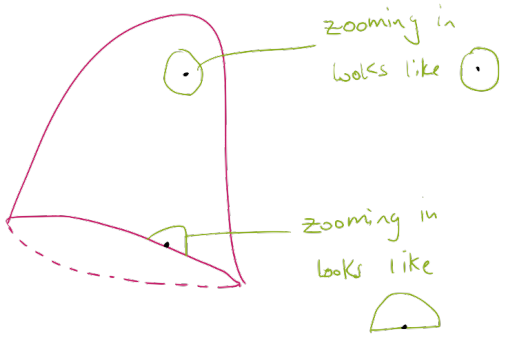
which suggests that $\mathbf{F} = \left(-\frac{y}{x^2+2}, \frac{x}{x^2+y^2}\right)$ should actually NOT be conservative over the punctured plane $\mathbb{R}^2 \setminus \{(0,0)\}$. This is in fact true, as we can see from the fact that

$$\int_{\substack{\text{unit circle,} \\ \text{counterclockwise}}} \mathbf{F} \cdot \mathbf{T}\, ds = 2\pi$$

The fact that $\mathbf{F}$ is conservative over $\mathbb{R}^2 \setminus \{\text{nonnegative } x\text{-axis}\}$ but not $\mathbb{R}^2 \setminus \{(0,0)\}$ is crucial in many applications of vector calculus, and in particular is related to the fact in complex analysis that there is no version of the complex log function which is differentiable on the set of all nonzero complex numbers; if you want a differentiable log function in complex analysis, you must delete half of an axis or more generally a ray emanating from the origin in $\mathbb{C}$. For this given field $\mathbf{F}$, the line integral over any closed simple curve can only have one of three values:

$$\int_C \left(-\frac{y}{x^2+y^2}, \frac{x}{x^2+y^2}\right) \cdot \mathbf{T}\, ds = \begin{cases} 2\pi & \text{if } C \text{ encircles the origin counterclockwise} \\ -2\pi & \text{if } C \text{ encircles the origin clockwise} \\ 0 & \text{if } C \text{ does not encircle the origin.} \end{cases}$$

Manifold boundary. Before talking about Stokes' Theorem, we must clarify a new use of the word "boundary" we will see. Given a surface S, we define its *manifold boundary* as follows. Given a point in S, if we zoom in on S near this point we will see one of two things: either S near this point will look like an entire disk, or S near this point will look like a half-disk:



The points near which S looks like a half-disk are called the *manifold boundary* points of S, and the manifold boundary ∂S of S is the curve consisting of the manifold boundary points. Intuitively, the manifold boundary is the curve describing where there is an "opening" into S. A *closed* surface is one which has no manifold boundary; for instance, a sphere is a closed surface. Notationally, we denote manifold boundaries using the same ∂ symbol as we had for topological boundaries, and which type of boundary we mean should be clear from context.

To distinguish this from our previous notion of boundary, we might refer to the previous notion as the *topological boundary* of a set. In general, the manifold and topological boundaries of a surface are quite different; they agree only for a surface fully contained in the xy-plane viewed as a subset of $\mathbb{R}^2$. Similar notions of manifold boundary can be given for objects other than surfaces: the "manifold boundary" of a curve will be the set of its endpoints, and the "manifold boundary" of a three-dimensional solid in $\mathbb{R}^3$ is actually the same as its topological boundary.

An orientation on S induces a corresponding orientation on ∂S, which we call the *positive* orientation on ∂S. The simplest way to describe this orientation is likely what you saw in a previous course: if you stand on ∂S with your head pointing in the direction of the normal vector

determined by the orientation on S, the positive orientation on ∂S corresponds to the direction you have to walk in in order to have S be on your "left" side. It is possible to give a more precise definition of this positive orientation, but this would require, as usual, a more precise definition of orientation in terms of linear algebra and differential geometry.

Stokes' Theorem. Suppose that S is a piecewise smooth oriented C^2 surface whose boundary ∂S is a piecewise smooth C^1 curve, oriented positively. Stokes' Theorem says that if $\mathbf{F} = (P, Q, R) : S \to \mathbb{R}^3$ is a C^1 vector field, then

$$\int_{\partial S} \mathbf{F} \cdot \mathbf{T}\, ds = \iint_S \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS$$

where

$$\operatorname{curl} \mathbf{F} = (R_y - Q_z, P_z - R_x, Q_x - P_y)$$

is the *curl* of $\mathbf{F}$, which is a continuous vector field $\operatorname{curl} \mathbf{F} : S \to \mathbb{R}^3$. Thus, Stokes' Theorem relates line integrals on one side to surface integrals of curls on the other.

We note that Green's Theorem is the special case where S is a surface fully contained in the xy-plane. Indeed, here the normal vector $\mathbf{n}$ is $\mathbf{k} = (0, 0, 1)$ and

$$\operatorname{curl} \mathbf{F} \cdot \mathbf{n} = Q_x - P_y,$$

so the surface integral in Stokes' Theorem becomes simply $\iint_S (Q_x - P_y)\, dA$ as in Green's Theorem. One may ask: if Green's Theorem is a consequence of Stokes' Theorem, why prove Green's Theorem first instead of simply proving Stokes' Theorem and deriving Green's Theorem from it? The answer, as we'll see, is that the proof of Stokes' Theorem uses Green's Theorem in a crucial way.

Geometric meaning of curl. Suppose that $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a C^1 vector field, so that $\operatorname{curl} \mathbf{F}$ is defined, and let $\mathbf{n}$ be a unit vector in $\mathbb{R}^3$. Fix $p \in \mathbb{R}^3$ and let $D_r(p)$ be the disk of radius r centered at p in the plane which is orthogonal to $\mathbf{n}$ at p. Then we have

$$\int_{\partial D_r(p)} \mathbf{F} \cdot \mathbf{T}\, ds = \iint_{D_r(p)} \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS$$

and

$$\lim_{r \to 0^+} \frac{1}{\pi r^2} \iint_{D_r(p)} \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS = \operatorname{curl} \mathbf{F}(p) \cdot \mathbf{n}$$

since the function $q \mapsto \operatorname{curl} \mathbf{F}(q) \cdot \mathbf{n}$ is continuous. Thus we get

$$\operatorname{curl} \mathbf{F}(p) \cdot \mathbf{n} = \lim_{r \to 0^+} \frac{1}{\pi r^2} \int_{\partial D_r(p)} \mathbf{F} \cdot \mathbf{T}\, ds,$$

giving us the interpretation the $\operatorname{curl} \mathbf{F}(p)$ measures the "circulation" of $\mathbf{F}$ around the point p; in particular, $\operatorname{curl} \mathbf{F}(p) \cdot \mathbf{n}$ measures the amount of this circulation which occurs on the plane orthogonal to $\mathbf{n}$. This is the geometric interpretation of curl you might have seen in a previous calculus course, and the point is that it follows from Stokes' Theorem and the homework problem which tells us how to compute the limit above.

Proof of Stokes' Theorem. The proof of Stokes' Theorem is in the book, and the basic strategy is as follows. As in the proof of Green's Theorem, we look at a special case first and then piece

together these special surfaces to build up the general statement; the special case in this case is that of a surface S given by the graph $z = f(x, y)$ of a C^2 function.

Let E be the portion of the xy-plane lying below S, and let $(x(t), y(t)), t \in [a, b]$ be parametric equations for ∂E. Then

$$\phi(t) = (x(t), y(t), f(x(t), y(t))), t \in [a, b]$$

are parametric equations for ∂S. Using this we can express the left-hand side of Stokes' Theorem as

$$\int_{\partial S} \mathbf{F} \cdot \mathbf{T}\, ds = \int_a^b \mathbf{F}(\phi(t)) \cdot \phi'(t)\, dt.$$

After computing $\phi'(t)$ and writing out this dot product, the resulting expression can be written as another dot product of the form:

$$(\text{some two-dimensional field}) \cdot (x'(t), y'(t)),$$

which is the type expression you get in two-dimensional lines integrals. Indeed, this will write the line integral over ∂S as a line integral over ∂E instead:

$$\int_{\partial S} \mathbf{F} \cdot \mathbf{T}\, ds = \int_{\partial E} (\text{some two-dimensional field}) \cdot \mathbf{T}\, ds.$$

To this we can now apply Green's theorem, which will express this line integral over ∂E as a double integral over E:

$$\int_{\partial E} (\text{some two-dimensional field}) \cdot \mathbf{T}\, ds = \iint_E (\text{some function})\, d(x, y).$$

Finally, the integrand in this double integral can be written as a dot product which looks like:

$$(R_y - Q_z, P_z - R_x, Q_x - P_y) \cdot (\text{normal vector to } S),$$

so that the double integral describes what you get when you use the parametrization

$$\psi(x, y) = (x, y, f(x, y))\, (x, y) \in E$$

to compute the surface integral of $\operatorname{curl} \mathbf{F}$ over S:

$$\iint_E (\text{some function})\, d(x, y) = \iint_S \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS.$$

To recap, the process is: write the line integral of $\mathbf{F}$ over ∂S as a line integral over ∂E instead, apply Green's Theorem to get a double integral over E, rewrite this double integral as the surface integral of $\operatorname{curl} \mathbf{F}$ over S.

Check the book for full details. The tricky thing is getting all the computations correct. For instance, in the first step you have to compute $\phi'(t)$ for

$$\phi(t) = (x(t), y(t), f(x(t), y(t))),$$

where differentiating the third component $z(t) = f(x(t), y(t))$ here requires a chain rule:

$$z'(t) = f_x x'(t) + f_y y'(t).$$

Thus

$$(P, Q, R) \cdot (x'(t), y'(t), f_x x'(t) + f_y y'(t)) = (P + f_x, Q + f_y) \cdot (x'(t), y'(t)),$$

so that the "some two-dimensional field" referred to in the outline above is the field

$$(P + RRf_x, Q + f_y).$$

This is the field to which Green's Theorem is applied, in which we need to know:

$$\frac{\partial}{\partial x}(Q + Rf_y) - \frac{\partial}{\partial y}(P + Rf_x).$$

Each of these derivatives again require chain rules since Q and R are dependent on x in two ways:

$$Q = Q(x, y, f(x, y)) \qquad R = R(x, y, f(x, y))$$

and P and R are dependent on y in two ways:

$$P = P(x, y, f(x, y)) \qquad R = R(x, y, f(x, y)).$$

We have:

$$\frac{\partial}{\partial x}Q(x, y, f(x, y)) = Q_x + Q_z f_x \quad \text{and} \quad \frac{\partial}{\partial y}P(x, y, f(x, y)) = P_y + P_z f_y,$$

and similar expressions for the derivatives of R. This all together gives the "some function" referred to in the outline above. Note that in this step there will be some simplifications in the resulting expression using the fact that $f_{yx} = f_{xy}$ since f is a C^2 function.

We'll stop here and leave everything else to the book, but notice how the curl of $\mathbf{F}$ naturally pops out of this derivation, in particular in the "some function" in the outline. Indeed, it was through this derivation that the curl was first discovered: historically, it's not as if someone randomly wrote down the expression

$$(R_y - Q_z, P_z - R_x, Q_x - P_y)$$

and then wondered what interesting properties this might have, but rather this expression showed up in the proof of Stokes' Theorem and only after that was it identified as an object worth of independent study. The name "curl" historically came from the geometric interpretation in terms of circulation we gave earlier.

Important. Stokes' Theorem relates the surface integral of the curl of a field over a surface to the line integral of the "uncurled" field over the (manifold) boundary of that surface. The power of Stokes' Theorem in practice comes from relating one-dimensional objects (line integrals) to two-dimensional objects (surface integrals), and in giving geometric meaning to curl $\mathbf{F}$.

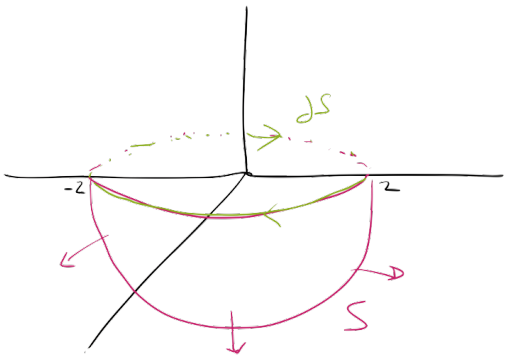
Lecture 24: Gauss's Theorem

Today we spoke about Gauss's Theorem, the last of the "Big Theorems of Vector Calculus". Gauss's Theorem relates surface integrals and triple integrals, and should be viewed in the same vein as the Fundamental Theorem of Calculus, Green's Theorem, and Stokes' Theorem as a theorem which tells us how an integral over the boundary of an object can be expressed as an integral over the entire object itself.

Warm-Up 1. We verify Stokes' Theorem (i.e. compute both integrals in Stokes' Theorem to see that they are equal) for the vector field

$$\mathbf{F}(x, y, z) = (2y - z, x + y^2 - z, 4y - 3x)$$

and the portion S of the sphere $x^2 + y^2 + z^2 = 4$ where $z \leq 0$, with outward orientation.



First we compute

$$\iint_S \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS.$$

We have

$$\operatorname{curl} \mathbf{F} = (5, 2, -1) \cdot \mathbf{n}\, dS.$$

At a point (x, y, z) on S, the vector going from the origin to (x, y, z) itself is normal to S, so $\mathbf{n} = \frac{1}{2}(x, y, z)$ is a unit normal vector to the S at (x, y, z). Thus

$$\iint_S \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS = \iint_S \frac{1}{2}(5x + 2y - z)\, dS.$$

The integral of $5x + 2y$ over S is zero due to symmetry (the integrand is odd with respect to a variable the surface is symmetric with respect to), so

$$\iint_S \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS = -\frac{1}{2} \iint_S z\, dS.$$

Parametrizing S using

$$\psi(\phi, \theta) = (2 \sin\phi \cos\theta, 2 \sin\phi \sin\theta, 2 \cos\phi),\ (\phi, \theta) \in [\frac{\pi}{2}, \pi] \times [0, 2\pi],$$

we have

$$\begin{aligned} \psi_\phi \times \psi_\theta &= (2 \cos\phi \cos\theta, 2 \cos\phi \sin\theta, -2 \sin\phi) \times (-2 \sin\phi \sin\theta, 2 \sin\phi \cos\theta, 0) \\ &= (4 \sin^2\phi \cos\theta, 4 \sin^2\phi \sin\theta, 4 \sin\phi \cos\phi), \end{aligned}$$

so $\|\psi_\phi \times \psi_\theta\| = 4 \sin\phi$. Hence

$$-\frac{1}{2} \iint_S z\, dS = -\frac{1}{2} \int_0^{2\pi} \int_{\pi/2}^{\pi} (2 \cos\phi)(4 \sin\phi)\, d\phi\, d\theta = -\frac{1}{2} \int_0^{2\pi} -4\, d\theta = 4\pi$$

is the value of $\iint_S \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS$.

Now we compute $\int_{\partial S} \mathbf{F} \cdot \mathbf{T}\, ds$. The (manifold) boundary of S is the circle of radius 2 in the xy-plane with clockwise orientation. We parametrize ∂S using

$$\phi(t) = (2 \cos t, -2 \sin t, 0),\ 0 \le t \le 2\pi,$$

we have

$$\int_{\partial S} \mathbf{F} \cdot \mathbf{T}\, ds = \int_0^{2\pi} \mathbf{F}(\phi(t)) \cdot \phi'(t)\, dt$$

$$
\begin{aligned}
&= \int_0^{2\pi} (-4\sin t, 2\cos t + 4\sin^2 t, 8\sin t - 6\cos t) \cdot (-2\sin t, -2\cos t, 0)\, dt \\
&= \int_0^{2\pi} (8\sin^2 t - 4\cos^2 t - 8\sin^2 \cos t)\, dt \\
&= 4\pi
\end{aligned}
$$

after using $\sin^2 t = \frac{1}{2}(1 - \cos 2t)$ and $\cos^2 t = \frac{1}{2}(1 + \cos 2t)$. Thus $\int_{\partial S} \mathbf{F} \cdot \mathbf{T}\, ds = \iint_S \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS$ as claimed by Stokes' Theorem.

Warm-Up 2. Suppose that $\mathbf{G}$ is a continuous vector field on $\mathbb{R}^3$. We say that surface integrals of $\mathbf{G}$ are *surface-independent* if $\iint_{S_1} \mathbf{G} \cdot \mathbf{n}\, dS = \iint_{S_2} \mathbf{G} \cdot \mathbf{n}\, dS$ for any oriented smooth surfaces S_1 and S_2 with the same boundary and which induce the same orientation on their common boundary. We show that surface integrals of $\mathbf{G}$ are surface-independent if and only if the surface integral of $\mathbf{G}$ over any closed surface is zero.

Before doing so, here's why we care. First, this is the surface integral analog of the fact for line integrals we showed in a previous Warm-Up that line integrals of a field $\mathbf{G}$ are path-independent if and only if the line integral of $\mathbf{G}$ over any closed curve is zero. Second, fields of the form $\mathbf{G} = \operatorname{curl} \mathbf{F}$ where $\mathbf{F}$ is a C^1 field always have these properties as a consequence of Stokes' Theorem. Indeed, in the setup above, Stokes' Theorem gives

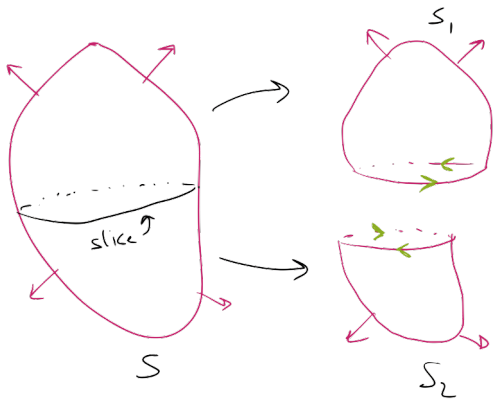
$$\iint_{S_1} \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS = \int_{\partial S_1 = \partial S_2} \mathbf{F} \cdot \mathbf{T}\, ds = \iint_{S_2} \operatorname{curl} \mathbf{F} \cdot \mathbf{n}, dS,$$

so surface integrals of $\operatorname{curl} \mathbf{F}$ are surface-independent. Also, if S is closed, then $\partial S = \emptyset$ so

$$\iint_S \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS = \int_{\partial S} \mathbf{F} \cdot \mathbf{T}\, ds = 0$$

since the latter takes place over a region of volume zero, empty in fact. This is one of the many reasons why curls play a similar role in surface integral theory that conservative fields do in line integral theory. (We'll see a deeper reason why this analogy holds next time when we talk about differential forms.)

Suppose that surface integrals of $\mathbf{G}$ are surface-independent and let S be a closed oriented surface. Slice through S to create two surfaces S_1 and S_2 with the same boundary, which is the curve where the "slicing" occurred:



Note that the given orientations on S_1 and S_2 actually induce opposite orientations on their common boundary. Thus S_1 and $-S_2$ induce the same orientation on their common boundary, so surface-independent gives

$$\iint_{S_1} \mathbf{G} \cdot \mathbf{n}\, dS = \iint_{-S_2} \mathbf{G} \cdot \mathbf{n}\, dS.$$

Since $\iint_{-S_2} \mathbf{G} \cdot \mathbf{n}\, dS = -\iint_{S_2} \mathbf{G} \cdot \mathbf{n}\, dS$, this in turned gives

$$\iint_{S=S_1+S_2} \mathbf{G} \cdot \mathbf{n}\, dS = \iint_{S_1} \mathbf{G} \cdot \mathbf{n}\, dS + \iint_{S_2} \mathbf{G} \cdot \mathbf{n}\, dS = 0$$

as desired.

Conversely suppose that the surface integral of $\mathbf{G}$ over any closed oriented surface is zero, and let S_1 and S_2 be two oriented surfaces with the same boundary and which induce the same orientation on their common boundary. Glue S_1 and $-S_2$ along their common boundary to get a closed surface S. Since we switched the orientation on S_2, this closed surface $S = S_1 \cup (-S_2)$ has a consistent orientation, and so is itself oriented. Our assumption gives

$$\iint_{S_1 \cup (-S_2)} \mathbf{G} \cdot \mathbf{n}\, dS = 0,$$

so

$$\iint_{S_1} \mathbf{G} \cdot \mathbf{n}\, dS - \iint_{S_2} \mathbf{G} \cdot \mathbf{n}\, dS = 0,$$

which in turns gives surface-independence.

Gauss's Theorem. Suppose that $E \subseteq R^3$ is a three-dimensional solid region whose boundary ∂E is a piecewise smooth C^1 surface, and let $\mathbf{F} = (P, Q, R) : E \to \mathbb{R}^3$ be a C^1 vector field on E. Gauss's Theorem (also called the *Divergence Theorem*) says that

$$\iint_{\partial E} \mathbf{F} \cdot \mathbf{n}\, dS = \iiint_E \operatorname{div} \mathbf{F}\, dV$$

where $\operatorname{div} \mathbf{F} = P_x + Q_y + R_z$ and where we give ∂S the outward orientation, meaning the orientation consisting of normal vectors which point away from E. Thus, surface integrals over a closed surface can be expressed as triple integrals over the region enclosed by that surface.

Geometric meaning of divergence. Before talking about the proof of Gauss's Theorem, we give one application, which justifies the geometric meaning behind divergence you likely saw in a previous course. The claim is that for a C^1 field $\mathbf{F}$ on $\mathbb{R}^3$ and a point $p \in \mathbb{R}^3$,

$$\operatorname{div} \mathbf{F}(p) = \lim_{r \to 0^+} \frac{1}{\operatorname{Vol}(B_r(p))} \iint_{\partial B_r(p)} \mathbf{F} \cdot \mathbf{n}\, dS.$$

Here, $\partial B_r(p)$ is the sphere of radius r centered at p. The integral on the right measures the flow of $\mathbf{F}$ across this sphere, and so this equality says that $\operatorname{div} \mathbf{F}(p)$ measures the "infinitesimal" flow of $\mathbf{F}$ at p, where $\operatorname{div} \mathbf{F}(p) > 0$ means that there is a net flow "away" from p while $\operatorname{div} \mathbf{F}(p) < 0$ means there is a net flow "towards" p.

By Gauss's Theorem we have

$$\iint_{\partial B_r(p)} \mathbf{F} \cdot \mathbf{n}\, dS = \iiint_{B_r(p)} \operatorname{div} \mathbf{F}\, dV.$$

Thus

$$\lim_{r\to 0^+} \frac{1}{\text{Vol}(B_r(p))} \iint_{\partial B_r(p)} \mathbf{F}\cdot\mathbf{n}\,dS = \lim_{r\to 0^+} \frac{1}{\text{Vol}(B_r(p))} \iiint_{B_r(p)} \operatorname{div}\mathbf{F}\,dV = \operatorname{div}\mathbf{F}(p)$$

since $\operatorname{div}\mathbf{F}$ is a continuous function. This gives the required equality.

Proof of Gauss's Theorem. The proof of Gauss's Theorem is in the book, and follows the same strategy as the proof of Green's Theorem or Stokes' Theorem: prove a special case where ∂E is given by the graphs of C^1 functions, and then "glue" these special surfaces together.

Assuming that E is the region between the graphs of two C^1 functions:

$$z = g_2(x,y) \quad \text{and} \quad z = g_2(x,y),$$

the boundary of E is then given by these two graphs. After setting up parametric equations for each of these and using these to compute

$$\iint_{\partial S} (0,0,R)\cdot\mathbf{n}\,dS,$$

at some point we get an expression of the form

$$R(x,y,g_2(x,y)) - R(x,y,g_1(x,y)),$$

which, due to the continuity of R_z, we can write as

$$R(x,y,g_2(x,y)) - R(x,y,g_1(x,y)) = \int_{g_1(x,y)}^{g_2(x,y)} R_z(x,y,t)\,dt$$

using the Fundamental Theorem of Calculus. This is the same type of thing we did in the proof of Green's Theorem, and is the step which transforms the two-dimensional surface integral on one side of Gauss's Theorem to the three-dimensional triple integral on the other side. This is also the step that explains where the R_z term in the divergence comes from. Doing the same for the field $(P,0,0)$ gives the P_x term and $(0,Q,0)$ gives the Q_y term.

Check the book for full details, and next time we will see that these similarities in the proofs of Gauss's and Green's Theorem are no accident, but are a suggestion of a deeper connection between the two.

Important. Gauss's Theorem relates the surface integral of a field over the boundary of some three-dimensional solid to the ordinary triple integral of the divergence of that field over the entire solid. Thus, as with the other "Big" theorems, it relates the behavior of an object over a boundary to the behavior of a type of derivative of that object over the region enclosed by that boundary.

Lecture 25: Differential Forms

Today we spoke about differential forms, which provides a framework which unifies basically all the material we've seen these past few weeks. In particular, it shows that the Big Theorems of Vector Calculus all reflect the same idea by showing they are consequences of a single theorem phrased in terms of differential forms. This material will not be on the final, and is purely meant to give a sense as to what vector calculus is "really" all about.

Warm-Up. A C^2 function $u(x, y, z)$ is called *harmonic* over a region if $u_{xx} + u_{yy} + u_{zz} = 0$ on that region. The expression $u_{xx} + u_{yy} + u_{zz}$ is called the *Laplacian* of u and is usually denoted by Δu, so harmonic functions are ones which have zero Laplacian.

Suppose that E is a closed Jordan region in $\mathbb{R}^3$, and that u is a harmonic function on E which is zero on ∂E. We show that $u = 0$ everywhere on E. To do so, we apply Gauss's Theorem to the C^1 vector field $u\nabla u$:

$$\iint_{\partial E} u\nabla u \cdot \mathbf{n}\, dS = \iiint_E \operatorname{div}(u\nabla u)\, dV.$$

On the one hand, since $u = 0$ on ∂E, $u\nabla u = 0$ on ∂E so the surface integral on the left is zero. On the other hand, we compute:

$$\operatorname{div}(u\nabla u) = \operatorname{div}(uu_x, uu_y, uu_z) = u_x u_x + uu_{xx} + u_y u_y + uu_{yy} + u_z u_z + uu_{zz} = \|\nabla u\|^2 + u\Delta u.$$

Thus

$$\iiint_E (\|\nabla u\|^2 + u\Delta u)\, dV = 0.$$

Since u is harmonic, $\Delta u = 0$ so the integral above becomes

$$\iiint_E \|\nabla u\|^2 \, dV = 0.$$

Since $\|\nabla u\|^2$ is a continuous nonnegative expression, the only way it can have integral zero is if $\|\nabla u\|^2 = 0$ on E. This gives $\nabla u = 0$ on E, so u is constant on E. Since $u = 0$ on $\partial E \subseteq E$, the constant which u equals must be zero, so $u = 0$ on E as claimed.

Point behind the Warm-Up. The Warm-Up shows that if a harmonic function is zero over the boundary of some region, it must be zero everywhere on that region. This in turn implies that the behavior of a harmonic function over a boundary determines its behavior everywhere.

Thinking back to last quarter, we saw a similar type of result for *analytic functions*: if f is analytic on $\mathbb{R}$ and is zero on some interval (a, b), then f is zero on all of $\mathbb{R}$. Thus, the behavior of an analytic function on a small interval determines its behavior everywhere. (We called this the *Identity Theorem* last quarter.)

The fact that harmonic functions and analytic functions both have this type of property is no accident: it is a reflection of a deep fact in *complex analysis*. Recall (from the 10 minute introduction to complex analysis I gave at one point last quarter) that, in the complex setting, differentiable functions are automatically analytic. Any complex function $f : \mathbb{C} \to \mathbb{C}$ can be written as $f = u + iv$ where u and v are real-valued, and the basic fact is the following: if $f = u + iv$ is complex analytic (i.e. differentiable), then u and v must be harmonic! Thus, harmonic functions play a significant role in complex analysis as well as real analysis, and the result of the Warm-Up is a glimpse of this. Note however that the proof we gave involved no complex analysis, and was a direct application of Gauss's Theorem.

Differential forms. Differential forms give a unified approach towards describing all types of integrals you've ever done in your lives. In particular, if you've ever wondered what the "dx" in the notation for a single-variable integral actually means, or what the "$dx\, dy$" and "$dx\, dy\, dx$" in double and triple integrals mean, the answer is given by the language of differential forms. We'll give definitions which are goood enough for our purposes, but rest assured that everything we'll do can be given very precise definitions.

A *differential* 0*-form* on $\mathbb{R}^3$ is simply a smooth (i.e. C^∞) function on $\mathbb{R}^3$. A *differential* 1*-form* on $\mathbb{R}^3$ is an expression of the form

$$P\,dx + Q\,dy + R\,dz$$

where each of P, Q, R are smooth functions on $\mathbb{R}^3$. You might object that we're defining a 1-form in this way without saying what dx, dy, or dz mean, but again we are only giving definitions which are good enough for what we want to do here. If you want the precise definition of a 1-form, it is the following: a differential 1-form on $\mathbb{R}^3$ is a smooth section of the cotangent bundle $T^*\mathbb{R}^3$ of $\mathbb{R}^3$ over $\mathbb{R}^3$. Of course, none of this will make any sense unless you've had a serious differential geometry course, and we won't elaborate on this further. Again, the definition we gave is suitable for our purposes, even though more rigorous approaches can be given.

A *differential* 2*-form* on $\mathbb{R}^3$ is an expression of the form

$$A\,dx\,dy + B\,dy\,dz + C\,dz\,dx$$

where A, B, C are smooth functions. At this point we can talk about the "algebra" underlying the theory of differential forms. In particular, one might ask why there is no $dx\,dx$ term in the expression above, or why there is no $dx\,dz$ term explicitly given? The answer is that differential forms obey the following algebraic rules:

- $dx_i\,dx_j = -dx_j\,dx_i$ for any coordinates x_i, x_j, and
- $dx_k\,dx_k = 0$ for any coordinate x_k.

The second condition follows from the first since taking $x_i = x_j$ gives $dx_i\,dx_i = -dx_i\,dx_i$, which implies $dx_i\,dx_i = 0$, but it is worth mentioning the second property on its own. So, $dx\,dx$, $dy\,dy$, and $dz\,dz$ are all zero, which is why these terms don't show up, and terms with $dy\,dx$, $dz\,dy$, or $dx\,dz$ in them can be written by flipping orders and putting in a negative to be of the types given in the 2-form expression above. Why do differential forms have these algebraic properties? The answer, again, depends on the more formal definition of differential forms.

A *differential* 3*-form* on $\mathbb{R}^3$ is an expression of the form

$$F\,dx\,dy\,dz$$

where F is a smooth function. Any expression such as $dy\,dx\,dz$ or with some other rearrangement of x, y, z can be written as one which involves $dx\,dy\,dz$ using the algebraic properties above, and any expression such as $dx\,dy\,dy$ or something which involves two of the same coordinate will be zero. There are no nonzero 4-forms (or higher-order forms) on $\mathbb{R}^3$ since any expression such as $dx\,dy\,dz\,dx$ will necessarily repeat a coordinate since there are only three coordinates to choose from, and so will be zero. In general, $\mathbb{R}^n$ will have nonzero differential k-forms only for $k = 0, 1, \ldots, n$.

Pullbacks of differential forms. We can easily determine what happens to a differential form under a change of variables using the algebraic properties given above and the interpretation of df where f is a function as a *differential*. For instance, suppose we want to see what $dx\,dy$ becomes in polar coordinates $x = r\cos\theta, y = r\sin\theta$. Then

$$dx\,dy = d(r\cos\theta)\,d(r\sin\theta) = (\cos\theta\,dr - r\sin\theta\,d\theta)(\sin\theta\,d\theta + r\cos\theta\,d\theta),$$

where $d(r\cos\theta)$ and $d(r\sin\theta)$ are computed according to the definition of a differential as:

$$df = \frac{\partial f}{\partial x_1}\,dx_1 + \cdots + \frac{\partial f}{\partial x_n}\,dx_n$$

for variables $x_1, \dots, x_n$. When we multiply out the resulting expression, the $dr\,dr$ and $d\theta\,d\theta$ terms are zero, so we get:

$$(\cos\theta\,dr - r\sin\theta\,d\theta)(\sin\theta\,d\theta + r\cos\theta\,d\theta) = r\cos^2\theta\,dr\,d\theta - r\sin^2\theta\,d\theta\,dr.$$

Using $d\theta\,dr = -dr\,d\theta$, this final expression can be simplifed to

$$r(\cos^2\theta + \sin^2\theta)\,dr\,d\theta = r\,dr\,d\theta,$$

so we get that

$$dx\,dy = r\,dr\,d\theta.$$

The fact that $dx\,dy$ becomes the usual expression you get when converting to polar coordinates in double integrals is no accident, and is a reflection of the fact that differential forms give a very convenient way to express the change of variables integration formula in general. If you convert the 3-form $dx\,dy\,dz$ into spherical coordinates using the same procedure as above, you will indeed get

$$dx\,dy\,dz = \rho^2\sin\phi\,d\rho\,d\phi\,d\theta,$$

as expected when rewriting triple integrals in spherical coordinates.

To make this precise, think of the polar change of coordinates as coming from the change of variables function

$$\phi(r,\theta) = (r\cos\theta, r\sin\theta).$$

The calculation we went through above for $dx\,dy$ defines what is called the *pullback* of $dx\,dy$ by ϕ and is denoted by $\phi^*(dx\,dy)$, so we showed that

$$\phi^*(dx\,dy) = r\,dr\,d\theta.$$

The term "pullback" comes from the fact that we started with a differential form on the "target' side of ϕ and "pulled it back" to rewrite it as a form on the "domain" side: ***FINISH***Pullbacks in general are computed in a similar way, where we substitute in for our coordinates the expressions they equal under the given change of variables, and use differentials and the algebraic properties of differential forms to rewrite the result. Under the spherical change of variables

$$\psi(\rho,\phi,\theta) = (\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)$$

the pullback of $dx\,dy\,dz$ is

$$\psi^*(dx\,dy\,dz) = \rho^2\sin\phi\,d\rho\,d\phi\,d\theta.$$

In general, pulling back a differential form will give an expression involving the Jacobian determinant of the change of variables, which is to say that differential forms give a convenient way of encoding these Jacobians. With this notation, the change of variables formula for integration becomes:

$$\int_E \phi^*\omega = \int_{\phi(E)} \omega$$

where $\phi : E \to \mathbb{R}^n$ is the change of variables function, ω is a differential form on $\phi(E)$, and, as mentioned already, the pullback $\phi^*\omega$ encodes the Jacobian of this transformation. Notice how "pretty" this formula is as compared to the usual way of writing out the change of variables formula!

Integrating differential forms. We can now talk about what it means to integrate a differential form, which is where the connection to vector calculus and vector fields starts to show up. In general, k-forms are integrated over k-dimensional objects.

To start with, the integral of a 1-form $P\,dx + Q\,dy + R\,dz$ over a curve $C \subseteq \mathbb{R}^3$ is denoted by

$$\int_C P\,dx + Q\,dy + R\,dz.$$

If this notation seems familiar, it is because you probably saw it in a previous course as an alternate notation for vector line integrals; in particular, the above integral is *defined* to be the integral of the vector field (P, Q, R) over C:

$$\int_C P\,dx + Q\,dy + R\,dz = \int_C (P, Q, R) \cdot \mathbf{T}\,ds.$$

To be clear, if we rewrite the integrand on the left-hand side as a "dot product"

$$P\,dx + Q\,dy + R\,dz = (P, Q, R) \cdot (dx, dy, dz),$$

then (dx, dy, dz) represents the tangent vector along C: if $(x(t), y(t), z(t))$ are parametric equations for C, then

$$(dx, dy, dz) = (x'(t), y'(t), z'(t))\,dt$$

and so $(P, Q, R) \cdot (dx, dy, dz)$ gives the usual expression for the integrand you get when rewriting $\int_C (P, Q, R) \cdot \mathbf{T}\,ds$ using parametric equations. Thus, from this point of view, the differential form $P\,dx + Q\,dy + R\,dz$ is a convenient way to express $(P, Q, R) \cdot$ (tangent vector).

Similarly, integrals of 2-forms simply give an alternate way of expressing vector surface integrals. Given a 2-form $P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy$, its integral over a surface is defined to be the vector surface integral of the vector field (P, Q, R) over that surface:

$$\int_S P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy = \iint_S (P, Q, R) \cdot \mathbf{n}\,dS.$$

Writing the integrand on the left as a dot product

$$(P, Q, R) \cdot (dy\,dz, dz\,dx, dx\,dy),$$

the point is that $(dy\,dz, dz\,dx, dx\,dy)$ is meant to represent normal vectors to S. Indeed, given parametric equations $(x(u, v), y(u, v), z(u, v))$ for S, we have:

$$dy\,dz = d(y(u, v))\,d(z(u, v)) = (y_u\,du + y_v\,dv)(z_u\,du + z_v\,dv) = (y_u z_v - z_u y_v)\,du\,dv,$$

and a similar computation gives

$$dz\,dx = (z_u x_v - x_u z_v)\,du\,dv \quad \text{and} \quad dx\,dy = (x_u y_v - x_v y_u)\,du\,dv).$$

Thus

$$(dy\,dz, dz\,dx, dx\,dy) = (y_u z_v - z_u y_v, z_u x_v - x_u z_v, x_u y_v - x_v y_u)\,du\,dv,$$

and the point is that the vector $(y_u z_v - z_u y_v, z_u x_v - x_u z_v, x_u y_v - x_v y_u)$ is precisely what you get when you compute the normal vector determined by the given parametric equations. Hence the 2-form expression

$$P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy$$

produces the integrand you get when writing a vector surface integral in terms of parametric equations, so that integrals of 2-forms are simply vector surface integrals as claimed.

An integral of a 3-form $f\,dx\,dy\,dz$ over a three-dimensional solid in $\mathbb{R}^3$ is just an ordinary triple integral, and an integral of a 2-form $f\,dx\,dy$ on $\mathbb{R}^2$ over a two-dimensional region in $\mathbb{R}^2$ is an ordinary double integral. We'll say what it means to integrate a 0-form later.

Exterior derivatives. Before talking about how the language of differential forms unifies the Big Theorems of Vector Calculus, we need to define one more operation on differential forms: the *exterior derivative* $d\omega$ of a differential form ω. As with forms in general, this can be given a fully precise definition which explain "what" exterior differentiatoin actually means, but here we'll give definitions which are good enough for us.

For a 0-form f, df is simply the ordinary differential of f, which we already used above:

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy + \frac{\partial f}{\partial z}\,dz.$$

Thinking of the right-hand side as characterizing the vector field (f_x, f_y, f_z) (which gives the same value as the form when integrating over a curve), df is just a way to represent the ordinary gradient field ∇f of f.

The exterior derivative of a 1-form is defined by applying the previous definition on 0-forms to each component function of the 1-form; for instance:

$$d(P\,dx) = dP\,dx = (P_x\,dx + P_y\,dy + P_z\,dz)\,dx = P_z\,dz\,dx - P_y\,dx\,dy$$

since $dx\,dx = 0$. In general we have:

$$\begin{aligned} d(P\,dx + Q\,dy + R\,dz)) &= dP\,dx + dQ\,dy + dR\,dz \\ &= (P_y\,dy + P_z\,dz)\,dx + (Q_x\,dx + Q_z\,dz)\,dy + (R_x\,dx + R_y\,dy)\,dz \\ &= (R_y - Q_z)\,dy\,dz + (P_z - R_x)\,dz\,dx + (Q_x - P_y)\,dz. \end{aligned}$$

The components of the resulting 2-form should look familiar, as they are precisely the components of $\operatorname{curl}(P, Q, R)$! The point is that if we identity the 1-form $\omega = P\,dx + Q\,dy + R\,dz$ with the vector field (P, Q, R), then $d\omega$ is the 2-form which characterizes $\operatorname{curl}(P, Q, R)$ in the sense that integrating *omega* and $\operatorname{curl}(P, Q, R)$ over a surface gives the same value. Thus, the exterior derivative of a 1-form encodes the curl operation.

Finally, if $\omega = A\,dy\,dz + B\,dz\,dx + C\,dx\,dy$, then its exterior derivative is

$$d\omega = dA\,dy\,dz + dB\,dz\,dx + dC\,dx\,dy.$$

Since $dx_i\,dx_i = 0$ for any coordinate, only the dx term from dA, the dy term from dB, and the dz term from dC will give nonzero contributions to the expression above, so:

$$d\omega = A_x\,dx\,dy\,dz + B_y\,dy\,dz\,dx + C_z\,dz\,dx\,dy = (A_x + B_y + C_z)\,dx\,dy\,dz$$

where in the second equality we use $dx_i\,dx_j = -dx_j\,dx_i$ to say that

$$dx\,dy\,dz = dy\,dz\,dx = dx\,dy\,dz.$$

The resulting coefficient of $dx\,dy\,dz$ is precisely the divergence of the field (A, B, C), and so the exterior of a 2-form encodes divergences.

To summarize, all three main "differentiation" operations in vector calculus—gradient, curl, divergence—can be described in one shot using exterior differentiation of differential forms: the

derivative of a 0-form gives the gradient, the derivative of a 1-form gives the curl, and the derivative of a 2-form gives the divergence.

Generalized Stokes' Theorem. We can finally state the single fact which explains all of the Big Theorems of Vector Caclulus in a unified way. To distinguish this result from what we have previously called Stokes' Theorem, often this new result is referred to as the *Generalized Stokes' Theorem*. We'll give the statement in full generality in any dimension.

One last notion we need is that of a *smooth manifold*, which is meant to be some sort of geometric object suitable for performing integration over it. We won't give the definition here, but will simply note that a 0-dimensional manifold is simply a collection of points, a 1-dimensional manifold is a smooth curve, a 2-dimensional manifold is a smooth surface, and a 3-dimensional manifold is a three-dimensional solid. In general, an n-dimensional manifold will be some sort of n-dimensional geometric object.

Here, then, is the *Generalized Stokes' Theorem*:

> Let M be an n-dimensional oriented smooth manifold whose boundary ∂M has the induced orientation and let ω be a differential $(n-1)$-form on M. Then
>
> $$\int_{\partial M} \omega = \int_M d\omega.$$

And that's it! Note that the dimensions/orders in the integrals match up: on the left we are integrating an $(n-1)$-form over an $(n-1)$-dimensional object, and on the right an n-form over an n-dimensional object.

Now using the interpretations we derived previously in terms of gradient, curl, and divergence, we can easily see how the Generalized Stokes' Theorem encodes all vector calculus theorems. When ω is a 0-form f and $\dim M = 1$ so that $M = C$ is a curve, this gives the Fundamental Theorem of Line Integrals:

$$f(\text{end point}) - f(\text{start point}) = \int_C \nabla f \cdot \mathbf{T}\, ds$$

where the left side is taken to be the definition of the "integral" of f over the finite set ∂C consisting of the end point and start point of C. When ω is a 1-form characterizing a vector field $\mathbf{F}$ and $\dim M = 2$ so that $M = S$ is a surface, this gives Stokes' Theorem (of which Green's Theorem is a special case):

$$\int_{\partial S} \mathbf{F} \cdot \mathbf{T}\, ds = \iint_S \operatorname{curl} \mathbf{F} \cdot \mathbf{n}\, dS.$$

When ω is a 2-form characterizing a vector field $\mathbf{F}$ and $\dim M = 3$ so that $M = E$ is a solid, this gives Gauss's Theorem:

$$\iint_{\partial E} \mathbf{F} \cdot \mathbf{n}\, dS = \iiint_E \operatorname{div} \mathbf{F}\, dV.$$

Thus all of the Big Theorems express the same idea, namely that integrating a form over the boundary of an object relates to the integral of its derivative over the entire object, and the only differences come in the dimensions to which this general fact is applied.

One final thing to comment on: the proof of the Generalized Stokes' Theorem works in the same way as the proof of Green's or Gauss's Theorem, in that you prove some special cases and then glue. When proving the special cases, you start computing $\int_{\partial M} \omega$ using parametric equations,

and at some point you get a difference which you can rewrite as an integral using the single-variable Fundamental Theorem of Calculus:

$$(\text{something evaluated at } g_2(x_i)) - (\text{something evaluated at } g_1(x_i)) = \int_{g_1(x_i)}^{g_2(x_i)} \frac{\partial(\text{something})}{\partial x_i}\, dx_i.$$

As usual, this is the step which transforms the original $(n-1)$-dimensional integral into an n-dimensional one, and explains why the exterior derivative $d\omega$ shows up. ***FINISH***

Important.