Introduction

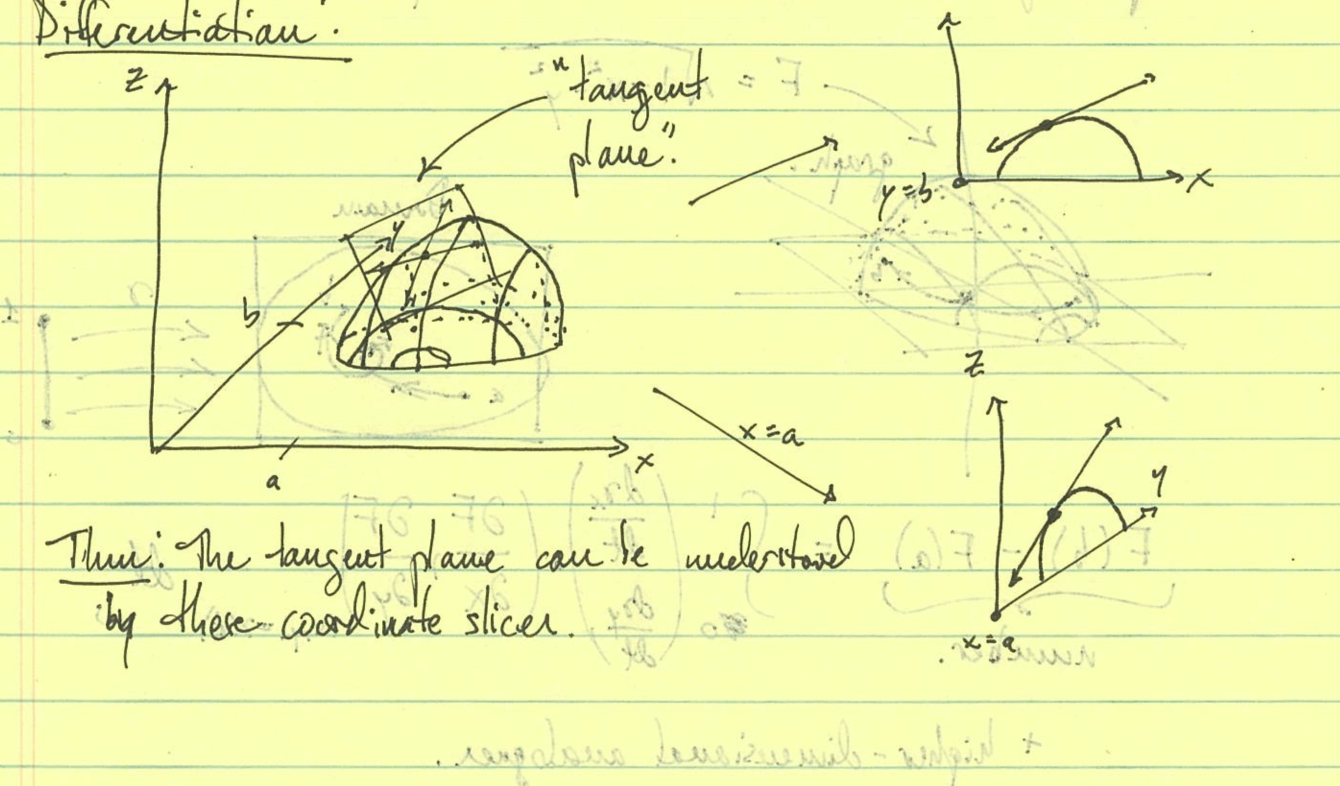
Our goal this semester is to explore the geometry of functions of many variables (and many outputs). In the single-variable case, this is called calculus, which consists of three points:

- differentiation (local approximation by standard functions),
- integration (average value of a function on some domain)
- their interaction (the FTC: $\frac{d}{dx}\int_a^x f(t)\,dt = f(x)$, $\int_a^b f(x)\,dx = F(b) - F(a)$.)

These are simplified because of the simple geometry of $\mathbb{R}$: there aren't many 1-dim^{l} shapes, and $\mathbb{R}$ itself is ordered: either $a \leq b$ or $a > b$.

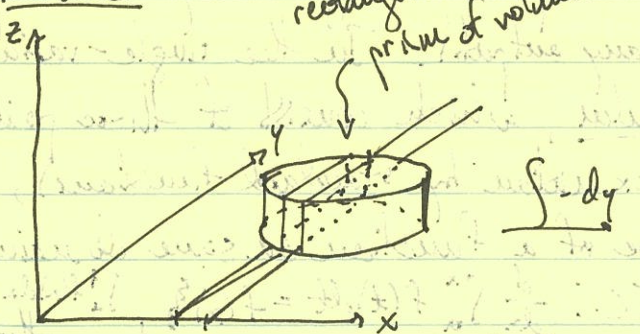
In more dimensions, essentially everything becomes more complicated.

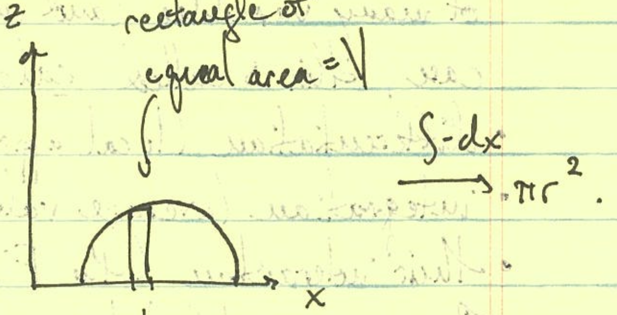
Differentiation:



Thus: The tangent plane can be understood by these coordinate slices.

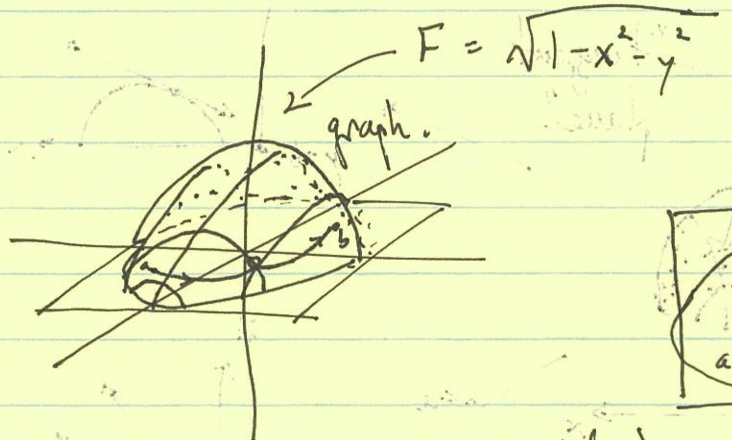
Integration:

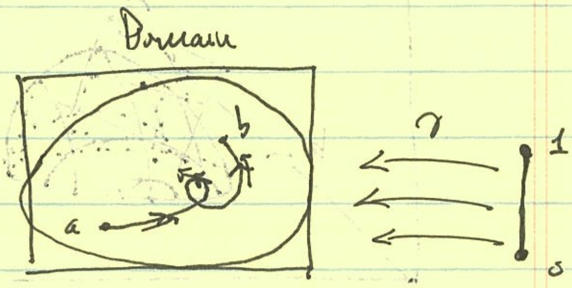




Thm: The method of slicing works.

Interaction: This is the least obvious one, since the types do not seem to line up at all: differentiation was giving us planes, but integration can only swallow normal $f^{\underline{ns}}$ + give back numbers, and picking different domains is not as easy as "$\int_a^x$".





$$\underbrace{F(b)-F(a)}_{\text{number.}} = \int_0^1 \begin{pmatrix} \frac{d\gamma_x}{dt} \\ \frac{d\gamma_y}{dt} \end{pmatrix} \cdot \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}\right)_{(x,y)=\gamma(t)} dt,$$

+ higher-dimensional analogues.

We have our work cut out for us.

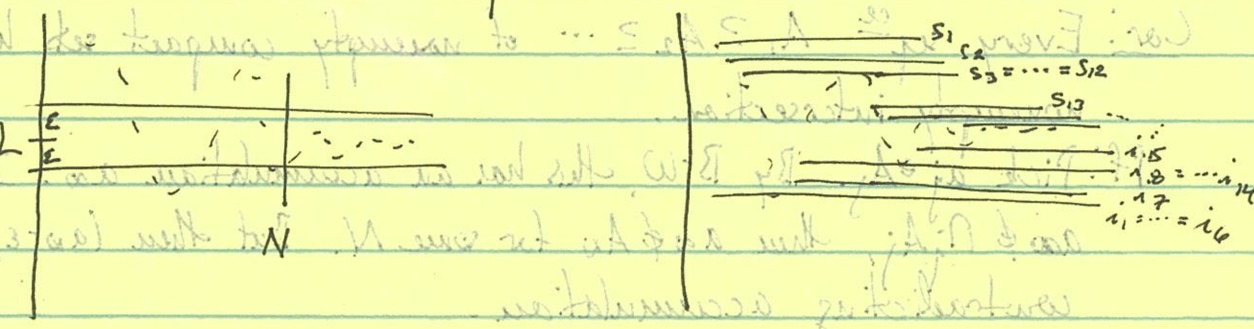
Topological properties of $\mathbb{R}$ (1.2)

Thm/Def: The real numbers are the (unique) complete ordered Archimedean field.

"total"

- **Ordered:** There is a "$<$" compatible with $+, -, *, /$.
- **Complete:** This admits many reformulations. One is: every subset has a unique greatest lower bound "inf" and least upper bound "sup". Here's another: every Cauchy sequence ($\forall \varepsilon > 0\ \exists N$ s.t. $\forall m > N$, $|x_N - x_m| < \varepsilon$) converges to a limit ($\exists! L\ \forall \varepsilon > 0\ \exists N$ s.t. $\forall m > N$, $|x_m - L| < \varepsilon$), ~~$L = \limsup x_n = \liminf x_n$~~).
- **Archimedean:** for any $x \in \mathbb{R}_{\geq 0}$ there's an $n \in \mathbb{N}$ with $n \leq x \leq n+1$.

Here's another way to think about convergence: $s_n = \sup\{x_m \mid m \geq n\}$ and $i_n = \inf\{x_m \mid m \geq n\}$ are decreasing/increasing respectively. Set $\limsup(x_n) = \inf\{s_n\}$ and $\liminf(x_n) = \sup\{i_n\}$. Then L is the limit of x_n iff $\limsup x_n = \liminf x_n = L$.



The order gives an important tool for breaking $\mathbb{R}$ up into regions.

Def: A set $U \subseteq \mathbb{R}$ is open if for all $u \in U$ there is an $\varepsilon > 0$ with $(u-\varepsilon, u+\varepsilon) \subseteq U$. ("$U$ has no edges.") A set $A \subseteq \mathbb{R}$ is closed if its complement is open. ("U has all its edges, or limit points.")

Def: An <u>open cover</u> $\{U_\alpha\}$ of $X \subseteq \mathbb{R}$ is a set of open subsets $U_\alpha \subseteq \mathbb{R}$ such that $X \subseteq \bigcup_\alpha U_\alpha$. X is <u>compact</u> if out of <u>any</u> open cover of X, only finitely many of the U_α are needed ("finite subcover").

<u>Thm</u>: (Heine-Borel): The closed interval $[a,b] = X$ is compact. (In fact, every compact is the finite union of such.)

Pf: Let $\mathcal{O}$ be an open cover and let $A = \{x \in X \mid [a,x] \text{ belongs to a finite subcover}\}$.

① Set $s = \sup A$. In fact, $s \in A$: s belongs to some $U \in \mathcal{O}$, there is some $x \in A \cap U \cap [a,s]$, $[a,x]$ is finitely covered and $[x,s] \subseteq U$.

② Also, $s = b$: if $s < b$, then add any $U \in \mathcal{O}$ containing s, since $(s-\varepsilon, s+\varepsilon) \subseteq U$. □

<u>Thm</u> (Bolzano-Weierstrass): Every bdd seq^{ce} has a convergent subseq^{ce}.

Pf: Subdivide the bounding interval into successive halves containing ∞'^y many seq^{ce} terms. The remaining points is an accumulation point. □

<u>Cor</u>: Every seq^{ce} $A_1 \supseteq A_2 \supseteq \cdots$ of nonempty compact sets has nonempty intersection.

Pf: Pick $a_j \in A_j$. By B-W, this has an accumulation a_∞. If $a_\infty \notin \bigcap_j A_j$, then $a_\infty \notin A_N$ for some N. But then $(a_\infty - \varepsilon, a_\infty + \varepsilon) \subseteq A_N^c$, contradicting accumulation. □

<u>Rem</u>: Most of these <u>definitions</u> can be generalized to metric spaces: e.g., open <u>balls</u> are $B_\varepsilon(x) = \{y \mid d(x,y) < \varepsilon\}$, and a set is <u>open</u> if each point is contained in an open ball.

Continuous functions (1.3)

We are interested in functions $f: A \to \mathbb{R}^m$ for $A \subseteq \mathbb{R}^n$.
These are built out of output components: $f(a) = (f_1(a), \dots, f_m(a))$.

<u>Def</u>: $\lim_{x \to a} f(x) = b$ means $\forall \varepsilon > 0 \; \exists \delta > 0$ s.t. $0 < |x - a| < \delta$ implies $|f(x) - b| < \varepsilon$.

<u>Or</u>: $\lim_{x \to a} f(x) = b$ iff for every $x_n \to a$, $f(x_n) \to b$.

f is said to be <u>continuous at a</u> if $\lim_{x \to a} f(x) = f(a)$, and f is merely <u>continuous</u> if it is continuous at every $a \in A$.

Continuity admits a reformulation without limits. It's the first of many instances where a <u>global property</u> takes on a very different guise.

<u>Lem</u>: $f: A \to \mathbb{R}^m$ is continuous iff $\forall \, \mathcal{U} \subseteq \mathbb{R}^m$ open, $f^{-1}(\mathcal{U}) \subseteq A$ is <u>relatively open</u>: there is an open $V \subseteq \mathbb{R}^n$ with $f^{-1}(\mathcal{U}) = V \cap A$.

Pf: Given $\mathcal{U}$, let $a \in A$ lie in $f^{-1}(\mathcal{U})$. Then $B_\varepsilon(f(a)) \subseteq \mathcal{U}$ for some ε, hence there is a $B_\delta(a) \cap A \subseteq f^{-1}(\mathcal{U})$. Taking the union over a gives V with $V \cap A = f^{-1}(\mathcal{U})$. Conversely, for $B_\varepsilon(fa) \subseteq \mathbb{R}^m$, $f^{-1}(B_\varepsilon(fa)) \subseteq A$ is open, so $B_\delta(a) \subseteq f^{-1}(B_\varepsilon(fa))$. $\square$

This makes some properties of continuous f^ns super clear.

<u>Lem</u>: For $X \subseteq A$ compact, $f(X)$ is compact.

Pf: A cover $\mathcal{U}_*$ of $f(X)$ preimages to a cover $f^{-1}(\mathcal{U}_*)$ of X. Then, use compactness of X to reduce to a finite subcover. $\square$

Similarly, connected $\Rightarrow$ connected image. The extreme value theorem follows.

The original analytic definition is good for measuring the <u>failure</u> of continuity.

no ≥ 0!!.

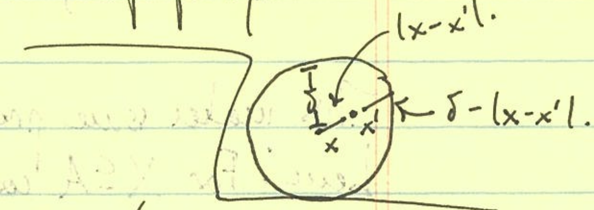
Def: Fixing $f: A \to \mathbb{R}$ and $a \in A$, set $M(\delta) = \sup\{f(x) \mid |x-a| < \delta\}$ and $m(\delta) = \inf\{ \text{—''—} \}$. The <u>oscillation</u> is $\lim_{\delta \to 0} (M(\delta) - m(\delta))$.

Lem: f is continuous at a iff its oscillation is zero.

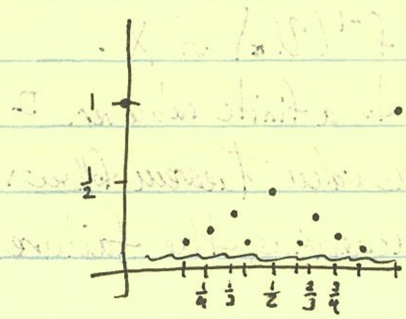
Pf: If f is continuous, then for any ε_0 we pick δ_c such that $0 < |x-a| < \delta_c$ implies $|f(x) - f(a)| < \varepsilon_c = \frac{1}{3}\varepsilon_0$. Then $|(M(\delta_c) - m(\delta_c)) - 0| \leq \frac{2}{3}\varepsilon_0 < \varepsilon_0$, hence we can take $\delta_0 = \delta_c$ to show the limit. The converse is similar: pick δ_0 so that $M(\delta_0) - m(\delta_0) < \varepsilon_0 = \varepsilon_c$. Then $|f(x) - f(a)| < \varepsilon_0 = \varepsilon_c$, since $m(\delta_0) \leq f(x), f(a) \leq M(\delta_0)$. □

Lem: If $A \subseteq \mathbb{R}^n$ is <u>closed</u> and $f: A \to \mathbb{R}$ is bounded, then $B_\varepsilon = \{x \in A \mid \text{oscillation of } f \text{ at } x \text{ is} \geq \varepsilon\}$ is closed as well.

Pf: We'll show $\mathbb{R}^n \setminus B_\varepsilon$ is open. If $x \in \mathbb{R}^n \setminus B_\varepsilon$ has $x \notin A$, then there is an open ball around x on which f is not defined (which we exclude from the "$\geq \varepsilon$" condition). Otherwise, if $x \in \mathbb{R}^n \setminus B_\varepsilon$ and $x \in A$, then there is a δ with $M(\delta) - m(\delta) < \varepsilon$ at x. For x' with $|x - x'| < \delta$, $\delta' = \delta - |x - x'|$ also has this property. □



Ex: $f(x) = \begin{cases} \text{if } x = \frac{p}{q} \text{ reduced, then } \frac{1}{q}, \\ 0 \text{ otherwise.} \end{cases}$



This is cts at irrationals, (= oscillation 0)
<u>not</u> cts at rationals, (= oscillation $1/q$)
and $B_{1/q}$ = (rat^{ls} with denom$^s \geq q$).

Differentiation (2.1)

Remember that a f^n $f: \mathbb{R} \to \mathbb{R}$ is called differentiable at a (with derivative $f'(a)$) when $\lim_{h\to 0} \frac{f(a+h)-f(a)}{h} = f'(a)$. Later on, this is modified to say that f and g "agree to n^{th} order" when $\lim_{h\to 0} \frac{f(a+h)-g(a+h)}{h^n} = 0$, and in the case of the derivative, f and $g(a+h) = f(a) + h\cdot f'(a)$ agree to 1^{st} order.

$g(x) = f(a) + (x-a)f'(a)$.

Our definition in the case of $f: \mathbb{R}^n \to \mathbb{R}^m$, ~~$A \subseteq \mathbb{R}^n$~~, will be similar: f is differentiable at $a \in A$ (with derivative a linear transformation L) when

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

Rem: The type signature of L is worth considering: it consumes displacement vectors h off of a + outputs displacement vectors off of $f(a)$.

Rem: Often L is written as $D_a f$ (or, in the book, as $Df(a)$).

Lem: When the derivative exists, it is unique.

Pf: Take K to be another linear transformation satisfying the limit property $\lim_{h\to 0} \frac{f(a+h) - (f(a) + K(h))}{|h|} = 0$. Then

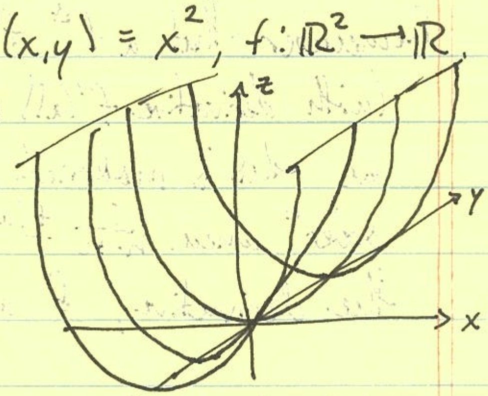
$$\lim_{h\to 0} \frac{K(h) - L(h)}{|h|} = \lim_{h\to 0} \frac{K(h) + f(a) - f(a+h) + f(a+h) - (f(a) + L(h))}{|h|} = 0.$$

Now pick $x \in \mathbb{R}^n$ and set $h = t\cdot x$, $t \to 0$. Then

$$0 = \lim_{t\to 0^+} \frac{K(tx) - L(tx)}{|tx|} = \lim_{t\to 0^+} \frac{K(x) - L(x)}{|x|} = \frac{K(x)-L(x)}{|x|}. \quad \Box$$

Ex: Consider a simple function like $f(x,y) = x^2$, $f: \mathbb{R}^2 \to \mathbb{R}$.
We claim $D_{a,b} f$ is the linear transform

$$D_{(a,b)} f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2a \cdot x \\ \cancel{0} \end{pmatrix} = \underbrace{\begin{pmatrix} 2a & 0 \\ \cancel{0} & \cancel{0} \end{pmatrix}}_{\text{"Jacobian"}} \begin{pmatrix} x \\ y \end{pmatrix}.$$



Check: $$\lim_{h \to 0} \frac{\left| f\left(\begin{pmatrix} a \\ b \end{pmatrix} + h\right) - \left(f\begin{pmatrix} a \\ b \end{pmatrix} + D_{a,b} f\begin{pmatrix} h \end{pmatrix}\right)\right|}{|h|} = \lim_{h \to 0} \frac{(a + h_x)^2 - (a^2 + 2ah_x)}{\sqrt{h_x^2 + h_y^2}}$$

$$= \lim_{h \to 0} \frac{h_x^2}{\sqrt{h_x^2 + h_y^2}}.$$ Since $\lim_{h \to 0} \frac{h_x^2}{\sqrt{h_x^2}} = 0$ and $\sqrt{h_x^2 + h_y^2} \geq |h_x|$, this $\to 0$ too.

More generally, f will be called _differentiable_ if it's differentiable at all $a \in A$, its domain; it's differentiable _at $a \in A$_ if it can be defined on an open neighborhood of a and is differentiable under that definition.

Our goals for the next while will amount to understanding how to compute $D_a f$ efficiently, analogous to the next steps in single-variable calculus.

Properties of the derivative

The most important property of the derivative (+ the one least clearly expressed in ordinary calculus texts) is the chain rule:

Thm: Take $f:\mathbb{R}^n \to \mathbb{R}^m$ diff^{ble} at $a \in \mathbb{R}^n$ and $g:\mathbb{R}^m \to \mathbb{R}^l$ diff^{ble} at $f(a)=b$. Then $g\circ f$ is diff^{ble} at a with

$$D_a(g\circ f) = (D_{f(a)}g)\circ(D_a f).$$

Pf: We introduce a lot of notation:

obvious error-terms, non-obvious

$$\begin{cases} \varepsilon_f(x) = (f(x)-f(a)) - (D_a f)(x-a), \\ \varepsilon_g(y) = (g(y)-g(b)) - (D_b g)(y-b), \\ \varepsilon_{gf}(x) = (gf(x)-gf(a)) - (D_b g)(D_a f)(x-a). \end{cases}$$

By def^n we have $\lim_{x\to a} \frac{|\varepsilon_f(x)|}{|x-a|} = 0$, $\lim_{y\to b} \frac{|\varepsilon_g(x)|}{|y-b|} = 0$, and we want to conclude $\lim_{x\to a} \frac{|\varepsilon_{fg}(x)|}{|x-a|} = 0$.

$$\begin{aligned} \varepsilon_{gf} &= (gf(x)-gf(a)) - (D_b g)(f(x)-f(a)-\varepsilon_f(x)) \\ &= (gf(x)-gf(a)) - (D_b g)(f(x)-f(a)) + (D_b g)(\varepsilon_f(x)) \\ &= \varepsilon_g(f(x))\text{①} + (D_b g)(\varepsilon_f(x))\text{②} \end{aligned}$$

②: $\lim_{x\to a} \frac{|(D_b g)(\varepsilon_f(x))|}{|x-a|} \le \lim_{x\to a} \frac{M\cdot|\varepsilon_f(x)|}{|x-a|} = 0.$

①: We chain continuity properties. For any $\varepsilon>0$, we can find a $\delta'>0$ s.t. $|f(x)-b|<\delta \Rightarrow |\varepsilon_g(f(x))| < \varepsilon\cdot|f(x)-b|$. Then, for $\varepsilon' := \delta'$, we can find a $\delta>0$ s.t. $|x-a|<\delta \Rightarrow |f(x)-b|<\varepsilon'$. Altogether, in this range we have $|\varepsilon_g(f(x))| < \varepsilon\cdot|f(x)-b|$
$= \varepsilon\cdot|\varepsilon_f(x) + (D_a f)(x-a)| \le \varepsilon\cdot|\varepsilon_f(x)| + \varepsilon\cdot M'\cdot|x-a|.$

Dividing through gives $\frac{|\varepsilon_g(f(x))|}{|x-a|} \le \varepsilon\cdot\frac{|\varepsilon_f(x)|}{|x-a|} + \varepsilon\cdot M'.$

As $x\to a$, both terms go to zero. $\square$

More properties:

1. If f is constant, then $D_a f = 0$ for any a.
2. If f is linear, then $D_a f = f$.
3. ~~If~~ For $f: \mathbb{R}^n \to \mathbb{R}^m$, f is diff'ble iff each f_i is, and in that case $D_a f = (D_a f_1, \dots, D_a f_m)^T$.
   (This is a matter of checking that the guess for $D_a f$ is correct.)
4. For $s: \mathbb{R}^2 \to \mathbb{R}$, $s(x,y) = x+y$, $D_{a,b} s = s$. $(\Leftarrow 2)$
5. For $p: \mathbb{R}^2 \to \mathbb{R}$, $p(x,y) = x \cdot y$, $(D_{a,b} p)(x,y) = bx + ay$, or $D_{a,b} p = (b \;\; a)$.

Pf of 5: We again check this guess to be correct.

$$\lim_{(h,k)\to 0} \frac{|p(a+h, b+k) - p(a,b) - (bh + ak)|}{|(h,k)|} = \lim_{(h,k)\to 0} \frac{|hk|}{|(h,k)|} \le \lim_{(h,k)\to 0} \frac{h^2+k^2}{\sqrt{h^2+k^2}} = 0.$$

For $f, g: \mathbb{R}^n \to \mathbb{R}$,

Cor: i) $D_a(f+g) = D_a f + D_a g$.

ii) $D_a(f \cdot g) = (D_a f) \cdot g(a) + f(a) \cdot (D_a g)$.

iii) $D_a(f/g) = [(D_a f) \cdot g(a) - f(a) \cdot (D_a g)] / (g(a))^2$.

Pf of ii): The function $f \cdot g$ can be thought of as $p \circ (f, g)$.

$$D_a(p \circ (f,g)) = (D_{f(a),g(a)} p) \circ (D_a(f,g)) = (g(a) \;\; f(a)) \begin{pmatrix} D_a f \\ D_a g \end{pmatrix}$$

$$= g(a) \cdot D_a f + f(a) \cdot D_a g. \qquad \square$$

Partial derivatives

Today we define one of the objects discussed in the introductory lecture: the "planar slice derivatives", or partial derivative.

Def: Take $f:\mathbb{R}^n \longrightarrow \mathbb{R}$, a function $f(x_1,\dots,x_n)$ of n variables. The partial derivative against x_i at $a=(a_1,\dots,a_n)$ is given by:

$$\left.\frac{\partial f}{\partial x_i}\right|_{x=a} = \lim_{h\to 0} \frac{f(a_1,\dots,a_i+h,\dots,a_n) - f(a_1,\dots,a_n)}{h}.$$

(Setting $g(x)=f(a_1,\dots,x,\dots,a_n)$, this is a/k/a $\left. dg/dx\right|_{x=a_i}$.)

Ex: $f(x,y)=e^{x^2y} \rightsquigarrow \partial f/\partial x = 2xy e^{x^2y}$, $\partial f/\partial y = x^2\cdot e^{x^2 y}$.

We'll also be interested in second-order derivatives:

$$\frac{\partial^2 f}{\partial x^2} = 2ye^{x^2y}+4x^2y^2e^{x^2y}, \quad \frac{\partial^2 f}{\partial y^2} = x^4 e^{x^2y},$$

$$\frac{\partial^2 f}{\partial x\partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = 2xe^{x^2y}+2x^3y\cdot e^{x^2y}, \quad \frac{\partial^2 f}{\partial y \partial x} = 2xe^{x^2y}+2x^3ye^{x^2y}.$$

Remarkably, this is generic:

Thm: If $\partial^2 f/\partial x\partial y$ and $\partial^2 f/\partial y \partial x$ are continuous in an open set containing a, then they agree at a.

Pf: Not for a few weeks. $\boxed{\text{TBC}}$

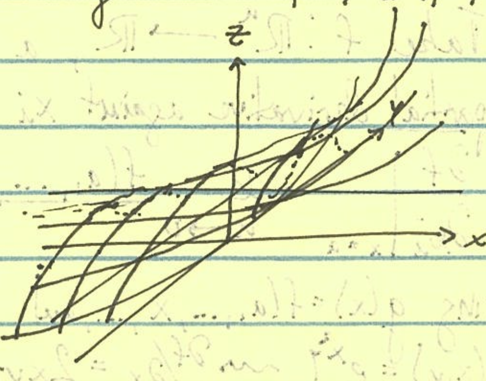
Def: A function with continuous partial derivatives of all orders (i.e., $\partial^n f/\partial x_{i_1}\cdots\partial x_{i_n}$) is called a C^∞-function.

Before using partial derivatives to calculate D_af, we take the opportunity to investigate some classical phenomena: local extrema.

Lem: If a is a local extremum of a diff^{ble} f^n f, then $\partial f/\partial x_i|_{x=a} = 0$ for any index i.

Pf: a_i is also an extremum of $g(x) = f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$. □

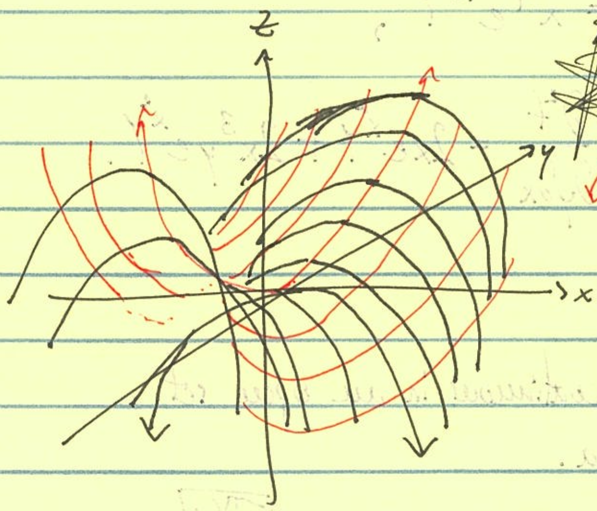
Ex: $f = e^{x^2 y}$ as before.

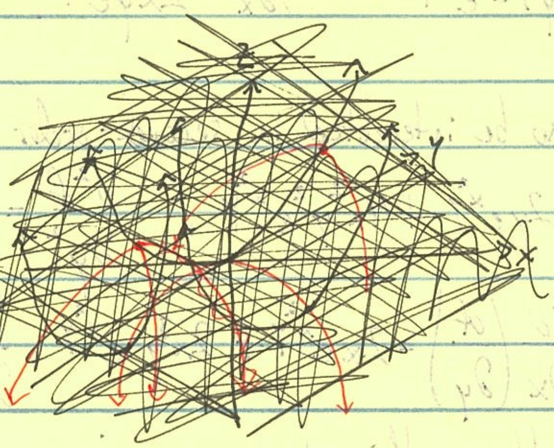


$$\frac{\partial f}{\partial x} = 2xye^{x^2 y} = 0 \text{ when } x = 0 \text{ or } y = 0.$$

$$\frac{\partial f}{\partial y} = x^2 e^{x^2 y} = 0 \text{ when } x = 0.$$

Ex: $g = y^2 - x^2$





$$\frac{\partial g}{\partial x} = -2x = 0 \text{ when } x = 0.$$

$$\frac{\partial g}{\partial y} = 2y = 0 \text{ when } y = 0.$$

Ex: $h = (1 - xy)^2 + x^2$ ~ bounded below poly^l f^n with no minimum.

Lem: If $\partial f/\partial x_i = 0$, then f is independent of x_i.
If $\partial f/\partial x_i = 0 \ \forall i$, then f is constant. □

From $\partial/\partial x_j$ to D

Today we prove one of the main theorems from our introduction:

Thm: If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a, then $\partial/\partial x_j (f^i)$ exists at a for every i, j, and $D_a f$ is presented by the matrix they form: $D_a f = \left(\partial/\partial x_j f^i (a) \right)_{i,j}$.

Pf: We construct an auxiliary function to read off a particular matrix entry: $h(x) = (a_1, \dots, x, \dots, a_n)$ j^{th} spot. Then $D_{a_j}(f \circ h) = D_a f \cdot D_{a_j} h = D_a f \cdot e_j$. The matrix $D_{a_j}(f \circ h)$ is, definitionally, the column of partials $\left(\partial/\partial x_j (f \circ h)(a_j) \right)_i$. □

The converse is harder and more useful:

Thm: For $f : \mathbb{R}^n \to \mathbb{R}^m$, if each $\partial f^i/\partial x_j$ exists in an open set containing a & it is continuous there, then $D_a f$ exists.

Pf: We take $m = 1$ for sanity. Consider the staggered differences:

$$[f(a_1+h_1, a_2, \dots, a_n) - f(a_1, \dots, a_n)] + [f(a_1+h_1, a_2+h_2, a_3, \dots, a_n) - f(a_1+h_1, a_2, \dots, a_n)] + \dots = f(a+h) - f(a).$$

The mean value theorem for derivatives guarantees b_j with

$$f(a_1+h_1, \dots, a_j+h_j, a_{j+1}, \dots, a_n) - f(a_1+h_1, \dots, a_{j-1}+h_{j-1}, a_j, \dots, a_n) = h_j \cdot \frac{\partial f}{\partial x_j} \underbrace{(a_1+h_1, \dots, a_{j-1}+h_{j-1}, b_j, a_{j+1}, \dots, a_n)}_{c_j}.$$

We use them to perform our approximation test:

$$\lim_{h \to 0} \frac{\left| f(a+h) - f(a) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) \cdot h_i \right|}{\|h\|} = \lim_{h \to 0} \frac{\left| \sum_{i=1}^n \left[\frac{\partial f}{\partial x_i}(c_i) - \frac{\partial f}{\partial x_i}(a) \right] \cdot h_i \right|}{\|h\|}$$

C-S + triangle $\rightsquigarrow$ $$\le \lim_{h \to 0} \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(c_i) - \frac{\partial f}{\partial x_i}(a) \right| \cdot \frac{|h_i|}{\|h\|} \le \lim_{h \to 0} \sum_{i=1}^n \left| \text{—}''\text{—} \right|,$$

and this is zero b/c each $\partial f/\partial x_j$ is cts at a. □

Congratulations: you can now compute derivatives. :)

Cor: For $g_1, \dots, g_m : \mathbb{R}^m \to \mathbb{R}$ cts'ly diff'ble and $f : \mathbb{R}^m \to \mathbb{R}^n$ diff'ble, the composite $F = f \circ (g_1, \dots, g_m)$ has derivatives

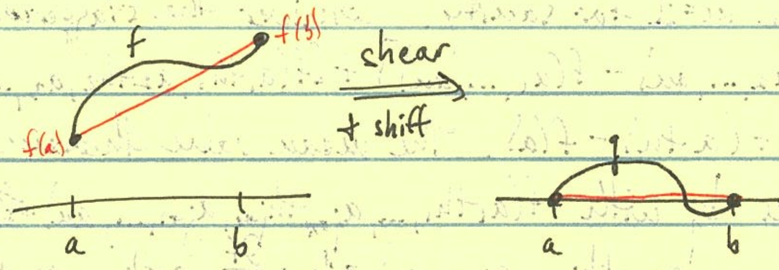
$$\frac{\partial F}{\partial x_i}(a) = \sum_{j=1}^{m} \frac{\partial f}{\partial x_j}\big(g_1(a), \dots, g_m(a)\big) \cdot \frac{\partial g_j}{\partial x_i}(a).$$

Pf: Since each g_j is cts'ly diff'ble, $(g_1, \dots, g_m)$ is diff'ble. Apply chain rule. □

Ex: Consider $F(x,y) = f(g(x,y), h(x), k(y))$. Then

$$\frac{\partial F}{\partial x}(a,b) = \frac{\partial f}{\partial x_1}\big(g(a,b), h(a), k(b)\big) \cdot \frac{\partial g}{\partial x}(a,b) +$$
$$\frac{\partial f}{\partial x_2}\big(g(a,b), h(a), k(b)\big) \cdot \frac{\partial h}{\partial x}(a) +$$
$$\frac{\partial f}{\partial x_3}\big(g(a,b), h(a), k(b)\big) \cdot \cancel{\frac{\partial k}{\partial x}}^{\,0}.$$

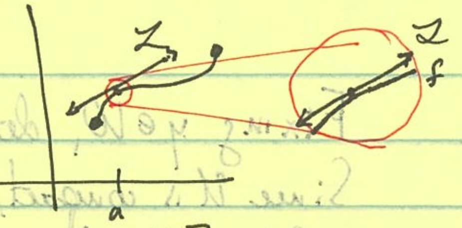
Pf of MVT:



$\overset{\text{Rolle's}}{\Longrightarrow}$ $\exists\, c$ with $a \le c \le b$ and $g'(c) = 0$. □

Pf of Rolle's: $[a,b]$ is compact $\Rightarrow g$ achieves its max + min. Local extrema have vanishing derivatives. □

The Inverse Function Theorem



Recall that if $f'(a)>0$ and f is cts'ly diff'ble, then f is increasing in a nbhd of a, f^{-1} exists, and $(f^{-1})'(b) = 1/f'(f(a))$. What about $f:\mathbb{R}^n \to \mathbb{R}^m$?

(2-10) **Lem:** $A \subseteq \mathbb{R}^n$ a rectangle, $f: A \to \mathbb{R}^m$ cts'ly diff'ble. If $\exists M$ with $|\partial f_i/\partial x_j|(A) \le M$ $\forall i, j$, then $|f(x)-f(y)| \le n^2 M |x-y|$.

Pf: Anything else would violate the MVT + the entry bound on $(\partial f_i/\partial x_j)$. □

Thm: Suppose $f:\mathbb{R}^n \to \mathbb{R}^n$ is cts'ly diff'ble in an open set containing a, and say $\det(D_a f) \neq 0$. Then $\exists\, V \ni a$ open, $W \ni f(a)$ open, $f^{-1}: W \to V$ defined here, and $D_y(f^{-1}) = (D_{f^{-1}(y)} f)^{-1}$.

Pf: The determinant condition $\Longrightarrow$ $D_a f$ is nonsingular. The chain rule $\Longrightarrow D_a\left((D_a f)^{-1} \circ f\right) = (D_a f)^{-1} \circ D_a f = id$. If the theorem is true for $(D_a f)^{-1} \circ f$ it is true for f, so we may assume $D_a f = id$. Now consider an h with $f(a+h) = f(a)$ $\leftarrow$ non-invertibility.

Then $\dfrac{\|f(a+h) - f(a) - (D_a f)(h)\|}{\|h\|} = \dfrac{\|h\|}{\|h\|} = 1$, but $\lim_{h\to 0}(\text{—''—}) = 0$.

So, there exists a ball (+hence a rectangle $= \mathcal{U}$) containing a on which ~~[illegible]~~ $f(x) \neq f(a)$ for any x in the rectangle. Continuous diff'ble lets us also assume $\det(D_x f) \neq 0$ and $\|\partial f_i/\partial x_j(x) - \partial f_i/\partial x_j(a)\| < \frac{1}{2n^2}$ on the rectangle. The difference $g(x) = f(x) - x$ has $D_a g = 0$, hence

$\frac{1}{2}\|x_1 - x_2\| \ge \|g(x_1) - g(x_2)\| \overset{\Delta}{\ge} \|x_1 - x_2\| - \|f(x_1) - f(x_2)\|$, and rearranging gives $\|x_1 - x_2\| \le 2\|f(x_1) - f(x_2)\|$. ~~Since ∂(rect) is compact~~

Lem.

Break: ~~[illegible]~~

Shelve that. Since $\partial(\text{rect})$ is compact, $\exists\, d$ with $\|f(a) - f(\partial(\text{rect}))\| \ge d$. Set W to be the ball of radius $d/2$ about $f(a)$. We claim ~~[illegible]~~ $\exists\, f^{-1}: W \to \mathcal{U}^\circ$.

$g: \mathcal{U} \to \mathbb{R}$

Fixing $y \in \mathcal{W}$, define $g(x) = \|f(x) - y\|^2 = \sum_{i=1}^n \|f_i(x) - y_i\|^2$. Since $\mathcal{U}$ is compact, g attains a minimum, x. This x cannot be in $\partial\mathcal{U}$, since $|y - f(a)| < |y - f(x)|$ is true on $\partial\mathcal{U}$. $D_x g = 0$, but $\partial g/\partial x_j = \sum_i 2\|f_i(x) - y_i\| \cdot \partial f_i/\partial x_j$. This is the result of applying Df to something, but $\det Df \neq 0$, so $f_i(x) = y_i$ $\forall i$. Uniqueness follows from our pause point.

To see f^{-1} is cts, write the shelf point as $\overbrace{\|f^{-1}(y_1) - f^{-1}(y_2)\|}^{\varepsilon} \leq \overbrace{2\|y_1 - y_2\|}^{\delta}$.

To see diff^{ble} let $f(x') = f(x) + (D_x f)(x' - x) + \varepsilon_f(x' - x)$,
where $\lim_{x' \to x} \|\varepsilon_f(x' - x)\| / |x' - x| = 0$. Take inverse:

$$(D_x f)^{-1}(f(x') - f(x)) = x' - x + (D_x f)^{-1}(\varepsilon_f(x' - x)).$$

Trade x's for y's:

$$(D_x f)^{-1}(y' - y) = f^{-1}(y') - f^{-1}(y) + (D_x f)^{-1}(\varepsilon_f(f^{-1}(y') - f^{-1}(y)))$$

$$f^{-1}(y') = f^{-1}(y) + (D_x f)^{-1}(y' - y) - \underbrace{(D_x f)^{-1}(\varepsilon_f(f^{-1}(y') - f^{-1}(y)))}.$$

So, we need $\lim_{y' \to y} \dfrac{\| \quad \|}{\|y' - y\|} = 0.$

$$\Longleftarrow \lim_{y' \to y} \frac{\|\varepsilon_f(f^{-1}(y') - f^{-1}(y))\|}{\|y' - y\|} = 0, \text{ by cty of linear f}^{\underline{ns}}.$$

$$\|$$

$$\lim_{y' \to y} \underbrace{\frac{\|\varepsilon_f(f^{-1}(y') - f^{-1}(y))\|}{\|f^{-1}(y') - f^{-1}(y)\|}}_{\to 0.} \cdot \underbrace{\frac{\|f^{-1}(y') - f^{-1}(y)\|}{\|y' - y\|}}_{\to \leq 2, \text{ by cty.}} \quad \square$$

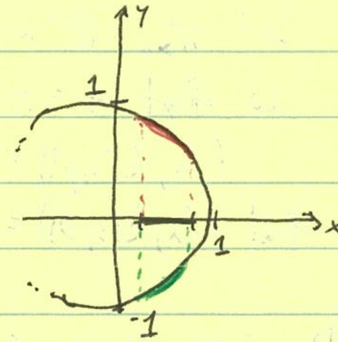
Implicit function theorem

I. Oftentimes we describe a geometric object as the set of solutions to some equation.

<u>Def</u>: Let $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ be some function. The <u>level set</u> of f at $c \in \mathbb{R}$ is the set $f^{-1}(c) \subseteq \mathbb{R}^n$.

<u>Ex</u>: Set $f(x,y) = x^2 + y^2$. Then the level curve $f^{-1}(1)$ is the circle of radius 1 (and generally $f^{-1}(r^2)$ is of radius r).

II. The other tool we have to define geometric objects is as <u>graphs</u>: for a function $g: \mathbb{R}^{n-1} \longrightarrow \mathbb{R}$, we get a subset $\{(x, g(x)) \mid x \in \mathbb{R}^{n-1}\} \subseteq \mathbb{R}^n$.



<u>Ex</u>: $g_+(x) = \sqrt{1-x^2}$ and $g_-(x) = -\sqrt{1-x^2}$

<u>Q</u>: When are these approaches interchangeable?

<u>A</u>: II $\Rightarrow$ I: Given g, we can form an f by $f(\vec{x}, y) = y - g(x)$, and then the graph appears as $f^{-1}(0)$.

<u>A</u>: I $\Rightarrow$ II: The implicit function theorem, a corollary of last time.

<u>Cor</u>: (Implicit function theorem) Suppose $f: \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^m$ is continuously diff$^{\text{ble}}$ on an open nbhd of (a,b) and $f(a,b) = 0$, and let $M = \left(\partial f_i / \partial x_{n+j}(a,b)\right)_{i,j=1}^m = D_b f(a,-)$ be the derivative of the restriction of f. If $\det M \neq 0$, then $\exists$ an open nbhd $A \ni a$ ($\subseteq \mathbb{R}^n$) + an open nbhd $b \in B \subseteq \mathbb{R}^m$ such that $\forall x \in A$ $\exists!$ $g(x) \in B$ with $f(x, g(x)) = 0$. Moreover, g is diff$^{\text{ble}}$.

Pf: Set $F:\mathbb{R}^n\times\mathbb{R}^m \longrightarrow \mathbb{R}^n\times\mathbb{R}^m$ to be $F(x,y)=(x,f(x,y))$.
Then $D_{a,b}F = \left(\begin{array}{c|c} I & 0 \\ \hline 0 & M\end{array}\right)$, so $\det D_{a,b}F = \det M \neq 0$. We can apply the inverse function theorem: $\exists\ F(a,b)=(a,0)\in W\subseteq\mathbb{R}^n\times\mathbb{R}^m$ open and $(a,b)\in V\subseteq\mathbb{R}^n\times\mathbb{R}^m$ open, which WLOG we may assume to be $V=A\times B$, such that $f|_{A\times B}: A\times B\longrightarrow W$ has diff^{ble} inverse $h: W\rightarrow A\times B$. We know what happens to the first n coordinates, so $h(x,y)=(x,k(x,y))$ for some diff^{ble} f^n $k:\mathbb{R}^n\times\mathbb{R}^m\longrightarrow\mathbb{R}^m$. Set $\pi_{>n}:\mathbb{R}^n\times\mathbb{R}^m\rightarrow\mathbb{R}^m$ be the projection to the last m coord^s, so that $f=\pi\circ F$. We use this to extract what we want from k:

$$f(x,k(x,y)) = f\circ h(x,y) = \pi\circ F\circ h(x,y) = \pi(x,y) = y.$$

It follows that $f(x,k(x,0))=0$, which is the level set eq^n, so we set $g(x)=k(x,0)$. $\square$

Rem: To get the derivatives of g, use the defining rel^n $f(x,g(x))=0$, hence

$$\frac{\partial}{\partial x_j}\Big(f_i(x,g(x))\Big) = \frac{\partial f_i}{\partial x_j}(x,g(x)) \overset{\cdot\,\mathrm{Id}}{+} \sum_{k=1}^{m}\frac{\partial f_i}{\partial x_{n+k}}(x,g(x))\cdot\frac{\partial g_k}{\partial x_j}(x).$$

You can solve these for $\partial g_k/\partial x_j$ — in terms of $(x,g(x))$, and you can't do better than that. (Note $g_+' \neq g_-'$ in the Ex.)

Thm: $f:\mathbb{R}^n\rightarrow\mathbb{R}^p$, $p\le n$, cts^{ly} diff^{ble} in a nbhd of a. If $f(a)=0$ and D_af has full rank, then $\exists$ open nbhd $A\ni a$ + diff^{ble} $h:A\rightarrow\mathbb{R}^n$ w/ diff^{ble} inverse s.t. $f\circ h(x^1,\dots,x^n)=(x^{n-p+1},\dots,x^n)$.

Pf: If $f|_{\mathbb{R}^p\subseteq\mathbb{R}^{n-p}\oplus\mathbb{R}^p}$ has full rank, use the previous proof. Otherwise, shuffle the coordinates and try again. $\square$

Darboux integrals (3-1)

We're going to start finding the volumes under ~~mu~~ multi-dimensional graphs. There are *many* ways to set this theory up, so if you take more classes in analysis you're likely to encounter other approaches.

Important Remark: Nothing we're doing today has any mathematical complexity in moving from 1-dim'l to n-dim'l setting. So, feel free to set $n=1$ to ease your understanding.

Def: Let $R = \prod_{i=1}^{n} [a_i, b_i]$ be a closed rectangle in $\mathbb{R}^n$. A partition P of R is a set of cut points $a_i \le s_{i,1} < \cdots < s_{i,N_i-1} < b_i$, cutting each direction of the rectangle into N_i pieces, so that $R = \bigcup_{j_1,\dots,j_n} \prod_{i=1}^{n} [s_{i,j_i}, s_{i,j_i+1}]$ decomposes into subrectangles.

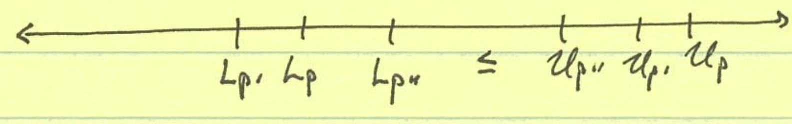
Def: Fix $f: R \to \mathbb{R}$ bounded. For P a partition + S a subrectangle of P, we set $m_S(f) = \inf\{f(x) \mid x \in S\}$ and $M_S(f) = \sup\{f(x) \mid x \in S\}$. The upper and lower sums are $L_P(f) = \sum_S m_S(f) \cdot \mathrm{vol}(S)$ and $U_P(f) = \sum_S M_S(f) \cdot \mathrm{vol}(S)$ respectively.

Lem: P' is said to refine P when P' has all the cut points of P (and maybe more!). If P' refines P, then $L_P \le L_{P'} \le U_{P'} \le U_P$.

Pf: Infima grow + suprema shrink as you restrict attention to smaller sets. $\square$

Cor: $L_{P'} \le U_P$ even if $P + P'$ do not refine one another.

Pf: We can form P'', the union of all the cut points of $P + P'$, and P'' refines both P and P' individually. $\square$



We see that the two sets $\overset{\mathcal{L}}{\overbrace{\{L_P \mid P \text{ a partition}\}}}$ and $\overset{\mathcal{U}}{\overbrace{\{U_P \mid P \text{ a partition}\}}}$ are bounded each other above + below respectively. It follows that $\sup \mathcal{L} \leq \inf \mathcal{U}$.

Def: f is called (Darboux) <u>integrable</u> when $\sup \mathcal{L} = \inf \mathcal{U}$, in which case this number is called $\int_R f$, or sometimes $\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$.

Lem: f is integrable iff $\forall \varepsilon > 0 \, \exists P$ with $U_P - L_P < \varepsilon$.

Ex: If $f: A \to \mathbb{R}$ is constant, then f is integrable with $\int_A f = c \cdot \mathrm{vol}(A)$.

Ex: Consider $\chi_{\mathbb{Q}} : [0,1] \to \mathbb{R}$, $\chi_{\mathbb{Q}}(x) = \{ 1$ if $x \in \mathbb{Q}$, 0 o/w.
$U_P(\chi_{\mathbb{Q}}) = 1$ and $L_P(\chi_{\mathbb{Q}}) = 0$ for all partitions, so <u>not</u> integrable.

Ex: Recall Thomae's function, $f: [0,1] \to \mathbb{R}$:

$$f(x) = \begin{cases} 1 & \text{if } x = 0 \\ \frac{1}{q} & \text{if } x = p/q \text{ in l.c.f.} \\ 0 & \text{o/w.} \end{cases}$$

This <u>is</u> integrable!

Measure zero + content zero (3.2)

One of our short-term goals is to give a characterization of integrable functions (which, of course, can't involve computing any explicit integrals). We need some supporting vocabulary to do this.

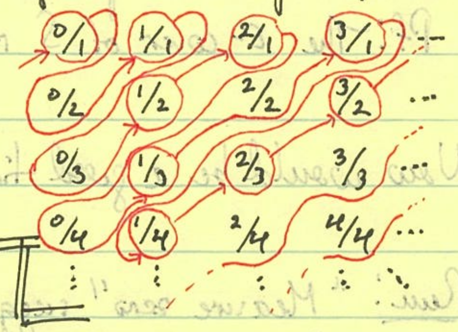
Def: A set $S \subseteq \mathbb{R}^n$ has measure zero if for any $\varepsilon > 0$ there is a cover of S by open rectangles $\{U_j\}_{j=1}^{\infty}$ such that $\sum_j \mathrm{vol}(U_j) \leq \varepsilon$.

Ex: Any finite set.

Ex: Any countable set, meaning $S = (s_n)_{n=1}^{\infty}$ for some seq^ce (s_n):
□ We take $U_n = (-\varepsilon/2^{n+1} + s_n, s_n + \varepsilon/2^{n+1})$ as the cover, so that $\sum_n \mathrm{vol}(U_n) = \sum_n \varepsilon/2^n = \varepsilon$.

Ex: The rationals form a countable set. There are many ways to see this. For instance, you can arrange them in a grid, or we can use an injection like

$$(-1)^n p/q \longmapsto 2^n 3^p 5^q$$
$$\mathbb{Q} \hookrightarrow \mathbb{N}.$$



Rem: This is a confusing collection of facts to consider all at once. Near each real number, there's an arbitrarily close rational number. But, if you use the ε-cover above, you can delete all the rational #'s while removing only ε of the real line's volume. It's normal if this makes you uncomfortable.

Cor: If $A = \bigcup_{j=1}^{\infty} A_j$ and each A_j has measure zero, so does A. □

Def: A set $S \subseteq \mathbb{R}^n$ has <u>content zero</u> if $\forall \varepsilon > 0$ there is a <u>finite cover</u> $\{U_1, \ldots, U_n\}$ of S by open rectangles with $\sum_{j=1}^{n} \mathrm{vol}(U_j) < \varepsilon$.

This is more dramatic a restriction, since ∞ is excluded.

Lem: For $a < b$, $[a,b] \subseteq \mathbb{R}$ does <u>not</u> have content zero.
In fact, if $\{U_j\}_{j=1}^{n}$ cover $[a,b]$, then $\sum_j \mathrm{vol}(U_j) \geq b-a$.
Pf: Passing to $U_j \cap [a,b]$, aggregate all their endpoints into a seq$^{\underline{ce}}$ $t_0 = a < t_1 < \cdots < t_{m-1} < b = t_m$. Each segment (t_j, t_{j+1}) must belong to (at least) one U_{k_j}, hence $\sum_i |t_{j+1} - t_j| \leq \sum \mathrm{vol}(U_j)$. $\square$

That's a relief. In fact $[a,b]$ is also not of measure zero.

Lem: If S is compact + has measure zero, then it has content zero.
Pf: The ε-cover of S reduces to a finite subcover of size $< \varepsilon$. $\square$

[Now would be a good time to re/visit <u>oscillation</u>.]

Rem: "Measure zero" suggests it's possible to have other values, like "measure one" and so on. This deciphering of the volume of subsets of $\mathbb{R}^n$ is a critical part of any advanced analysis course.

Integrability and Continuity (3.3)

Recall our notion of the oscillation of a function $f: \mathbb{R} \to \mathbb{R}$:

$$o(f,x) = \lim_{\delta \to 0}\left(\sup\{f(x) \mid |x-a|<\delta\} - \inf\{f(x) \mid |x-a|<\delta\}\right),$$

that f is cts at a iff $o(f,a)=0$, and that $\{x \mid o(f,x) \geq \varepsilon\}$ is closed.
This difference between upper and lower bounds smells like Darboux sums.

Lem: $A \subseteq \mathbb{R}^n$ a closed rectangle, $f: A \to \mathbb{R}$ bounded, $o(f,x) < \varepsilon$.
There then exists a partition P of A with $U_P(f) - L_P(f) < \varepsilon \cdot \mathrm{vol}(A)$.

Pf: For each $x \in A$, there is a closed rectangle A_x containing x with $\sup\{f(x) \mid x \in A_x\} - \inf\{f(x) \mid x \in A_x\} < \varepsilon$. Since A is compact, finitely many such cover A. Form a partition from their cut points, and use $M - m < \varepsilon$ in the def^{n} of L and U. □

Thm: Let $A \subseteq \mathbb{R}^n$ be a closed rectangle, $f: A \to \mathbb{R}$ a bounded f^{n}, and $B \subseteq A$ the set of points where f is not continuous. Then f is integrable iff B is of measure zero.

Pf: ($\Leftarrow$) Let $\{U_j\}_{j=0}^{\infty}$ be an ε-cover of B witnessing its measure zero, let $B_\varepsilon = \{x \mid o(f,x) \geq \varepsilon\}$, note that B_ε is compact hence content zero, and let $U_1, \dots, U_n$ be the finite ε-subcover. Take a fine enough partition P so that its subrectangles S either have (II) $S \cap B_\varepsilon = \emptyset$ or (I) $S \subseteq U_j$ for some j. For the type I rectangles, use the bound $|f(x)| \leq M$ to get

$$\sum_{S \text{ type I}} (M_S(f) - m_S(f)) \cdot v(S) < 2M \cdot \sum_{j=1}^{n} \mathrm{vol}(U_j) < 2M\varepsilon.$$

For a type II rectangle, $o(f,x) < \varepsilon$ couples to the Lemma to give a refinement P' of P with $\sum_{S' \subseteq S} (M_S(f) - m_S(f)) \cdot v(S') \leq \varepsilon \cdot \mathrm{vol}(S)$.
Altogether, $U_{P'}(f) - L_{P'}(f) < 2M\varepsilon + \sum_{S \text{ type II}} \varepsilon \cdot \mathrm{vol}(S) \leq 2M\varepsilon + \varepsilon \mathrm{vol}(A)$.
Since $\varepsilon > 0$ was arbitrary, f is integrable.

($\Rightarrow$) Take f to be integrable. We have $B = B_{1/2} \cup B_{1/3} \cup B_{1/4} \cup \cdots$, so it suffices to show $B_{1/n}$ of measure zero (and, indeed, content zero). For $\varepsilon > 0$, take a partition P so that $U_P(f) - L_P(f) < \varepsilon/n$, and let $\mathcal{S}$ be the subset of the rectangles intersecting with $B_{1/n}^\circ$; this will be our cover. Indeed:

$$\frac{1}{n}\sum_{S\in\mathcal{S}} \operatorname{vol}(S) \le \sum_{S\in\mathcal{S}} (M_S(f) - m_S(f))\operatorname{vol}(S) \le \sum_{\text{all } S} \text{---}\text{''}\text{---} < \frac{\varepsilon}{n}. \quad \square$$

To deal with other bounded sets $C \subseteq \mathbb{R}$, we extend $f : C \to \mathbb{R}$ by zero to a rectangle $A \supseteq C$ and compute its integral there.

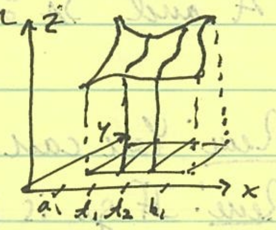
<u>Lem</u>: $\chi_C : A \to \mathbb{R}$ is integrable iff ∂C has measure zero ($\Leftrightarrow$ content zero).

Pf: χ_C is not continuous exactly at ∂C. $\square$

Such a set C is called <u>Jordan-measurable</u>, and its <u>content</u> is defined as $\mu(C) = \int_C 1$. This definition has its defects: It is possible for an <u>open set</u> to <u>fail</u> to be Jordan-measurable, so even for cts $f : C \to \mathbb{R}$ we may not be able to define $\int_C f$. (For instance, I think an ε-cover of $\mathbb{Q} \cap [0,1]$ causes trouble...)

Fubini's Theorem (3.4)

As explained in the introduction, there is a theorem connecting multi-dimensional integrals to iterated single-variable integrals by dealing with the slicing "one axis at a time", which suggests formulas like



$$\int_A f = \int_{a_1}^{b_1}\left(\int_{a_2}^{b_2} f(x,y)\,dy\right)dx.$$

Unfortunately, as we try to do this rigorously, f may not be continuous, and the failure of f to be continuous may only be "visible from certain trajectories" — we complained about similar things when studying partial derivatives. The theorem statement may thus not be recognizable...

Thm (Fubini): Let $A \subseteq \mathbb{R}^n$, $B \subseteq \mathbb{R}^m$ be closed rectangles and $f: A\times B \to \mathbb{R}$ be integrable. For $x \in A$, define $g_x(y) = f(x,y)$ and set $\mathcal{L}(x) = \mathbf{L}\int_B g_x = \mathbf{L}\int_B f(x,y)\,dy$, $\mathcal{U}(x) = \mathbf{U}\int_B g_x = \mathbf{U}\int_B f(x,y)\,dy$.
Then $\mathcal{L}$ and $\mathcal{U}$ are themselves integrable on A, and

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

Pf: A partition P_A of A and P_B of B give rise to a partition P of $A\times B$, and $L_P(f) = \sum_S m_S(f)v(S) = \sum_{S_A,S_B} m_{S_A\times S_B}(f)\cdot v(S_A\times S_B)$

$= \sum_{S_A}\left(\sum_{S_B} m_{S_A\times S_B}(f)\cdot v(S_B)\right)\cdot v(S_A)$. For $x\in S_A$, $m_{S_A\times S_B}(f) \le m_{S_B}(g_x)$,

hence $\sum_{S_B} m_{S_A\times S_B}(f)\cdot v(S_B) \le \sum_{S_B} m_{S_B}(g_x)\cdot v(S_B) \le \mathbf{L}\int_B g_x = \mathcal{L}(x)$.

This gives $\qquad\qquad$ similar

$$\overbrace{L_P(f) \le L_{P_A}(\mathcal{L})}\le U_{P_A}(\mathcal{L}) \le \overbrace{U_{P_A}(\mathcal{U}) \le U_P(f)}$$

$$\underbrace{L_{P_A}(\mathcal{L}) \le U_{P_A}(\mathcal{L}) \le U_{P_A}(\mathcal{U})}_{\text{trivial.}}$$

Since f is integrable, the outer terms sandwich the middle terms to all exist and agree. That is, $\mathcal{L}$ is integrable on A and $\int_A \mathcal{L} = \int_{A\times B} f$. (Repeat the proof with $\mathcal{U}$.) $\square$

Rem: You can switch the roles of x and y without harm.
Rem: If g_x is integrable ($\Leftarrow$ cts), then we recover the usual formula.
Rem: Often g_x fails to be integrable at only finitely many x.
Since this doesn't affect $\int_B \mathcal{L}$, we often define this away.
Rem: Fubini's theorem has an interesting interaction with irregular regions $C \subseteq \mathbb{R}^n$. For instance, take $C = \{\vec{x} \in \mathbb{R}^2 \mid \|x\| \leq 1\}$.
Then, $\chi_C(x,y) = \{1$ if $-\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2}$, 0 o/w, and hence

$$\int_C f = \int_{[-1,1]^{\times 2}} f \cdot \chi_C = \int_{-1}^{1}\int_{-1}^{1} f \cdot \chi_C \, dy\, dx = \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f \, dy\, dx.$$

$\int_{[0,1]^2} \chi_C$, $C = \{(x,y) \in [0,1]^2 \mid x+y \leq 1\}$.

Partitions of Unity (3.5)

Today is another (very important) technical utility day.

<u>Thm</u>: Let $A \subseteq \mathbb{R}^n$ be some set + let $\mathcal{O}$ be an open cover of A. There is a collection $\Phi = \{\varphi\}$ of C^∞-f$^{\underline{ns}}$ defined in an open set containing A with the following properties:

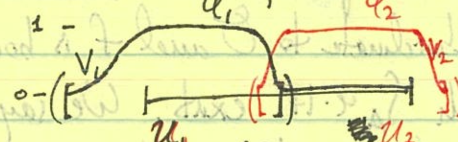
(i) For each $x \in A$, $0 \leq \varphi(x) \leq 1$.

(ii) For each $x \in A$, there is an open nbhd $V \ni x$ such that all but finitely many $\varphi \in \Phi$ are 0 on V.

(iii) Among the nonzero φ, $\sum_\varphi \varphi(x) = 1$.

(iv) For each $\varphi \in \Phi$ there is a $\mathcal{U} \in \mathcal{O}$ and a closed $V \subseteq \mathcal{U}$ such that $\varphi|_{\mathbb{R}^n \setminus V} = 0$, i.e., φ vanishes off of V, or $\varphi = \chi_V \cdot \varphi$.

<u>Ex</u>: Two open sets in $\mathbb{R}$:



<u>Ex</u>: These are used to break things up as follows:

$$\int_A f = \int_A f \cdot \sum_{\varphi \in \Phi} \varphi = \sum_{\varphi \in \Phi} \int_A f \cdot \varphi = \sum_{\varphi \in \Phi} \int_A f \cdot \varphi \cdot \chi_{V_\varphi} = \sum_{\varphi \in \Phi} \int_{V_\varphi \subseteq \mathcal{U}_\varphi} f \cdot \varphi.$$

<u>Pf of Thm</u>: We prove this initially under assumptions on A.

① Suppose that A is <u>compact</u>, and let $\{\mathcal{U}_1, \ldots, \mathcal{U}_n\}$ be a finite subcover of $\mathcal{O}$ covering A. We shrink each $\mathcal{U}_i$ to a compact set as $\leftarrow D_k$ follows: set $C_k = A \setminus (D_1^\circ \cup \cdots \cup D_{k-1}^\circ \cup \mathcal{U}_{k+1} \cup \cdots \cup \mathcal{U}_n)$, and invoke the homework problem to find a compact D_k with $C_k \subseteq D_k^\circ \subseteq D_k \subseteq \mathcal{U}_k$. By another homework problem, we can find ψ_k, a non-negative C^∞-f$^{\underline{n}}$ which is positive on D_k and 0 outside of some closed set $\subseteq \mathcal{U}_k$. We

pos. on some $\mathcal{U} \supseteq A$ $\rightsquigarrow$ set $\xi_k = \psi_k / (\psi_1 + \cdots + \psi_n)$, so that for $f : \mathcal{U} \to [0,1]$ a C^∞-f$^{\underline{n}}$ which is 1 on A + 0 outside of some closed set $\supseteq \mathcal{U}$, we may finally set $\varphi_j = f \cdot \xi_j = f \psi_j \cdot (\psi_1 + \cdots + \psi_n)^{-1}$.

② Suppose $A = A_1 \cup A_2 \cup A_3 \cup \cdots$, where each A_j is compact and $A_j \subseteq \mathring{A}_{j+1}$.
We deal with the "differences": for each j, set $\mathcal{O}_j = \{ U \cap (\mathring{A}_{j+1} \setminus A_{j-2}) \mid U \in \mathcal{O} \}$, which is an open cover of $B_j = A_j \setminus \mathring{A}_{j-1}$. Each B_j is compact, so we apply ① to get a partition of unity Φ_j'. For each $x \in A$, $\sigma(x) = \sum_j \sum_{\varphi \in \Phi_j'} \varphi(x)$ is a finite sum, so set $\varphi^* = \varphi / \sigma$ and take all these φ to get the desired partition.

③ A open: set $A_j = \{ x \in A \mid |x| < j \text{ and } x \text{ is distance} \geq 1/j \text{ from } \partial A \}$.
This presents A as an ascending union of compact sets.

④ In general, we may replace A with $\bigcup_{U \in \mathcal{O}} U$, an open. □

Take $A \subseteq \mathbb{R}^n$ open and $\mathcal{O}$ <u>admissible</u>, meaning $\bigcup_{U \in \mathcal{O}} U \overset{\text{not } \supseteq!!}{=} A$. If Φ is subordinate to $\mathcal{O}$ and f is bounded in a nbhd of any point in A, and the set of discontinuities has meas. zero!! then each $\int_A \varphi \cdot |f|$ exists. We say f is <u>integrable</u> if $\sum_{\varphi \in \Phi} \int_A \varphi \cdot |f|$ converges ($\Rightarrow \sum_i \left| \int_A \varphi \cdot |f| \right|$ converges $\Rightarrow \sum_i \int_A \varphi \cdot f$ converges $=: \int_A f$).

Lem: ① If Ψ is some other partition subordinate to $\mathcal{O}$, then $\sum_{\psi \in \Psi} \int_A \psi \cdot |f|$ also converges and $\sum_{\varphi \in \Phi} \int_A \varphi f = \sum_{\psi \in \Psi} \int_A \psi f$.

② If A + f are bounded, ~~this agrees with the old integral~~.
then f is integrable in this new sense.

③ If A is Jordan-measurable and f is bounded, then this definition agrees with the old one.

Change of Coordinates (3.6)

Recall u-substitution from calculus: $\int_{u(a)}^{u(b)} f(x)dx = \int_a^b f(u(t))u'(t)dt$.
Once you have the FTC, this is an easy corollary: write $F'=f$, then $(F\circ u)' = (F'\circ u)\cdot u'$, and integrate. The geometry of this situation is encoded in a Mathematica demo.

Thm: Let $A\subseteq \mathbb{R}^n$ be open, and $g: A\to\mathbb{R}^n$ an injective C^1 diff'ble f^n with $D_a g$ invertible $\forall a\in A$. If g is integrable, then
$$\int_{g(A)} f = \int_A (f\circ g)\cdot |\det Dg|.$$

Reductions: ① If $\mathcal{O}$ is an admissible cover of A + the theorem is true for each $U\in\mathcal{O}$, then it's true on A by adding the integrals up on a partition of unity. ② Then, it also suffices to show the theorem just for the constant functions $f=1$, since arbitrary integrals are built out of patchworks of constant functions. ③ If two functions g_1 and g_2 satisfy the theorem, then it holds for $g_1\circ g_2$ too, since det and D are both multiplicative. ④ The theorem is true if $f=1$ and g is linear, by a homework problem. ⑤ So, in general, we need only find a small open where the theorem holds, and for any one point $a\in A$ we may assume $D_a g = \mathrm{id}$.

Pf: Since we have the case $n=1$ + Fubini's theorem, we induct.
We carve g into two pieces: set $h = (g_1(x), \ldots, g_{n-1}(x), x_n)$ and $k = (x_1, \ldots, x_{n-1}, g_n(h^{-1}(x)))$, so that $g = h\circ k$. ($D_a h = I$, so we can find a nbhd of a where h^{-1} exists.)
We want to show the theorem at a for h + at $h(a)$ for k; we'll do just h, since k is similar but easier.

The goal in this definition is to capture the square upper-left block of $Dh = \left(\begin{array}{c|c} Dh_{x_n} & 0 \\ \hline \cdots & 1 \end{array}\right)$.

$(n-1)$-dim'l rectangle

Let $W \subseteq U$ now be a rectangle of the form $D \times [a_n, b_n]$. Fubini says $\int_{h(W)} 1 = \int_{a_n}^{b_n} \left(\int_{h(D \times \{x_n\})} 1 \, dx_1 \cdots dx_{n-1} \right) dx_n$. For any $x_n \in [a_n, b_n]$, we define $h_{x_n}(x_1, \ldots, x_{n-1}) = (g_1(x_1, \ldots, x_n), \ldots, g_{n-1}(x_1, \ldots, x_n))$. Since h was injective, so is each h_{x_n}, and $\det D_{(x_1, \ldots, x_{n-1})} h_{x_n} = \det D_{(x_1, \ldots, x_n)} h \neq 0$. The $(n-1)$-dim'l case of the theorem thus applies, giving

$$\int_{h(W)} 1 = \int_{a_n}^{b_n} \left(\int_{h_{x_n}(D)} 1 \, dx_1 \cdots dx_{n-1} \right) dx_n \overset{\text{ind.}}{=} \int_{a_n}^{b_n} \left(\int_D |\det D_{(x_1, \ldots, x_{n-1})} h_{x_n}| \, dx_1 \cdots dx_{n-1} \right) dx_n$$

h_{x_n} preserves the last coord, can check this manually.

$$= \int_{a_n}^{b_n} \left(\int_D |\det D_{(x_1, \ldots, x_n)} h| \, dx_1 \cdots dx_{n-1} \right) dx_n \overset{\text{Fub.}}{=} \int_W |\det D h|. \quad \square$$

The $\det D g \neq 0$ condition can be eliminated using Sard's theorem:

We saved all this for the next day.

Thm (Sard): Let $g : A \to \mathbb{R}^n$ be continuously diff'ble, $A \subseteq \mathbb{R}^n$ open, and set $B = \{x \in A \mid \det D_x g = 0\}$. Then $g(B)$ has measure zero.

Rem: There is a version of this for $f : A \to \mathbb{R}^m$.

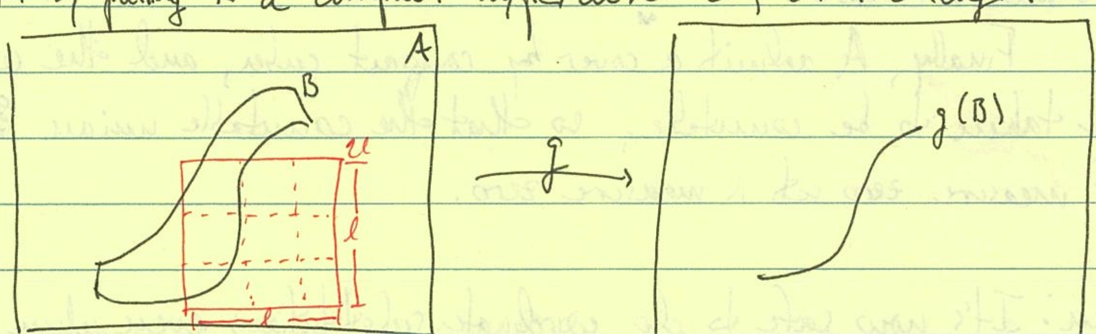
Pf: Start with a cube $U \subseteq A$ of side length l, and take $\varepsilon > 0$. For $N \gg 0$, the N-subdivision of U into N^n pieces S satisfy the error bound $\|(D_x g)(y-x) - g(y) - g(x)\| < \varepsilon \|x - y\| \leq \varepsilon \sqrt{n} \cdot l/N$ for all $x, y \in S$. If S meets B, pick an $x \in S$ with $\det D_x g = 0$; the set $\{D_x g(y-x) : y \in S\}$ lies in an $(n-1)$-dim'l subspace $V \subseteq T_x A$, and $\{g(y) - g(x) \mid y \in S\}$ lies within $\varepsilon \sqrt{n} \cdot l/N$ of $g(x) + V$. Our old dispersion bound gives an M with $\|g(x) - g(y)\| < M \|x - y\| \leq M \sqrt{n} \cdot l/N$. Thus, $\{g(y) : y \in S\}$ is contained in a cylinder of ~~height~~ radius $< M \sqrt{n} (l/N)$ and height $< 2 \varepsilon \sqrt{n} \, l/N$, hence volume $C (l/N)^n \varepsilon$ for some constant C. Since there are at most N^n rectangles, the total cylindrical volume is $< C l^n \cdot \varepsilon$. $\square$

Sard's Theorem

Thm (Sard): Let $A \subseteq \mathbb{R}^n$ be open, $f : A \to \mathbb{R}^n$ continuously differentiable, and $B = \{a \in A \mid \det D_a g = 0\}$. Then $g(B)$ has measure zero.

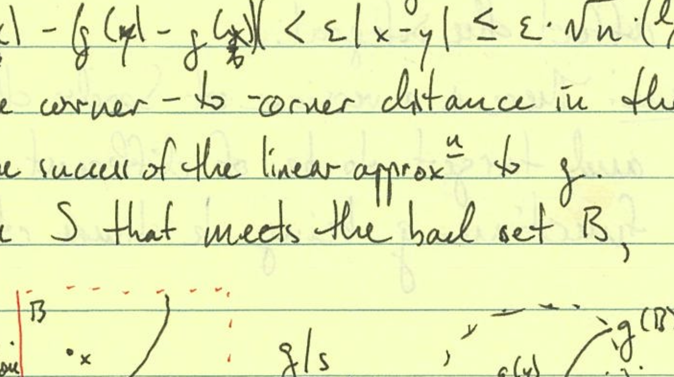
Rem: This is a beautiful example of geometric analysis, so we're going to do it slowly. Intuitively, it makes some sense: B is exactly the places where g "collapses" some direction, so we expect $g(B)$ itself to be collapsed — i.e., measure zero.

Pf: Start by passing to a compact hypercube $\mathcal{U}$, of side length l.



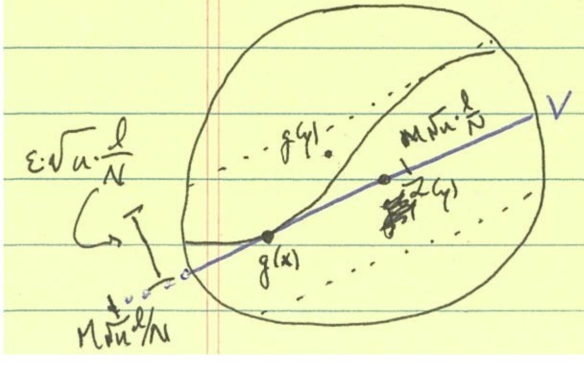
Because g is diff^{ble}, $\forall \varepsilon > 0\ \exists N > 0$ s.t. subdividing $\mathcal{U}$ into N^n subcubes guarantees $|D_x g(y-x) - (g(y) - g(x))| < \varepsilon |x-y| \le \varepsilon \cdot \sqrt{n} \cdot (l/N)$, where the last inequality is the corner-to-corner distance in the subcube. This is a measure of the success of the linear approx^n to g.

Restrict attention to an S that meets the bad set B, and pick $x \in S \cap B$. The approximant $D_x g$ drops dimension; pick a proper subspace $V \subseteq T_{g(x)}\mathbb{R}^n$ containing $\mathrm{im}(D_x g)$. We have two bounds on the behaviors of $g(y)$ and $\mathcal{L}(y)$:



① we know $\varepsilon\sqrt{n}\, l/N$ bounds the error of $\mathcal{L}(y)$, so $g(y)$ can't be more than $\varepsilon\sqrt{n}\, l/N$ off of V.

② we also know, by the dispersion bound, that $\exists M > 0$ with $\|g(x) - g(y)\| < M\|x-y\| \le M\sqrt{n} \cdot l/N$.



The trick is not to apply ② directly to $g(y)$ — instead, this is a bound on the disk about the origin in V that we need to consider. So, in all, $g(y)$ must lie in a cylinder with base the sphere in V of radius $< M\sqrt{n}\cdot l/N$ + of height $< 2\varepsilon\sqrt{n}\cdot l/N$, which has total volume $< C\cdot(l/N)^n\varepsilon$ for some nasty constant C, dependent on M and n (but also involving π's and so on). ← Rem. This actually bounds $g(S)$!!!

There are at most N^n subcubes S meeting B, so in all $g(\mathcal{U}\cap B)$ lies in a set of volume $< C\cdot(l/N)^n\cdot\varepsilon\cdot N^n = C\cdot l^n\cdot\varepsilon$. Since we ended up with just an ε-factor and nothing else, $g(B\cap\mathcal{U})$ has measure zero.

Finally, A admits a cover by compact cubes, and the cover can be taken to be countable, so that the countable union $B = \bigcup_S (B\cap S)$ of measure zero sets is measure zero. $\square$

Rem: It's now safe to do coordinate substitutions even when $\det g = 0$ sometimes, since the value of the function on $g(B)$ will not affect the integral.

Rem: There is a version of Sard's theorem that allows the source and target to be of different dimensions, in trade for the function g being k-times ctsly diff^{ble}, $k>1$.

Multivariate Taylor Polynomials

In single-variable calculus (and on your homework!), you learn the following

Thm (Taylor): For $U \subseteq \mathbb{R}$ open and $f: U \to \mathbb{R}$ k-fold continuously diff^{ble}, $P^k_{f@a}(x) = \sum_{j=0}^{k} \frac{1}{j!} f^{(j)}(a)(x-a)^j$ is the unique polynomial of degree $\le k$ satisfying $\lim_{h\to 0} (f(a+h) - P^k_{f@a}(a+h)) h^{-k} = 0$. $\square$

(In fact, there's a version of this theorem that bounds the limit in terms of $\sup f^{(k+1)}(x)$, which we will elect not to think about.)

Today we're going to build a multivariate analogue of this, $f: \mathbb{R}^n \to \mathbb{R}$.

Ex: As in the single variable case, the local linear approximation to f is supposed to form $P^1_{f@a}(x) = f(a) + (D_a f)(x-a) = f(a) + \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}\Big|_a \cdot (x_j - a_j)$.

Ex: As a silly guess, we can try to throw in all the degree 2 terms in a multivariate polynomial into its P^2. For instance, $f(x) = x^3 + y^6 + 3z^2 \rightsquigarrow P^2_{f@0} = 3z^2$. What's special about this? Well, $\frac{\partial^2}{\partial z^2}(x^3 + y^6 + 3z^2)\Big|_0 = 6$ and $\frac{\partial^2}{\partial z^2}(3z^2)\Big|_0 = 6$. In fact, these two f^{ns} agree on $\partial/\partial x_j|_0$ and $\partial^2/\partial x_j^2|_0$ for all j. :)

Ex: What about $f(x,y) = xy$? Should this belong to degree 2?

Yes: if we want our approximation condition to be independent of coordinates, then, e.g., after setting $x = u - v$ and $y = u + v$, we also want $\partial^2/\partial u^2$ and $\partial^2/\partial v^2$ of $f(u,v) = u^2 - v^2$ to be zero. This is tantamount to including xy in P^2 — or to considering *mixed partials*

Lem: For $I = (i_1, \ldots, i_n)$, set $\frac{\partial^{I_+}}{\partial x^I} = \frac{\partial^{i_1}}{\partial x_1^{i_1}} \cdots \frac{\partial^{i_n}}{\partial x_n^{i_n}}$. There is a unique monomial M with $(\partial^{I_+}/\partial x^I)(M)\big|_0 = 1$: $M = \frac{1}{I!} \cdot x^I$, where $I! = i_1! \cdots i_n!$ and $x^I = x_1^{i_1} \cdots x_n^{i_n}$. $\square$

This gives a recipe for $P^n_{f@a}(x) = \sum_{I_+ \leq n} \left.\frac{\partial^{I_+} f}{\partial x^I}\right|_a \cdot \frac{1}{I!} \cdot (x-a)^I$.

<u>Thm</u>: $P^n_{f@a}(x)$ is the unique polynomial of degree $\leq n$ with $\lim_{h\to 0} \|f(a+h) - P^n_{f@a}(a+h)\| \cdot \|h\|^{-n} = 0$ where $f: \mathcal{U} \to \mathbb{R}$ is n-fold continuously diff$^{\underline{ble}}$ on an open $\mathcal{U} \subseteq \mathbb{R}^k$.

<u>Lem</u>: If $f: \mathcal{U} \to \mathbb{R}$ is a n-fold ctsly diff$^{\underline{ble}}$ $f^{\underline{n}}$, $\mathcal{U} \subseteq \mathbb{R}^k$, and <u>all</u> partials up to order n of f at $a \in \mathcal{U}$ vanish, then $\lim_{h\to 0} f(a+h) \cdot \|h\|^{-n} = 0$.

<u>Pf of Lem</u>: This is done by induction, using a familiar scheme. Break h into components $(h_1, \dots, h_k)$ and consider $a \to a + h_1e_1 \to a + h_1e_1 + h_2e_2 \to \cdots \to a+h$. The difference between $f(a + h_{<j}e_{<j})$ and $f(a + h_{\leq j}e_{\leq j})$ lies in the $j^{\underline{th}}$ direction alone, where we apply the MVT to produce a c_j with $f(a+h_{\leq j}e_{\leq j}) - f(a + h_{<j}e_{<j}) = h_j \frac{\partial f}{\partial x_j}(a + h_{<j}e_{<j} + c_je_j)$. We thus have

$$\lim_{h\to 0} \frac{f(a+h)}{\|h\|^n} = \lim_{h\to 0} \sum_{j=1}^{k} \frac{h_j}{\|h\|} \cdot \frac{\partial f/\partial x_j(a+h_{<j}e_{<j} + c_je_j)}{\|h\|^{k-1}}.$$

Replacing the $\|h\|^{k-1}$ in the denom. by the smaller value $\|h_{<j}e_{<j} + c_je_j\|^{k-1}$, we induct. □

<u>Pf of Thm</u>: We have designed $P^n_{f@a}$ to have identical partials to f through order n. Now we take the difference and apply the Lemma. □

<u>Rem</u>: The usual rules apply for sum, product, + composition of Taylor polynomials. For instance, $f(x,y) = \sin(x+y^2)$ has

$$\begin{aligned} P^3_{f@0} &= (x+y^2) - ((x+y^2)^3) \cdot \tfrac{1}{6} + o(4) \\ &= x + y^2 - \tfrac{1}{6}x^3 + o(4) \end{aligned}$$

$= \frac{1}{6!} \cdot \partial^3 f / \partial x^3$. :)

Quadratic forms

Today we have an algebraic interlude. We are after a generalization of

Thm: $f:\mathbb{R}\to\mathbb{R}$ twice continuously diff^{ble} in a nbhd of $a\in\mathbb{R}$, and $f'(a)=0$. Then $f''(a)>0 \Rightarrow a$ is a local min. + $f''(a)<0 \Rightarrow a$ is a local max. □

In the multivariate setting, we would also like to discern the behavior of critical points based on "second order" information. Last time we showed that the appropriate analogue of this is $\frac{1}{2}\sum_j\sum_i \partial^2 f/\partial x_i \partial x_j \cdot x_i x_j$.

Def: A quadratic form $Q:\mathbb{R}^n\to\mathbb{R}$ is a polynomial of homogeneous degree 2 (i.e., exactly of the form above).

Ex: If $p(t)$ is any polynomial, then $\int_0^1 (p(t))^2\,dt$ is a quadratic form in its coeff^s. For example, set $p(t)=p_0+p_1t+p_2t^2$. Then

$$\int_0^1 \left(p_0^2+2p_0p_1t+(p_1^2+2p_0p_2)t^2+2p_1p_2t^3+p_2^2t^4\right)dt = p_0^2+p_0p_1+p_1^2/3 + 2p_0p_2/3+p_1p_2/2+p_2^2/5$$

Ex: The classification of quadratic forms / $\mathbb{Z}$ + their images is a famous topic.

Thm: For any quadratic form Q on $\mathbb{R}^n$, there exist m l.i. linear functions $a_1,\dots,a_m$ such that $Q=(a_1)^2+\cdots+(a_k)^2-(a_{k+1})^2-\cdots-(a_m)^2$. The functions are not unique, but the index pair $(k,m-k)$ is, called the signature of Q. □

Ex: $x^2+xy=(x+y/2)^2-(y/2)^2$.

Ex: $xy-xz+yz \xrightarrow[x=u+y]{u=x-y} y^2+uy-ux = (y^2+u/2)^2-(u/2)^2-ux$

$$=(y^2+u/2)^2-(u/2+z)^2+z^2.$$

$$\leftsquigarrow (x/2+y/2)^2-(x/2-y/2+z)^2+z^2.$$

Doing this in generality is basically the proof of the first half of the Thm.

Def: A q.f. is positive definite if $x\neq 0\Rightarrow Q(x)>0$,

negative definite if $x\neq 0 \Rightarrow Q(x)<0$.

Lem: If Q has signature (k, l), then the largest v.s. subspace on which Q is pos. def. is dim. k, and ditto for neg. def. and dim l.

Pf idea: The entries a_j give directions in which Q behaves as positive or negative definite. Too many directions contradict ~~the theorem~~ a given decomp., and a decomposition as in the theorem can be used to extract a subspace. □

Ex: For $Q = xy = \frac{1}{4}\left((x+y)^2 - (x-y)^2\right)$, vectors (x, y) with $x = y$ behave positive-definitely + those with $x = -y$ behave negative definitely.

Def: Q is called non-degenerate if $m = n$ in the Theorem statement — i.e., Q has n terms in its expansion, $Q : \mathbb{R}^n \to \mathbb{R}$. Otherwise, it is called degenerate.

Lem: If Q is p.d. nondegenerate, then $\exists\, C > 0$ with $Q(x) > C\|x\|^2$.

Pf: Decompose $Q(x) = (a_1(x))^2 + \cdots + (a_n(x))^2$, and recognize this as $Q(x) = \|Tx\|^2$, where $T : \mathbb{R}^n \to \mathbb{R}^n$ acts by a_j on the j^{th} output component. The Lem then follows by the existence of the operator norm/singular value decomposition. □

Local extrema of multivariate functions

Again, we are in pursuit of a theorem recognizing local extrema in the multivariate context, using data analogous to the second derivative/ 2^{nd} order Taylor polynomial in the single variable case.

Def: A critical point of a function $f: \mathbb{R}^n \to \mathbb{R}$ is a point $a \in \mathbb{R}^n$ with $D_a f = 0$.

Rmk: You might like to think of this as $D_a f$ not being full rank, which is only possible in the codomain $= \mathbb{R}$ case if $D_a f = 0$ outright.

Idea: If a function behaves locally like its 2^{nd} order Taylor polynomial, then at a critical point we have $f(a+h) - f(a) = \frac{1}{2}\sum_{i,j} \partial^2 f/\partial x_i \partial x_j|_a \, x_i x_j = Q$. From last time we saw that Q admits expression as $Q = (a_1)^2 + \cdots + (a_k)^2 - (b_1)^2 - \cdots - (b_l)^2$, which behaves as positive-definite on some subspace, negative-definite on some other, and degenerate on a third.

Thm: Take $U \subseteq \mathbb{R}^n$ open, $f: U \to \mathbb{R}$ twice continuously diff^{ble}, and $a \in U$ a critical point of f. If Q is nondegenerate and negative definite, a is a local max; if Q is nondegen. and pos. def, a is a local min; if the signature is (k, l) then $k > 0 \Rightarrow$ not a local max and $l \Rightarrow 0$ not a local min.

Pf: Start by decomposing $f(a+h) = f(a) + Q(h) + r(h)$, where $\lim_{h \to 0} r(h) \cdot \|h\|^{-2} = 0$. If Q is pos. def., then we can find a C with $C\|x\|^2 \leq Q(x)$, so that $(f(a+h) - f(a)) \cdot \|h\|^{-2} = (Q(h) + r(h))\|h\|^{-2} \geq C + r(h) \cdot \|h\|^{-2}$. It follows that the RHS is positive for small h, so so is the LHS, except at 0 where $f(a+0) - f(a) = 0$. This is the local min.

The second clause about local maxima follows identically — just replace f by $-f$. To see the claim about signature, restrict the behavior of f to the plane cut out by the basis for the negative definite part. The same proof witnesses a as a local max here, hence it cannot also be a local min as a function on the parent space $\mathcal{U}$. $\square$

<u>Thm</u>: For a critical point $a \in \mathcal{U}$ of $f : \mathcal{U} \to \mathbb{R}$, if Q has signature $(k>0, l>0)$ then a is called a <u>saddle</u>. In this case, in <u>every</u> nbhd of a there are points b with $f(b) > f(a)$ <u>and</u> points c with $f(c) < f(a)$.

<u>Rem</u>: If Q is degenerate, we can't really say anything; consider x^2+y^4, x^2-y^4, and x^2+y^3.

<u>Rem</u>: x^3-2xy^2 "goes up" in 3 directions and "goes down" in 3 others. Its quadratic form vanishes identically.

~~Alternating tensors (4.3)~~ The road ahead

Today we begin to pursue a generalization of the fundamental theorem of calculus. This will evidently be some work: the classical FTC says that differentiation + integration are inverse processes, but we have constructed $D: \{(\text{nice f}^{\text{ns}}\ \mathbb{R}^n \to \mathbb{R}^m, \text{points in } \mathbb{R}^n)\} \to \{\text{linear op}^{\text{rs}}\ T_a\mathbb{R}^n \to T_{f(a)}\mathbb{R}^m\}$ and $\int: \{(\text{nice functions } \mathbb{R}^n \to \mathbb{R}, \text{regions in } \mathbb{R}^n)\} \to \mathbb{R}$. These don't look like inverses at all.

Let's recast the classical story. An integral $\int_a^b f(t)\,dt$ has two components: a function $f(t)$ and a region $[a,b]$. I would like to think of this as a kind of pairing, à la last semester: $\langle f, [a,b]\rangle = \int_a^b f(t)\,dt$.

(One of) Our main theorem(s) from last semester was that a sufficiently nice pairing induces an iso$^{\text{m}}$ from one side to the dual of the other. The FTC can be interpreted as saying that f can be recovered from $\langle f, [0,x]\rangle$, evaluated at all the different values of x (like ℓ needed to be evaluated on a basis of V). (This story needs fudging: diff$^{\text{ble}}$ f$^{\text{ns}}$ do not form a f.d. v.s., so Riesz doesn't apply; "regions" don't form a v.s. at all, ...)

In the general situation, what might we find to mimic this? Start small, where a "region" is a curve in $\mathbb{R}^n$: $\gamma: \mathbb{R} \to \mathbb{R}^n$. This immediately gives complications: the same curve may be traced out by different parametrizing functions, and we would like not to depend on this. (i.e., at different speeds) Our solution is that, what the mystery object "M" is in the expression $\int_\gamma M$, it should be gifted both $\gamma(t)$ and $D_t\gamma$, as input, and then we will set $\int_\gamma M = \int_{\mathbb{R}} M(\gamma(t), \gamma'(t))\,dt$. "Currying" this, we have that for any point $x \in \mathbb{R}^n$, $M(x): T_x\mathbb{R}^n \to \mathbb{R}$.

<u>Def:</u> A <u>differential (1) form</u> on $\mathbb{R}^n$ is a function assigning $x \in \mathbb{R}^n$ to a linear, <u>functional</u> $T_x\mathbb{R}^n \to \mathbb{R}$. The v.s. of such is written $\Omega^1(\mathbb{R}^n)$.

Ex: Let's write dx for the functional dual to $\partial/\partial x$ and dy for the dual of $\partial/\partial y$. We can then think of $\omega = 3x\,dy + e^x dx$ as a kind of fitness function or diff^{l} eq^{n}: given a curve $\gamma(t) = (3t^2, 5\sin t)$, this data assembles into $\omega_{\gamma(t)}\gamma'(t) = (9t^2 dy + e^{3t^2} dx)\begin{pmatrix} 3t \\ 5\cos t \end{pmatrix} =$

$$= 3te^{3t^2} + 45t^2\cos t,$$

which is a function $\mathbb{R} \to \mathbb{R}$ that we can integrate. The vector $\begin{pmatrix} e^x \\ 3x \end{pmatrix}$ dual to ω is what ω "wants" to receive, and it's measuring how well $\gamma'(t)$ fits the bill — i.e., solves this PDE.

Ex: We also want the "top" case of the theory to be familiar. What about regions $\sigma: \mathbb{R}^n \to \mathbb{R}^n$ — i.e., volumes? We already have a theory of integration here: $\int_\sigma M = \int_{\mathbb{R}^n} M_{\sigma(x)} \cdot |\det M|\, dx_1 \cdots dx_n$. What does this mean for M? It means the action of M on $D_x\sigma$ must be by the determinant.

~~Def: A differential n-form on $\mathbb{R}^n$ is a multi-linear alternating function~~

Def: A differential n-form on $\mathbb{R}^n$ is an assignment from $x \in \mathbb{R}^n$ to multi-linear alternating functions $(T_x\mathbb{R}^n)^{\times n} \to \mathbb{R}$. The vector space of such is written $\Omega^n(\mathbb{R}^n)$.

Idea: A differential k-form on $\mathbb{R}^n$ is an assignment from $x \in \mathbb{R}^n$ to multilinear alternating functions $(T_x\mathbb{R}^n)^{\times k} \to \mathbb{R}$.

We are thus motivated to think hard about multilinear algebra for a bit before moving on. Eventually, we will return to the project of generalizing the FTC.

The alternating algebra

As promised, today we think about multi-linear algebra.

Def: A k-tensor on V is a multi-linear function $V^{\times k} \longrightarrow \mathbb{R}$, and the set of k-tensors is written $T^k(V)$.

Def: The tensor product $f \otimes g$ of ~~two~~ a k-tensor f + an l-tensor g is the $(k+l)$-tensor $(f \otimes g)(v_1, \ldots, v_{k+l}) = f(v_1, \ldots, v_k) \cdot g(v_{k+1}, \ldots, v_{k+l})$.

Lem: If $\varphi_1, \ldots, \varphi_n$ is a basis for V^*, then $\varphi_{i_1} \otimes \cdots \otimes \varphi_{i_k}$ is a basis for $T_k(V)$.

Pf: Take $v_1, \ldots, v_n$ dual to $\varphi_1, \ldots, \varphi_n$, and consider some input vectors $w_j = \sum_{i=1}^n c_{ij} v_i$. Then $t(w_1, \ldots, w_n) = t(\sum_{i_1=1}^n c_{i_1 1} v_{i_1}, \ldots, \sum_{i_k=1}^n c_{i_k k} v_{i_k})$
$= \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n c_{i_1,1} \cdots c_{i_k,k}\, t(v_{i_1}, \ldots, v_{i_k})$, so that $t = \sum \cdots \sum c_{--}\, \varphi_{i_1} \otimes \cdots \otimes \varphi_{i_k}$

Ex/Reminder: The determinant is an element of $T^{\dim V}(V)$. An inner product is a special element of $T^2(V)$. These both have special properties: alternating, symmetric, positive definite, ... — all from 25a. Lastly, $T^1(V)$ is just V^*.

Recalling our brief discussion of k-forms, we are interesting in alternating tensors.

Lem: For $\tau \in T^k(V)$, set $\mathrm{Alt}(\tau)(v_1, \ldots, v_k) = \sum_{\sigma \text{ perm.}} \frac{1}{k!} \cdot \mathrm{sign}(\sigma) \cdot \tau(v_{\sigma 1}, \ldots, v_{\sigma k})$. Alt is idempotent, has image the alternating tensors.

Pf: For ρ the permutation that swaps two indices, we have
$\mathrm{Alt}(\tau \circ \rho)(v_1, \ldots, v_k) = \frac{1}{k!} \sum_\sigma \mathrm{sign}(\sigma) \cdot \tau(v_{\sigma\rho 1}, \ldots, v_{\sigma\rho k}) = \frac{1}{k!} \sum_\sigma \mathrm{sign}(\sigma\rho^{-1})\, \tau(v_1, \ldots, v_k)$
and $\mathrm{sign}(\sigma\rho^{-1}) = \mathrm{sign}(\sigma) \cdot (-1)$. Then, if τ is already alternating, this sum collapses. Idempotence follows, as does image. $\square$

We'd like an analogue of the first lemma. For this, we need a product.

Def: For $\tau \in T^k V$, $\eta \in T^l V$, define $\tau \wedge \eta = \frac{(k+l)!}{k!\, l!} \mathrm{Alt}(\tau \otimes \eta)$.

This def$^{\underline{n}}$ has many reasonable properties: it's bilinear, alternating, respects the dual map $T^k(W) \longrightarrow T^k(V)$ induced by a map $f: V \longrightarrow W$, and it's associative (kinda hard, see Theorem 4-4.)

Lem: The set of all $\varphi_{i_1} \wedge \cdots \wedge \varphi_{i_k}$, $1 \leq i_1 < \cdots < i_k \leq n$, is a basis for $\Omega^k(V)$.

Pf: Write $\tau \in \Omega^k(V) \subseteq T^k(V)$ as $\tau = \sum_I c_I \varphi_{i_1} \otimes \cdots \otimes \varphi_{i_k}$. Then $\mathrm{Alt}(\tau) = \tau$, but $\mathrm{Alt}(\sum_I c_I \varphi_{i_1} \otimes \cdots \otimes \varphi_{i_k}) = \sum_I c_I \varphi_{i_1} \wedge \cdots \wedge \varphi_{i_k}$. $\square$

Cor: $\dim(V) = n$ means $\dim \Omega^k(V) = \binom{n}{k}$. $\square$

Cor: For any $\omega \in \Omega^n(V)$, and $u_j = \sum_i c_{ij} v_i$ vectors expressed in a basis (v_i),
$\omega(u_1, \ldots, u_n) = \det(c_{ij}) \cdot \omega(v_1, \ldots, v_n)$.

Pf: $\dim T^n(V) = \binom{n}{n} = 1$, so $\omega = k \cdot \det$ for some k. $\square$

Def/Cor: A choice of $\omega \in \Omega^n(V)$ partitions ordered bases into two sets: those with $\omega > 0$ and those with $\omega < 0$. A choice of preferred sign is called an orientation of V w/rt ω. We write $[v_1, \ldots, v_n] \in \{\pm 1\}$ for the orientation to which a given basis belongs under ω.

Rem: If V has an inner product, there is a unique ω for any ordered orthonormal basis $(v_1, \ldots, v_n)$ so that $\omega(v_1, \ldots, v_n) = \mu$, where $\mu \in \{\pm 1\}$ is some preferred orientation. The ω constructed in this way is called the volume element. Ex: $[e_1, \ldots, e_n] = 1$ in $\mathbb{R}^n$ determines det.

Rem: For $v_1, \ldots, v_{n-1} \in V$, there is a unique $v_n \in V$ such that $\langle w, v_n \rangle = \det(v_1 | \cdots | v_{n-1} | w)$. This vector v_n is often called the $\times$-product of $(v_1, \ldots, v_{n-1})$. For $n = 3$, there is the usual mnemonic

$$v_1 \times v_2 = \det \begin{pmatrix} "e_1" & "e_2" & "e_3" \\ \varphi_1 v_1 & \varphi_2 v_1 & \varphi_3 v_1 \\ \varphi_1 v_1 & \varphi_2 v_2 & \varphi_3 v_2 \end{pmatrix}.$$

Differentiation + k-forms (4.2)

Where do tensors arise in the theory of calculus? Consider a diff'ble f$^{\underline{n}}$ $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ and a point $a \in \mathbb{R}^n$ — then $D_a f : T_a\mathbb{R}^n \longrightarrow T_{f(a)}\mathbb{R} \cong \mathbb{R}$ is a linear f$^{\underline{n}}$, i.e., an element of $\Omega^1(T_a\mathbb{R}^n)$. We write $dx_j|_a$ for the functional dual to $e_j \in T_a\mathbb{R}^n$, so that $D_a f = \Sigma_{j=1}^n \partial f/\partial x_j|_a \cdot dx_j|_a$, and we write $df|_a \in \Omega^1(T_a\mathbb{R}^n)$ for this f$^{\underline{n}}$ considered as a tensor.

the coll$^{\underline{n}}$ of 1-forms we denote Ω^1.

Def: A 1-form ω is an assignment $(a \in \mathbb{R}^n) \longmapsto (\underset{\in \Omega^1(T_a\mathbb{R})}{\omega|_a} : T_a\mathbb{R}^n \to \mathbb{R})$, which can be expressed as $\omega|_a = \Sigma_{j=1}^n g_j(a) \cdot dx_j|_a$ for some f$^{\underline{ns}}$ $g_j : \mathbb{R}^n \longrightarrow \mathbb{R}$. The 1-form is cts/diff'ble/... if each g_j is.

Before moving on, it will be useful to have a toolbox of constructions on 1-forms related to this definition.

Def: For $g: \mathbb{R}^n \longrightarrow \mathbb{R}$, we set $(g \cdot \omega)|_a = g(a) \cdot \omega|_a$, the scaling.

Def: For $u: \mathbb{R}^m \longrightarrow \mathbb{R}^n$, we set $u^*\omega \in \Omega^1(T_*\mathbb{R}^m)$, the pullback or precomposite, by $u^*\omega|_b = \omega|_{u(b)}(D_b u(-))$.

Lem: (i) $u^*(dx_i) = \Sigma_{j=1}^m \partial u_i/\partial x_j \cdot dx_j$.

(ii) $u^*(\omega_1 + \omega_2) = u^*\omega_1 + u^*\omega_2$. (iii) $u^*(s \cdot \omega) = (s \circ u) \cdot (u^*\omega)$.

Pf of i: We can detect the dx_j component by feeding $u^*(dx_i)|_b$ the tangent vector e_j: $u^*(dx_i)|_b(e_j) = dx_i|_{u(b)}(D_b u(e_j)) = dx_i|_{u(b)}(\Sigma_{k=1}^n \partial u_k/\partial x_j|_{u(b)} \cdot e_k)$ $= \partial u_i/\partial x_j|_{u(b)}$, as claimed. □

the coll$^{\underline{n}}$ of k-forms we denote Ω^k.

Proceeding from here, it is also useful to have a notion of a k-form: these are assignments $(a \in \mathbb{R}^n) \longmapsto (\omega|_a \in \Omega^k(T_a\mathbb{R}^n))$, as well as products $\Omega^k \times \Omega^l \xrightarrow{\wedge} \Omega^{k+l}$. As noted last time, there is an especially interesting n-form:

Lem: If $f: \mathbb{R}^n \to \mathbb{R}^n$ is diff^{ble}, then $f^*(h\,dx_1 \wedge \cdots \wedge dx_n) = (h \circ f) \cdot (\det Df) \cdot dx_1 \wedge \cdots \wedge dx_n$.

Pf: $f^*(h \cdot dx_1 \wedge \cdots \wedge dx_n) = (h \circ f) \cdot f^*(dx_1 \wedge \cdots \wedge dx_n) = (h \circ f) \cdot (f^* dx_1 \wedge \cdots \wedge f^* dx_n)$

$$= (h \circ f) \cdot \left(\left(\sum_{j_1} \frac{\partial f_1}{\partial x_{j_1}} dx_{j_1} \right) \wedge \cdots \wedge \left(\sum_{j_n} \frac{\partial f_n}{\partial x_{j_n}} dx_{j_n} \right) \right).$$

Now note that $dx_j \wedge dx_j = 0$ and $dx_i \wedge dx_j = (-1) dx_j \wedge dx_i$.
The only surviving terms in this sum expansion are the permutation formula for the matrix $(\partial f_i / \partial x_j)_{i,j} = Df$. $\square$

If ω is itself diff^{ble}, we record its derivative as

$$d\omega = d\left(\sum_J g_J \cdot dx_{J_1} \wedge \cdots \wedge dx_{J_k} \right) = \sum_J (dg_J) \wedge dx_{J_1} \wedge \cdots \wedge dx_{J_k},$$

a $(k+1)$-form.

Lem: (i) $d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2$. (ii) $d(\omega_1 \wedge \omega_2) = d\omega_1 \wedge \omega_2 + (-1)^k \omega_1 \wedge d\omega_2$, for $\omega_1 \in \Omega^k$. (iii) $d(d\omega) = 0$. (iv) For $f: \mathbb{R}^n \to \mathbb{R}^m$ diff^{ble}, $f^*(d\omega) = d(f^*\omega)$

Pf: 1-3 are unpleasant derivations from the definition. At least 4 is nice: $f^*(d(\omega \wedge dx_i)) = f^*(d\omega \wedge dx_i + (-1)^k \omega \wedge ddx_i)$
$= f^*(d\omega \wedge dx_i) = f^*(d\omega) \wedge f^*(dx_i) = d(f^*\omega \wedge f^* dx_i) = d(f^*(\omega \wedge dx_i))$. $\square$

Rem: (iii) is a generalization of an old observation of ours.
Suppose that f is twice continuously diff^{ble}, say $\mathbb{R}^2 \to \mathbb{R}$.
Then $0 = ddf = d\left(\frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy \right) =$

$$= \left(\frac{\partial}{\partial x} \frac{\partial f}{\partial x} dx + \frac{\partial}{\partial y} \frac{\partial f}{\partial x} dy \right) \wedge dx + \left(\frac{\partial}{\partial x} \frac{\partial f}{\partial y} dx + \frac{\partial}{\partial y} \frac{\partial f}{\partial y} dy \right) \wedge dy$$

$$= -\frac{\partial^2 f}{\partial y \partial x} dx \wedge dy + \frac{\partial^2 f}{\partial x \partial y} dx \wedge dy.$$ So, this is equivalent to stating that mixed partials commute.

The Fundamental Theorem of Calculus

We've been using the classic FTC off-and-on, and we're about to use it in a really big way. It seems irresponsible not to prove it.

Thm: Let $f : [a,b] \to \mathbb{R}$ be integrable, and define $F(x) = \int_a^x f(t)\,dt$.
If f is cts at c then F is diff^{ble} at c, and $F'(c) = f(c)$.

Pf: For simplicity, take $a < c < b$, and consider $F'(c) = \lim_{h \to 0} \frac{F(c+h) - F(c)}{h}$.

We have a bound on this difference quotient:

provided $h > 0$ $\rightsquigarrow$ $$\frac{1}{h} \cdot h \cdot m_{f|[c,c+h]} \leq \frac{F(c+h) - F(c)}{h} \leq \frac{1}{h} \cdot h \cdot M_{f|[c,c+h]}.$$

Because f is ~~integrable~~ continuous, m_f and M_f approach $f(c)$ (since $\lim_{h \to 0} (M_f - m_f)$ is the oscillation of f at c).
A similar proof works for $h < 0$. $\square$

Rem: The theorem/proof also works for $\int_x^b f(t)\,dt$, whose derivative at c is $-f(c)$. (You can also use $\int_x^b = \int_a^b - \int_a^x$.)

Thm: Suppose f is continuously diff^{ble} on $[a,b]$; then
$$\int_a^b \left(\frac{df}{dt}\right) dt = f(b) - f(a).$$

Pf: Consider the function $\varphi(x) = \int_a^x (df/dt)\,dt$. By the previous theorem, $\varphi'(x) = f'(x)$; and hence the difference $\varphi - f$ is a constant function. $\varphi(a) = 0$, but $f(a) = f(a)$, and hence $\varphi(b) - f(b) = f(a)$. $\square$

In fact, we can get away with less — f need not be continuously diff^{ble}.

<u>Thm</u>: Let f be diff^{ble} on $[a,b]$ and let f' be <u>integrable</u> on $[a,b]$.
Then $\int_a^b (df/dx)\,dx = f(b) - f(a)$.

Pf: Partition $[a,b]$ into $a = t_0 < t_1 < \cdots < t_n = b$. For each subrectangle $S_j = [t_{j-1}, t_j]$, there's a point $x_j \in S_j$ such that $f\cancel{g}(t_j) - \cancel{g}f(t_{j-1}) = f'(x_j)(t_j - t_{j-1})$, by the MVT. As usual, set m_j to be the inf of f' on S_j and M_j to be the sup.

Then $m_j(t_j - t_{j-1}) \le f'(x_j)(t_j - t_{j-1}) = f(t_j) - f(t_{j-1}) \le M_j(t_j - t_{j-1})$.

Sum: $\sum_j m_j \,\mathrm{vol}(S_j) \le f(b) - f(a) \le \sum_j M_j \,\mathrm{vol}(S_j)$,

a/k/a $\mathcal{L}_P(f') \le f(b) - f(a) \le \mathcal{U}_P(f)$.

Hence, if f' is integrable, $f(b) - f(a)$ is its integral. $\square$

[The] Poincaré Lemma

[Las]t time we left off with the observation that $d d\omega = 0$ is a [ge]neralization of our theorem that partials commute for nice f^ns. [W]e also showed a small converse to this: if $\partial g_1/\partial y = \partial g_2/\partial x$ and [g_1], $g_2 : \mathbb{R}^2 \to \mathbb{R}$ are sufficiently nice, then $\exists f : \mathbb{R}^2 \to \mathbb{R}$ satisfying [∂]$f/\partial x = g_1$ and $\partial f/\partial y = g_2$.

[E]x: If $g_1 + g_2$ are not nice, this fails. For instance, $g_1 = \frac{-y}{x^2+y^2}$ and $g_2 = \frac{x}{x^2+y^2}$, defined on $\mathbb{R}^2 \setminus 0$. The associated diff^l form is $d\theta$, so if $\exists f$ with $df = d\theta$, then f would be independent of r and $f = \theta +$ constant, which is nonsense; the essential problem is that θ jumps from 0 to 2π near, say, the x-axis.

[L]et's take $\omega = df$ and use this strategy to reconstruct f:

$$f(x) = \int_0^1 \frac{d}{dt} f(tx)\,dt = \int_0^1 \sum_{j=1}^n \frac{\partial f}{\partial x_j}(tx) \cdot x_j\,dt = \int_0^1 \sum_{j=1}^n \omega_j(tx) \cdot x_j\,dt.$$

[Thi]s works fine when the line $0 \to tx$ is contained in A, which is open [an]d f is continuously diff^ble on A. This "$0 \to tx$" property is called **star-shaped.**

[T]hm: If $A \subseteq \mathbb{R}^n$ is an open star-shaped set, then every **closed** form (i.e., $d\omega = 0$) on A is **exact** (i.e., $\omega = df$ for some f).

[Pr]f: We will define a function I (for "integral") $\Omega^k \to \Omega^{k-1}$ such that $I(0) = 0$ and $\omega = I(d\omega) + d(I\omega)$. If $d\omega = 0$, we are thus done.

Let $\omega = \sum_J \omega_J\, dx_{J_1} \wedge \cdots \wedge dx_{J_k}$. Using A's star-shaped-ness, set

$$I\omega(x) = \sum_J \sum_{\alpha=1}^k (-1)^{\alpha-1} \left(\int_0^1 t^{k-1} \omega_J(tx)\,dt \right) x_{J_\alpha}\, dx_{J_1} \wedge \cdots \wedge \widehat{dx_{J_\alpha}} \wedge \cdots \wedge dx_{J_k}$$

where "$\widehat{dx_{J_\alpha}}$" means we skip this index.

we're in business.

$$) = d\left(\sum_J \sum_{a=1}^{k} (-1)^{a-1}\left(\int_0^1 t^{k-1}\omega_J(tx)\,dt\right) x_{J_a}\; dx_{J_1}\wedge\cdots\wedge\widehat{dx_{J_a}}\wedge\cdots\wedge dx_{J_k}\right)$$

$$= \sum_{a=1}^{k}\sum_{b=1}^{n} (-1)^{a-1}\frac{\partial}{\partial x_b}\left[\int_0^1 t^{k-1}\omega_J(tx)\,dt\cdot x_{J_a}\right] dx_b\wedge dx_{J_1}\wedge\cdots\wedge\widehat{dx_{J_a}}\wedge\cdots\wedge dx_{J_k}$$

$$= \left(k\int_0^1 t^{k-1}\omega_J(tx)\,dt\; dx_{J_1}\wedge\cdots\wedge dx_{J_k}\right.$$ ← second part of the product rule, only matches $b = J_a$.

$$\left. + \sum_{a=1}^{k}\sum_{b=1}^{n}(-1)^{a-1}\int_0^1 t^k\frac{\partial\omega_J}{\partial x_b}(tx)\,dt\cdot x_{J_a}\; dx_b\wedge dx_{J_1}\wedge\cdots\wedge\widehat{dx_{J_a}}\wedge\cdots\wedge dx_{J_k}\right)$$

$$\sum_J\sum_{b=1}^{n}\frac{\partial\omega_J}{\partial x_b}\cdot dx_b\wedge dx_{J_1}\wedge\cdots\wedge dx_{J_k}.$$

$$= \sum_J\sum_{b=1}^{n}\left(\int_0^1 t^k\frac{\partial\omega_J}{\partial x_b}(tx)\,dt\right)x_b\; dx_{J_1}\wedge\cdots\wedge dx_{J_k}$$

$$- \sum_J\sum_{b=1}^{n}\sum_{a=1}^{k}(-1)^{a-1}\left(\int_0^1 t^k\frac{\partial\omega_J}{\partial x_b}(tx)\,dt\right)x_{J_a}\; dx_b\wedge dx_{J_1}\wedge\cdots\wedge\widehat{dx_{J_a}}\wedge\cdots\wedge dx_{J_k}.$$

m cancel in the sum, giving

$$\omega) = \sum_J k\left(\int_0^1 t^{k-1}\omega_J(tx)\,dt\right)dx_{J_1}\wedge\cdots\wedge dx_{J_k}$$

$$+ \sum_J\sum_{b=1}^{n}\left(\int_0^1 t^k x_b\frac{\partial\omega_J}{\partial x_b}(tx)\,dt\right)dx_{J_1}\wedge\cdots\wedge dx_{J_k}$$

$$= \sum_J\left(\int_0^1\frac{d}{dt}\left(t^k\omega_J(tx)\right)dt\right)dx_{J_1}\wedge\cdots\wedge dx_{J_k}$$

$$\overset{\text{FTC}}{=} \sum_J \omega_J\, dx_J = \omega. \quad \ddot\smile \qquad \square$$

Stokes's Theorem

Today we realize our dream of k-forms as subjects of integration, and we prove one of the major theorems in this framework.

Def: A *singular k-cube* is a f^n $c : [0,1]^k \longrightarrow \mathbb{R}^n$, always cts + often diff^{ble}.
A *k-chain* is a formal sum of such: $k_1 c_1 + \cdots + k_\ell c_\ell$, $k_j \in \mathbb{R}$.
These form the vector space of domains of integration from before.

Ex: The cube $c(x) = x$ is called the *standard n-cube*. Any 1-cube is often called a *curve*.

Def: If $\omega = f\, dx_1 \wedge \cdots \wedge dx_k$ is a k-form on $[0,1]^{\times k}$, we define $\int_{[0,1]^k} \omega := \int_{[0,1]^{\times k}} f\, dx_1 \cdots dx_k$. More generally, if $\sigma = \sum_i k_j c_j$ is a *chain*, we set $\int_\sigma \omega = \sum_i k_j \int_{c_j} \omega = \sum_i k_j \int_{[0,1]^k} c^* \omega$; this last step builds u-substitution into the definition of integration.
Def: In the exceptional case of a 0-chain ω, we set $\int_{\{a\}} \omega = \omega(a)$.

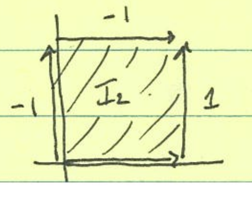
Ex: Take $c : [0,1] \longrightarrow \mathbb{R}^2$ a curve and $\omega = f\,dx + g\,dy$ a 1-form.
Then $\int_c \omega = \int_c (f\,dx + g\,dy) = \int_0^1 f(c(t)) \frac{\partial c_x}{\partial t} dt + \int_0^1 g(c(t)) \cdot \frac{\partial c_y}{\partial t} dt$.

It turns out that this RHS can be written as $\lim \sum_{i=1}^{n} \big[f(c(t_i)) \cdot (c_x(t_i) - c_x(t_{i-1}))$ $+ g(c(t_i))(c_y(t_i) - c_y(t_{i-1}))\big]$ with the limit taken over partitions.

Recall that we have recently proved $\int_a^b \frac{df}{dx} \cdot dx = f(b) - f(a)$ in the one-variable case. Set $a = 0, b = 1$, and $\omega = df = \frac{df}{dx} \cdot dx$, then this can be rewritten as $\int_{[0,1]} df = \int_{\{0\} + \{1\}} f$. — we traded "$d$" for something acting on chains.

Def: The *boundary* of the standard n-cube I_n is the $(n-1)$-chain

$$\partial I_n = \sum_{i=1}^{n} \sum_{a=0}^{1} (-1)^{i+a} \underbrace{\left(I_n(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n) \right)}_{I_{(i,a)}}.$$



like the alternating sum gave $dd\omega = 0$, $\partial\partial c = 0$.

like there was a partial converse $d\omega = 0 \Rightarrow \omega = df$, there is a

converse $\partial c = 0 \Rightarrow c = \partial c'$. An exhaustive analysis of this is alg. top.

es's Thm: If ω is a $(k-1)$-form on an open set $A \subseteq \mathbb{R}^n$ and

a k-chain on A, then $\int_c d\omega = \int_{\partial c} \omega$.

ng of $\int_c f$ as $\langle f, c \rangle$, these are adjoints: $\langle df, c \rangle = \langle f, \partial c \rangle$.

finitions immediately reduce us to the case $\omega = f\,dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k$,

ted over the standard n-cube. Note that such an ω restricts

on most of the faces of the standard n-cube: $I_{k,j,a}^{*}\omega = 0$

$i = j$, in which case $\int_{[0,1]^{k-1}} I_{k,j,a}^{*}\omega = \int_0^1 \cdots \int_0^1 f(x_1, \dots, a, \dots, x_k)\,dx_1 \cdots dx_k$.

all these up, $\displaystyle\int_{\partial I_k} f\,dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k = \sum_{j=1}^{k} \sum_{a=0}^{1} (-1)^{j+a} \int_{[0,1]^{k-1}} I_{k,j,a}^{*}\omega$

$$1)^{i+1} \int_{[0,1]^k} f(x_1, \dots, 1, \dots, x_k)\,dx_1 \cdots dx_k + (-1)^i \int_{[0,1]^k} f(x_1, \dots, 0, \dots, x_k)\,dx_1 \cdots dx_k.$$

$$^{i-1} \int_{[0,1]^k} \big(f(x_1, \dots, 1, \dots, x_k) - f(x_1, \dots, 0, \dots, x_k)\big)\,dx_1 \cdots dx_k.$$

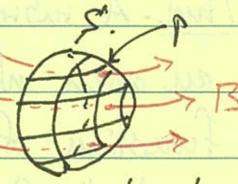
to single out the i^{th} coord. + use the FTC to get

$$^{i-1} \int_{[0,1]^k} \left(\frac{\partial f}{\partial x_i}(x_1, \dots, x_k)\right) dx_1 \wedge \cdots \wedge dx_k.$$

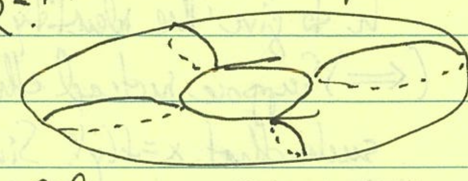
$$]^k \; d\left(f\,dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k\right). \qquad \square$$

Manifolds (5-1)

A common problem in basic electromagnetics is to compute the flux of something (typically an electric or magnetic field) through some surface. We can almost model this situation with differential forms: the "flux at a point" sounds like an inner product of B_p with a normal vector to S at p, and these pairings are the kinds of things forms are good at. What's less clear is how to add these up into an integral: S is not an open set in $\mathbb{R}^3$.



It looks "more like $\mathbb{R}^2$", but it is also easy to produce examples of surfaces S that decidedly do not "look like $\mathbb{R}^2$". Today we give a name to these objects.



Def: A subset $M \subseteq \mathbb{R}^n$ is a k-dim'l manifold when $\forall\, m \in M$ there is an open $U \ni m$, an open $V \subseteq \mathbb{R}^n$, and a diffeomorphism (diff'ble, invertible, diff'ble inverse) $h: U \to V$ s.t. $h(U \cap M) = V \cap (\mathbb{R}^k \times 0)$. Furthermore M is k-dim'l with boundary when $\forall\, m \in M$ either the above holds or $h(U \cap M) = V \cap (\mathbb{R}_{\geq 0} \times \mathbb{R}^{k-1} \times 0)$.

Slogan: Manifolds look locally like $\mathbb{R}^k$ (or $\mathbb{R}^{k-1} \times \mathbb{R}_{\geq 0}$).

Back when we were studying differential calculus, we proved a theorem (the "local coordinates theorem") that grants us easy access to these objects.

Cor: Let $A \subseteq \mathbb{R}^n$ be open + $g: A \to \mathbb{R}^p$ diff'ble with $D_a g$ of full rank p whenever $g(a) = 0$. Then $g^{-1}(0)$ is an $(n-p)$-dim'l manifold. $\square$

Ex: S^n is an n-dim'l manifold in $\mathbb{R}^{n+1}$, using $g: x \mapsto \|x\| - 1$.

There is a kind of local converse to this theorem, — not all manifolds arise this way, but locally they do.

<u>Thm</u>: A subset $M \subseteq \mathbb{R}^n$ is a k-dim^l manifold iff $\forall x \in M$ there is an open nbhd $\mathcal{U} \ni x$, an open $W \subseteq \mathbb{R}^k$, and an injective diff^{ble} function $f: W \longrightarrow \mathbb{R}^n$ such that $f(W) = M \cap \mathcal{U}$, $D_y f$ has rank k for each $y \in W$, and $f^{-1}: f(W) \longrightarrow W$ is continuous.

Pf: ($\Longrightarrow$) If M is known to be a manifold, then the diffeo^s provide a recipe for f: set $f(y) = h^{-1}(y, 0)$. Because f composes with h to give the identity, it can't be any other case than $\operatorname{rk} D_y f = k$.

($\Longleftarrow$) Suppose instead that we have an $f: W \longrightarrow \mathbb{R}^n$, and select y such that $x = f(y)$. Since $D_y f$ has rank k, pick k rows witnessing this — may as well ~~be~~ assume them to be the first k, so that $(\partial f_i/\partial x_j)_{i,j=1}^k$ is of rank k. The idea is to fix the distortion in f to get exactly what we want.

Set $g(a,b) = f(a) + (0,b)$, with block U.T. derivative, so that $\det D_{(a,b)} g = \det (\partial f_i/\partial x_j)_{i,j=1}^k \neq 0$. We apply the inverse f^n theorem: there are open $V_1' \ni (y,0)$ and $V_2' \ni g(y,0) = x$ such that $g|_{V_1'}$ has diff^{ble} inverse $h: V_2' \longrightarrow V_1'$. We need to check that h carries $V_2' \cap \mathcal{U} \rightsquigarrow V_2 \cap M$ into the standard hyperplane. Note $V_2 \cap M = \{f(a) \mid (a,0) \in V_1 := g^{-1}(V_2)\} = \{g(a,0) \mid (a,0) \in V_1\}$, so that $h(V_2 \cap M) = V_1 \cap (\mathbb{R}^k \times 0)$.

$\square$

<u>Rem</u>: Given two coordinate patches, the composite has everywhere nonsingular Jacobian.

Forms on manifolds + orientations (5-2)

We want to speak of forms on a manifold, which is an assignment from (tuples of) tangent vectors to $\mathbb{R}$ — and so we need to make sense of the space of tangent vectors to a point $x \in M$, M a mfld.

For a point $x \in M$, let $f: W \longrightarrow \mathbb{R}^n$ be a coordinate patch and let $g: V \longrightarrow \mathbb{R}^n$ be another. The idea is to set $T_x M$ to be the image of $T_{f^{-1}(x)} W$ under $D_{f^{-1}(x)} f$ — but we are worried this depends upon f. The chain rule gets us out of trouble:

$$D_{f^{-1}(x)} f = D_{f^{-1}(x)} (g \circ g^{-1} \circ f) = D_{g^{-1}(x)} g \circ D_{f^{-1}(x)} g^{-1} \circ f,$$

and this second component is a linear iso $\underline{\underline{m}}$, so the image of $D_{g^{-1}(x)} g$ is the same.

Essentially the same considerations give us access to diff$\underline{l}$ forms on M and to functions valued in $(T_* M)^{\otimes k}$ ("vector fields"). A form defined on an open nbhd of M restricts to a unique form defined on the source of a coordinate patch, and its value on a tangent vector is invariant under the choice of coordinate patch.

Lem: The derivative of a k-form ω can be calculated in any coordinate patch $f: W \to M$: $d(f^* \omega) = f^*(d\omega)$. □

(Again, there is an obnoxious invariance check to be done here that amounts to investigating the behavior of $g^{-1} \circ f$ for another patch g.)

Recall that an orientation of $\Omega^k(\mathbb{R}^k)$ was a choice of basis vector for this top exterior power. Since manifolds are only locally modeled on $\mathbb{R}^k$, we may have to make this choice many times.

Def: A choice of orientation $\mu_x \in \Omega^k(T_x M)$ for each point $x \in M$ is called consistent when for any coordinate patch $f: W \to M$ and pair of points $x, y \in im(f)$, $\mu_{f(x)} = [f_*(e_{1,x}), \dots, f_*(e_{k,x})]$ iff $\mu_{f(y)} = [g_*(e_{1,y}), \dots, g_*(e_{k,y})]$, where $e_{j,w}$ denotes the j^{th} standard basis vector in $T_w \mathbb{R}^k$ and $[\dots]$ denote the standard orientation on $\mathbb{R}^k$.

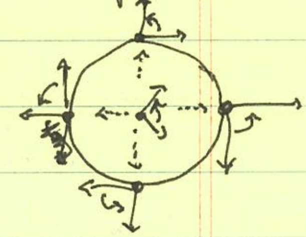
Def: f is orientation-preserving when μ_x agree with the std orientation on $\mathbb{R}^k$ for all $x \in im\, f$.

Lem: If M is consistently oriented, it is enough to check this for any $x \in im\, f$. □

Lem: If M is consistently oriented + f is not orientation-preserving it can be made orientation-preserving by composing with T, det T = −1.
$\Longrightarrow$ Always $\exists$ a system of coord charts, all orientation-preserving. □

Rem: Not all manifolds admit consistent orientations! Ex: Möbius band.

This can be adapted to the boundary case by noting that the "extra" vector in the orientation form gives a choice of "outward-facing-normal".



Rem: An orientation on M w/ ∂ also induces an orientation on ∂M.

Stokes's Thm on manifolds (5-3)

Today we hook our theory of integration up to manifolds. We were already defining our integrals through funny pullbacks, so the definitions work basically the same way.

Def: For $c: [0,1]^p \to M$ a singular p-cube in a k-manifold M and ω a p-form on M, we define $\int_c \omega = \int_{[0,1]^k} c^*(\omega)$ as before.

Rem: In the case $p=k$, we will assume that all our cubes c can be enlarged to coordinate patches $[0,1]^k \xrightarrow{c} M$ for $V \subseteq \mathbb{R}^k$ open. ($[0,1]^k \subseteq V \xrightarrow{f} M$)

(5-4) Lem: Integrals computed this way are independent of choice of coordinate: suppose c_1, c_2 are orientation-preserving k-cubes + ω is a k-form such that $\omega = 0$ off of $\operatorname{im} c_1 \cap \operatorname{im} c_2$. Then $\int_{c_1} \omega = \int_{c_2} \omega$.

"Pf": This is a computation with the u-substitution theorem. □

Cor: If ω vanishes outside of some singular k-cube c, we may as well define $\int_M \omega = \int_c \omega$. □

Def: Let $\mathcal{O}$ be a cover of M such that each member U of the cover lies inside the image of some singular k-cube c_U, and let Φ be a partition of unity subordinate to this cover. Then we define
$$\int_M \omega = \sum_{\varphi \in \Phi} \int_M \omega \cdot \varphi$$
when this sum converges (e.g., if M is compact).

Rem: Again, this can all be done to handle manifolds w/ ∂.

<u>Thm</u>: Let M be a compact, oriented k-manifold w/ ∂, and let ω be a $(k-1)$-form on M. Then $\int_M d\omega = \int_{\partial M} \omega$.

<u>Pf</u>: Our intention is to use partitions of unity to break into two cases: singular cubes that touch the boundary + those that live in the bulk.

(<u>Bulk</u>:) Let c be a k-cube in $M \setminus \partial M$, and say $\omega = 0$ off of c. Then the $\mathbb{R}^k$ version of Stokes' says $\int_c d\omega = \int_{\partial c} \omega$, but $\int_{\partial c} \omega = 0$ because $\omega|_{\partial c} = 0$, and lastly $\int_M \omega = \int_{\partial c} \omega = 0$ by vanishing off of c.

(<u>Boundary</u>:) Take c to be a singular k-cube with face $c_{k,0}$ in ∂M, and no others, and again let ω be a k-form vanishing off of c. We again have $\int_M d\omega = \int_c d\omega = \int_{\partial c} \omega = \int_{\partial M} \omega$.

(<u>General</u>:) Pick a cover carving M up into these two types, and pick a partition of unity Φ subordinate to the cover. The key observation is that Φ does not distort closedness by too much:

$$\sum_{\varphi \in \Phi} d(\varphi \cdot \omega) = \sum_{\varphi \in \Phi} \left(d\varphi \wedge \omega + \cancel{\varphi \cdot d\omega} \right) = d\left(\sum_{\varphi \in \Phi} \varphi \right) \wedge \omega = d(1) \wedge \omega = 0.$$

We couple this to the above calculations to get

$$\int_M d\omega = \sum_{\varphi \in \Phi} \int_M \varphi \cdot d\omega = \sum_{\varphi \in \Phi} \int_M d(\varphi \cdot \omega) \overset{B+B}{=} \sum_{\varphi \in \Phi} \int_{\partial M} \varphi \cdot \omega = \int_{\partial M} \omega. \quad \square$$

Volume elements (5-4)

Suppose $M \subseteq \mathbb{R}^n$ is an oriented k-dim^l manifold. There's a natural k-form on M defined by μ (and the inner product on $T_\cdot \mathbb{R}^n$), called the volume form. It's often denoted "dV", because mathematicians are fond of equations like "$V = \int_M dV$", and not because of a connection to the exterior derivative.

It's a fun exercise to make this concrete. Suppose that M is a curve embedded in $\mathbb{R}^2$. The orientations of M and of $\mathbb{R}^2$ determine an outward normal $n(x) \in T_x \mathbb{R}^2$ at each point $x \in M$. This gives a formula for dV: $dV(v) = \det\begin{pmatrix} n(x) \\ v \end{pmatrix}$, since this functional $T_x M \to \mathbb{R}$ has the right kernel and the right normalization. Equivalently, we have $dV(v) = \|v\|$ for any v with $\mu_x(v) = 1$.

To compute the arclength of a curve, we need to understand how to compute $\int_0^1 c^*(dA)$ for singular 1-cubes $c : [0,1] \to M$. Suppose that M itself is traced out by a map $\gamma : [0,1] \to \mathbb{R}^2$, and that we are integrating over this same 1-cube. Then:

$$\int_0^1 \gamma^*(dV) = \int_0^1 \gamma^*\left(\left\langle \left(\frac{\partial \gamma_x}{\partial t}, \frac{\partial \gamma_y}{\partial t}\right), - \right\rangle \Big/ \sqrt{\left(\frac{\partial \gamma_x}{\partial t}\right)^2 + \left(\frac{\partial \gamma_y}{\partial t}\right)^2}\right)$$

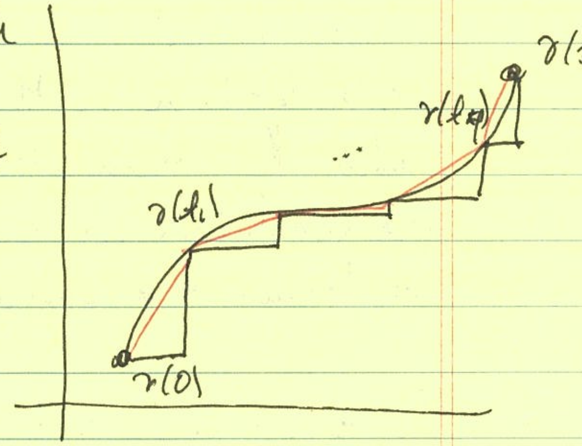
$$= \int_0^1 \gamma^*\left(\left(\frac{\partial \gamma_x}{\partial t} \cdot dx + \frac{\partial \gamma_y}{\partial t} dy\right) \Big/ \sqrt{\left(\frac{\partial \gamma_x}{\partial t}\right)^2 + \left(\frac{\partial \gamma_y}{\partial t}\right)^2}\right)$$

$$= \int_0^1 \left(\frac{\partial \gamma_x}{\partial t} \cdot \frac{\partial \gamma_x}{\partial t} dt + \frac{\partial \gamma_y}{\partial t} \cdot \frac{\partial \gamma_y}{\partial t} \cdot dt\right) \Big/ \sqrt{\left(\frac{\partial \gamma_x}{\partial t}\right)^2 + \left(\frac{\partial \gamma_y}{\partial t}\right)^2}$$

$$= \int_0^1 \sqrt{\left(\frac{\partial \gamma_x}{\partial t}\right)^2 + \left(\frac{\partial \gamma_y}{\partial t}\right)^2}\, dt.$$

In general, this is pretty hard, since integrating square roots is pretty hard. There's a similarly nasty formula for 2-manifolds in $\mathbb{R}^3$.

Rem: This last expression has geometric content. A "more obvious" method of defining the arclength of a curve is to break the curve into segments, as at right, and to note that the lengths of the segments tend to $\sqrt{\left(\frac{\partial x}{\partial t}\right)^2+\left(\frac{\partial y}{\partial t}\right)^2}$ as mesh(P of t) $\longrightarrow 0$.

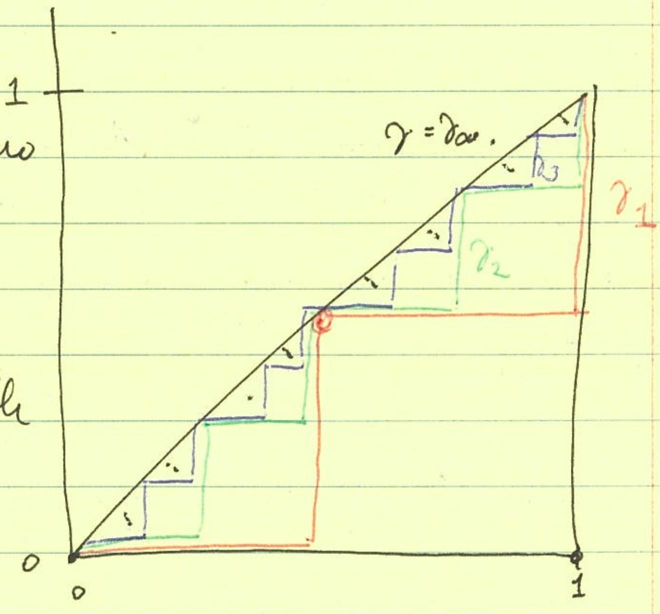


If γ is sufficiently differentiable, it can be shown that these two methods indeed give the same answer.

Instructive warning: If γ is not very diff^{ble}, arclength (and dV generally) is very poorly behaved. Consider the following family of piecewise diff^{ble} curves γ_n, which approach γ_∞, a straightline, pointwise.

The arclength of γ_∞ is $\sqrt{2}$, but the arclengths of each γ_n, no matter the value of n, is 2.

(It is fun to think about why it is OK to build integration with "flat approximations", but not arclength. ☺)



Vector calculus in $\mathbb{R}^3$ (5-5)

As a finishing stroke for this class, we illustrate how this last batch of theorems degenerates to the classical theorems of vector calculus. The main bonus to working in $\mathbb{R}^3$ is that 2-forms admit a very easy expression: Riesz pairs 1-forms with vector fields, the cross-product pairs 2-forms with vector fields and 3-forms pair with f^{ns} by the orientation form.

0-forms	1-forms	2-forms	3-forms
f^{ns}	$\varphi_x dx + \varphi_y dy + \varphi_z dz$	$\varphi_x dy \wedge dz + \cdots$	$\varphi\, dx\wedge dy \wedge dz$
$\updownarrow$	$\updownarrow$ Riesz	$\updownarrow$	$\updownarrow$
f^{ns}	$\left\langle \begin{pmatrix}\varphi_x\\ \varphi_y\\ \varphi_z\end{pmatrix}, - \right\rangle$ vector fields	$\left\langle \begin{pmatrix}\varphi_x\\ \varphi_y\\ \varphi_z\end{pmatrix}, (-)\times(-)\right\rangle$ vector fields	f^{ns}

The differentiation operators give assignments between these.

$$\begin{array}{ccccccc}
f^{ns} & \xrightarrow{\nabla} & \text{vector fields} & \xrightarrow{\nabla\times} & \text{vector fields} & \xrightarrow{\nabla\cdot} & f^{ns} \\
\cong & & \cong & & \cong & & \cong \\
\text{0-forms} & \xrightarrow{d} & \text{1-forms} & \xrightarrow{d} & \text{2-forms} & \xrightarrow{d} & \text{3-forms} \\
& \text{FTC} & & \text{Stokes'/Green's Thm} & & \text{Divergence Thm} & \\
\text{points} & \xleftarrow{\partial} & \text{lines} & \xleftarrow{\partial} & \text{surfaces} & \xleftarrow{\partial} & \text{solids}
\end{array}$$

The three maps on top are given by

$$f \overset{\nabla}{\longmapsto} \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right) \updownarrow \quad f \overset{d}{\longmapsto} \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz,$$

"gradient"

$$\varphi_x dx + \varphi_y dy + \varphi_z dz \overset{d}{\longmapsto} \left(\frac{\partial \varphi_y}{\partial x} - \frac{\partial \varphi_x}{\partial y}\right) dx\wedge dy + \left(\frac{\partial \varphi_z}{\partial y} - \frac{\partial \varphi_y}{\partial z}\right) dy \wedge dz + \left(\frac{\partial \varphi_x}{\partial z} - \frac{\partial \varphi_z}{\partial x}\right) dz \wedge dx$$

"curl" OR "rotation"

$$(\varphi_x, \varphi_y, \varphi_z) \overset{\nabla\times}{\longmapsto} \text{"}\det\begin{pmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ \varphi_x & \varphi_y & \varphi_z \\ \partial/\partial x & \partial/\partial y & \partial/\partial z\end{pmatrix}\text{"},$$

and

$$\varphi_x dy\wedge dz + \varphi_y dz \wedge dx + \varphi_z dx \wedge dy \overset{d}{\longmapsto} \left(\frac{\partial \varphi_x}{\partial x} + \frac{\partial \varphi_y}{\partial y} + \frac{\partial \varphi_z}{\partial z}\right) dx\wedge dy \wedge dz$$

"divergence"

$$(\varphi_x, \varphi_y, \varphi_z) \overset{\nabla\cdot}{\longmapsto} \frac{\partial \varphi_x}{\partial x} + \frac{\partial \varphi_y}{\partial y} + \frac{\partial \varphi_z}{\partial z}.$$

Our Stokes's theorem then specializes to three separate theorems:

<u>Thm</u> (Fundamental Theorem of Calc. for Line Integrals): Let $\gamma_1 : [0,1] \to \mathbb{R}^3$ be a smooth map and ~~(f_x, f_y, f_z) be a vector field~~ f be a function, diff'ble, on $\mathbb{R}^3$. Then $\displaystyle\int_{\gamma_1} \nabla f = f(\gamma(1)) - f(\gamma(0))$. $\square$

<u>Thm</u> ("Stokes's Theorem"): Let M_2 be a 2-manifold with boundary in $\mathbb{R}^3$, and let (f_x, f_y, f_z) be a diff'ble vector field on $\mathbb{R}^3$. Then $\displaystyle\int_{M_2} \nabla \times \begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} = \int_{\partial M_2} \begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix}$. $\square$

<u>Thm</u> (Divergence Theorem): Let M be a 3-manifold with boundary in $\mathbb{R}^3$, and let (f_x, f_y, f_z) be a diff'ble vector field in $\mathbb{R}^3$. Then $\displaystyle\int_{M_3} \nabla \cdot f = \int_{\partial M_3} f$. $\square$

These 3 theorems have had <u>profound</u> effect on mathematics and science. Here's one major one (out of literally thousands): <u>Maxwell's equations</u> in electromagnetics says that the magnetic field B and the electric field E satisfy $dB = 0$ and $dE = -\partial B/\partial t$. <u>All</u> the basic theorems about electrostatics (i.e., $\partial E/\partial t = 0$ and $\partial B/\partial t = 0$) then amount to applications of Stokes theorem, by observing that because $dB = 0$, there is a <u>magnetic potential</u>, or by observing that the charge in a region can be measured using the <u>divergence theorem</u>.

Postlude

My goodness, we have done a *lot* of mathematics!! One of the (not-so) secret goals of this course was to introduce you to so much math that you could jump off to any other 100-level course (or at least reg$^{\underline{ch}}$) that the department has to offer.

- Algebra: We spent all of 25a wrapped up in linear algebra, mostly over $\mathbb{C}$ or $\mathbb{R}$. There are lots of ways to make this "more interesting": for instance, you can work over $\mathbb{F}_p$, or over a ring without division, and build up an associated theory of vector spaces (123). Another direction to take this in is representation theory. A way to motivate this is that the basic object of a group arises as the symmetries (i.e., automorphisms) of a given object. For example, $\mathbb{R}^n$ has symmetry group $\mathcal{L}(\mathbb{R}^n)^\times = M_{n\times n}(\mathbb{R})^\times$. Finding recognizable subgp$^{\underline{s}}$ (e.g., $C_2^{\times n}$) is an interesting geometric enterprise with lots of consequences.
- Analysis: In 25b we focused primarily on analysis, and we left many loose ends unexplored. For instance, the measure of a set remains mysterious, but obviously central to a robust theory of integration. (114) Theory of diff$^{\underline{ble}}$ $\mathbb{C}$-valued f$^{\underline{ns}}$ turns out to be amazingly important, and it set the course of mathematics through much of the 19^{th} (+ even 20^{th}) century. (113). We also brushed up against the need to solve an occasional PDE, and their general theory is bottomless and very important to any kind of extra-mathematical interaction (110?).
- Geometry: Toward the end of 25b, we got involved with manifolds and their geometry. The idea of investigating manifolds using Stokes's theorem as the main tool is basically the subject of differential topology (131-2). There are also more "nonlinear" geometric objects on manifolds that form the study of differential geometry. (136).

There are also some classes that blend these together.

- Number theory we saw twice: at the very end of 25a, while discussing Dirichlet's theorem, and on 25b homework assignments, where the ζ-f^{n} made a brief appearance. Number theory draws from all sorts of fields, attempts to answer some very "basic" questions in math, and winds up being a driving force at producing some of the most abstract math around.
- Algebraic geometry is, at its core, about describing "level sets" of functions defined over "algebraic" rings, like $\mathbb{F}_p$ rather than $\mathbb{R}$. This clearly involves plenty of algebra, but a great number of our geometric tools — including diff^{n} and diff^{l} forms — appear as well (137).

There are also lots of applications of these ideas!

↪ Science is all about modeling situations, and mathematics is the language of modeling. The closer you get to statistics the more widely applicable the methods are, and this covers both linear algebra ("classifiers" + linear regression) + analysis (any kind of nonlinear model).

↪ Computation thus also becomes super important, once you agree that modeling is useful. There are lots of results about how to efficiently compute even the basic objects from these classes, including matrix products, determinants, derivatives, integrals, zero loci, ... !