General Remarks

(1) About this file:

   (a) Who – you know who.

   (b) What – lecture notes for USC's MATH 425b, taken real time in LaTeX.

   (c) Where – on Zoom classes, of course...

   (d) When – spring 2020, in the midst of COVID-19.

   (e) Why – to keep this as an memory of one of my favorite courses taken so far (up to freshman year).

(2) I did not proofread every sentence. Typos and mistakes may most likely appear. I will fix every issue I spot, but there's no guarantee I've found all of them.

(3) It was my intention to not use `hyperref` too much. Occasionally, you will see hyperlinks contained in red boxes. Before clicking on it, make sure you remember your current page number because there is no hyperlink to direct you back. (Or you could use keyboard shortcuts to jump back; on Mac's Preview this can be easily done by ⌘+[.)

(4) **Notations**. See below. Red ones are those that are more likely to cause confusion (and different notations were used in lectures).

$\mathbb{K}$	field a vector space is on; assumed to be $\mathbb{R}$ or $\mathbb{C}$	$\equiv$	identical to, e.g., $\cos(x - \pi/2) \equiv \sin(x)$
$\mathcal{F}_b(X, Y)$	set of bounded functions from $X \to Y$	$C_b(X, Y)$	set of continuous & bounded functions $X \to Y$
$\{x_n\}_{n\geqslant 1}, \{x_n\}$	sequence $(x_1, x_2, \ldots)$	$\mathcal{S} := \{s_n\}_{n\geqslant 1}$	enumeration of set
$e^{(i)}$	$(0, \ldots, 0, 1, 0, \ldots)$ with $e^{(i)}_j := e_{ij}$, Kronecker delta	Ω	open subset of $\mathbb{C}$
$\mathcal{L}(X, Y)$	space of linear operators $X \to Y$	$\mathcal{B}(X, Y)$	space of bounded operators $X \to Y$
$F^{-1}(\{z\})$	often-times used as "pre-image of $\{z\}$", see e.g. §5.4 (implicit function theorem)	$B(x, r)$	ball centered at x with radius r

Contents

Chapter 3

Functions of a Real Variable

Beginning of Jan. 15, 2021

Remark. Why 425ab?

(1) *Calculus done rigorously*, with definitions and proofs, *originally on vague foundations*, but later supplemented with solid foundations:

  (I) 1800s: $\epsilon - \delta$ language,

  (II) late 1900s / early 1900s: set theory, metric spaces, etc.

(2) Intro to a large field of modern math, "*analysis as a research field*". We will be learning theorems — *phased in abstract and modern language*, often proved in the 20th century — that *go beyond* what Newton & Leibniz knew. For example:

  (1) 425a: point-set topology of metric spaces

  (2) 425b: Arzelà-Ascoli theorem; Stone-Weierstraß theorem; Picard's theorem on existence and uniqueness for ODEs; implicit and inverse function theorems; exploration of Fourier series; and differential forms.

(3) Closely related to *complex analysis*; useful perspective in power series and so on.

3.2 Integration Techniques

Integration by Substitution

Theorem 3.2.1: Integration by Substitution

If $f \in \mathcal{R}[a,b]$ (i.e., R.I. on $[a,b]$; $f:[a,b] \to \mathbb{R}$ or $f:[a,b] \to \mathbb{C}$) and $g:[c,d] \to [a,b]$ is a C^1*diffeomorphism* (i.e., g is a C^1 function with an inverse also C^1) with $g'(0) > 0$ (so $g(c) = a$ and $g(d) = b$), then

$$\int_a^b f(y)\,\mathrm{d}y = \int_c^d f(g(x))g'(x)\,\mathrm{d}x.$$

Proof. The LHS is well-defined by assumption. For the RHS, $f(g(x))$, a function composed with a C^1 diffeomorphism is R.I. Since $g'(x)$ is continuous, the entire integrand of RHS is R.I.

Easier proof if f is continuous: let $F(y) := \int_a^b f(t)\,\mathrm{d}t$. By FTC, since f is continuous, F is differentiable on $[a,b]$ with $F' = f$. By chain rule, $F \circ g$ is differentiable with $(F \circ g)' = (F' \circ g)g'$, i.e.,

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

Therefore $\int_a^b f(y)\,\mathrm{d}y = F(b) - \underbrace{F(a)}_{=0} = F(g(d)) - F(g(c)) = \int_c^d (F \circ g)'(x)\,\mathrm{d}x = \int_c^d f(g(x))g'(x)\,\mathrm{d}x$.

Beginning of Jan. 20, 2021

Now we deal with the case where $f \in \mathcal{R}[a,b]$ where f is not necessarily continuous. We let $(P^m, T^m)_{m \geqslant 1}$ be a sequence of partition pairs on $[a,b]$ such that $\lim_{n\to\infty} \mathrm{mesh}(P^m) \to 0$. For each m we suppose the acquired subintervals are with endpoints $a = x_1^m < x_2^m < \cdots < x_n^m = b$.

Now we apply the MVT to the differentiable function g on each subinterval (from the partition) such that, in each subinterval there exists some $t_k^m \in [x_{k-1}^m, x_k^m]$ with

$$g(x_k^m) - g(x_{k-1}^m) = g'(t_k^m)(x_k^m - x_{k-1}^m).$$

Now let $T^m := \{t_k^m\}_{k=1}^n$. Then (P^m, T^m) is a partition pair of $[c,d]$. Applying g to (P^m, T^m) we get $(g(P^m, g(T^m))$, a partition pair of $[a,b]$. We then have

$$\begin{aligned} R(f, g(P^m), g(T^m)) &= \sum_{k=1}^n f(g(t_k^m))[g(x_k^m) - g(x_{k-1}^m)] \\ &= \sum_{k=1}^n f(g(t_k^m))g'(t_k^m)(x_k^m - x_{k-1}^m) \\ &= R((f \circ g)g', P^m, T^m) \\ &\to \int_c^d f(g(x)g'(x)\,\mathrm{d}x \text{ as } \mathrm{mesh}(P_m) \to 0. \end{aligned}$$

Now it remains to notice that $\lim_{m\to\infty} \mathrm{mesh}(g(P^m)) = 0$: since $g \in C^1$ by assumption, g is continuous, and thus it is bounded on $[c,d]$ and there exists C with $|g'(t)| \leqslant C$ for all $t \in [c,d]$. By MVT there exists some $s \in [c,d]$ such that

$$|g(x_k^m) - g(x_{k-1}^m)| = |g'(s)||x_k^m - x_{k-1}^m| \leqslant C|x_k^m - x_{k-1}^m|.$$

Hence $\int_a^b f(y)\,\mathrm{d}y = \lim_{m\to\infty} R(f, g(P^m), g(T^m)) = \lim_{m\to\infty} R((f \circ g)g', P^m, T^m) = \int_c^d f(g(x))g'(x)\,\mathrm{d}x$. □

Integration by Parts

Theorem 3.2.2: Integration by Parts

Let $f, g : [a,b] \to \mathbb{R}$ be differentiable with $f', g' \in C^0$ (in fact, $\in \mathcal{R}$ suffices and the proof is exactly the same). Then

$$\int_a^b f(x)g'(x)\,\mathrm{d}x = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\,\mathrm{d}x.$$

Proof. By Leibniz product rule, since f, g are differentiable,

$$(fg)' = f'g + fg'.$$

Integrating both sides from a to b gives

$$\int_a^b (fg)'(x)\,\mathrm{d}x = \int_a^b f'(x)g(x)\,\mathrm{d}x + \int_a^b f(x)g'(x)\,\mathrm{d}x$$

whereas the LHS by FTC II is the same as $f(b)g(b) - f(a)g(a)$. Hence the claim follows. □

Improper Riemann Integrals

Notice that there's *absolute vs. conditional convergence* for these integrals.

Definition 3.2.3

Let $f : [a, \infty) \to \mathbb{R}$ (or $\mathbb{C}$) be a function.

(1) If $\int_a^b |f(x)|\,\mathrm{d}x$ exists for all $b \geqslant a$ and $\lim_{b\to\infty} \int_a^b |f(x)|\,\mathrm{d}x$ exists, then we say that the improper Riemann integral $\int_a^\infty |f(x)|\,\mathrm{d}x$ converges, i.e., $\int_a^\infty f(x)\,\mathrm{d}x$ **converges absolutely**, i.e., f *is* ***absolutely Riemann integrable on*** $[a, \infty)$.

(2) If $\int_a^b f(x)\,\mathrm{d}x$ exists for all $b \geqslant a$ and $\lim_{b\to\infty} \int_a^b f(x)\,\mathrm{d}x$ exists, then we say $\int_a^\infty f(x)\,\mathrm{d}x$ **converges**, i.e., f *is* ***Riemann integrable on*** $[a, \infty)$.

A theorem to be proven in HW2: absolute convergence implies convergence for improper Riemann integrals.
There exists a powerful theory of integration (**Lebesgue integration, 525a**):

Lebesgue integrable $\iff$ *absolutely* Lebesgue integrable.

Absolutely convergent improper Riemann integrals are "proper" Lebesgue integrals, but conditionally convergent improper Riemann integrals may not make sense in Lebesgue theory.

Back to 425a: if $f(0)$ is undefined then we can treat $\int_0^b$ as $\lim_{a\downarrow 0} \int_a^b$. Likewise, $\int_{-\infty}^\infty$ can be split into sum of two improper integrals.

3.3 Series

Series (infinite sums) are an important context for ch.4 (uniform convergence), especially when we look at functions defined as series. *Sometimes we cannot differentiate series term by term.* For now (ch.3) we will be looking at series of numbers.

Definition 3.3.1

Let $\{a_k\}_{k=1}^{\infty}$ be a sequence of real (or complex) numbers. The **series** $\sum_{k=1}^{\infty} a_n$ converges if the **sequence of partial sums**

$$\left\{\sum_{k=1}^{n} a_k\right\}_{n=1}^{\infty} \text{ (where } k \text{ is a dummy variable)}$$

converges. If so, and the limit is $L \in \mathbb{R}$ (or $\mathbb{C}$), then we write $\sum_{k=1}^{\infty} a_k := L$.

Beginning of Jan. 22, 2021

Example 3.3.2. Let $x \in \mathbb{R}$ and consider $\sum_{k=0}^{\infty} x^k$ (a very simple case of **power series**, but with x fixed). This is called the **geometric series**; it converges if $|x| < 1$ and diverges otherwise.

Proof. Indeed, if $|x| < 1$ then $\sum_{k=0}^{n} x^k = \frac{1 - x^{n+1}}{1 - x}$. Thus the partial sum converges to $1/(1-x)$ since

$$\lim_{n\to\infty} \sum_{k=0}^{n} x^k = \lim_{n\to\infty} \frac{1 - x^{n+1}}{1 - x} = \frac{1}{1 - x}.$$

If $|x| \geqslant 1$ then $\lim_{k\to\infty} x^k \neq 0$. (*We'll show soon that this implies divergence of series.*) □

Theorem 3.3.3: Cauchy Convergence Criterion, CCC

Let $\left\{\sum_{k=1}^{n} a_k\right\}_{n=1}^{\infty}$ be a seuqence of real (or complex) numbers. The CCC states that the sequence converges if and only if it's Cauchy, i.e.,

> $\forall\ \epsilon > 0$, there exists $N \in \mathbb{N}$ such that $n_1, n_2 \geqslant N \implies \left|\sum_{k=1}^{n_1} a_k - \sum_{k=1}^{n_2} a_k\right| < \epsilon$. In other words, there exists N such that whenever $n \geqslant m \geqslant N$ we have $\left|\sum_{k=m}^{n} a_k\right| < \epsilon$.

Now recall the example above. Recall we stated that if $\sum_{k=1}^{\infty} a_k$ converges then $\lim_{k\to\infty} a_k = 0$. With CCC now introduced, this becomes obvious: given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $m = n \geqslant N$,

$$\left|\sum_{k=n}^{m} a_k\right| < \epsilon \implies |a_n| < \epsilon \text{ for all } n \geqslant N.$$

Definition 3.3.4

The series $\sum_{k=1}^{\infty} a_k$ **converges absolutely** if $\sum_{k=1}^{\infty} |a_k|$ converges.

Theorem 3.3.5

An absolutely convergent sequence converges.

Proof. It suffices to show that the partial sum, $\left\{\sum_{k=1}^{n} a_k\right\}_{k=1}^{\infty}$ is Cauchy. Since the sequence converges absolutely, for $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that if $n \geqslant m \geqslant N$, we have

$$\left|\sum_{k=m}^{n} |a_k|\right| < \epsilon.$$

On the other hand, by triangle inequality,

$$\left|\sum_{k=m}^{n} a_k\right| \leqslant \sum_{k=m}^{n} |a_k| = \left|\sum_{k=m}^{n} |a_k|\right| < \epsilon$$

and we are done by CCC. □

Definition 3.3.6

The series $\sum_{k=1}^{\infty} a_k$ **converges conditionally** if it converges but doesn't converge absolutely, i.e., $\sum_{k=1}^{\infty} a_k$ converges but $\sum_{k=1}^{\infty} |a_k|$ diverges.

Series Tests

Theorem 3.3.7: Comparison Test

If $|a_k| \leqslant b_k$ for all k *sufficiently large*, then the convergence of b_k implies the convergence of $\sum_{k=1}^{\infty} |a_k|$ and, in particular, of $\sum_{k=1}^{\infty} a_k$. We say that $\sum_{k=1}^{\infty} b_k$ **dominates** $\sum_{k=1}^{\infty} a_k$.

Proof. Again, we can bound sums of a_k by b_k and apply CCC. Indeed, for $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geqslant m \geqslant N$ we have

$$\left|\sum_{k=m}^{n} a_k\right| \leqslant \sum_{k=m}^{n} |a_k| \leqslant \sum_{k=m}^{n} b_k < \epsilon.$$

□

Theorem 3.3.8: Integral Test

Suppose $\int_0^\infty f(x)\,\mathrm{d}x$ is a given improper integral and $\sum_{k=1}^\infty a_k$ is a given series.

(1) (*The function is the bigger one*) Assume $|a_k| \leqslant f(x)$ for all $x \in (k-1, k)$ for sufficiently large k. Then if $\int_c^\infty f(x)\,\mathrm{d}x$ exists for some c (recall "sufficiently large"), then $\sum_{k=1}^\infty |a_k|$ converges.

(2) (*The series is the bigger one*) Now we assume $|f(x)| \leqslant a_k$ for all $x \in (k, k+1)$ for sufficiently large k. Then if $\sum_{k=1}^\infty a_k$ converges, so does the integral, i.e., (taking contrapositive) if the integral diverges then the series diverges.

Note that there is no assumption that f is decreasing. Also, we picked $(k-1, k)$ and $(k, k+1)$ respectively just for cleaner notation. It's only a matter of choice.

Proof.

(1) Note that for n, m sufficiently large,

$$\left|\sum_{k=m}^n |a_k|\right| \leqslant \sum_{k=m}^n \int_{k-1}^k f(x)\,\mathrm{d}x = \int_{m-1}^n f(x)\,\mathrm{d}x.$$

Now we apply the net $b \mapsto \int_c^d f(x)\,\mathrm{d}x$ with $\leq := \leqslant$. This converges as $b \to \infty$ by our assumption. Then given $\epsilon > 0$ there exists C such that if $a, b \geqslant C$ then

$$\int_a^b f(x)\,\mathrm{d}x < \epsilon.$$

Taking $N = \lceil C \rceil$ (ceiling) gives that, if $n \geqslant m-1 \geqslant N \geqslant C$ then

$$\int_{m-1}^n f(x)\,\mathrm{d}x < \epsilon.$$

Therefore for $n \geqslant m \geqslant N+1$ we have

$$\left|\sum_{k=m}^n |a_k|\right| \leqslant \int_{m-1}^n f(x)\,\mathrm{d}x < \epsilon$$

and the claim follows from CCC.

Beginning of Jan. 25, 2021

(2) Now we assume that for all sufficiently large k's, i.e., $k \geqslant c$, we have $|f(x)| \leqslant a_k$ for all $x \in (k, k+1)$. By HW2.1, it suffices to show that the net

$$b \mapsto \int_c^d |f(x)|\,\mathrm{d}x$$

on the directed set $([c, \infty), \leqslant)$ is Cauchy. Notice that, for d_1, d_2 sufficiently large,

$$\begin{aligned}\left|\int_c^{d_1} |f(x)|\,\mathrm{d}x - \int_c^{d_2} |f(x)|\,\mathrm{d}x\right| &= \left|\int_{d_1}^{d_2} |f(x)|\,\mathrm{d}x\right| \\ &= \int_{d_1}^{d_2} |f(x)|\,\mathrm{d}x \leqslant \int_{\lfloor d_1 \rfloor}^{\lceil d_2 \rceil} |f(x)|\,\mathrm{d}x \leqslant \sum_{k=\lfloor d_1 \rfloor}^{\lceil d_2 \rceil - 1} a_k\end{aligned}$$

and the claim follows from the convergence of the $\sum_{i=1}^{\infty} a_k$.

□

Corollary 3.3.9

Let $f : [1, \infty) \to \mathbb{R}$ be decreasing and $f : [1, b] \to \mathbb{R}$ is Riemann integrable for all $b \in \mathbb{R}$, and $\sum_{k=1}^{\infty} a_k$ is a series. Assume f and the series are both always positive. Furthermore, assume $f(k) = a_k$ for all integers k. Then

$$\sum_{k=1}^{\infty} a_k \text{ converges} \iff \int_1^{\infty} f(x)\,\mathrm{d}x \text{ converges.}$$

Proof. This is a direct application of the integral test. Since f is decreasing and $f(k) = a_k$, $f(x) \leqslant a_k$ for all $x \in (k, k+1)$. We also know that $a_k \leqslant f(x)$ for all $x \in (k-1, k)$. Now apply the theorem. □

Example 3.3.10. The "**p-series**" is defined as $\zeta(p) := \sum_{k=1}^{\infty} \frac{1}{k^p}$. We check convergence of $\zeta(p)$ for $p > 0$. Note that we can apply the corollary above between $a_k = 1/k^p$ and $f(x) = 1/x^p$, a decreasing function that agrees with the series at integer values. Therefore

$$\sum_{k=1}^{\infty} \frac{1}{k^p} \text{ converges} \iff \int_1^{\infty} \frac{1}{x^p}\,\mathrm{d}x \text{ converges.}$$

If $p \neq 1$, the RHS evaluates to $\lim_{b\to\infty} \int_1^b \frac{1}{x^p}\,\mathrm{d}x = \left[\frac{1}{1-p} x^{1-p}\right]_{x=1}^{b} = \frac{1}{1-p} \lim_{b\to\infty} (b^{1-p} - 1)$. From this we see that if $p > 1$, $1 - p$ is negative and so the limit is finite. Otherwise we have an infinite limit (so the integral does not converge). For $p = 1$ we get the log function which also diverges. See HW2.3.

Definition 3.3.11

The **exponential growth rate** of a series $\sum_{k=1}^{\infty} a_k$ is defined as $\alpha := \limsup_{k\to\infty} \sqrt[k]{|a_k|}$ (see HW2.2 for existence).

Remark. For existence: if $\sqrt[k]{|a_k|}$ is decreasing, then $\sup_{k'\geqslant n} \sqrt[k']{|a_{k'}|} = \sqrt[k]{|a_k|}$ and so $\limsup_{k\to\infty} \sqrt[k]{|a_k|} = \lim_{k\to\infty} \sqrt[k]{|a_k|}$, which exists by monotonicity.

Theorem 3.3.12: Root Test

Let α be the exponential growth rate of a series $\sum_{k=1}^{\infty} a_k$. If $\alpha < 1$ the series converges. If $\alpha > 1$ it diverges. If $\alpha = 1$ the test is inconclusive.

Proof. If $\alpha < 1$ then we can "squeeze" a β such that $\alpha < \beta < 1$. Then,

$$\lim_{k\to\infty} \underbrace{\sup_{k'\geqslant k} \sqrt[k']{|a_{k'}|}}_{\text{decreasing in } \mathbb{R}} = \alpha < \beta,$$

so there does exist some sufficiently large n such that $\sup_{k'\geqslant n} \sqrt[k]{|a_{k'}|} < \beta$ (since things like this converge to $\alpha < \beta$). This means $\sqrt[k']{|a_{k'}|} < \beta$ for all $k' \geqslant n$. Therefore $|a_{k'}| < \beta^{k'}$, and we can apply comparison test between $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=0}^{\infty} \beta^k$, where the second dominates the first and also converges.

If $\alpha > 1$ the we can pick β with $1 < \beta < \alpha$. Then we have

$$\lim_{k\to\infty} \underbrace{\sup_{k'\geqslant k} \sqrt[k']{|a_{k'}|}}_{\text{decreasing}} = \alpha > \beta.$$

In particular, the terms with underbrace (sup) never gets below β. This means for all k, there exists infinitely many $k' \geqslant k$ such that $\sqrt[k']{|a_{k'}|} > \beta$, i.e., $|a_{k'}| > \beta^{k'}$. Since $\lim_{k\to\infty} a_k \neq_0$ we know that the series diverges. It doesn't even converge conditionally. □

Beginning of Jan. 27, 2021

Remark. The root test is inconclusive if $\alpha = 1$. Indeed, all p-series have exponential growth rate 1; some converge but some don't. $1/k^p$ *grow or decay polynomially in* k*, while* r^k *grow / decay exponentially in* k. Indeed,

$$\sqrt[k]{1/k^p} = (1/k^p)^{1/k} = \exp\left(\frac{1}{k}\log\left(\frac{1}{k^p}\right)\right) = \exp\left(\frac{1}{k}(-p\log(k)\right),$$

and computing the limit of the exponent of rightmost term gives

$$\lim_{k\to\infty} \frac{-p\log(k)}{k} = \lim_{x\to\infty} \frac{-p\log(x)}{x} \overset{(H)}{=} \lim_{x\to\infty} \frac{1/x}{1} = 0.$$

(*Notice that we first replaced* k *by* x *since* k *takes discrete values while* x *is defined on entire* $\mathbb{R}^+$*, which L'Hop requires.*) Hence all p-series have exponential growth rate 1, and the claim follows.

Theorem 3.3.13: Ratio Test

Let $\{a_k\}_{k\geqslant 1}$ be a sequence. Assume $a_k \neq 0$ for large enough k. Let

$$\rho := \limsup_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| \text{ and } \lambda := \liminf_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right|.$$

Then if $\rho < 1$ the series converges and if $\lambda > 1$ the series diverges. Otherwise the ratio test is inconclusive.

Proof. If $\rho < 1$, we can again "squeeze" β between ρ and 1, i.e., $\rho < \beta < 1$. Then by the lim sup assumption, for all large enough k, say all $k \geqslant c$, we have $|a_{k+1}/a_k| < \beta$, so $|a_k| < \beta^{k-c}|a_c|$ and the claim follows from the comparison test with the series with ratio β. The other case is analogous. Again notice that for all p-series, $\rho = 1 = \lambda$. □

Remark. We will not be covering cases like $\rho < 1 \leqslant \lambda$ or $\rho \leqslant 1 < \lambda$. They are still inconclusive.

Theorem 3.3.14: Alternating Series Test

Let $\{a_k\}_{k\geqslant 1}$ be a *decreasing* sequence of real numbers with $a_k > 0$ for all k. Then its **alternating series**

$$\sum_{k=1}^{\infty}(-1)^k a_k = a_1 - a_2 + a_3 - \dots \text{ converges (to 0)} \iff \lim_{k\to\infty} a_k = 0.$$

Proof. The $\implies$ direction is obvious. We'll now show $\impliedby$. Since $\lim\limits_{k\to\infty} a_k = 0$, it's enough to show that $\lim\limits_{n\to\infty}\sum\limits_{k=1}^{2n}(-1)^{k-1}a_k$ converges. Indeed, $\sum\limits^{2n}$ differs from $\sum\limits^{2n-1}$ by a_{2n} which can be sufficiently small. Now,

$$\sum_{k=1}^{2n}(-1)^{k-1}a_k = \sum_{k=1}^{n}\underbrace{(a_{2k-1} - a_{2k})}_{>0}.$$

Since $(a_{2k-1} - a_{2k})$ is always positive, the RHS forms an increasing sequence in n. To show it's convergent, it suffices to show it's bounded. Indeed, since $a_{2k+1} \leqslant a_{2k+1} \implies -a_{2n+2} \geqslant -a_{2k+1}$,

$$\sum_{k=1}^{n}(a_{2k-1} - a_{2k}) \leqslant \sum_{k=1}^{n}(a_{2k-1} - a_{2k+1}) = a_0 - a_2 + a_2 - a_4 + \dots - \underbrace{a_{2n}}_{>0} \leqslant a_0.$$

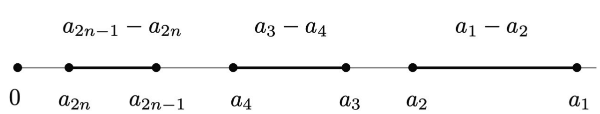


Figure 3.1: From Pugh, p.196

□

Finally, **power series**: series of functions of a variable x (terms of series depending on x). The general form is

$$\sum_{k=0}^{\infty} c_k(x - x_0)^k$$

where often times $x_0 = 0$ so the expression becomes $\sum\limits_{k=0}^{\infty} c_k x^k$ but generally x_0 can be arbitrary.

Some important questions:

(1) Does it converge?

(2) Is this example a differentiable function of x where x converges? – Yes, and even more: also C^∞ and also *holomorphic*. See *radius of convergence* below.

Beginning of Jan. 29, 2021

Theorem 3.3.15: Radius of Convergence

If $\sum_{k=0}^{\infty} c_k(x-x_0)^k$ is a power series (over $\mathbb{R}$ or $\mathbb{C}$) then there exists an *unique* $R \geqslant 0$, called the **radius of convergence** of the series, such that the series converges whenever $|x-x_0| < R$ and diverges whenever $|x-x_0| > R$. Again, the theorem is inconclusive on boundary of the disk, i.e., when $|x-x_0| = R$. (On $\mathbb{R}$ this reduces to the so-called **interval of convergence**). Furthermore, R is given by

$$R := \frac{1}{\limsup_{k\to\infty} \sqrt[k]{|c_k|}}$$

where the denominator is the e.g.r. of the series $\sum_{k=0}^{\infty} c_k(x-x_0)^k$ *evaluated at* $x - x_0 = 1$. *The e.g.r. is defined for numerical series, i.e., with fixed* x, *not for series where* x *varies.*

Proof. We use root test: for any x, we have

$$\sqrt[k]{|c_k(x-x_0)^k|} = \sqrt[k]{|c_k||x-x_0|^k} = |x-x_0|\sqrt[k]{|c_k|}$$

and taking lim sup gives $\limsup_{k\to\infty} \sqrt[k]{|c_k(x-x_0)^k|} = |x-x_0| \limsup_{k\to\infty} \sqrt[k]{|c_k|} = |x-x_0|/R$, and the claim follows.

For uniqueness: suppose R and R' are both radii of convergence. If $|x-x_0| = (R+R')/2$ our series converges and diverges at the same time, clearly a contradiction. □

End of Ch.3 of Pugh

3.4 Detour: Holomorphic Functions, Complex Logarithm

Appendix 5.c of Pugh says that if $z = x + iy$ then $f : \mathbb{C} \to \mathbb{C}$ can be written as

$$f(z) = u(x, y) + iv(x, y) \text{ where } u, v : \mathbb{R}^2 \to \mathbb{R}.$$

Definition 3.4.1

$f : \Omega \to \mathbb{C}$ is **holomorphic** (or *complex differentiable*) if, for all $z_0 \in \Omega$ we have $\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$ exists. If so, we define the **complex derivative** of z_0, written as $f'(z_0)$, to be that ratio.

Some facts from MATH 475 (not inducing circular reasoning here):

(1) If f is holomorphic on Ω then $f \in C^\infty$, i.e., being first-order differentiable implies being infinitely differentiable. *This is in stark contrast with real-valued functions*.

(2) (*Principle of analytic continuation*.) If f, g are two holomorphic functions on connected Ω and f, g agree on any small open ball in Ω, then they agree everywhere. *Even better: if f and g agree on any non-discrete subset (i.e., with a limit point) of Ω, then they agree everywhere*.

An application of this fact: if we have defined $\sin(x)$ for $x \in \mathbb{R}$ and we want to extend $\sin(x)$ to $\sin(z)$ for complex z. Then there exists *at most* one way to do so.

(3) How do we know if f is holomorphic?

(a) $f : \mathbb{C} \to \mathbb{C}$ is holomorphic if and only if f viewed as $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f(x, y) = (u(x, y), v(x, y))$, assuming first-order partials exist and are continuous, satisfies the *Cauchy-Riemann equations*:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \text{ and } \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

(b) Alternative perspective: one way to view the Cauchy-Riemann equations is to introduce operators

$$\frac{\partial}{\partial z} := \frac{\partial}{\partial x} - i\frac{\partial}{\partial y} \text{ and } \frac{\partial}{\partial \overline{z}} := \frac{\partial}{\partial x} + i\frac{\partial}{\partial y}.$$

Then,

$$\begin{aligned} \frac{\partial f}{\partial \overline{z}} &:= \frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y} \\ &= \left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)(u + iv) \\ &= \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + i\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) = 0 \end{aligned}$$

if and only if the Cauchy-Riemann equations hold for f! If so, $f'(z)$ exists, and it's equal to $\frac{\partial f}{\partial z}$.

(c) Physics heuristic: first pretend f (any complex function) depends on "two 'independent' complex variables z and $\overline{z}$" (e.g., writing $f(z) = |z|^2$ as $f(z) = z\overline{z}$). *Then we can take $\partial/\partial z$ and $\partial/\partial \overline{z}$ of such functions as if $z, \overline{z}$ were independent variables, even though they are not.* The reason that accounts for it is because

$$\frac{\partial}{\partial z}(\overline{z}) = \frac{\partial}{\partial \overline{z}}(z) = 0.$$

Then, f is holomorphic if and only if $\partial f/\partial\bar{z} = 0$ if and only if "*f depends only on z, not $\bar{z}$.*"

For example, $f(z) = z^3$ is holomorphic but $f(z) = |z^2| = z\bar{z}$ is not holomorphic (because we can express it in terms of $\bar{z}$).

This is good for algebraically defined functions, but how about *transcendental functions*, e.g., log?

Now, a fact about radius of convergence:

Theorem: (from MATH 475)

Let Ω be an open subset of $\mathbb{C}$ and let $f : \Omega \to \mathbb{C}$ be holomorphic. Let $z_0 \in \Omega$. Then the Taylor series of $f(z_0)$

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(z_0)}{k!}(z - z_0)^k$$

has radius of convergence $\geqslant \inf_{z\notin\Omega}|z - z_0|$, i.e., "distance to the boundary", and it converges to f.

Beginning of Feb. 1, 2021

Logical dependencies for holomorphic functions (*a sneak peak*):

(1) Define holomorphic functions and prove theorems like Cauchy-Riemann.

(2) Define line integrals $\int_C f(z)\mathrm{d}z$ with $C \in \Omega$ and $f : \Omega \to \mathbb{C}$ continuous. To define the line integral: write $\mathrm{d}z = \mathrm{d}x + i\mathrm{d}y$, the sum of two *differential 1-forms* and write $f = u + iv$ where $u, v : \Omega \to \mathbb{R}$. Then

$$\int_C f\mathrm{d}z = \int_C (u + iv)(\mathrm{d}x + i\mathrm{d}y) = \int_C u\mathrm{d}x - v\mathrm{d}y + i\int_C v\mathrm{d}y + u\mathrm{d}x.$$

Same thing for $\mathrm{d}\bar{z} = \mathrm{d}x - i\mathrm{d}y$. Note that we can express $\mathrm{d}x$ and $\mathrm{d}y$ in terms of $\mathrm{d}z$ and $\mathrm{d}\bar{z}$:

$$\mathrm{d}x = \frac{1}{2}(\mathrm{d}z + \mathrm{d}\bar{z}) \text{ and } \mathrm{d}y = \frac{1}{2i}(\mathrm{d}z - \mathrm{d}\bar{z}).$$

Then arbitrary line integrals of a complex vector field on a curve $\mathbb{R}^2 \approx \mathbb{C}$ can be written as $\int_C P\mathrm{d}z + Q\mathrm{d}\bar{z}$ where P, Q are complex functions.

(3) Prove (more general statements hold): if f is holomorphic on Ω and $E \subset \Omega$, then $\int_{\partial E} f(z)\mathrm{d}z = 0$. *This basically follows from general theory of differential forms*. The integral here is "basically like" the line integral on a closed curve of a conservative vector field. Indeed,

$$\mathrm{d}(f(z)\mathrm{d}z) = \left(\frac{\partial f}{\partial z}\mathrm{d}z + \underbrace{\frac{\partial f}{\partial \bar{z}}}_{=0}\mathrm{d}\bar{z}\right) \wedge \mathrm{d}z = \frac{\partial f}{\partial z}\underbrace{\mathrm{d}z \wedge \mathrm{d}z}_{=0} = 0.$$

(4) *** With above, prove *Cauchy integral formula*: if f is holomorphic on Ω, then for z inside a circle C in Ω (or more general curve),

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(\zeta)}{\zeta - z}\mathrm{d}\zeta,$$

i.e., values of f <u>inside</u> the circle are determined by the values of f <u>on</u> the circle! Here ζ is simply a dummy variable. *The proof uses calculation of $\int \frac{1}{z}\mathrm{d}z$ along the unit circle.*

(5) From here: expand $1/(\zeta - z)$ as a power series in z and get the theorem that Taylor series converge to f in any ball $B_r(z_0) \subset \Omega$, i.e., f is holomorphic $\implies$ f is complex analytic.

(6) Now use things from *this class* (425b): power series can be differentiated term by term in their disk of convergence; radius of convergence remains the same, see HW3. Then analytics functions and holomorphic functions are $\mathbb{C}^\infty$.

Transcendental Functions, $\log(x)$ & e^x in $\mathbb{C}$

Just like $\log(x) := \int_1^x \frac{1}{\tilde{t}}\,\mathrm{d}\tilde{t}$ for real logarithm (HW2), we want to define $\log(z) := \int_1^z \frac{1}{\tilde{z}}\,\mathrm{d}\tilde{z}$ for complex z. Furthermore, we want to make this integral path-independent for obvious reasons. It turns out this integral is "kind of" not path-independent so we can't.

Recall that on $\mathbb{R}^2$, $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$ would make $\langle P, Q\rangle$ conservative, but only on **simply connected** domain. Our domain, $\mathbb{C} \setminus \{0\}$ is not.

Basic topological fact: we can compute a "test integral" $\int_C \frac{1}{z}\mathrm{d}z$ where C is the curve along the unit circle. Then any two paths γ_1, γ_2 from $(1,0)$ to z differ by *some multiple* of that test integral. In fact this integral = $2\pi i$.

Standard approach to our problem: we "make a branch cut" E, for example the negative real axis, on the complex plane. Then $\mathbb{C} \setminus E$ is simply connected, and so we can define

$$\log(z) := \int_1^z \frac{1}{\tilde{z}}\,\mathrm{d}\tilde{z} \text{ for } z \in \mathbb{C} \setminus E.$$

Cleaner approach: have a "function" $\mathbb{C} \setminus \{0\} \to \mathbb{C}$ defined by

$$z \mapsto \int_1^z \frac{1}{\tilde{z}}\,\mathrm{d}\tilde{z}$$

but it's "multiple-valued": values at z differ by integer multiples of $2\pi i$ (recall the "test integral").

Now we introduce an equivalence relation $\sim$ on $\mathbb{C}$: $z \sim z'$ if $z - z' = 2\pi i k$ for some integer k. Then $\mathbb{C}$ gets "divided" by horizontal lines with complex axis coordinates $2\pi i k$. Then $\mathbb{C}/\sim$ is simply a cylinder with $0 = 2\pi i = 4\pi i = \dots$. Now we have a well-defined function:

$$C \setminus \{0\} \to \mathbb{C}/\sim \text{ defined by } z \mapsto \log(z) := \left\{ \int_1^z \frac{1}{\tilde{z}}\,\mathrm{d}\tilde{z} + 2\pi i k \right\} = \left[\int_1^z \frac{1}{\tilde{z}}\,\mathrm{d}\tilde{z} \right].$$

($\log(z)$ represents an equivalence class here.) This is holomorphic, bijective, and furthermore it's a *group homomorphism* from $(\mathbb{C} \setminus \{0\}, \cdot)$ to $(\mathbb{C}/\sim, +)$, i.e., the *quotient of* $(\mathbb{C}, +)$ *by the subgroup generated by the element* $2\pi i \in \mathbb{C}$, in the sense that

$$\log(xy) = \log(x) + \log(y).$$

Furthermore, the inverse defines the exponential function $\mathbb{C} \to \mathbb{C}/\sim \to \mathbb{C} \setminus \{0\}$ by:

$$z \mapsto [z] \mapsto \log^{-1}([z]) =: \exp(z).$$

This is also a group homomorphism so $\exp(xy) = \exp(x) + \exp(y)$ and furthermore $\exp(x + 2\pi i k) = \exp(x)$. (*By this way of defining log, we need to precompose z to $[z]$ before using its inverse, exp.*)

Beginning of Feb. 3, 2021

With $\log$ and $\exp$ defined, now we define exponentiation:

Definition 3.4.2

Define $a^b := \exp(b\log(a))$. *Assumption: we make branch cut and take* $a \in \mathbb{C} \setminus (-\infty, 0) \times \{0\}$ *since* $\exp$ *is* $2\pi i$ *periodic but the values of* $\log(a)$ *is "only defined" up to* $\mathbb{Z}$ *multiples of* $2\pi i b$.. Now let $e := \exp(1)$ so $\exp(z) = \exp(z \cdot 1) = \exp(z \cdot \log(e)) =: e^z$.

Example 3.4.3. Let $\alpha \in \mathbb{C}$. If $\alpha \in \mathbb{Z}_{\geqslant 0}$ then $(1+z)^\alpha$ is a polynomial in z, namely $\sum_{k=0}^{\alpha} \binom{\alpha}{k} z^k$, called the binomial formula. If $\alpha \notin \mathbb{Z}_{\geqslant 0}$ then

$$(1+z)^\alpha = e^{\alpha \log(1+z)}$$

which is defined on $\mathbb{C}$ excluding $(-\infty, -1] \times \{0\}$ (so that $1+z$ is the branch cut). From the sneak peak (5), holomorphic functions are analytic, so $(1+z)^\alpha$ equals its Taylor series centered at the origin if $|z| < 1$ (recall the MATH 475 Theorem on Jan. 29).

Remark. Last lecture we defined $\log$ with codomain $\mathbb{C}/\sim$ the quotient. How do we define $\log$ with codomain $\mathbb{C}$ instead, bypassing the branch cut method? The answer is by defining the *Riemann surface* of $\log(z)$ – take infinitely many copies of $C \setminus$branch cut and create an infinite "staircase" such that $\log$ from this surface to $\mathbb{C}$ is single-valued.

Trig Functions

Recall how we "defined" trig functions using diagrams with a unit circle and a right triangle with hypotenuse 1 and angle θ counterclockwise to the positive x-axis. How do we define trig functions analogously to $\log$?

Given (x, y) on the unit circle, we say $y = \sin\theta$. Hence we should define θ as the arc length on the unit circle from $(1, 0)$ to (x, y). (If $y < 0$ simply take the negative of arc length.) Since the unit circle is given by $x^2 + y^2 = 1$, we can parametrize the circle by $\mathbf{r}(t) := \langle \sqrt{1-t^2}, t \rangle$ and so the arc length θ is

$$\theta = \int_0^y \|\mathbf{r}'(t)\| \, dt = \int_0^y \|\langle -t/\sqrt{1-t^2}, 1\rangle\| \, dt = \int_0^y \sqrt{\frac{t^2}{1-t^2} + \frac{1-t^2}{1-t^2}} \, dt = \int_0^y \frac{1}{\sqrt{1-t^2}} \, dt.$$

Can we use the above equation to define $\arcsin(y) := \int_0^y \frac{1}{\sqrt{1-t^2}} \, dt$ for $-1 < y < 1$ (with some modification when $y < 0$)? The answer is yes, but with extra concepts – a Riemann surface for the integrand:

$$w^2 + z^2 = 1 \subset \mathbb{C}^2 \text{ with coordinates } z, w$$

and the integral becomes $1/w \, dz$.

Remark. If we used arc length of ellipses instead of circles, we get *elliptic integrals* with inverses *elliptic functions*. The Riemann surface $w^2 + z^2 = 1$ gets replaced by *elliptic curves*, e.g., $w^2 = z^3 + az + b$.

Chapter 4

Function Spaces

4.1 Uniform Convergence and $C^0[a, b]$

Our goals:

(1) Prove the very fact about differentiating power series, using uniform convergence.

(2) (A more modern perspective) intro to functional analysis.

A short history of why uniform convergence matters:

(1) (Cauchy, 1821) If $f_n : \mathbb{R} \to \mathbb{R}$ are continuous functions and $\sum_{n=1}^{\infty} f_n$ converges to f, then f is continuous.

(2) (Abel, 1826) Objection to above: consider Fourier series of the discontinuous function $f(x) = x/2$ on $(-\pi, \pi)$ but extended 2π-periodically to $\mathbb{R}$ with $f(2\pi) = 0$. It was known that

$$f(x) = \sin x - \frac{1}{2}\sin(2x) + \frac{1}{3}\sin(3x) - \cdots = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\sin(nx).$$

In $\epsilon - \delta$ language one can show that for any $x \in \mathbb{R}$, the series $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\sin(nx)$ converges to $f(x)$. *However*, each partial sum is continuous whereas the limit is discontinuous. We run into problems if we try to differentiate the series term by term: the $\sum$ becomes $\cos(x) - \cos(2x) + \cos(3x) - \ldots$ which diverges for most x, whereas the other side is $1/2$ plus a bunch of *Dirac delta functions* ($1/2$ nearly everywhere).

Beginning of Feb. 5, 2021

Definition 4.1.1

Let X be a set and (Y, d) a metric space. Let $\{f_n\}_{n\geqslant 1}$ be functions from X to Y, and let $f : X \to Y$ be another function.

(1) We say the sequence $\{f_n\}$ **converges pointwise to** f if for all $x \in X$, $\{f_n(x)\}$ converges to $f(x)$ as a sequence of elements of Y, and we write $f_n \to x$. To put formally:

> For all $x \in X$ and $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for any $n \geqslant N$ we have $d(f_n(x), f(x)) < \epsilon$.

(2) We say the sequence $\{f_n\}$ **converges uniformly** to f and write $f_n \to x$ uniformly if

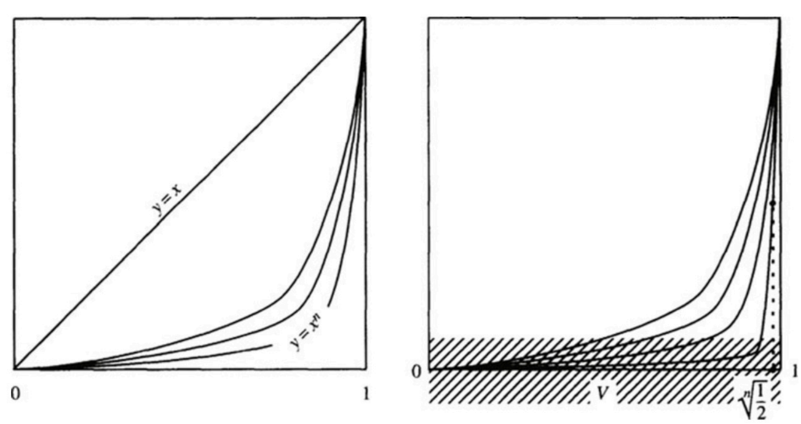
> For all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that, for all $n \geqslant N$, $d(f_n(x), f(x)) < \epsilon$ for all x.

Remark. Pointwise limits are unique (since limits in metric spaces are unique and if $f \equiv g$ then $f = g$).

Example 4.1.2: Pointwise but not uniform convergence. *This is one of the most classic examples that illustrates the difference between pointwise and uniform convergence.* Let $X = [0,1]$, $Y = \mathbb{R}$, and $f_n(x) = x^n$. It follows that

$$\begin{cases} \lim_{n\to\infty} f_n(x) = 0 & x < 1, \text{ and} \\ \lim_{n\to\infty} f_n(x) = 1 & x = 1. \end{cases}$$

Now define $f(x) :=$ the piecewise function $f(x) = 0$ for $x \in [0,1)$ and $f(x) :=$ if $x = 1$. Indeed, above shows that $f_n \to f(x)$. Notice that for $\epsilon < 1/2$ no $f_n(x)$ is entirely contained in the so-called ϵ-tube of f (because f_n always attains values in $(\epsilon, 1-\epsilon)$ but that lies outside the ϵ-tube of f when $\epsilon < 1/2$).



Time to fix Cauchy's 1821 "theorem":

Theorem 4.1.3

Let (X,d) and (Y,d') be two metric spaces. Let $\{f_n\}_{n\geqslant 1}$ be a sequence of functions from X to Y that are continuous at $x_0 \in X$ and let $f : X \to Y$ be another function such that $f_n \to f$ uniformly. Then f is continuous at x_0. In addition, since x_0 is arbitrary, this shows f is globally continuous.

Proof. This proof uses the famous $\epsilon/3$ trick. Let $\epsilon > 0$ be given. By uniform convergence, there exists $N \in \mathbb{N}$ such that for all $n \geqslant N$ and all $x \in X$, we have

$$d'(f(x), f_n(x)) < \epsilon/3. \quad (1)$$

Also, since f_n is continuous at x_0, there exists $\delta > 0$ such that whenever $d(x, x_0) < \delta$, we have

$$d'(f_n(x), f_n(x_0)) < \epsilon/3. \tag{2}$$

Therefore, if $d(x, x_0) < \delta$, applying (1) twice (on x and x_0) and (2) once gives

$$\begin{aligned} d'(f(x), f(x_0)) &\leqslant d'(f(x), f_n(x)) + d'(f_n(x), f_n(x_0) + d'(f_n(x_0), f(x_0)) \\ &\leqslant \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon, \end{aligned}$$

and the claim follows. □

The above theorem shows that our previous example of $f_n(x) = x^x$ cannot converge uniformly to $f(x)$ the piecewise function (since it's discontinuous).

Some questions that follow naturally:

(1) If (X, d) is a compact metric space and $f : X \to (Y, d')$ is continuous function, then f is uniformly continuous, i.e., for all $\epsilon > 0$ there exists $\delta > 0$ such that $d'(f(x), f(x')) < \epsilon/2$ whenever $d(x, x') < \delta$, where $x, x' \in X$ are chosen without further restriction.

   Proof. Let $\epsilon > 0$. For each $x_0 \in X$, f is continuous at x_0. It follows that if $d(x, x_0) < \delta_{x_0}$ then $d'(f(x), f(x_0)) < \epsilon/2$. Consider the set of open balls of (different) radii $\delta_{x_0}/2$ centered at different x_0,

$$\{B(x_0, \delta_{x_0}/2) : x_0 \in X\}.$$

   Let $\mathcal{S} := \{B(x_1, \delta_{x_1}/2), \ldots, B(x_n, \delta_{x_n})/2\}$ be a finite subcover of the above cover (exists by compactness). Now define

$$\frac{\delta}{2} := \frac{\min\{\delta_{x_1}, \ldots, \delta_{x_n}\}}{2}.$$

   It follows that if $x, x' \in X$ then $x \in B(x_i, \delta_{x_i}/2)$ for some i. Furthermore, if $d(x, x') < \delta/2$, then for this particular x_i,

$$d(x', x_i) \leqslant d(x', x) + d(x, x_i) < \frac{\delta_{x_i}}{2} + \frac{\delta}{2} \leqslant \frac{\delta}{2} + \frac{\delta}{2} = \delta,$$

   so

$$d'(f(x), f(x')) \leqslant d'(f(x), f(x_i)) + d'(f(x_i), f(x')) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

   and the claim follows. □

(2) If $\{f_n\}$ are functions on a compact domain (X, d) and $f_n \to f$ for some f, can we upgrade $f_n \to f$ to $f_n \to f$ uniformly? *No, consider our same old example where $[0, 1]$ is indeed compact. The compact domain upgrades continuity to uniform continuity but not convergence to uniform convergence!*

(3) If (X, d) is compact, $\{f_n\}$ and f are continuous, and $f_n \to f$, do we have uniform convergence? *Still no. Let f_n be the functions on $[0, 1]$ connecting $(0, 0)$ and $(1/2n, 1)$, $(1/2n, 1)$ to $(1/n, 0)$, and $(1/n, 0)$ to $(1, 0)$. Then $f_n \to f \equiv 0$ but not uniformly for obvious reasons ($\epsilon = 1/2$) for example.*

However, persistence does pay off.

Theorem 4.1.4: Dini's theorem, 1878

If (X,d) is compact, and $\{f_n\}, f$ are continuous from (X,d) to $\mathbb{R}$ with standard Euclidean metric, $f_n \to f$, and $f_n(x) \leqslant f_{n+1}(x)$ or $f_n(x) \geqslant f_{n+1}(x)$ for all x, n, then $f_n \to x$ uniformly. (*This combines (2) and (3) in the remark above*).

Beginning of Feb. 8, 2021

Today we'll view uniform convergence from a more modern perspective: *metric space of functions* and *functional analysis*. Though most examples of these require Lebesgue integration, there's one special case that does not gets Lebesgue involved: "uniform" metric for spaces of bounded functions.

Idea: let X be a set and let (Y,d) be a metric space. We want to build a metric space of functions from $X \to Y$, and we want the convergence in this space to capture uniform convergence of the functions.
Want: if $f, g : X \to Y$ are functions then $d(f,g)$ is small if and only if when f and g are uniformly close:

$$d(f,g) \coloneqq \sup_{x \in X} d_Y(f(x), g(x)).$$

Important part of the metric space axioms: codomain of d is $\mathbb{R}$, not $\mathbb{R} \cup \{\infty\}$, i.e., given any two points in the metric space are at finite distance from each other.

Issue: The supremum can still attain infinity though.
Fix: build a metric space out of bounded functions only.

Remark. Let (X,d) be a metric space and let $x_0 \in X$. TFAE:

(1) There exists $M \in \mathbb{R}$ such that $d(x, x_0) \leqslant M$ for all M.

(2) There exists $M \in \mathbb{R}$ and $x_1 \in X$ such that for all $x \in X$, $d(x, x_1) \leqslant M$.

(3) There exists $M \in \mathbb{R}$ such that for all $x, x' \in X$, $d(x, x') \leqslant M$.

(4) If (X,d) is a metric subspace of $\tilde{X}, \tilde{d})$ and $\tilde{x}_0 \in \tilde{X}$:

   (a) There exists $M \in \mathbb{R}$ such that $\tilde{d}(x, \tilde{x}_0) \leqslant M$ for all $x \in X$.

   (b) There exists $M \in \mathbb{R}$ and $\tilde{x}_1 \in \tilde{X}$ such that $\tilde{d}(x, \tilde{x}_1) \leqslant M$ for all $x \in X$.

Definition 4.1.5

(X,d) is **bounded** if any of the above conditions hold.

Definition 4.1.6

Let X be a set and let (Y,d) be a metric space. A function $f : X \to Y$ is **bounded** if its image $f(X) \subset Y$ is bounded as a metric subspace of Y. Equivalently, $f : X \to \mathbb{R}$ is bounded if and only if there exists $M \in \mathbb{R}$ such that $|f(x)| \leqslant M$ for all $x \in X$.

Having established the idea of boundedness, we now move to the supremum norm:

Definition 4.1.7

If X is a set and (Y,d) is a metric space, let $\mathcal{F}_b(X,Y)$ be the **set of bounded functions** (not necessarily continuous; not to be confused with $C_b(\Omega,\mathbb{K})$) from X to Y. *X is just any set, and these functions do not have to be necessarily continuous.* Define a **sup metric** d_{sup} (or $\|\cdot\|_\infty$) on $\mathcal{F}_b(X,Y)$ by

$$d_{\text{sup}}(f,g) := \sup_{x\in X} d(f(x),g(x)), \text{ as long as it's } \in \mathbb{R}.$$

Lemma 4.1.8

If $f,g \in \mathcal{F}_b(X,Y)$, then $\sup_{x\in X} d(f(x),g(x)) < \infty$.

Proof. Let $x_0 \in X$ and choose M,N such that $d(f(x),f(x_0)) \leqslant M$ for all $x \in X$ and $d(g(x),g(x_0)) \leqslant N$ for all $x \in X$. Then by triangle inequality, for any $x \in X$,

$$\begin{aligned} d(f(x),g(x)) &\leqslant d(f(x),f(x_0)) + d(f(x_0),g(x_0)) + d(g(x_0),g(x)) \\ &\leqslant M + d(f(x_0),g(x_0)) + N \end{aligned}$$

which is finite (recall the metric space axiom that says $d(f(x_0),g(x_0)) < \infty$ for any given $f(x_0),g(x_0)$). □

Now, a bunch of theorems connecting the sup metric to uniform convergence:

Theorem 4.1.9

If $\{f_n\} \subset \mathcal{F}_b(X,Y)$ be a sequence of bounded functions, then $f_n \to f$ uniformly if and only if $f_n \to f$ with respect to the sup metric.

Proof. The proof consists of a chain of "$\iff$"s:

(1) $f_n \to f$ uniformly.

(2) For all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(f_n(x),f(x)) < \epsilon/2$ for all $x \in X$ whenever $n \geqslant N$.

(3) For all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $\sup_{x\in X} d(f_n(x),f(x)) \leqslant \epsilon/2$.

(4) For all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $d_{\text{sup}}(f_n,f) < \epsilon$.

(5) $f_n \to f$ with respect to the sup metric.

□

Does this capture uniform convergence of bounded functions? What if we know f_n's are bounded and $f_n \to f$ uniformly but don't know if f is bounded?

Proposition 4.1.10

Let X be a set and (Y,d) a metric space. If $\{f_n\}$ are bounded functions from $X \to Y$ and $f_n \to f$ uniformly for some $f \in X \to Y$, then f is bounded.

Proof. Let $x_0 \in X$. Take $\epsilon = 1$. By uniform convergence of f_n there exists $N \in \mathbb{N}$ such that $d(f_n(x), f(x)) < 1$ for all $x \in X$ whenever $n \geqslant N$. Then

$$d(f(x), f(x_0)) \leqslant d(f(x), f_n(x)) + d(f_n(x), f_n(x_0)) + d(f_n(x_0), f(x_0)) \leqslant M + 2$$

which shows f is bounded. □

Corollary 4.1.11

If $\{f_n\} \subset \mathcal{F}_b(X, Y)$, then $\{f_n\}$ converges uniformly if and only if it converges in $(\mathcal{F}_b(X, Y), d_{\sup})$.

Definition 4.1.12

Let X be a set and (Y, d) a metric space. Let $\{f_n\}$, a sequence of functions from $X \to Y$, is **uniformly Cauchy** if, for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$d(f_m(x), f_n(x)) < \epsilon \text{ for all } x \in X, \text{ whenever } m, n \geqslant N.$$

Theorem 4.1.13

Let X be a set and (Y, d) a metric space. If (Y, d) is complete then any uniformly Cauchy sequence of functions from X to Y is uniformly convergent.

Beginning of Feb. 10, 2021

Proof. *"Uniform estimate derived by non-uniform means"*: let $\{f_n\}$ be a uniformly Cauchy sequence of functions $X \to Y$. We first show $f_n \to f$ pointwise for some $f : X \to Y$. Indeed, uniform Cauchy-ness implies pointwise Cauchy-ness which, by completeness of Y, implies pointwise convergence, and we define $\{f_n(x)\} \to f(x)$ for all $x \in X$.

Now we show $f_n \to f$ uniformly. Let $\epsilon > 0$ be given. There exists $N \in \mathbb{N}$ such that $d(f_{\tilde{n}}(x), f_{\tilde{m}}(x)) < \epsilon/2$ for all $x \in X$ as long as $\tilde{m}, \tilde{n} \geqslant N$. Then, for all $x \in X$, for m, n sufficiently large,

$$d(f_n(x), f(x)) \leqslant d(f_n(x), f_m(x)) + d(f_m(x), f(x))$$

where the first term can be made $< \epsilon/2$ by uniform Cauchy-ness and the second can also be made $< \epsilon/2$ by pointwise convergence. The claim then follows. □

Corollary 4.1.14

If Y is a complete metric space and X is any set, then $(\mathcal{F}_b(X, Y), d_{\sup})$ is complete.

Brief proof: any Cauchy sequence in $\mathcal{F}_b(X, Y)$ is uniformly Cauchy and thus uniformly convergent by above. Then the limit is also bounded – guaranteed by pointwise convergence in $(\mathcal{F}_b(X, Y), d_{\sup})$.

One more thing:

Assuming X and Y are both metric spaces, we can talk about continuous functions $X \to Y$.

Definition 4.1.15

We define $C_b(X,Y)$ as the set of **bounded continuous functions** from $X \to Y$. This is a subset of $\mathcal{F}_b(X,Y)$, and $(C_b(X,Y), d_{\sup})$ is a metric subspace of $(\mathcal{F}_b(X,Y), d_{\sup})$.

Proposition 4.1.16

$C_b(X,Y)$ is a closed subset of $\mathcal{F}_b(X,Y)$.

Proof. A sequence $\{f_n\}$ in $C_b(X,Y)$ that converges with respect to $d_{\sup}$ converges to some $f \in \mathcal{F}_b(X,Y)$. Then $f_n \to f$ uniformly. Since each f_n is continuous, so is f. Hence the closure. □

Corollary 4.1.17

If Y is a complete metric space and X is any metric space, $(C_b(X,Y), d_{\sup})$ is a complete metric space. *This follows from the fact that closed subsets of complete metric spaces are complete as metric subspaces.*

Normed Vector Spaces

So far, we've talked about uniform convergence for functions in a general metric space Y (no additions, no multiplications, no series, only sequences). For series, we look at functions into *metric spaces which are also vector spaces*. A natural structure is the **normed vector space** $(V, \|\cdot\|)$. Recall that a normed vector space over $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$ given our context) is a vector space equipped with a norm that is **non-degeneracy, absolute homogeneity, and triangle inequality**. This gives rise to a metric on V by $d(u,v) := \|u - v\|$.

Example 4.1.18. Some examples of normed spaces:

(1) $V = \mathbb{R}$ or $\mathbb{C}$ and $\|v\| := |v|$. More generally, $(V, \langle\cdot,\cdot\rangle)$ gives a inner product space where $\langle\cdot,\cdot\rangle$ satisfies

- (I) positive definiteness: $\langle v, v\rangle \geqslant 0$ and $\langle v, v\rangle = 0 \iff v = 0$,
- (II) antilineaity w.r.t. the first (or second) argument: $\langle u + cv, w\rangle = \langle u, w\rangle + \bar{c}\langle v, w\rangle$ (or $\langle u, v + cw\rangle = \langle u, v\rangle + \bar{c}\langle u, w\rangle$) depending on which argument is antilinear, and
- (III) conjugate symmetry: $\langle v, w\rangle = \overline{\langle w, v\rangle}$.

If this holds, we define $\|v\| := \sqrt{\langle v, v\rangle}$. Examples of inner product spaces include $\mathbb{R}^n$ with $\langle v, w\rangle = \sum v_i w_i$ and $\mathbb{C}^n$ with $\langle v, w\rangle = \sum \overline{v_i} w_i$ or $\sum v_i \overline{w_i}$ (antilinearity on different arguments).

Furthermore, $\|\cdot\|$ can be induced by an inner product if and only if it satisfies the parallelogram law

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

If so, the inner product is given by the **polarization identity**

$$4\langle x,y\rangle = \begin{cases} \|x+y\|^2 - \|x-y\|^2 & \mathbb{K}=\mathbb{R} \\ \|x+y\|^2 - \|x-y\|^2 + i(\|x-iy\|^2 - \|x+iy\|^2) & \mathbb{K}=\mathbb{C}, \text{ antilinear in first arg} \\ \|x+y^2\| - \|x-y\|^2 + i(\|x+iy\|^2 - \|x-iy\|^2) & \mathbb{K}=\mathbb{C}, \text{ antilinear in second arg} \end{cases}$$

(2) $V=\mathbb{R}^n$ or $\mathbb{C}^n$ for $p \in [1,\infty)$ and define the ***p*-norm**

$$\|v\|_p := \left(\sum_{k=1}^n |v_k|\right)^{1/p}.$$

For $p=\infty$ define $\|v\|_\infty := \max_{k\in[1,n]} |v_k|$. Among these normed spaces, only $p=2$ gives rise to an inner product. *Among the ℓ^p spaces only ℓ^2 is Hilbert.*

Beginning of Feb. 12, 2021

Definition 4.1.19

Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ are called **equivalent** or **comparable** if there exist constants c, c' such that

$$c\|x\|_1 \leqslant \|x\|_2 \leqslant c'\|x\|_1 \text{ for all } x \in X,$$

i.e., $(X, \|\cdot\|_1) \cong (X, \|\cdot\|_2)$ (isomorphic). *We say that equivalent norms define the same topology*.

> **Remark.** One thing not preserved by equivalence of norm is whether the norm is induced by an inner product. In fact, **all norms on a finite-dimensional space over $\mathbb{K}$ are equivalent.** See HW5/6. (Recall that among all p-norms on $\mathbb{K}^n$, only $p=2$ is induced by an inner product.)

Corollary 4.1.20

An immediate corollary arising from the remark above is that any finite-dimensional normed vector space is complete in the metric induced by the norm. Indeed, such finite-dimensional space X is isomorphic to $\mathbb{K}^n$. Therefore, we can understand Cauchyness or convergent sequences w.r.t. any norm by reference to the *usual norm* which gives a complete metric.

Definition 4.1.21

$(X, \|\cdot\|)$ is a **Banach space** (or ***X* is Banach**) if it's complete. A **Hilbert space** is an inner product space that is Banach.

> **Remark.** Infinite-dimensional Banach spaces may have various structures, but any **separable, infinite-dimensional Hilbert spaces are isometrically isomorphic**, with H (over $\mathbb{K}$) $\equiv \ell^2(\mathbb{K})$.

Proposition 4.1.22

X is Banach if and only if $\sum_{k=1}^{\infty}\|x_k\| < \infty \implies \left\{\sum_{k=1}^{n} x_k\right\}_{n\geqslant 1}$ converges in X, i.e., $\left\|\sum_{k=1}^{n} x_k - x\right\| \to 0$ for some $x \in X$. This says that, in Banach spaces, "absolute convergence" implies ordinary convergence.

Proof of $\implies$. Since X is complete, it suffices to show that the partial sums $\sum_{k=1}^{n} x_k$ is Cauchy. By the Cauchy-ness of the absolute series, given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that if $m > n \geqslant N$ then

$$\left\|\sum_{k=n}^{m} x_k\right\| \leqslant \sum_{k=n}^{m}\|x_k\| = \sum_{k=1}^{m}\|x_k\| - \sum_{k=1}^{n}\|x_k\| < \epsilon.$$

For $\impliedby$, I will copy & paste part of my previous notes. *Take some Cauchy sequence $\{y_n\} \subset X$. We'll find a convergent subsequence $\{y_{n_k}\} \to y$. This, along with $\{y_n\}$'s being Cauchy, suffices to show $\{y_n\}$'s convergence [$\epsilon/2$ proof].*
Let $n_0 = 1$ and for $k \geqslant 1$ let n_k be such that $n_k > n_{k-1}$ and

$$\|y_i - y_k\| \leqslant 2^{-k} \text{ for all } i, j \geqslant n_k.$$

(This is possible because $\{y_n\}$ is assumed to be Cauchy.) Now define a sequence $\{x_n\}$ such that $x_1 = y_{n_1}$ and $x_i = y_{n_i} - y_{n_{i-1}}$. Then

$$\sum_{i=1}^{\infty}\|x_i\| \leqslant \sum_{i=0}^{\infty} 2^{-i} = 2 < \infty,$$

and by assumption $\sum_{i=1}^{\infty} x_i$ converges. This finishes the proof since $\sum_{i=1}^{\infty} x_i = y_{n_i}$ by construction. □

Functions into a Normed Vector Space

Let X be a set and V a vector space. We define $\text{Fun}(X,V) :=$ the set of functions from $X \to V$. One can check this is a vector space with $(f+g)(x) := f(x) + g(x)$ and $(\lambda f)(x) := \lambda(f(x))$. Now assume V has a norm $\|\cdot\|$. It then makes sense to talk about boundedness of functions.

Then $\mathcal{F}_b(X,V)$ (the set of bounded functions) is well defined and is a vector subspace of $\text{Fun}(X,V)$. Sums and scalar multiples of bounded functions are still bounded.

In addition, if X is a metric space (not just a set), then $C^0(X,V) \subset \mathcal{F}_b(X,V)$ and it is also a vector subspace. Sums and scalar multiples of bounded continuous functions are bounded and continuous.

Definition 4.1.23

The sup metric on $\mathcal{F}_b(X,V)$ mentioned before now upgrades to the **sup norm**. Let X be a set and $(V, \|\cdot\|)$ a normed vector space. For $f \in \mathcal{F}_b(X,V)$, we define the sup norm as

$$\|f\|_{\sup} := \sup_{x\in X}\|f(x)\|.$$

One can easily verify that $\|\cdot\|_{\sup}$ is a norm on $\mathcal{F}_b(X,V)$ and $d_{\sup}(f,g) = \|f-g\|_{\sup}$.
Furthermore, if X is also a metric space then $(C_b(X,V), \|\cdot\|_{\sup})$ is also well-defined and normed.

Beginning of Feb. 17, 2021

Corollary 4.1.24

If X is a set and $(V, \|\cdot\|)$ is Banach, then $(\mathcal{F}_b(X,V), \|\cdot\|_{\sup})$ is also Banach.

Brief proof: by completeness, being Cauchy w.r.t. $\|\cdot\|_{\sup}$ is equivalent to uniform convergence.

An example of $\|\cdot\|_{\sup}$ where $\mathcal{F}_b(X,V)$ is finite dimensional:
Let X be a finite set, e.g., $\{1,\dots,n\}$, and take $V := \mathbb{K}$. Then any function $\{1,\dots,n\} \to \mathbb{K}$ carry the same data as vectors in $\mathbb{R}^n$ (think of vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$). Define

$$\|f\|_{\sup} = \max_{1\leqslant i\leqslant n}|f(i)|, \text{ in this case equivalent to } \|\mathbf{v}\|_\infty = \max_{1\leqslant i\leqslant n}|\mathbf{v}_i|,$$

from which we see that $\|\cdot\|_{\sup}$ is a generalization of the ∞-norm.

Series of Functions

Now that we have a normed vector space $(V, \|\cdot\|)$, we can talk about series, not just sequences, of fnctions $X \to V$:

$$\sum_{n=1}^\infty f_n, \text{ where } f_n : X \to V.$$

Definition 4.1.25

Let X be a set and $(V, \|\cdot\|)$ a normed vector space. A series $\sum_{n=1}^\infty f_n$ of functions with $f_n : X \to V$ **converges uniformly / pointwise** if the sequence of partial sums converges uniformly / pointwise.

Definition 4.1.26

Let X be a set and $(V, \|\cdot\|)$ a normed vector space. The series $\sum_{n=1}^\infty f_n$ **converges absolutely** if it converges "absolutely pointwise", i.e., for all $x \in X$, the series $\sum_{n=1}^\infty f_n(x)$ converges (as a series) in V.

*We say the series converges "**absolutely uniformly**" if the series $\sum_{n=1}^\infty \|f_n\|$, a function $X \xrightarrow{f_n} V \xrightarrow{\|\cdot\|} \mathbb{R}$, where $\|\cdot\|$ is a norm on V, not $\|f_n\|_{\sup}$.* This is a *nice* convergence in a Banach space:

Theorem 4.1.27

Let X be a set and $(V, \|\cdot\|)$ a Banach space. Assume $\sum_{n=1}^\infty f_n$ converges "absolutely uniformly" as above. Then it converges absolutely and uniformly.

Proof. Absolute (pointwise) convergence is automatic. To show uniform convergence, it suffices to show that the partial sums of $\sum f_n$ are uniformly Cauchy given $(V, \|\cdot\|)$ is Banach. Indeed, this holds because

$$\left\| \sum_{k=m}^n f_k(x) \right\| \leqslant \sum_m^n \|f_k(x)\| < \epsilon \text{ for } m,n \text{ large},$$

where the last $<$ comes from uniform Cauchyness of $\left\{ \sum_{k=1}^{n} \|f_k\| \right\}_{n\geqslant 1}$. □

Remark. Observe that if V is Banach then $(\mathcal{F}_b(X,V), \|\cdot\|_{\sup})$ is also Banach. Absolute convergence implies ordinary convergence for series in a Banach space. This gives the following theorem.

Theorem 4.1.28: Weierstraß M-test, "nice version"

Let X be a set and $(V, \|\cdot\|)$ a normed vector space. Let $\{f_n\} \subset \mathcal{F}_b(X,V)$. Then if $\sum_{n=1}^{\infty} \|f_n\|_{\sup}$ converges (as a sequence of real numbers), $\sum_{n=1}^{\infty} f_n$ converges as a series in $(\mathcal{F}_b(X,V), \|\cdot\|_{\sup})$, i.e., $\sum_{n=1}^{\infty} f_n$ converges uniformly.

Theorem 4.1.29: Weierstraß M-test

Let X be a set and $(V, \|\cdot\|)$ a Banach space. Let $\{f_n\} \subset \mathcal{F}_b(X,V)$ be a sequence of bounded sequences. If $\sum_{n=1}^{\infty} M_n$ is a convergent series with $\|f_n\| \leqslant M_n$, then $\sum_{n=1}^{\infty} f_n$ converges absolutely and uniformly.

Proof. The convergence of $\sum_{n=1}^{\infty} M_n$ implies convergence of $\sum_{n=1}^{\infty} \|f_n\|_{\sup}$. Then by above this means $\sum_{n=1}^{\infty} f_n$ converges absolutely uniformly, and finally since $(V, \|\cdot\|)$ is Banach, $\sum_{n=1}^{\infty} f_n$ converges absolutely and uniformly. □

Beginning of Feb. 19, 2021

In the next few days, we limit our conditions to domain $[a,b] \subset \mathbb{R}, \Omega \subset \mathbb{C}$ and codomain $\mathbb{R}$ or $\mathbb{C}$.

Recall from HW4 p3 that if $f_n : [a,b] \to \mathbb{R}$ are Riemann integrable functions with $f_n \to f$ uniformly, then f is Riemann integrable as well. Uniform convergence preserves limits of nets:

$$\lim_{P,T} R(f,P,T) \text{ exists and equals } \lim_{n\to\infty} \lim_{P,T} R(f_n,P,T) = \lim_{n\to\infty} \int_a^b f_n(\tilde{t})\, \mathrm{d}\tilde{t}.$$

In other words,

$$\int_a^b \lim_{n\to\infty} f_n(\tilde{t})\, \mathrm{d}\tilde{t} = \lim_{n\to\infty} \int_a^b f_n(\tilde{t})\, \mathrm{d}\tilde{t},$$

i.e., we can "exchange" limits and Riemann integrals given uniform convergence of f_n. *525a will show that this is a weaker condition than uniform convergence in Lebesgue integrals.*

Remark. Abstractly, note $\mathcal{R}[a,b] \subset \mathcal{F}_b([a,b],\mathbb{R})$ is a closed subset with respect to $\|\cdot\|_{\sup}$.

Proposition 4.1.30

Let $f_n : [a,b] \to \mathbb{R}$ (or $\mathbb{C}$) be Riemann integrable functions such that $f_n \to f$ uniformly. Let

$$F_n(x) := \int_a^x f_n(\tilde{t})\, \mathrm{d}\tilde{t} \text{ and } F(x) := \int_a^x f(\tilde{t})\, \mathrm{d}\tilde{t} \text{ for } x \in [a,b].$$

Then $F_n \to F$ uniformly on $[a,b]$.

Proof. Let $\epsilon > 0$ be given. We choose $N \in \mathbb{N}$ sufficiently large such that $|f_n(t) - f(t)| < \epsilon/(b-a)$ for all $t \in [a,b]$ when $n \geqslant N$. If this is true,

$$|F_n(x) - F(x)| = \left| \int_a^x f_n(\tilde{t}) - f(\tilde{t}) \, \mathrm{d}\tilde{t} \right| \leqslant \int_a^x |f_n(\tilde{t}) - f(\tilde{t})| \, \mathrm{d}\tilde{t} < \epsilon.$$

□

Now we apply the partial sums of series:

Corollary 4.1.31

Let $f_n, f : [a,b] \to \mathbb{R}$ (or $\mathbb{C}$) be Riemann integrable functions. If $\sum_{k=1}^{\infty} f_k = f$ (more conventional way to say partial sums converge uniformly to f) then

$$\sum_{k=1}^{\infty} \left(\int_a^x f_k(\tilde{t}) \, \mathrm{d}\tilde{t} \right) \text{ converges to } \int_a^x f(\tilde{t}) \, \mathrm{d}\tilde{t} \text{ uniformly on } [a,b].$$

This can be easily shown using the proposition above.

Remark. In complex plane we have line integrals, to which the above results apply, after parametrization.

If $f_n : [a,b] \to \mathbb{R}$ are differentiable and $f_n \to f$ uniformly, is f necessarily differentiable?

The answer is no. Counterexample from HW5: $f_n(x) := \sqrt{x^2 + 1/n}$ and $f := |x|$ on $[-1,1]$. Each f_n is differentiable but f has a cusp and thus not differentiable. This is because the derivatives converges pointwise but not uniformly. See theorem below .

In particular, Weierstraß in 1872 pointed out that it's possible to have differentiable f_n's with $f_n \to f$ uniformly but f is nowhere differentiable.

Theorem 4.1.32

Let $f_n : [a,b] \to \mathbb{R}$ (or $\mathbb{C}$) be differentiable on $[a,b]$, $f_n \to f$ pointwise, and $f_n' \to g$ uniformly for some $g : [a,b] \to \mathbb{R}$. Then f is differentiable on $[a,b]$ and $f' = g$.

Proof. Let $x \in [a,b]$. We want to show $\lim_{h\to 0} Q(h) := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$ exists and $Q = g$.

We define a net on the directed set $A := [a-x, b-x] \setminus \{0\}$, with ordering $\preceq$ as $\geqslant$, distance from 0, and codomain $\mathbb{R}$ (or $\mathbb{C}$).

We know each f_n is differentiable, so $\lim_{h \to 0} Q_n(h) := \lim_{h \to 0} \frac{f_n(x+h) - f_n(x)}{h}$ exists and equals $f_n'(x)$.

Claim: the nets Q_n converge uniformly (in h) to Q as $n \to \infty$.

If this claim holds, define $L_n = f_n'(x) = \lim_{h \to 0} Q_n(h)$ and $L = g(x)$. Then $f_n' \to g$ uniformly so $L_n \to L$. Also since each f_n is differentiable, for each n, the net $Q_n(h)$ converges to L_n. If it so happens that $Q_n \to Q$ uniformly, then by HW4 p3(b) we have $Q \equiv L$, i.e., $f' = g$.

Proof of claim. We know $Q_n(h) \to Q(h)$ pointwise in h; indeed, $f_n(x+h) \to f(x+h)$ and $f_n(x) \to f(x)$, so the quotient converges. It suffices to show $Q_n(h)$ are uniformly Cauchy (in h) as $n \to \infty$ since $\mathbb{R}$ (or $\mathbb{C}$) is complete.

Given $\epsilon > 0$, choose N sufficiently large that if $m, n \geqslant N$ then

$$|f_n'(\xi) - f_m'(\xi)| < \epsilon \text{ for all } \xi \in [a,b].$$

(This is possible since the derivatives indeed converge uniformly, in particular uniformly Cauchy by assumption.) If $m, n \geqslant N$ then by MVT

$$\begin{aligned}|Q_n(h) - Q(h)| &= \left|\frac{(f_n - f_m)(x+h) - (f_n - f_m)(x)}{h}\right| \\ &= \frac{|(f_n - f_m)'(\xi)||(x+h) - x|}{|h|} \\ &= |f_n'(\xi) - f_m'(\xi)| < \epsilon,\end{aligned}$$

and the main claim follows. ☐

Remark. Notice that in the theorem we only stated that $f_n \to f$ pointwise. Indeed, this upgrades to uniform convergence since the derivatives converge at least somewhere pointwise (cf. HW5.1).

Beginning of Feb. 22, 2021

A nice generalization of the theorem above:

Theorem 4.1.33

Let $\Omega \subset \mathbb{C}$ be open and let $f_n : \Omega \to \mathbb{C}$ be holomorphic functions. Assume $f_n \to f$ pointwise ($f : \Omega \to \mathbb{C}$), and $f_n' \to g$ uniformly for some $g : \Omega \to \mathbb{C}$ (actually suffices to assume locally uniform convergence / uniform on compact subsets on Ω, cf. HW5.3). Then f is holomorphic on Ω and $f' = g$.

Proof. Apply the real-variable proof to $x \in \Omega$ with $0 < |h| < r$ where $B(x,r) \subset \Omega$.
As in previous proof, it suffices to prove that $Q_n(h) \to Q(h)$ uniformly (then again apply the limits of net stuff) where $Q_n(h) := \dfrac{f_n(x+h) - f_n(x)}{h}$ and $Q(h) := \dfrac{f(x+h) - f(x)}{h}$.
The "crucial step" to prove the claim: given $\epsilon > 0$, choose $N \in \mathbb{N}$ such that if $m, n \geqslant N$, the uniform Cauchy-ness of f_n' gives

$$|f_m'(\zeta) - f_n'(\zeta)| < \epsilon \text{ for all } \zeta \in B(x,r).$$

(Compare this with the real-valued MVT.) Then if $m, n \geqslant N$, and $0 < |h| < r$, we have

$$\begin{aligned}|Q_n(h) - Q(h)| &= \left|\frac{(f_n - f_m)(x+h) - (f_n - f_m)(x)}{h}\right| \\ &\leqslant \sup|f_m'(\zeta) - f_n'(\zeta)| \leqslant \epsilon,\end{aligned}$$

where the supremum takes the supremum of $|\cdot|$ on all ζ on the line segment between x and $x+h$. See Pugh's Theorem 5.11 later, *multivaraible MVT*. ☐

Remark. Since $\mathbb{R}^2 \cong \mathbb{C}$, the derivative operator $(Df)_p$ for $p \in \Omega \subset \mathbb{R}^2$ can be viewed as a linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$. (The transformation is multiplication by $f'(z)$ as a map $\mathbb{C} \to \mathbb{C}$.) Its *operator norm* is $|f'(z)|$.

Then the MV inequality gives

$$|f(p+v) - f(p)| \leqslant \sup\left(\|(Df)_p\|\right)|v|$$

where the supremum is again taken over all points on the segment between p and $p+v$.

Corollary 4.1.34

(1) If f_n's are differentiable on $[a,b]$, $\sum_{n=1}^{\infty} f_n$ converges pointwise to f on $[a,b]$, and $\sum_{n=1}^{\infty} f_n'$ converges uniformly to g on $[a,b]$, then f is differentiable on $[a,b]$ with $f' = g$. In particular,

$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{n=1}^{\infty} f_n(x) = \sum_{n=1}^{\infty}\frac{\mathrm{d}}{\mathrm{d}x} f_n(x),$$

summation and derivative operator commute.

(2) If f_n's are holomorphic on Ω, $\sum_{n=1}^{\infty} f_n$ converge pointwise to f on Ω, and $\sum_{n=1}^{\infty} f_n'$ converge locally uniformly to some $g : \Omega \to \mathbb{C}$, then f s holomorphic on Ω and $f' = g$. Once again commutativity holds.

4.2 Power Series

Now we apply what's previously shown to power series. First, a consequence of Weierstraß M-test:

Theorem 4.2.1

Let $\sum_{n=0}^{\infty} c_n(z-z_0)^n$ be a complex power series, and let (Hadamard's formula)

$$R := \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|c_n|}}$$

be the radius of convergence. Then the series converges uniformly on any compact subset $K \subset B(z_0, R)$.

Proof. *The main idea is to bound the series by some convergent geometric series.* Consider the continuous function $z \mapsto d(z, z_0) = |z - z_0|$. It clearly attains a maximum $r < R$ on a compact domain K.

Now define $\tilde{r} = r + \epsilon$ (slightly bigger) such that $K \subset B(z_0, \tilde{r})$.

Now pick β with $\tilde{r} < \beta < R$ (sounds familiar?) so $1/R = \limsup_{k\to\infty} \sqrt[k]{|c_k|} < 1/\beta$. Therefore all late enough terms of $\sqrt[k]{|c_k|} < 1/\beta$, i.e., there exists k_0 such that, for $k \geqslant k_0$ we have

$$\sqrt[k]{|c_k|} \leqslant \frac{1}{\beta} \implies |c_k| \leqslant \frac{1}{\beta^k}.$$

Now, for all $z \in B(z_0, \tilde{r})$, we have

$$|c_k(z-z_0)^k| \leqslant |c_k||z-z_0|^k \leqslant \frac{r^k}{\beta^k}.$$

Taking sup of all $z \in B(z_0, \tilde{r})$ and applying the M-test with $M_k := r^k/\beta^k$, we obtain uniform convergence for z in $B(z_0, \tilde{r})$, hence the compact convergence. □

How about derivatives of the terms $c_k(z-z_0)^k$, namely $kc_k(z-z_0)^{k-1}$? HW3.2(f) says they have the same radius of convergence. Therefore the series of series also converge locally uniformly on $B(z_0, R)$. It follows that we can iterative take derivatives and always obtain the same result without ever shrinking the radius of convergence [!!] See these results below.

Corollary 4.2.2

$\sum_{k=0}^{\infty} c_k(z-z_0)^k$ is holomorphic on $B(z_0, R)$ as summation and derivative operator commute. For real version, $\sum_{k=0}^{\infty} c_k(x-x_0)^k$ is differentiable on $(x_0 - R, x_0 + R)$ with the same formula.

Corollary 4.2.3

Real analytic functions are **smooth** (i.e., C^∞): derivatives of all orders exist. Complex analytic functions are holomorphic and in fact analytic $\iff$ holomorphic [!!]

Remark. (Real) smooth functions are not necessarily analytic. For example

$$f(x) := \begin{cases} e^{-1/x} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

is smooth but not analytic at 0.

Beginning of Feb. 24, 2021

Besides differentiation, we can also integrate power series to get new power series. For example, the geometric series

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \text{ for } x \in \mathbb{C} \text{ with } |x| < 1$$

can be integrated term-by-term inside the unit disk. For $|x| < 1$,

$$\int_0^x \frac{1}{1-\tilde{t}}\, \mathrm{d}\tilde{t} = \sum_{k=0}^{\infty} \int_0^x \tilde{t}^k\, \mathrm{d}\tilde{t} = \sum_{k=0}^{\infty} \frac{x^{k+1}}{k+1}$$

whereas the LHS is

$$\int_0^x \frac{1}{1-\tilde{t}}\, \mathrm{d}\tilde{t} = -\log(1-x),$$

so the series converge to $-\log(1-x)$ for $|x| < 1$, i.e.,

$$\log(1+x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{k+1}}{k+1}.$$

Remark. We gave a "general machinery" to use power series for functions like $e^z, \sin(z), \cos(z)$, etc., so the above $\log(1+x)$ is not just a "lucky case".

For example, if we believe these three are holomorphic (where e^x is the inverse function to $\log(z) = \int_1^z \frac{1}{\tilde{t}}\, \mathrm{d}\tilde{t}$ and $\sin(x), \cos(x)$ are defined to be inverse functions of $\sin^{-1}(z) = \int_0^z \frac{1}{\sqrt{1-\tilde{t}^2}}\, \mathrm{d}\tilde{t}$ and $\cos^{-1}(z)$; see notes on Feburary

3), then theorems from complex analysis say that these functions have convergent power series expansions (Taylor series). For example Taylor series for $\log(z)$ at $z = 1$ has radius of convergence 1, so we can take the open ball centered at 1 and get

$$\log^{(k+1)}(z) = (-1)^k k! \implies \log(z) = \sum_{k=0}^{\infty} (-1)^k \frac{k!}{(k+1)!}(z-1)^{k+1}$$
$$\implies \log(1+z) = \sum_{k=0}^{\infty} (-1)^k \frac{z^{k+1}}{k+1}.$$

More generally, we can apply this idea to $e^z, \sin(z), \cos(z)$ and use Taylor series at $z = 0$, which has radius of convergence ∞. Computing the derivatives just like above gives

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} \qquad \sin(z) = \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k+1}}{(2k+1)!} \qquad \cos(z) = \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k}}{(2k)!}$$

so instead of saying "e^z" is <u>defined</u> to be that power series, we have shown <u>why</u>. Once we have these series,

$$e^{i\theta} = \sum_{k=0}^{\infty} \frac{(i\theta)^k}{k!} = \sum_{k=0}^{\infty} \frac{(i\theta)^{2k}}{(2k)!} + \sum_{k=0}^{\infty} \frac{(i\theta)^{2k+1}}{(2k+1)!} = \cos(\theta) + i\sin(\theta),$$

The famous *Euler's equation*. Valid for all $e^{i\theta}$ (in fact valid for all e^{iz}, $z \in \mathbb{C}$). *Wow*.

> **Remark**. One last remark: another way to know that these series agree with $e^z, \cos(z), \sin(z)$ is by using *Picard's theorem* on existence and uniqueness for solutions to ODEs. We'll get to this later.

4.3 Compactness and Equicontinuity in $C^0[a,b]$

Previously we have shown just how useful uniform convergence is. Now we will focus on a more modern perspective.

Recall that Heine-Borel states that $K \subset \mathbb{R}^n$ if and only if E is closed and bounded. The $\implies$ is true by definition, but we've already shown that $\impliedby$ is not in general: $(\mathbb{N}, d_{\text{sup}})$ was the example given previously.

What about function spaces of form $(C_b(X,Y), d_{sup})$ where Y is normed? Consider the <u>closed</u> unit ball

$$\mathcal{B} := \{f \in C_b(X,Y) : \|f\|_{sup} \leqslant 1\}.$$

In a finite-dimensional function space $\mathcal{B}$ is compact, whereas if $C_b(X,Y)$ is infinite-dimensional then $\mathcal{B}$ is <u>not</u> compact. (In fact this is an iff statement: *a normed space is finite-dimensional if and only if the closed unit ball is compact.*)

For example, let $X := [0,1]$, $Y = \mathbb{R}$ and consider our same old example $f_n(x) := x^n$ on $[0,1]$. If $\mathcal{B}$ is compact then we should be able to extract a convergent subsequence of $\{f_n\}$ (w.r.t. $\|\cdot\|_{sup}$), but we know this is impossible since f_n converges pointwise to a discontinuous f, and there is no way how such a function can be approximated by a continuous function in $\|\cdot\|_{\text{sup}}$.

Beginning of Feb. 26, 2021

Definition 4.3.1

Let (X, d) and (Y, d') be metric spaces. Let $\mathcal{E}$ be a set of functions $X \to Y$. $\mathcal{E}$ is called (uniformly) **equicontinuous** if, for $\epsilon > 0$, there exists $\delta > 0$ such that all functions $f \in E$ are *uniformly continuous in the same way*, i.e., for all $x, x' \in X$ with $d(x, x') < d$ and for all $f \in \mathcal{E}$ we have $d'(f(x), f(x')) < \epsilon$.

Remark. It becomes immediate that

(1) this definition works just well for a sequence $\{f_n\}$ of functions, and

(2) if $\mathcal{E}' \subset \mathcal{E}$ and $\mathcal{E}$ is equicontinuous, then so is $\mathcal{E}'$.

Example 4.3.2. An example: any finite set of uniformly continuous functions is automatically equicontinuous (see HW6; "*picking the smallest* δ" argument).

A non-example: let $X := [0,1]$, $Y = \mathbb{R}$, and $f_n(x) := x^n$ on $[0,1]$ (same example again).

Brief proof. For $\epsilon = 1/2$, assume there exists $\delta > 0$ satisfying the equicontinuity condition (assuming $\delta < 1$). We know that

$$\lim_{n\to\infty} (1-\delta)^n = 0,$$

so there exists n such that $(1-\delta)^n < 1/2$. Then $|f_n(1-\delta) - f_n(1)| > 1/2$, contradicting the assumption.

This example gave non-compactness of unit ball in $(C^0([0,1], \mathbb{R}), \|\cdot\|_{\sup})$ *(and non-equicontinuity, of course).*

We will soon show that equicontinuity will *fix* Heine-Borel theorem in function spaces (i.e., compact iff closed, bounded, and equicontinuous).

For example, let $X = \{1, \dots, n\}$ be equipped with the discrete metric. Then $C^0(X, \mathbb{R})$ is <u>all</u> functions from X to $\mathbb{R}$, which then has a natural bijection with $\mathbb{R}^n$. It follows that any $\mathcal{E} \subset C^0(X, \mathbb{R})$ is equicontinuous, and so by the bijection $\mathcal{E} \subset (\mathbb{R}^n, \|\cdot\|_\infty)$ is compact if and only if it's closed and bounded, i.e., the non-generalized Heine-Borel.

Intuition. Functions in a non-equicontinuous set $\mathcal{E}$ become "*unboundedly steep*". See formal definitions below.

Definition 4.3.3

Let X, Y be metric spaces.

(1) If $f: X \to Y$, then $L \in \mathbb{R}$ is a **Lipschitz constant** for f if, for all $x, x' \in X$,

$$d'(f(x), f(x')) \leqslant L d(x, x').$$

If this happens, f is said to be **Lipschitz** or Lipschitz continuous. Think of this as the "global steepness". *This is stronger than uniform continuity: given* ϵ *simply let* $\delta := \epsilon/L$.

(2) If $\mathcal{E}$ is a set of functions $X \to Y$, then $L \in \mathbb{R}$ is a **uniform Lipschitz constant** if it's a Lipschitz constant for all $f \in \mathcal{E}$.

Some immediate result following the Lipschitz constant:

Proposition 4.3.4

If a uniform Lipschitz constant exists for $\mathcal{E}$, then $\mathcal{E}$ is equicontinuous. (*Given* $\epsilon > 0$, *simply take* $\delta = \epsilon/L$.)

Corollary 4.3.5

If in addition $\mathcal{E}$ is a set of differentiable functions from $[a,b] \to \mathbb{R}$ (allowing $a, b = \infty$), and there exists a global bound $M \in \mathbb{R}$ for all derivatives of all f, then $\mathcal{E}$ is equicontinuous with uniform Lipschitz constant M.

The Arzelá-Ascoli Theorem

We'll split the long proof into two parts, the **Arzeá-Ascoli Propogation Theorem**, and the main **Arzelá-Ascoli Theorem**. Some definitions for the remainder of this lecture, and we'll start Arzelá-Ascoli next lecture.

Definition 4.3.6

A metric space (X, d) is **totally bounded** if, for all $\delta > 0$, there exists a finite covering of X by open balls of radii δ.

Theorem 4.3.7

For a metric space (X, d), TFAE:

(1) (X, d) is compact (the open-cover definition),

(2) (X, d) is sequentially compact (every sequence admits a convergent subsequence), and

(3) (X, d) is totally bounded and complete. Note that (1) $\implies$ (3) is trivial.

Definition 4.3.8

A subset $A \subset (X, d)$ is **dense** in X if any of the following (equivalent) conditions holds:

(1) For all $x \in X$ and $\delta > 0$, there exists $a \in A$ with $a \in B(x, \delta)$.

(2) Each point of x is a limit of sequence in A.

(3) $\overline{A} = X$.

Definition 4.3.9

A metric space (X, d) is **separable** if it has a countable dense subset.

Proposition 4.3.10

If X is a compact metric space then X is separable.

Beginning of March 1, 2021

Proof. Let $\delta = 1, 1/2, 1/3, \dots$. The total boundedness of X implies that a finite subset $A_n \subset X$ such that

$$\bigcup_{a \in A_n} B(a, 1/n) \text{ covers } X.$$

Letting n vary and taking the union of all such A_n's, we obtain a countable subset $A \subset X$ that is dense! Reason: given $\epsilon > 0$, there exists n large enough that $1/n < \epsilon$. Then each point in X can be approximated by something in A_n with distance $< \epsilon$, which finishes the proof. □

Theorem 4.3.11: Arzeá-Ascoli Propogation Theorem

Let (X, d) be a compact metric space and let A be a countable dense subset of X. Let (Y, d') be a complete metric space. Let $\{f_n\}$ be an equicontinuous functions $X \to Y$ such that, for all $a \in A$, the sequence $\{f_n(a)\}$ converges in Y. *In other words, $\{f_n\}$ converge pointwise on this dense subset*. Then there exists $f : X \to Y$ such that $f_n \to f$ uniformly. *Pointwise convergence to some function on A propagates to uniform convergence on all of X to some (possibly the same) function.*

Remark. The some function above refers to the fact that, if $f_n \to \tilde{f}$ pointwise on A for some $\tilde{f}$, it does not mean does not mean the same f is the one that f_n converges uniformly to on all of X. We can manipulate the $\tilde{f}$ and alter the value of $\tilde{f}(x)$ for some $x \in X \setminus A$, obviously.

Proof of Propagation Theorem. Let $\epsilon > 0$ be given. We choose $\delta > 0$ such that, if $x, x' \in X$ and $d(x, x') < \delta$ then for all n we have $d'(f_n(x), f_n(x')) < \epsilon/3$ (by equicontinuity).
Let $\{a_1, a_2, \dots\}$ be an enumeration of A. Given our δ, there exsts a subset $J = \{a_1, \dots, a_j\} \subset A$ that is "δ-*dense*" in X (capable of approximating any $x \in X$ with distance $< \delta$; see HW7 p1). From now on we will only look at these points. For $1 \leqslant i \leqslant j$, the sequence $\{f_n(a_i)\}$ is assumed to be convergent in Y and in particular Cauchy. Therefore, there exists a N_i such that $d'(f_n(a_i), f_m(a_i)) < \epsilon/3$ whenever $m, n \geqslant N_i$. Since J is finite, it is well-defined to let $N = \max\{N_1, \dots, N_J\}$.
Now for $x \in X$, we choose $m, n \geqslant N$ and $a_i \in J$ with $d(x, a_i) < \delta$ (recall the δ-density). Then,

$$\begin{aligned} d'(f_n(x), f_m(x)) &\leqslant d'(f_n(x), f_n(a_i)) + d'(f_n(a_i) + f_m(a_i)) + d'(f_m(a_i), f_m(x)) \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Therefore $\{f_n\}$ is uniformly Cauchy and, since Y is complete, uniformly convergent. □

Theorem 4.3.12: Arzelá-Ascoli Theorem, "Traditional Version"

Let X be a compact metric space and let $\{f_n\}$ be a sequence of functions $X \to \mathbb{R}$ (or $\mathbb{K}$ in general). Assume $\{f_n\}$ is equicontinuous is equicontinuous and pointwise bounded (for each $x \in x$, $\{|f_n(x)|\}$ is bounded, a stronger result then uniform boundedness as presented in Pugh's book), then

(1) $\{f_n\}$ is uniformly bounded, and

(2) $\{f_n\}$ has a uniformly convergent subsequence, i.e., a subsequence that converges in $\|\cdot\|_{\sup}$.

A stronger version: $\mathcal{E} \subset (C^0(X,\mathbb{K}), \|\cdot\|_{\sup})$ *is precompact if and only if it is bounded and equicontinuus.*

Proof of (1). Let A be a countable dense subset of X. The equicontinuity implies that for some $\delta > 0$ and all n, if $d(x,x') < \delta$ then $|f_n(x) - f_n(x')| < \epsilon = 1$, say. By the δ-density lemma, there exists a finite subset $J \subset A$ that is δ-dense in X. Then for all x, there exists some $a_i \in J$ we have $d(x,a_i) < \delta$.
On the other hand, $\{f_n\}$ is pointwise bounded and so $\{|f_n(a_i)|\}$ is bounded, say by M. Therefore $\{|f_n(x)|\}$ is bounded by $M+1 < \infty$. □

Beginning of March 3, 2021

Proof os (2). Since X is compact, we are able to subtract $A = \{a_1, a_2, \dots\}$ a countable dense subset of it. Our goal is to show that there exists a subsequence of $\{f_n\}$ that converges pointwise on all a_i, which then by the Propagation theorem implies $f_n \to f$ uniformly for some f.
To start, consider $f_n(a_i)$, a bounded sequence of real (or complex) numbers. Thus (by Bolzano-Weierstraš) it admits a convergent subsequence. Therefore there exists a subsequence $\{f_{1,i}\}$ that converges pointwise at a_1 ($\{f_{1,i}\}_{i\geqslant 1} = \{n_1, n_2, \dots\}$ which form a subsequence of $\{n\}_{n\geqslant 1}$). Using this fact again, since $\{f_{1,n}(a_2)\}$ is a bounded sequence, there exists a convergent subsequence $f_{2,i}$ that converges pointwise at a_2. Notice that $\{f_{2,i}\}$ also converges at a_1! To generalize this notion, we can define recursively and get $f_{m,i}$ to be a subsequence of $f_{m-1,i}$ that converges pointwise at a_m (and thus $a_1, \dots, a_{m-1}$ as well).

Recall **Cantor's diagonalization** which we previously encountered when showing $\mathbb{R}$ is uncountable. Here we apply the same idea. After enumerating $f_{m,i}$, we can extract a *diagonal sequence* $g_n := f_{n,n}$, a subsequece of <u>all</u> $f_{m,i}$'s (ignoring finitely many terms if necessary) which therefore converges pointwise at <u>all</u> $a_i \in A$.
Now we use the Propagation theorem and claim $g_n \to g$ for some $g : X \to \mathbb{R}$ (or $\mathbb{K}$) uniformly on all of X, not A, and we have proven the claim. □

Applications of Arzelá-Ascoli Theorem

Theorem 4.3.13: Generalized Heine-Borel Theorem

If X is compact, then $\mathcal{E} \subset (C_b(X,\mathbb{R}), \|\cdot\|_{\sup})$ is compact if and only if it is closed, bounded, and equicontinuous.

Proof. For $\Longleftarrow$, if $\mathcal{E}$ is closed, bounded, and equicontinuous, then we want to show that an arbitrary sequence $\{f_n\} \subset \mathcal{E}$ has a convergent subsequence.
The assumption that $\mathcal{E}$ is a bounded subset implies there exists M such that $\{f\}_{\sup} \leqslant M$ for all $f \in \mathcal{E}$. Therefore $\mathcal{E}$ is uniformly bounded. Therefore $\{f_n\}$ is uniformly bounded. It is also equicontinuous since $\mathcal{E}$ is. By Arzeá-Ascoli, $\{f_n\}$ has a uniformly convergent subsequence, whose limit f is also guaranteed to be continuous, i.e., $f_n \to f$ in $\|\cdot\|_{\sup}$. Since $\{f_n\} \subset \mathcal{E}$ and $\mathcal{E}$ is closed, we know $f \in \mathcal{E}$. Therefore every sequence in $\mathcal{E}$ has a convergent subseequence, thus $\mathcal{E}$ is compact using the definition of sequential compactness.

For $\implies$, assume $\mathcal{E}$ is compact and thus totally bounded. This immediately implies $\mathcal{E}$ is bounded. $\mathcal{E}$ is compact so $\mathcal{E}$ is closed. For equicontinuity, given $\epsilon > 0$, the total boundedness implies that there exists finitely many $\epsilon/3$-balls covering $\mathcal{E}$ with centers $f_1, \dots, f_n$. Then if $f \in \mathcal{E}$ we know $\|f - f_k\|_{\sup} < \epsilon/3$ for some $k \in [1, n]$.
Since each $f \in \mathcal{E}$ is continuous on a compact domain, there exists a $\delta_k > 0$ such that

$$d(x, x') < \delta_k \implies |f_k(x) - f_k(x')| < \frac{\epsilon}{3}.$$

Now let δ vary between 1 and n and take the minimum. Then if $f \in E$ and $x, x' \in X$ with $d(x, x') < \delta$, we can choose $k \in [1, n]$ with $\|f - f_k\|_{\sup} < \epsilon/3$, and

$$\begin{aligned} |f(x) - f(x')| &\leqslant |f(x) - f_k(x)| + |f_k(x) - f_k(x')| + |f_k(x') - f(x')| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \end{aligned}$$

where the first and third comes from $\|f - f_k\|_{\sup} < \epsilon/3$ and the middle one from equicontinuity. □

More applications?

Theorem 4.3.14: Montel's Theorem

If $\Omega \subset \mathbb{C}$ is open and $\mathcal{E}$ is a set of holomorphic functions $\Omega \to \mathbb{C}$ that is locally uniformly bounded, then any sequence in $\mathcal{E}$ has a subsequence that converges locally uniformly / uinformly on compact subsets and the limit is holomorphic. This leads to the proof of Riemann Mapping Theorem.

Beginning of March 5, 2021

Now we show one more application of Arzelá-Ascoli, returning to the uniform convergence of derivativees. Some propositions first.

Proposition 4.3.15

Let $f_n : [a, b] \to \mathbb{R}$ be differentiable functions. Assume $f_n' \to g$ uniformly to some g and f_n is bounded somewhere, i.e., for some $x_0 \in [a, b]$, the sequence $\{f_n(x_0)\}$ is bounded sequence of real numbers. Then $\{f_n\}$ has a subsequence that converges uniformly to f with $f' = g$.
Proof is immediate since a bounded sequence of real numbers admits a convergent subsequence and this subsequence converges pointwise somewhere with a uniformly convergent derivative, which from HW5 + some previous theorem implies the existence and differentiability of f.

What if we weaken one of the hypotheses, now we require instead that $\{f_n'\}$ are uniformly bounded rather than uniformly convergent? (*Left as exercise to show that this is indeed a weaker condition.*)

Corollary 4.3.16

Suppose $f_n : [a, b] \to \mathbb{K}$ is differentiable for each $n \geqslant 1$, and there exists M such that $|f_n'(x)| \leqslant M$ for all $n \geqslant 1$ and $x \in [a, b]$ (*this replaces the uniform convergent derivatves*). Also assume that f_n is bounded somewhere, say, by C ($|f_n(x_0)| \leqslant C$ for all $n \geqslant 1$). Then $\{f_n\}$ has a uniformly convergent subsequence (no claims on differentiability is made).

Proof. Notice that M is a uniform Lipschitz constant for $\{f_n\}$ and therefore $\{f_n\}$ is equicontinuous (with $\delta := \delta/M$ in the proof). Also, if $x \in [a,b]$, then $|f_n(x)| \leqslant |f_n(x) - f_n(x_0)| + |f_n(x_0)| \leqslant M|x - x_0| + C \leqslant M(b-a) + C$, independent of n (even x). Therefore $\{f_n\}$ is uniformly (in particular pointwise) bounded. It remains to apply Arzelá-Ascoli theorem, and the claim follows. ☐

4.4 Uniform Approximation in $C^0[a,b]$

Recall that the set of polynomials is the span of $\{1, x, x^2, x^3, \dots\}$, i.e., the set of finite linear combinations of these elements. Having discussed the compactness of $\mathcal{E} \subset (C^0(X,\mathbb{K}))$, the next question to consider is: *is $\mathcal{E} \subset (C^0(X,\mathbb{K}))$ dense, in particular for $\mathcal{E} \subset (C^0[a,b],\mathbb{R})$?* We already know that any analytic functions can be uniformly approximated by polynomials (Taylor series), but what about all continuous functions in general?

The answer is yes; any $f \in C^0[a,b]$ is the uniform limit of polynomials!

*For another example, consider X the unit circle $S^1 \subset \mathbb{C}$. It is compact (with $S^1 = \{e^{2\pi i\theta} : \theta \in \mathbb{R}\}$). Let $\mathcal{A}$ be the span of $\{e_n(\theta) := e^{2\pi in\theta}\}$. Note that $e^{2\pi in\theta} = \cos(2\pi n\theta) + i\sin(2\pi in\theta)$. Because of this, $\mathcal{A}$ is called the set of **trigonometric polynomials**. In fact, $\mathcal{A}$ is dense in $C^0(S^1,\mathbb{C})$ (which can be bijectively mapped to the set of 2π-periodic continuous functions from $\mathbb{R}$ to $\mathbb{C}$). This is by Stone-Weierstraß Theorem. In particular, for Lipschitz continuous functions $f : S^1 \to \mathbb{C}$, such polynomials is simply its Fourier series, of which the partial sums are trigonometric polynomials.*
One more remark: partial sums of Fourier series of f are closest trigonometric polynomials to f in $\|\cdot\|_2$ (to be shown in HW), but not necessarily in $\|\cdot\|_{\sup}$.

Theorem 4.4.1: Stone-Weierstraß Theorem(s)

Let X be compact, and let $\mathcal{A} \subset C^0(X,\mathbb{R})$ satisfy the following 3 properties:

(1) $\mathcal{A}$ is a *function algebra* (sub-$\mathbb{R}$ algebra of $C^0(X,\mathbb{R})$), i.e., $\mathcal{A}$ is closed under sums, $\mathbb{R}$-scalar multiples, and function multiples (if $f, g \in \mathcal{A}$ then $fg \in \mathcal{A}$),

(2) $\mathcal{A}$ "vanishes nowhere", i.e., if $x \in X$ then $f(x) \neq 0$ for some $f \in \mathcal{A}$, and

(3) $\mathcal{A}$ "separates points", i.e., if $x \neq y$ then $f(x) \neq f(y)$ for some $f \in \mathcal{A}$.

Then the real version of **Stone-Weierstraß** states that $\mathcal{A}$ is dense in $(C^0(X,\mathbb{R}), \|\cdot\|_{\sup})$.

Theorem 4.4.2

The complex version of **Stone-Weierstraß** is analogous, but $\mathcal{A} \subset C^0(X,\mathbb{C})$ and (1) becomes *$\mathcal{A}$ is a $\mathbb{C}$-function algebra: closed under sums, $\mathbb{C}$-scalar multiples, function multiplication and complex conjugation, i.e., if $f \in \mathcal{A}$ then $\overline{f} \in \mathcal{A}$ where $\overline{f}(x) := \overline{f(x)}$.*
Then $\mathcal{A}$ is dense in $(C^0(X,\mathbb{C}), \|\cdot\|_{\sup})$.

Beginning of March 8, 2021

Say we have $f \in C^0([0,1],\mathbb{R})$ and we want to uniformly approximate f by polynomials.

Probabilistic interpretation: f depends on "the probability x that an unfair coin lands heads on a single flip." We want to compute "f at the true probability value x" but we don't know x. Like in real life, we can only do a finite number n of coin flips to guess x.

Strategy: do as said above; say we get heads k times. Then our guess is that $x = k/n$. Then evaluate f at $f(k/n)$. As long as we have enough flips (for large n), we are confident that the expected value should be sufficiently close to the actual x.

(I) What is the expected outcome of this strategy, given the true probability of x? If we do n flips (each with a probability x of getting heads), what is the probability of getting k heads? The answer is given by the **Bernstein basis polynomial**

$$p(k \mid n) = \binom{n}{k} x^k (1-x)^{n-k}.$$

(II) The expected value of f given this strategy:

$$\sum_{k=0}^{n} f(k/n) p(k \mid n) = \sum_{k=0}^{n} f(k/n) \binom{n}{k} x^k (1-x)^{n-k}.$$

This is called the **Berstein polynomial**. Notice that the Berinstein polynomial is a polynomial of x of degree n, a *much* nicer condition than f, which we know nothing more beyond its continuity.

Intuition: take $n = 1$. If it lands on heads we guess the probability to be 1; if tails, we assume $x = 0$. Then the $n = 1$ Berstein polynomial of f is simply $f(0)(1-x) + f(1)(x) = f(0) + x(f(1) - f(0))$. This clearly is a bad guess but at least it's a polynomial. We'll soon show that as n gets larger, the guess becomes more accurate, and eventually we can uniformly approximate f using a Bernstein polynomial.

Let $r_k(x) = \binom{n}{k} x^k (1-x)^{n-k}$. For example, if $n = 1$, $r_0(x) = 1 - x$ and $r_1(x) = x$. For $n = 2$, we have $r_0(x) = (1-x)^2, r_1(x) = 2x(1-x)$, and $r_2(x) = x^2$.

In general, for a given n, there are $n+1$ Berstein basis polynomials $r_k(x)$. Also notice that the set $\{1, x, x^2, \dots, x^n\}$ contains exactly $n+1$ elements [basis!], and they serve as the standard basis for vector space of polynomials of degree $\leqslant n$. In fact, these Berinstein basis polynomials also form a basis for same vector space.

(III) Next sub-goal: we will prove some properties of $r_k(x)$, which can be viewed as binomial distributions over possible k values given x. Then we will use these properties to prove Weierstraß approximation.

(0) $r_k(x) \geqslant 0$ for all n, k, and $x \in [0,1]$. Obvious.

(1) For any x, $\sum_{k=0}^{n} r_k(x) = 1$. After all, the sum of all probabilities of all possible outcomes should be 1, as the $r_k(x)$'s give a probability distribution of k. Indeed,

$$1 = (x + (1-x))^n = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = \sum_{k=0}^{n} r_k(x).$$

Beginning of March 10, 2021

Before moving to the next equation, some definitions from probability theory first.

(2) If, for $k \in \{0, \dots, 1\}$ we have $p_k \geqslant 0$ and $\sum_{k=0}^{n} p_k = 1$, we say $(p_0, p_1, \dots, p_n)$ forms a **probability distribution** on the finite set $\{0, 1, \dots, n\}$.

(3) For $c \in \mathbb{R}$ and $m \geqslant 0$ and integer, the $\boldsymbol{m}^{\mathbf{th}}$ **moment** of the probability distribution $(p_0, p_1, \ldots, p_n)$ is defined to be

$$\sum_{k=0}^{n} (k-c)^m p_k.$$

(a) For example, if $m = 0$ and $c =$ anything then the corresponding 0^{th} moment is $\sum_{k=0}^{n} 1 \cdot p_k = 1$.

(b) If $m = 1, c = 0$, we get the **expected value** $\sum_{k=0}^{n} k p_k$ of the distribution. This equals nx because

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

$$\text{take } \frac{\partial}{\partial x} \text{ of both sides} \implies n(x+y)^{n-1} = \sum_{k=1}^{n} k \binom{n}{k} x^{k-1} y^{n-k}$$

$$\text{multiply both sides by } x \implies nx(x+y)^{n-1} = \sum_{k=1}^{n} k \binom{n}{k} x^k y^{n-k}.$$

Therefore, letting $y = 1 - x$ gives

$$nx = \sum_{k=1}^{n} k \binom{n}{k} x^k (1-x)^{n-k} = \sum_{k=1}^{n} k r_k(x) = \sum_{k=0}^{n} k r_k(x).$$

(c) For $m \geqslant 2$, given $(p_0, p_1, \ldots, p_n)$ a probability distribution on $\{0, 1, \ldots, n\}$, the $\boldsymbol{m}^{\mathbf{th}}$ **central moment** of the distribution is the (m, c) moment at $c =$ the expected value E (i.e., the $m = 1, c = 0$ moment). This gives rise to the 2^{nd} central moment, $\sum_{k=0}^{n} (k-E)^2 r_k(x)$, known as the **variance**. It follows that, if $p_k = r_k(x)$ for some $x \in [0,1]$ (binomial distribution), then the variance is $nx(1-x)$. The square root of variance is known as the **standard deviation**. Indeed, from above we have

$$n(x+y)^{n-1} = \sum_{k=1}^{n} k \binom{n}{k} x^{k-1} y^{n-k}$$

$$\text{take } \frac{\partial}{\partial x} \text{ of both sides} \implies n(n-1)(x+y)^{n-2} = \sum_{k=2}^{n} k(k-1) \binom{n}{k} x^{k-2} y^{n-k}$$

$$\text{multiply both sides by } x^2 \implies n(n-1)x^2(x+y)^{n-2} = \sum_{k=2}^{n} k(k-1) \binom{n}{k} x^k y^{n-k}.$$

Then setting $y = 1 - x$ gives

$$n(n-1)x^2 = \sum_{k=2}^{n} k(k-1) \binom{n}{k} x^k (1-x)^{n-k} = \sum_{k=0}^{n} k(k-1) r_k(x).$$

Splitting the parenthesis gives

$$n(n-1)x^2 = \sum_{k=0}^{n} k^2 r_k(x) - \sum_{k=0}^{n} k r_k(x) = \sum_{k=0}^{n} k^2 r_k(x) - nx.$$

Therefore,

$$\begin{aligned}
\sum_{k=0}^{k} (k-nx)^2 r_k(x) &= \sum_{k=0}^{n} (k^2 - 2nxk + n^2x^2) r_k(x) \\
&= \sum_{k=0}^{n} k^2 r_k(x) - 2nx \sum_{k=0}^{n} k r_k(x) + n^2x^2 \sum_{k=0}^{n} r_k(x) \\
&= \left[n(n-1)^2 + nx \right] - 2nx(nx) + n^2x^2(1) \\
&= nx(1-x).
\end{aligned}$$

(IV) **Main proof of the theorem** (Berinstein 1912-1913).

Theorem 4.4.3

Let $f \in (C^0[0,1],\mathbb{R})$ (or replace $\mathbb{R}$ with $\mathbb{C}$). The polynomial functions

$$p_n(x) = \sum_{k=0}^{n} f(k/n)\binom{n}{k}x^k(1-x)^{n-k}$$

converge uniformly to f on $[0,1]$.

Proof. Let $\epsilon > 0$ be given. Since f is continuous on the compact domain $[0,1]$, it is uniformly continuous and bounded. Therefore, uniform continuity says there exists some $\delta > 0$ such that $|f(x) - f(x')| < \epsilon/2$ for all $x, x' \in [0,1]$ with $|x - x'| < \delta$, and boundedness says there exists there exists M such that $|f(x)| \leqslant M$ for all $x \in [0,1]$.

Let $N \geqslant M/(\epsilon\delta^2)$ be a sufficiently large integer. Claim: if $n \geqslant N$ then $|p_n(x) - f(x)| < \epsilon$ for all $x \in [0,1]$.

Notice that we can re-write $f(x)$ as $f(x)\sum_{k=0}^{n} r_k(x)$. Then we can distribute the sum.

$$\begin{aligned}
|p_n(x) - f(x)| &= \left|\sum_{k=0}^{n} f(k/n)\binom{n}{k}r_k(x) - \sum_{k=0}^{n} f(x)r_k(x)\right| \\
&= \left|\sum_{k=0}^{n}(f(k/n) - f(x))r_k(x)\right| \\
&\leqslant \sum_{\substack{k=0 \\ |x-\frac{k}{n}<\delta|}}^{n}\left|(f(k/n) - f(x))r_k(x)\right| + \sum_{\substack{k=0 \\ |x-\frac{k}{n}\geqslant\delta|}}^{n}\left|(f(k/n) - f(x))r_k(x)\right| && \text{for convenience, denote as } \sum_1 \,\&\, \sum_2 \\
&< \sum_1(\epsilon/2)r_k(x) + \sum_2 2Mr_k(x) && \epsilon/2 \text{ by unif. cont; } 2M \text{ by boundedness} \\
&\leqslant \frac{\epsilon}{2} + \sum_2 2M\cdot\frac{|k-nx|^2}{(n\delta)^2} && \text{since } 1 \leqslant \left(\frac{|k/n - x|}{\delta}\right)^2 = \frac{|k-nx|^2}{(n\delta)^2} \\
&\leqslant \frac{\epsilon}{2} + \frac{2M}{(n\delta)^2}\sum_2|k-nx|^2 r_k(x) \\
&\leqslant \frac{\epsilon}{2} + \frac{2M}{n\delta^2}x(1-x) && \text{since } \sum_2 \leqslant \sum_{k=0}^{n}|k-nx|^2 r_k(x) = nx(1-x) \\
&\leqslant \frac{\epsilon}{2} + \frac{2M}{n\delta^2}\cdot\frac{1}{4} = \frac{\epsilon}{2} + \frac{M}{2n\delta^2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \text{ for sufficiently large } n.
\end{aligned}$$

□

Remark. For Stone-Weierstraß theorem, we will apply the classical one (above) to $f(x) = |x|$ on $[-1,1]$ plus some more lemmas. Then compactness will finish the remaining job. To be continued in the near future.

Beginning of March 15, 2021

Stone-Weierstraß Theorem (Stone, 1937)

Before proving the Stone-Weierstraß Theorem, we need some lemmas first.

Lemma 4.4.4

If $\mathcal{A} \in C^0(X, \mathbb{R})$ is a function algebra, then so is its closure $\overline{\mathcal{A}}$ (i.e., closed).

Proof. $\overline{\mathcal{A}}$ is closed under sums: if $f, g \in \overline{\mathcal{A}}$ such that there exist $\{f_n\}, \{g_n\} \subset \mathcal{A}$ with $f_n \to f$ and $g_n \to g$ uniformly, then $f_n + g_n \to f + g$ uniformly. Similarly, $\overline{\mathcal{A}}$ is closed under scalar multiples and function multiples. (Proofs are not hard.) □

Lemma 4.4.5

Approximation of continuous functions by polynomials no only work for $f \in C^0([0,1], \mathbb{R})$ but also for $f \in C^0([a,b], \mathbb{R})$. *We'll focus on* $f(x) = |x|$ *on* $[-1, 1]$ *later, thanks to this lemma.*

Proof. We simply need to linearly rescale the functions by $\Phi : C^0([a,b], \mathbb{R}) \to C^0([0,1], \mathbb{R})$ with

$$f(x) \mapsto f((x-a)/(b-a)) \text{ and } f^{-1}(y) =\mapsto f^{-1}(a + y(b-a))$$

(where $x \in [a,b]$ and $y \in [0,1]$). Notice that these operations are linear transformations of vector spaces (isomorphism, in particular). Thus they preserve $\|\cdot\|_{\text{sup}}$.
Therefore, $f_n \to f$ uniformly in $C^0([a,b], \mathbb{R})$ if and only if the corresponding $\Phi(f_n) \in C^0([0,1], \mathbb{R})$ converge uniformly to $\Phi(f) \in C^0([0,1], \mathbb{R})$.
Furthermore, these linear transformations preserve the subset of polynomial functions, i.e., polynomials in $C^0([a,b], \mathbb{R})$ gets mapped to polynomials in $C^0([0,1], \mathbb{R})$ and vice versa. This proves the lemma. □

Now two slightly more difficult lemmas.

Lemma 4.4.6

Let X be compact and let $\mathcal{A} \subset C^0(X, \mathbb{R})$ be a function algebra (and $\overline{\mathcal{A}}$ is also a function algebra). Then if $f \in \overline{\mathcal{A}}$, $|f| \in \overline{\mathcal{A}}$, i.e., $\overline{\mathcal{A}}$ also has closure under absolute values.

Proof. Main idea: $|f|$ is a composition of f and the usual absolute function: $X \xrightarrow{f} \mathbb{R} \xrightarrow{|\cdot|} \mathbb{R}$. We will approximate the second absolute value function by a polynomial (on some compact interval containing the image of f as Weierstraß approximation works for compact domain). Ideally, we want to approximate $|x|$ by $a_1 x + a_2 x^2 + \cdots + a_n x^n$. Then,

$$(|\cdot| \circ f)(x) = a_1 f + a_2 f^2 + \cdots + a_n f^n,$$

where each f^k is obtained from function multiplication and are in $\overline{\mathcal{A}}$, and $|f|$ is a linear combination of elements of $\overline{\mathcal{A}}$. Thus $f \in \overline{\overline{\mathcal{A}}} = \overline{\mathcal{A}}$.

Formally: note that $\|f\|_{\text{sup}}$ is finite since X is compact. Consider the absolute value function

$$[-\|f\|_{\text{sup}}, \|f\|_{\text{sup}}] \to \mathbb{R} \text{ defined by } y \mapsto |y|.$$

This is continuous, and by the lemma involving "transformations" of Weierstrß approximation there exists a polynomial $p(y)$ such that $|p(y) - |y|| < \epsilon/2$ for all y in the domain. In parciular, $|p(0) - |0|| < \epsilon/2$, i.e., $|p(0)| < \epsilon/2$.

Therefore, there exists a polynomial $q(y)$ such that $q(0) = 0$ (i.e., the constant term vanishes!!) and $|q(y) - |y|| < \epsilon$. To see this, simply let $q(y) := p(y) - p(0)$. We know $|p(0) - |0|| < \epsilon/2$. Therefore for all y in the domain,

$$|q(y) - |y|| = |p(y) - p(0) - |y|| \leqslant |p(y) - |y|| + |p(0)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Now we expand q as $q(y) = a_1 y + a_2 y^2 + \cdots + a_n y^n$, without constant term. Let $g : X \to \mathbb{R}$ be defined by $g = q \circ f$. Then

$$g = a_1 f + a_2 f^2 + \cdots + a_n f^n \in \overline{\overline{\mathcal{A}}} = \overline{\mathcal{A}}.$$

We want to show that g is a nice approximation of $|f|$, i.e., $\|g - |f|\|_{\sup}$ is small. Indeed,

$$\begin{aligned}\|g - |f|\|_{\sup} &= \sup_{x \in X} |q(f(x)) - |f(x)|| \\ &= \sup_{y \in \operatorname{im}(f)} |q(y) - |y|| \leqslant \epsilon\end{aligned}$$

since the image of f is a subset of $[-\|f\|_{\sup}, \|f\|_{\sup}]$. Therefore $|f|$ can be approximated arbitrarily closely by elements of $\overline{\mathcal{A}}$. Thus $|f| \in \overline{\overline{\mathcal{A}}} = \overline{\mathcal{A}}$.

□

Immediately following the above lemma, we have the following:

Lemma 4.4.7

Let X be compact and let $\mathcal{A} \subset C^0(X, \mathbb{R})$ be a function algebra. If $f, g \in \overline{\mathcal{A}}$ then $\max(f,g), \min(f,g)$ (pointwise maximum / minimum) are also in $\overline{\mathcal{A}}$.

Proof. Indeed,

$$\max(a,b) = \frac{a+b}{2} + \frac{|a-b|}{2} \text{ and } \min(a,b) = \frac{a+b}{2} - \frac{|a-b|}{2}.$$

We either get a or b when evaluating the max or min. Therefore if $f, g \in \overline{\mathcal{A}}$ then so are $\max(f,g), \min(f,g)$.

□

One more lemma to go! To be shown next lecture.

Lemma 4.4.8

Let X be compact. Assume $\mathcal{A} \subset C^0(X, \mathbb{R})$ satisfies all hypotheses of the Stone-Weierstraß theorem (i.e., a function algebra that vanishes nowhere and separates points), then we have a stronger result: if $x_1, x_2 \in X$ with $x_1 \neq x_2$ and given $c_1, c_2 \in \mathbb{R}$, then there exists $f \in \mathcal{A}$ with $f(x_1) = c_1$ and $f(x_2) = c_2$. In other words, we can prescribe the values at 2 points $\in X$ and get $f \in \mathcal{A}$ which satisfies this prescription.

Beginning of March 17, 2021

Proof. Define $\Phi : \mathcal{A} \to \mathbb{R}^2$ defined by $\Phi(f) := (f(x_1), f(x_2))$. Clearly as $\mathcal{A}$ is infinite-dimensional, Φ cannot be injective. We are, and will, instead show that Φ is surjective (which is what the lemma claims).

Notice that Φ is a linear transformation of vector spaces:

$$\begin{aligned}\Phi(f+\lambda g) &= ((f+\lambda g)(x_1), (f+\lambda g)(x_2)) \\ &= (f(x_1)+\lambda g(x_1), f(x_2)+\lambda g(x_2)) \\ &= \Phi(f)+\lambda\Phi(g).\end{aligned}$$

(*Manion doesn't like this $f+\lambda g$ notion! He prefers to show $f+g$ and λg separately.*) Therefore the image of Φ is a vector subspace of $\mathbb{R}^2$, whose dimension can be $0, 1$, or 2. It remains to show that 0 and 1 both give contradictions.

Suppose $\dim(\mathrm{im}(\Phi)) = 0$, i.e., $\Phi(f) = (0,0)$ for all $f \in \mathcal{A}$. This is clearly a contradiction as a function algebra is assumed to vanish nowhere.

Now suppose $\dim(\mathrm{im}(\Phi)) = 1$. Then $\mathrm{im}(\Phi)$ can be any line through the origin in $\mathbb{R}^2$. Clearly it cannot be the x- or y-axis as $\mathcal{A}$ is assumed to vanish nowhere. However, even for other lines, if $(c_1, c_2) \in \mathrm{im}(\Phi)$ then $c_1 = 0 \iff c_2 = 0$, i.e., $f(x_1) = 0 \iff f(x_2) = 0$.

Our goal now is to construct some $f \in \mathcal{A}$ with $f(x_1) = 0 \neq f(x_2)$. Since $\mathcal{A}$ separates points, there exists some $f \in \mathcal{A}$ with $f(x_1) \neq f(x_2)$. *It's natural to think about $g(x) := f(x) - f(x_1)$ but g may or may not still be in $\mathcal{A}$.* The fix: let $g(x) = f(x)[f(x) - f(x_1)] = f(x)^2 - f(x)f(x_1)$. It becomes clear that $g \in \mathcal{A}$ by closure of function addition, multiplication, and scalar multiplication. Indeed,

$$g(x_1) = f(x_1)^2 - f(x_1)^2 = 0 \text{ and } g(x_2) = f(x_2)[f(x_2) - f(x_1)] \neq 0 \text{ by assumption.}$$

□

Finally, **proof of the Stone-Weierstraß Theorem**.

Recall that $\mathcal{A} \subset C^0(X, \mathbb{R})$ is a function algebra that vanish nowhere and separates points. Let $F \in C^0(X,\mathbb{R})$ and let $\epsilon > 0$ be given and fixed. We want to find $G \in \overline{\mathcal{A}}$ such that $\|F - G\|_{\sup} < \epsilon$. If we could show this then $F \in \overline{\overline{\mathcal{A}}} = \overline{\mathcal{A}}$ and we are done.

Idea: given a $p \in X$, we first find $G_p \in \overline{\mathcal{A}}$. It doesn't necessarily have to satisfy $\|F - G\|_{\sup} < \epsilon$, but we want to ensure $G_p(x) > F(x) - \epsilon$ for all x, i.e., "half of our desired condition". We also want G_p to be such that $G_p(x) < F(x) + \epsilon$ for x in an open neighborhood of p. This is possible because, given $p, q \in X$, there exists $H_{pq} \in \mathcal{A}$ such that

$$H_{pq}(x) > F(x) - \epsilon \text{ for x} \in \text{an open neighborhood of } q$$

and

$$H_{pq}(x) < F(x) + \epsilon \text{ for x} \in \text{an open neighborhood of } p.$$

Then we use the "q-neighborhoods" to cover X with open sets and extract a finite subcover, from which we can take the maximum.

Then we take the finite cover of all such covers (recall X is compact, in particular covering compact). This gives $G_{p_1}, \ldots, G_{p_n}$. The claim then follows from taking $G(x) := \min_{1 \leqslant i \leqslant n} \{G_{p_1}(x), \ldots, G_{p_n}(x)\}$.

Proof of Stone-Weierstraß. For any $p \neq q$ in X, the *prescription* lemma above (the latest one) implies that there exists a function $H_{pq} \in \mathcal{A}$ such that $H_{pq}(p) = F(p)$ and $H_{pq}(q) = F(q)$. Recall that $\mathcal{A}$ is a subset of the continuous functions. Therefore, since H_{pq} and F agree at p and q, there indeed exists a neighborhood of p in which $H_{pq} > F - \epsilon$ and a (maybe another) neighborhood of q, U_{pq}, in which $H_{pq} < F + \epsilon$. (If $q = p$ then this claim is also automatically true, as any of the other H_{pq}'s with distinct p, q already implies that there exists $H \in \mathcal{A}$ such that $H(p) = F(p)$.)

For a fixed p, the compactness of X admits a finite subcovering consisting of U_{p,q_n}'s. Define

$$G_p(x) := \max_{1 \leqslant i \leqslant n} H_{p,q_i}(x).$$

This is the $G_p(x)$ we are looking for. On one hand, $G_p(x) > F(x) - \epsilon$ for all $x \in X$; on the other hand, the continuity of G_p implies that, for any $p \in X$, $G_p < F + \epsilon$ on some (open) neighborhood V_p of p. Letting p vary, we obtain another open cover of X by $\{V_p : p \in X\}$ which admits a finite subcovering $\{V_{p_1}, \ldots, V_{p_m}\}$. Therefore we can define

$$G(x) := \min_{1 \leqslant i \leqslant m} G_{p_i}(x).$$

The min/max lemma gives $G \in \overline{\mathcal{A}}$. We know $G_{p_i}(x) > F(x) - \epsilon$ for all x so $G(x) > F(x) - \epsilon$ for all x. On the other hand, if $x \in X$ then $x \in V_{p_i}$ for some i, so indeed $G(x) < F(x) + \epsilon$ for all $x \in X$. Therefore

$$F(x) - \epsilon < G(x) < F(x) + \epsilon \text{ for all } x \in X, \text{i.e., } \|F - G\|_{\sup} < \epsilon.$$

This proves the claim! □

Remark. **Complex version.** If $\mathcal{A}$ is closed under complex conjugation, then if $f \in \mathcal{A}$ we have

$$\mathfrak{Re}(f) = \frac{f + \overline{f}}{2} \in \mathcal{A} \text{ and } \mathfrak{Im}(f) = \frac{f - \overline{f}}{2i} \in \mathcal{A}.$$

Given $F \in C^0(X, \mathbb{C})$, we have $\mathfrak{Re}(F), \mathfrak{Im}(F) \in C^0(X, \mathbb{R})$. By the real version of Stone-Weierstraß we can find sequences $\{P_n\}, \{Q_n\}$ that converge uniformly to $\mathfrak{Re}(F)$ and $\mathfrak{Im}(F)$. It follows then that

$$\|F - (P_n + iQ_n)\|_\infty \leqslant \|\mathfrak{Re}(F) - P_n\|_\infty - \|\mathfrak{Im}(F) - Q_n\|_\infty \to 0.$$

Beginning of March 19, 2021

Before moving to the next topic, we introduce another application of Stone-Weierstraß:

Theorem 4.4.9

Let D^2 be the unit disk in $\mathbb{R}^2$. Let $F : D^2 \to \mathbb{R}^2$ be continuous ("vector field on D^2"). Then, for all $\epsilon > 0$, there exists some "other vector field" $G : D^2 \to \mathbb{R}^2$ continuous such that

$$d_{\sup}(F, G) < \epsilon \text{ w.r.t. Euclidean norm on } \mathbb{R}^2.$$

and G vanishes at most finitely many points.

Proof. Let $\mathcal{A}$:= the set of *two-variable polynomials* $\sum_{i,j=0}^{n} c_{ij} x^i y^j = \mathbb{R}[x, y]$, the *polynomial ring of two variables over* $\mathbb{R}$. We will view these polynomials as functions on D^2 (not all of $\mathbb{R}^2$). If two such polynomials agree on D^2 then they agree on $\mathbb{R}^2$. Notice that D^2 is compact and $\mathcal{A} \subset C^0(D^2, \mathbb{R})$ also satisfies all the hypotheses of

Stone-Weierstraß.

If we write $F(x,y) = (F_1(x,y), F_2(x,y))$, then $F_1, F_2 \in C^0(D^2, \mathbb{R})$. Therefore there exist $P, Q \in \mathcal{A}$ with $d_{\sup}(P, F_1) < \epsilon/\sqrt{2}$ and $d_{\sup}(Q, F_2) < \epsilon/\sqrt{2}$. It follows that

$$d_{\sup}((P,Q),(F_1,F_2)) = \sup_{x \in D^2} \sqrt{(P-F_1)(x)^2 + (Q-F_2)(x)^2} < \sqrt{\epsilon^2/2 + \epsilon^2/2} = \epsilon.$$

These polynomials vanish at most finitely many points.

An algebraic brief proof: *$R[x,y]$ is a "unique factorizatizion[sic] domain, UFD and thus Q has finitely many irreducible factors. Consider the polynomials $P + \delta$ where δ is any small constant δ. If $\delta \neq \delta'$ then $P+\delta, P+\delta'$ have no irreducible factors in common (if so, $\delta - \delta'$ would be an irreducible polynomial of degree > 0, contradiction). Therefore there exists small δ such that $P' := P + \delta$ and Q have no irreducible factors in common, and (P', Q) is still uniformly close to F.*

Let $P := P'$. Fact: Res$(P,Q) \in \mathbb{R}[x]$ (the resultant). It vanishes for some x if and only if for some y such that $P(x,y) = Q(x,y) = 0$.

If Res$(P,Q) = 0$ then as a polynomial in $\mathbb{R}[x]$ then P, Q have a common polynomial factor of degree > 0, contradiction.

If Res(P,Q) is a nonzero polynomial in x then it vanish at finitely many x (for all but finitely many x, there does not exist y with $P(x,y) = Q(x,y) = 0$). For the rest of points x (finitely many), $P(x,-)$ and $Q(x,-)$ are polynomials in y and so they have finitely many zeros. This proofs the claim.

□

4.5 Contractions and ODEs

As far as we are concerned, there are two kinds of differential equations, Partial (PDEs) and ordinary (ODEs). For example, the wave equation

$$\frac{\partial^2 f}{\partial f^2} = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

is an equation for an unknown function f of variables x, y, t involving partial derivatives of f.

For ODEs, things look simpler, for example

$$f''(t) - 3f'(t) + 2f(t) = g(t) \text{ with a given } g.$$

Here we need to solve for f considering only one variable, t. There are all kinds of "tricks" and "recipes", for example

$$f''(t) - 3f'(t) + 2f(t) = 0$$

can be viewed as

$$\left(\frac{\partial^2}{\partial t^2} - 3\frac{\partial}{\partial t} + 2\right)(f) = 0 \implies \left(\frac{\partial}{\partial t} - 1\right)\left(\frac{\partial}{\partial t} - 2\right)(f) = 0.$$

The trick (haven't justified): the derivative operators commute so one of them must be 0. This gives solutions $A_1 e^t$ and $A_2 e^{2t}$. This gives the general solution

$$f(t) = A_1 e^t + A_2 e^{2t}.$$

In a linear algebra sense, this space of solutions is a vector space and it is 2-dimensional: a basis is given by $\{e^t, e^{2t}\}$. The next question arises: why does this trick find all solutions? We will show this soon.

First task: **reduce to first-order systems**. For example, consider again $f'' - 3f' + 2f = 0$. Since $f : \mathbb{R} \to \mathbb{R}$ we can define $u(t) := f(t)$ and $v(t) := f'(t)$. Therefore f solves an ODE if and only if u, v satisfy

$$u' = v \text{ and } v' = 3v - 2u.$$

Now we have reduced the second-order ODE (involving f'') to a system of two first-order ODEs. This gives

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

for a vector-valued function $\left[u(t), v(t)\right]^T : \mathbb{R} \to \mathbb{R}^2$. More generally, we have $x = x(t) : \mathbb{R} \to \mathbb{R}^n$ where we can consider the equations of form $x'(t) = F(x(t))$ where F is a function $\mathbb{R} \to \mathbb{R}$, i.e., a vector field on $\mathbb{R}^n$.

Limitations:

(1) We could really have $x'(t) = F(t, x(t))$ where F is continuous on $\mathbb{R} \times \mathbb{R}^n$.

(2) F might be a defined on an open subset of $\mathbb{R}^n$, not the entire $\mathbb{R}^n$.

Definition 4.5.1

Let $U \subset \mathbb{R}^n$ be open. Let $F : U \to \mathbb{R}$ be continuous. A solution to the system of ODEs $x'(t) := F(x(t))$ with an "initial value" p at $t_0 \in (a, b)$ is a differentiable function $x : (a, b) \to \mathbb{R}^n$ for some (a, b) such that

$$x'(t) = F(x(t)) \text{ for all } t \in (a, b) \text{ and } x(t_0) = p.$$

Theorem 4.5.2: Picard-Lindelöf Theorem, "autonomous case"

Let $U \subset \mathbb{R}^n$ be open and let $F : U \to \mathbb{R}^n$ be locally Lipschitz. Then, given $p \in U$, there exists a solution to $x'(t) = F(x(t))$ defined in some neighborhood (a, b) of t_0 that satisfies $x(t_0) = p$. (The solution is a function $(a, b) \to U$). Furthermore, any two solutions satisfying these properties (defined on $(a, b), (a', b')$ containing t_0) must agree on some smaller interval (a'', b'') contained in both (a, b) and (a', b'). This gives local uniqueness of solutions. *Cf. HW9; in fact they must agree on all of the intersection.*

Beginning of March 22, 2021

The Picard-Lindelöf Theorem can be generalized into the following form (that can be applied to differential forms):

Theorem 4.5.3: Picard-Lindelöf Theorem, "time-dependent case"

(Picard-Lindelöf, ~ 1880; Lipschitz 1876.) Let $\Omega \subset \mathbb{R} \times \mathbb{R}^n$ be open (the first $\mathbb{R}$ represent *time*). Let $F : \Omega \to \mathbb{R}^n$ a "*time-varying vector field*". Let elements of Ω be of form (t, y) where $t \in \mathbb{R}$ (time) and $y \in \mathbb{R}^n$ (spatial coordinates). Assume

(1) F is continuous on Ω, "jointly continuous in t and in y, and

(2) F is locally Lipschitz in y and locally uniformly in t, i.e., for all $(t_0, y_0) \in \Omega$, there exists an open neighborhood V of $(t_0, y_0) \in \Omega$ and $L \geqslant 0$ such that, whenever $(t, y), (t, y') \in V$, the Lipschitz condition holds. *Must compare t with the same t!*

Then for all $(t_0, y_0) \in \Omega$, there exists an open neighborhood (a, b) of t_0 and a differentiable function $\gamma : (a, b) \to \mathbb{R}^n$ such that

(1) For all $t \in (a, b)$, the pair $(t, \gamma(t)) \in \Omega$,

(2) $\gamma'(t) = F(t, \gamma(t))$ for all $t \in (a, b)$, and

(3) $\gamma(t_0) = y_0$.

Local uniqueness holds as before.

Remark. Differential equations like $f''(t) - 3'(t) + 2f(t) = 0$ is an autonomous system (now and also after reduction of order). On the other hand, something like $f''(t) = 3f'(t) + 2f(t) = \cos(t)$ is non-autonomous as now the RHS is of form $g(t)$. This corresponds to the more general time-varying case.

Contraction

Now we talk about fixed points of functions $f : X \to X$.

Definition 4.5.4

Let X be a set and $f : X \to X$ a function. We say $x \in X$ is called a **fixed point** of f if $f(x) = x$. The **orbit** of x under f is given by $\{x, f(x), f^2(x), \dots\}$. Thus equivalently x is a fixed point of f if and only if its orbit is just $\{x\}$.

Example 4.5.5: Brouwer Fixed-Point Theorem. Any continuous $f : D^n \to D^n$ has at $\geqslant 1$ fixed point.

We will be especially concerned with fixed points for contraction mappings:

Definition 4.5.6

If (X, d) is a metric and $f : X \to X$ a function, we say f is a **contraction** (or contractive mapping) if there exists $k < 1$ such that $d(f(x), f(y)) \leqslant k(x, y)$ for all $x, y \in X$. If such $k < 1$ does not exist and $d(f(x), f(y)) < d(x, y)$ for all x, y, f is called a **weak contraction**.

Example 4.5.7. Let $X := [1, \infty)$. $f : X \to X$ defined by $f(x) = x + 1/x$ is a weak contraction but not a contraction. It does not admit a fixed point as $f(x) > x$ for all x.

Theorem 4.5.8: Banach Contraction Mapping Theorem

Let (X, d) be a complete metric space and $f : X \to X$ a contraction. Then f has a unique fixed point.

Beginning of March 24, 2021

Proof. Uniqueness is clear (though it should be mentioned only after the existence has been shown). Suppose x, y are fixed points then $f(x) = x$ and $f(y) = y$. Hence $d(f(x), f(y)) = d(x, y)$, contradicting the contraction. Define iteratively $x_1 = f(0)$ and $x_n = f(x_{n-1})$ so that $x_n = f^n(x_0)$. Let $k < 1$ be the "contraction constant" for f (the Lipschitz constant). Claim: $\{x_n\}$ forms a Cauchy sequence with $d(x_n, x_{n+1}) \leqslant k^n d(x_0, x_1)$. Indeed, by (informal) induction

$$d(x_n, x_{n+1}) \leqslant kd(x_{n-1}, x_n) \leqslant \cdots \leqslant k^n d(x_0, x_1).$$

The remainder of the proof is simply calculating the partial sums of this geometric series and picking the correct δ accordingly ($\sum_{i=n}^{\infty} k^i d(x_0, x_1) = k^n d(x_0, x_1)/(1-k)$ and so we can bound any $d(x_n, x_m)$ with $m \geqslant n$ by this). Relatively simple, thus omitted. Then, given the limit exists,

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} f(x_n) \implies \lim_{n\to\infty} x_n =: x_\infty = f(x_\infty).$$

□

Before the proof of Picard-Lindelöf, we need a few lemmas relating to the "integral equations".

Definition 4.5.9

Let $F : [a, b] \to \mathbb{R}^m$ be defined by $F := \begin{bmatrix} F_1 & \cdots & F_m \end{bmatrix}^T$. We say F is Riemann integrable if each F_i is. If so,

$$\int_a^b F(\tilde{t})\, \mathrm{d}\tilde{t} := \begin{bmatrix} \int_a^b F_1(\tilde{t})\, \mathrm{d}\tilde{t} & \cdots & \int_a^b F_m(\tilde{t})\, \mathrm{d}\tilde{t} \end{bmatrix}.$$

Analogous to functions $F : [a, b] \to \mathbb{C}^m$.
Done rigorously, for a partition pair $(P, T) \in A$ (the directed set defined for nets), define

$$R(F, P, T) := \sum_{i=1}^{n} F(t_i)(x_i - x_{i-1}) = \begin{bmatrix} R(F_1, P, T) & \cdots & R(F_m, P, T) \end{bmatrix}.$$

Then, the assignment $(P, T) \mapsto R(F, P, T)$ is a net from $(A, \leq)$ to $\mathbb{R}^m$.

Lemma 4.5.10

(1) F is Riemann integrable if and only if the net defined as above converges. If so, the integral = the limit of the net.

(2) Continuous functions $F : [a, b] \to \mathbb{R}^m$ are Riemann integrable.

(3) If $F'(x)$ exists then f is Riemann integrable on $[a, b]$ with $\int_a^b F'(x) = F(b) - F(a)$.

Lemma 4.5.11

Just like the MVT, we have a generalized version:

$$\left\| \int_a^b F(\tilde{t})\, \mathrm{d}\tilde{t} \right\| \leqslant M(b-a) \text{ where } M = \sup_{x\in[a,b]} \|F(x)\|.$$

Proof. First notice that
$$\int_a^b F(\tilde{t})\, \mathrm{d}\tilde{t} = \lim_{(P,T)} R(F,P,T).$$
Since the norm function $\|\cdot\| : \mathbb{R}^m \to \mathbb{R}$ is continuous, it sends convergent nets to convergent nets! Thus
$$\left\| \int_a^b F(\tilde{t})\, \mathrm{d}\tilde{t} \right\| = \lim_{(P,T)} \|R(F,P,T)\|.$$
For any (P,T), we have
$$\|R(F,P,T)\| \leqslant \sum_{i=1}^n \|F(t_i)\|(x_i - x_{i-1}) \leqslant M \sum_{i=1}^n (x_i - x_{i-1}) = M(b-a).$$
Thus the limit also $\leqslant M(b-a)$. □

Lemma 4.5.12

Let $\Omega \subset \mathbb{R} \times \mathbb{R}^m$ be open. Let $F : \Omega \to \mathbb{R}^m$ be continuous. Let $\gamma : (a,b) \to \mathbb{R}^m$ be continuous such that
$$(t, \gamma(t)) \in \Omega \text{ for all } t \in (a,b).$$
Let $t_0 \in (a,b)$. The following are equivalent:

(1) γ is a solution to the differentiable equation, i.e., γ is differentiable on (a,b) and $\gamma'(t) = F(t, \gamma(t))$.

(2) No differenentiability assumed, γ solves the *integral equation*
$$\gamma(t) = \gamma(t_0) + \int_{t_0}^t F(s, \gamma(s))\, \mathrm{d}s.$$

Proof. Assuming (1), γ' is continuous (since F is) and thus Riemann integrable. By the anti-derivative theorem
$$\int_{t_0}^t \gamma'(\tilde{t})\, \mathrm{d}\tilde{t} = \gamma(t) - \gamma(t_0).$$
Assuming (2), then γ and F are both continuous and so is $F(s, \gamma(s))$. Then by the FTC
$$\int_{t_0}^t F(s, r(s))\, \mathrm{d}s \text{ is differentiable with deriv. at } t = F(t, \gamma(t)).$$
Thus,
$$\gamma(t) = \gamma(t_0) + \int_{t_0}^t F(s, \gamma(s))\, \mathrm{d}s \text{ is differentiable and } \gamma'(t) = F(t, \gamma(t)).$$
□

We now begin the proof of the Picard-Lindeöf Theorem (autonomous case):

Proof of the Picard-Linidelöf Theorem. We could let $t_0 \in \mathbb{R}$ but for convenience we assume $t_0 = 0$. We want to solve
$$\gamma'(t) = F(\gamma(t)) \text{ and } \gamma(t_0) = y_0.$$
WLOG, assume F is Lipschitz on U (an open neighborhood of y_0). Let L be the corresponding Lipschitz constant. We need some data about F to choose an interval $[-\tau, \tau]$ on which the solutions exist and are unique. We will set up these data first then finish the proof next lecture.

Since U is open, there exists $r > 0$ such that the closed ball $N := \overline{B(y_0, r)} \subset U$. Notice that N is closed and bounded in $\mathbb{R}^m$. Therefore it is compact. Therefore $F(N)$ is compact; in particular, for some M we have $\|F(x)\| \leqslant M$ for all $x \in N$.

Beginning of March 26, 2021

Now we prove the following claims:

(1) Local existence: for any $\tau < \min(r/M, 1/L)$, there exists $\gamma : (-\tau, \tau) \to N$ differentiable with $\gamma'(t) = F(\gamma(t))$ for all $t \in (-\tau, \tau)$.

(2) Local uniqueness holds.

We first show local existence. Let $\mathcal{C} := C^0([-\tau, \tau], N)$. Since N is compact, $(\mathcal{C}, \mathrm{d}_{\sup})$ is complete. If we are able to find a contraction mapping from $\mathcal{C} \to \mathcal{C}$ then we can find a fixed point by the Banach contraction mapping theorem. We want to define $\Phi : \mathcal{C} \to \mathcal{C}$ to be a such that $\gamma \in \mathcal{C}$ is a fixed point of Φ if and only if γ solves

$$\gamma(t) = \gamma(0) + \int_0^t F(\gamma(s))\, \mathrm{d}s.$$

The method to define such Φ is called the *Picard iteration*: for $\gamma \in \mathcal{C}$, simply define (compare to above)

$$(\Phi(\gamma))(t) := y_0 + \int_0^t F(\gamma(s))\, \mathrm{d}s.$$

First notice that if $\gamma \in \mathcal{C}$ then $\Phi(\gamma) \in \mathcal{C}$, so Φ is a function from $\mathcal{C} \to \mathcal{C}$: indeed, continuity (even differentiability) is obvious, and

$$\|\Phi(r)(t) - y_0\| = \left\| \int_0^t F(\gamma(s))\, \mathrm{d}s \right\| \leqslant M|t-0| = M\tau = M \cdot \min(r/M, 1/L) < r$$

where the inequality follows from Lemma 4.5.11. Therefore $(\Phi(r))(t) \in N$ for all $t \in [-\tau, \tau]$.
Now we shall show that Φ is a contraction with constant $\tau L < 1$. Let $\gamma, \sigma \in \mathcal{C}$. We have

$$\begin{aligned} d(\Phi(\gamma), \Phi(\sigma)) &= \sup_{t \in [-\tau,\tau]} \left\| y_0 + \int_0^t F(\gamma(s))\, \mathrm{d}s - y_0 - \int_0^t F(\sigma(s))\, \mathrm{d}s \right\| \\ &= \sup_{t \in [-\tau,\tau]} \left\| \int_0^t F(\gamma(s)) - F(\sigma(s))\, \mathrm{d}s \right\| \\ &\leqslant |t-0| \sup_{s \in [-\tau,\tau]} \|F(\gamma(s)) - F(\sigma(s))\| \\ &\leqslant \tau \cdot L \cdot \sup_{s \in [-\tau,\tau]} \|\gamma(s) - \sigma(s)\| = \tau L \cdot \mathrm{d}_{\sup}(\gamma, \sigma). \end{aligned}$$

The uniqueness of Φ's fixed point implies that there exists only one $\gamma \in \mathcal{C}$ satisfying $\Phi(\gamma) = \gamma$. To put more formally, given $\gamma : (a,b) \to U$ and $\sigma : (a', b') \to U$, two solutions satisfying the boundary conditions, we still have the r, M, L chosen from the beginning the problem, and we can define τ accordingly and decrease it to ensure $[-\tau, \tau] \in (a,b) \cap (a', b')$ if needed. Then uniqueness on $[-\tau, \tau]$ follows. □

Remark. This result is local, but we can do better in a few ways:

(1) By HW9 p1, if $\gamma : (a,b) \to U$ and $\sigma : (a',b') \to U$ agree locally, they actually agree on all of $(a,b) \cap (a',b')$.

(2) "Global existence" (solutions $\gamma(t)$ defined for all time t) may fail if F only satisfies a local Lipschitz condition. However, we have the following proposition:

Proposition 4.5.13

Suppose $K \subset U$ is compact. Let $\gamma : (a,b) \to U$ be a solution to $\gamma' = F(\gamma)$ where F is (locally) Lipschitz such that $\gamma(t) \in K$ for $t \in (b-\delta, b)$ for some δ. Then there exists $b' > b$ and a solution $\tilde{\gamma} : (a,b') \to U$ with $\tilde{\gamma} \equiv \gamma$ on (a,b). An analogous statement can be made for a.

Corollary 4.5.14

If $F : \mathbb{R}^m \to \mathbb{R}^m$ is globally Lipschitz then solutions to initial value problems (IVPs)

$$\gamma' = F(\gamma) \text{ and } \gamma(t_0) = y_0$$

exist for all time.

Finally, we talk about **flows**. Assume $U = \mathbb{R}^m$ and F has global Lipschitz constants. Then for all $p \in \mathbb{R}^m$ there exists $\gamma : \mathbb{R} \to \mathbb{R}^m$ such that

$$\gamma(0) = p \text{ and } \gamma'(t) = F(\gamma(t)) \text{ for all } t.$$

Once we fix the initial condition, we can define a *flow on* $\mathbb{R}^m$ or a t*-advance map*:

Definition 4.5.15

For $t \in \mathbb{R}$, define $\varphi_t : \mathbb{R}^m \to \mathbb{R}^m$ by

$$\varphi_t(p) = \gamma(t) \text{ for the unique solution to the IVP.}$$

Key facts of φ_t:

(1) It is continuous. If further assumptions holds, e.g., F is smooth, then φ_t is also smooth. Smame holds for C^∞. This is called the *continuous / smooth dependence of solutions on initial conditions*.

(2) $\varphi_{s+t}(p) = \varphi_s(\varphi_t(p))$, i.e., the map sending t to φ_t is a group homomorphism from $(\mathbb{R}, t)$ to the *group of homeomorphisms* of $\mathbb{R}^m$ with function composition, i.e.,

$$(\mathbb{R}, +) \to (\text{Homeo}(\mathbb{R}^m), \circ).$$

Beginning of March 29, 2021

Today we will briefly mention one more application before moving to chapter 5.

Hamiltonian Formulation

Let $H : \mathbb{R}^{2n} \to \mathbb{R}$ be a smooth function. Write coordinates in $\mathbb{R}^n$ by $(q_1, \ldots, q_n, p_1, \ldots, p_n)$ where the q's are position coordinates and p's are momentum coordinates. (More generally this is called a *symplectic manifold*.) Here $\mathbb{R}^{2n}$ is a phase space: points in $\mathbb{R}^{2m}$ is a "state of the system". Observable quantities can be defined as functions from this phase space to $\mathbb{R}^m$, of which $H : \mathbb{R}^{2n} \to \mathbb{R}^n$ is one example. This should specify a physical system that evolves over time. How?

From H, we define a Hamiltonian vector field X_H :

$$X_H(x) := X_H(q_1, \ldots, q_n, p_1, \ldots, p_n) := \left[\frac{\partial H}{\partial p_1}(x) \quad \cdots \quad \frac{\partial H}{\partial p_n}(x) \quad -\frac{\partial H}{\partial q_1}(x) \quad \cdots \quad -\frac{\partial H}{\partial q_n}(x)\right]^T.$$

(Think of this as a symplectic analogue of gradient we've seen.) Then we have a system of ODEs

$$\gamma'(t) = X_H(\gamma(t)).$$

This defines a flow $\varphi_H : \mathbb{R} \to \text{Homeo}(\mathbb{R}^{2n})$.

Theorem 4.5.16

The mapping φ_H determines the state of a system over time in the following sense: if one starts in state $(q_1, \ldots, q_n, p_1, \ldots, p_n)$ at $t_0 = 0$, then one goes to state $[\varphi_H(t)](q_1, \ldots, q_n, p_1, \ldots, p_n)$ after t seconds.

Rest of chapter 4 which we will not cover:

4.6 real analytic functions.

4.7 continuous but nowhere differentiable functions. The *Baire category theorem* in fact implies that "most" functions are pathological just like the Weierstraß function.

4.8 d_{sup} for unbounded functionsl spaces of unbounded functions; generalization of the Arzelà-Ascoli Theorem.

We are "officially" done with chapter 4 in MATH 425b.

Chapter 5

Multivariable Calculus

Roadmap for the remainder of this semester:

The HWs will lead to "multilinear algebra" / tensor algebra and differential forms, but here we will primarily focus on linear transformation vs. linear transformations in §5.1. In particular, we will focus on the **operator norm**.

Once done with §5.1, we will talk about **total derivative** and **Jacobian**, e.g., "chain rule done right" in §5.2.

We will briefly skip §5.3 and move to §5.4: inverse and implicit function theorems.

5.1 Linear Algebra; Operator Norms

Recall that, if V, W are vector spaces over $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$), the natural notion of a "structure-preserving" map $V \to W$ is a **linear transformation** $T : V \to W$. (Think of the axioms.)

If V, W are normed, we could look at bounded functions. However, from linearity, it is immediately clear that if $T : (V, \|\cdot\|_V) \to (W, \|\cdot\|_W)$ is linear transformation then $T(x)$ is not bounded as we can scale x to make $T(x)$ arbitrarily large.

However, we can fix this by introducing the "scaling factor" or simply focus on all $x \in V$ with $\|x\|_V = 1$. This naturally leads to the notion of **operator norm**.

Definition 5.1.1

Let V, W be normed spaces over $\mathbb{K}$ $T : V \to W$ a linear transformation. We define the **operator norm** $\|T\|_{\text{op}}$ by

$$\|T\|_{\text{op}} = \sup_{\|v\|=1} \|T(v)\| = \sup_{v \neq 0} \frac{\|T(v)\|}{\|v\|} = \inf\{L > 0 : \|T(v)\| \leqslant L\|v\| \text{ for all } v \in V\}.$$

(Easy exercise to check that these are equivalent. *Hint: scaling.*)

It becomes immediate that $\|T(v)\| \leqslant \|T\|_{\text{op}}\|v\|$ for all v. Note that the inequality holds even if $v = 0$, in which case $0 \leqslant 0$ is vacuously true.

Question. Which T's have $\|T\|_{\text{op}} < \infty$? We have the following proposition.

Proposition 5.1.2

Let V, W be normed over $\mathbb{K}$ and let $T : V \to W$ be linear. Ridiculously as it sounds, TFAE:

(1) $\|T\|_{\text{op}} < \infty$,

(2) T is Lipschitz on V,

(3) T is uniformly continuous on V,

(4) T is continuous on V,

(5) T is continuous at $0 \in V$.

Proof. (1) $\implies$ (2): simply take $\|T\|_{\text{op}}$ as the Lipschitz constant:

$$\|T(v) - T(w)\| = \|T(v-w)\| \leqslant \|T\|_{\text{op}}\|v-w\|.$$

(2) $\implies$ (3) $\implies$ (4) $\implies$ (5) is trivial. Now we show (5) $\implies$ (1) and assume T is continuous at 0. Then, for $\epsilon = 1$, there exists $\delta > 0$ such that if $\|v\| < \delta$ then $\|T(v)\| < 1$. Therefore $\|T\|_{\text{op}} < 2/\delta$ since

$$\sup_{\|v\|=1} \|T(v)\| = \sup_{\|v\|=\delta/2} \frac{2\|T(v)\|}{\delta} < \frac{2}{\delta}.$$

□

Definition 5.1.3

If $T : V \to W$ satisfies any of above, it is called a **bounded** (or continuous) operator / linear transformation. In terms of notation, I prefer to define $\mathcal{L}(V, W)$ to be the space of linear operators $V \to W$ and $\mathcal{B}(V, W)$ to be that of bounded operators $V \to W$ (this is different from the notations used in lectures).

Beginning of March 31, 2021

Proposition 5.1.4

Operator norms under composition satisfy triangle inequality. To put formally, let V, W, Z be normed. Let $S \in \mathcal{L}(V, W)$ and $T :\in \mathcal{L}(W, Z)$. Then

$$\|T \circ S\|_{\text{op}} \leqslant \|T\|_{\text{op}}\|S\|_{\text{op}}.$$

Proof. By definition,

$$\|T \circ S\|_{\text{op}} = \sup_{\|v\| \neq 0} \frac{\|T(S(v))\|}{\|v\|} = \sup_{\substack{\|v\| \neq 0 \\ S(v) \neq 0}} \frac{\|T(S(v))\|}{\|v\|} = \sup_{\substack{\|v\| \neq 0 \\ S(v) \neq 0}} \frac{\|T(S(v))\|}{\|S(v)\|} \cdot \frac{\|S(v)\|}{\|v\|}$$

and the claim follows since supremum of product $\leqslant$ product of supremums and

$$\|T \circ S\|_{\text{op}} \leqslant \sup_{\substack{\|v\| \neq 0 \\ S(v) \neq 0}} \frac{\|T(S(v))\|}{\|S(v)\|} \sup_{\substack{\|v'\| \neq 0 \\ S(v') \neq 0}} \frac{\|S(v)\|}{\|v\|} = \sup_{S(v) \neq 0} \frac{\|T(S(v))\|}{\|S(v)\|} \|S\|_{\text{op}} \leqslant \|T\|_{\text{op}}\|S\|_{\text{op}}.$$

□

Remark. If we let $V = W = Z$, then $\mathcal{B}(V, V)$ is not just a vector space; it's a ring (with composition as multiplication). Since $\mathcal{B}(V, V)$ is both a ring and a vector space over $\mathbb{K}$, it is called an $\mathbb{R}$-algebra (or $\mathbb{C}$-algebra, respectively). The norm is given by $\|\cdot\|_{\text{op}}$.

Definition 5.1.5

An algebra $\mathcal{A}$ over $\mathbb{K}$ is a **Banach algebra** if it has a complete norm $\|\cdot\|$ such that

$$\|ab\| \leqslant \|a\|\|b\| \text{ for all } a, b \in \mathcal{A}.$$

If V is Banach then $\mathcal{B}(V, V)$ is a Banach algebra.

A special case of $(\mathcal{B}(V, W), \|\cdot\|_{\text{op}})$: given V, $\mathcal{B}(V, \mathbb{K})$ consists of all bounded operators from V to $\mathbb{K}$. They are called **linear functional** on (or **dual vectors** for) V. Then $\mathcal{B}(V, \mathbb{K})$ is called the (continuous) **dual space** to V, written V^*.

For example, we can let $V = \mathbb{R}^n$ and define a linear functional on V by a linear map $V \to \mathbb{R}$ via matrix multiplication with a "row vector" (i.e., a $1 \times n$ matrix). So, if V is the space of "column vectors", its dual space consists of "row vectors". If V is finite-dimensional and if we dualize V twice, we get

$$V^{**} \cong V \text{ if } V \text{ is finite-dimensional (not an "only if" statement).}$$

Otherwise, V^{**} might be bigger. If $V^{**} \cong V$ we say V is *reflexive*.

Now we look at operator norms for linear operators between finite-dimensional vector spaces.

Proposition 5.1.6

(From HW10) All norms on a finite-dimensional vector space are equivalent.

Corollary 5.1.7

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous in Euclidean norm if and only if f is continuous given any norms on $\mathbb{R}^n, \mathbb{R}^m$.

Proposition 5.1.8

Any n-dimensional vector space is isometrically isomorphic to $\mathbb{R}^n$ by picking a basis. To put formally, if $\dim(V) = n$ and $\{v_1, \dots, v_n\}$ is a basis, then

$$T\Big(\sum_{i=1}^{n} \alpha_i v_i\Big) := \left(\sum_{i=1}^{n} \alpha_i^2\right)^{1/2}$$

defines an isometry $V \to \mathbb{R}^n$.

If $\mathbb{R}^n, \mathbb{R}^m$ have usual Euclidean metrics, then any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is continuous. More generally, any linear transformation between finite-dimensional vector spaces is continuous given any norms on the vector spaces.

Therefore, if $\dim(V) = n$ and $\dim(W) = m$, after picking bases for V, W, we can relate V to $\mathbb{R}^n$ and W to $\mathbb{R}^m$. Then the transformation matrix, a $n \times m$ real (or complex) matrix, can again be related to (mn)-tuples of real (or complex) numbers in $\mathbb{R}^{nm}$ or $\mathbb{C}^{nm}$.

Thus, we have many norms on $\mathcal{B}(\mathbb{R}^n, \mathbb{R}^m)$. This includes operator norm, various norms on $\mathbb{R}^{nm}$, and more. And they are all equivalent[!]

Corollary 5.1.9

Now we present a corollary about matrix-valued function: let $U \subset \mathbb{R}^k$ be open and let $A : U \to M_{m\times n}(\mathbb{K})$ (space of $m \times n$ functions, which can be identified as $\mathcal{B}(\mathbb{R}^n, \mathbb{R}^m)$) be a matrix-valued function. Then TFAE:

(1) each entry $A_{i,j}(x)$ of $A(x)$ is a continuous function of $x \in U$, and

(2) A is a continuous function $U \to (M_{m\times n}(\mathbb{K}), \|\cdot\|_{\text{op}})$, i.e., for all $x_0 \in U$, given $\epsilon > 0$, there exists $\delta > 0$ such that $\|A(x) - A(x_0)\|_{\text{op}} < \epsilon$ whenever $\|x - x_0\| < \delta$ (in Euclidean norm, say, for convenience).

Before moving to §5.2, we provide some more intuition for $\|\cdot\|_{\text{op}}$ on $M_{m\times n}(\mathbb{R})$:

(1) If $n = m$ and $A \in M_{n\times n}(\mathbb{R})$ is symmetric then the **spectral theorem** gives an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors for A. In this case, $\|A\|_{\text{op}}$ is the max of $|\lambda|$ of A.

(2) In the more general case, if $A \in M_{m\times n}(\mathbb{R})$ then it has a **singular value decomposition** (SVD) $A = U\Sigma V^T$ (where $U_{m\times m}, V_{n\times n}$ are orthogonal and Σ is diagonal consisting of the singular values σ_i's of A with larger σ's closer to the top-left). In this case, $\|A\|_{\text{op}} = \sigma_1$, the largest singular value.

5.2 Differential Multivariable Calculus; Total Derivatives

For $\mathbb{R}^n \to \mathbb{R}^m$, at $p = (p_1, \ldots, p_n) \in \mathbb{R}^n$, the *total derivative* of $F = (f_1, \ldots, f_m)$ at p, should it exist, will be a linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ (or equivalently an $n \times m$ matrix). It is *usually* the *Jacobian matrix*

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(p) & \cdots & \frac{\partial f_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(p) & \cdots & \frac{\partial f_m}{\partial x_n}(p) \end{bmatrix}$$

Beginning of April 2, 2021

We will set stage by taking **smooth manifolds** as a black box.

In a nutshell, a *smooth manifold* is supposed to generalize curves, surfaces, and their analogue in higher dimensions: curves in $\mathbb{R}^2, \mathbb{R}^3, \mathbb{R}^n$ are called 1-*dimensional manifold*, surfaces in $\mathbb{R}^2, \mathbb{R}^3, \mathbb{R}^n$ are called 2-*dimensional manifold*, and higher-dimensional analogues exist.

Here, we mainly consider open subsets of $\mathbb{R}^n$ as n-dimensional manifolds. Within this context, these open subsets cannot have boundaries and cannot have singular points. We want *smooth manifolds*.

Definition 5.2.1

A **smooth manifold** is a *topological space* that is *second-countable*, *Hausdorff*, and equipped with "extra data": *equivalence class of smooth structures.*

Key construction: if M is a smooth manifold and $p \in M$, then there exists an *abstract vector space* T_pM, the **tangent space** to M at p, an n-dimensional vector space if M is n-dimensional. In particular, if $M \subset \mathbb{R}^n$, this can be identified with the "usual" tangent space seen in Calculus III, for example the tangent plane of a surface in $\mathbb{R}^3$.

Some extra black box definitions:

(1) T_pM is the set of *equivalence class of smooth curves in* m *through* p*; equal to* $(m/m^2)^*$ *(whatever that means)*.

With theory of smooth manifolds, one can look at functions

$$F : M \to N \qquad (M, N \text{ being smooth manifolds}).$$

"Differentiable" functions will be those $F : M \implies N$ that "can be <u>linearly</u> approximated subtlety" by a <u>linear transformation</u> $T : T_pM \to T_{F(p)}N$. If such T exists and "approximates F appropriately", we say that F is differentiable with T being its **total derivative** (or simply Jacobian in this context), written $(DF)_p$.

What is $(DF)_p$ good for? For example, the *inverse function theorem*. If $F : M \to N$ and $G : N \to M$ are inverses to each other, and if $p \in M$ with $F(p) = q \in N$, then

$$T_pM \underset{(DG)_q}{\overset{(DF)_p}{\rightleftarrows}} T_qN$$

and chain rule gives $(DF)_p, (DG)_q$ are inverse to each other. In addition, *F is nonlinearly invertible near p*: there exists an open neighborhood U of p and V of $q := F(p)$ such that

$$F : U \to V \text{ is invertible, a } \textit{diffeomorphism}.$$

Now we focus on $F : U \to V$ where U, V are open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively (so they are *smooth n-manifold* and *smooth m-manifold*, respectively).

Key fact: given $p \in U$, we have a canonical identification $T_pU \cong \mathbb{R}^n$ *because U has a canonical choice of coordinates*. (Heuristically, think of $\mathbb{R}^2$; the tangent plane at each point in $U \subset \mathbb{R}^2$ (open) is just all of $\mathbb{R}^2$.) Likewise $T_qV \cong \mathbb{R}^m$. Then $(DF)_p$ should be a linear transformation $\mathbb{R}^n \to \mathbb{R}^m$.

Definition 5.2.2

Let $U \subset \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}^m$ be a function (or to $V \subset \mathbb{R}^m$ open). Let $p \in U$, and let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. We say F is **differentiable at p** (heuristically, T approximates F at p) if

$$\lim_{\substack{v \to 0 \\ v \in \mathbb{R}^n \setminus \{0\}}} \frac{F(p+v) - F(p) - T(v)}{\|v\|} = 0.$$

Intuitively, $F(p+v)$ is the *real value of F at $p+v$* and $F(p) + T(v)$ is the "approximate value".

For convenience we denote the denominator above as $R(v)$ the "error term for approximation". This condition shows that T is a good "first-order" approximation to F near p: it should encode data of first-order partials of F.

Beginning of April 5, 2021

Proposition 5.2.3

If such T exists, it is unique.

Proof. If T and S both satisfy the limit condition, then

$$\lim_{v\to\infty} \frac{S(v)-T(v)}{\|v\|} = 0 = \lim_{v\to\infty} \frac{(S-T)(v)}{\|V\|}$$

by directly subtracting one from another. Assume $S-T$ is not the zero linear transformation, so there exists $v_0 \neq 0$ such that $(S-T)(v_0) \neq 0$. On one hand, we should have $\lim_{c\to\infty} cv_0 = 0$, whereas

$$\lim_{c\to 0+} \frac{(S-T)(cv_0)}{\|cv_0\|} = \frac{(S-T)(v_0)}{\|v_0\|},$$

contradiction. □

Thanks to this uniqueness, if F is differentiable at p then $(\mathrm{D}F)_p$ is a well-defined linear transformation.

Example 5.2.4. Let $m = n = 1$. Any linear transformation $T : \mathbb{R}^1 \to \mathbb{R}^1$ is given by multiplication by a 1×1 matrix $A = [a]$, i.e., $T(v) = av$ for all $v \in \mathbb{R}$.
If $F : U \to \mathbb{R}$ is a function (with $U \subset \mathbb{R}$ open) and $p \in U$, then (assuming $v \neq 0$)

$$\frac{F(p+v)-F(p)-T(v)}{\|v\|} = \frac{F(p_v)-F(p)-av}{|v|} = \begin{cases} \dfrac{F(p+v)-F(p)}{v} - a & v > 0 \\ a - \dfrac{F(p+v)-F(p)}{v} & v < 0. \end{cases}$$

Therefore, F is differentiable at p if and only if the above quotient $\to 0$ as $v \to 0$, i.e.,

$$\lim_{v\to\infty} \frac{F(p+v)-F(p)}{v} = a.$$

If so, a is the derivative $(\mathrm{D}F)_p$. Obvious enough that we overkilled the single-variable version!

More generally, given $n = 1, m \geqslant 1$, any linear transformation $T : \mathbb{R} \to \mathbb{R}^m$ has an $m \times 1$ matrix:

$$T(v) = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}^T v = \begin{bmatrix} a_1 v & \cdots & a_m v \end{bmatrix}^T \text{ for } v \in \mathbb{R}.$$

Since functions from open subsets of $\mathbb{R}$ to $\mathbb{R}^m$ are often viewed as paths in $\mathbb{R}^m$, we will use γ instead of F to help strengthen the memory. If $\gamma : (a,b) \to \mathbb{R}^m$ is a function and $t_0 \in (a,b)$, then γ is differentiable at t_0 with derivative $T(v) = [a_1 \cdots a_m]^T v$ if and only if

$$\lim_{t\to 0} \frac{\gamma(t_0+t)-\gamma(t_0)-[a_1\cdots a_m]^T t}{|t|} = 0.$$

Similar to the $\mathbb{R} \to \mathbb{R}$ case, if so we say

$$\lim_{t\to\infty} \frac{\gamma(t_0+t)-\gamma(t_0)}{t} = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}^T.$$

Therefore each coordinate $\gamma_i(t)$ is differentiable at t_0 with derivative $a_i(t_0)$.

Upshot: $m > 1$ "doesn't cause troubles" except with MVT.

Proposition 5.2.5

Let $U \subset \mathbb{R}^n$ be open, let $F : U \to \mathbb{R}^m$, and let $p \in U$. Write F in coordinates as

$$F(x) = (F_1(x), \dots, F_n(x)) \text{ for } x \in U$$

(so each $F_i : U \to \mathbb{R}$). Then F is differentiable at p with matrix derivative $T(v) = A_{m\times n}v$ if and only if each F_i is differentiable at p with derivative $T_i(v) = [a_{i,1} \cdots a_{i,n}]v$.

Proof. Call the first statement (1) and the second (2). By definition, (1) is equivalent to saying

$$\lim_{v\to 0} \frac{F(p+v) - f(p) - T(v)}{\|v\|} = 0,$$

a limit of vectors in $\mathbb{R}^m$. Therefore the component-wise limit is the limit of the component, and hence (2). □

Proposition 5.2.6

We have the following properties:

(1) Linearity: if $F, G : U \to \mathbb{R}^m$ with $U \in \mathbb{R}^n$ an open subset and F, G are differentiable at $p \in U$, then for $c \in \mathbb{R}$ we have

$$(\mathrm{D}(F + cG))_p = (\mathrm{D}F)_p + c(\mathrm{D}G)_p.$$

(2) If F is a constant function $U \to \mathbb{R}^m$ then F is differentiable at all $p \in U$ with $(\mathrm{D}F)_p = 0$.

(3) If F is affine linear, $F(v) = c + Av$ for $v \in \mathbb{R}^n$, then F is differnetiable at any p with $(\mathrm{D}F)_p v = Av$.

Derivatives and Compositions

Theorem 5.2.7: Chain Rule

Suppose $U \in \mathbb{R}^n$ and $V \in \mathbb{R}^m$ are open, and suppose $U \xrightarrow{F} V \xrightarrow{G} \mathbb{R}^k$. Assume F is differentiable at p with $(\mathrm{D}F)_p(v) = A_{m\times n}v$ and likewise G is differentiable at $q := F(p)$ with $(\mathrm{D}G)_q(v) = B_{k\times m}v$. Then

$$G \circ F \text{ is differentiable at } p \text{ and } (\mathrm{D}(G \circ F))_p(v) = (BA)_{k\times n}(v)$$

where BA represents matrix multiplication (which corresponds to composition of linear transformations).

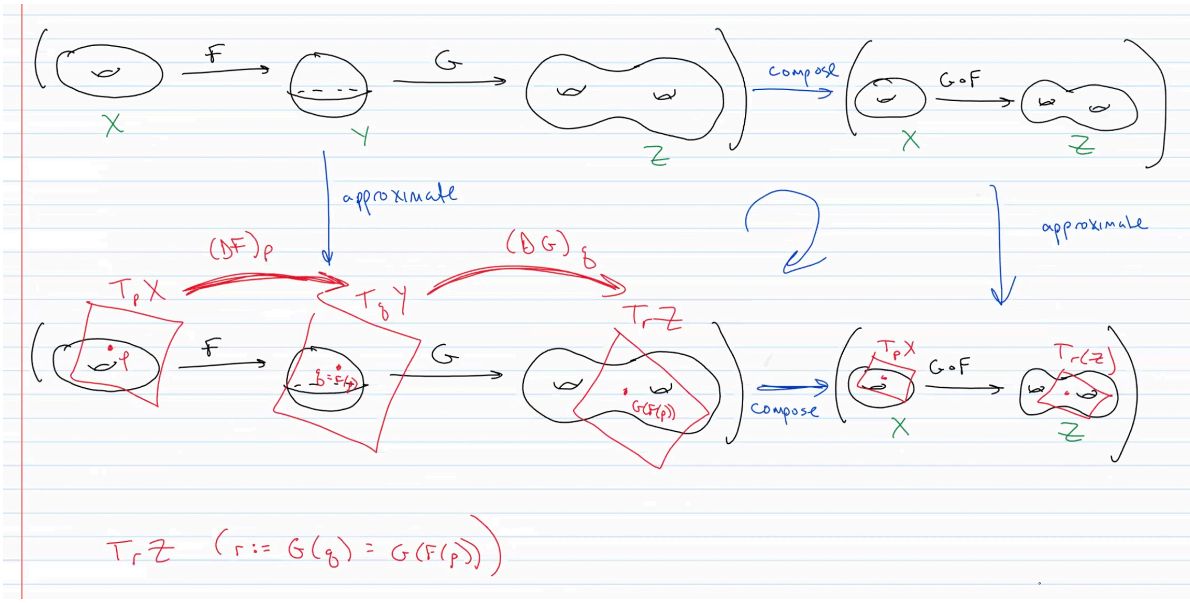


Figure from lecture: a little commutative diagram showing (compose) ∘ (approx) = (approx) ∘ (compose)

Before the chain rule, we first prove a seemingly obvious fact:

Proposition 5.2.8

Let $U \in \mathbb{R}^n$ be open and let $F : U \to \mathbb{R}^m$. Let $p \in U$. If F is differentiable with derivative T at p then F is continuous at p.

Proof. One say to show continuity is $\lim\limits_{v\to 0} F(p+v) = F(p)$ (sequential continuity). Notice that $F(p+v) = F(p) + T(v) + R_F(v)$ where $R_F(v)$ is the "remainder" term defined to be $R_F(v) := F(p+v) - F(p) - T(v)$ (the numerator of the differentiability limit equation). Clearly as $v \to 0$, $R_F(v) \to 0$ by the assumption that F is differentiable at p. Note that as $v \to 0$, $T(v) \to 0$ (any linear transformation between finite-dimensional normed spaces is continuous). For $\|v\| \leqslant 1$, we have

$$\|R_F(v)\| \leqslant \frac{\|R_F(v)\|}{\|v\|} \text{ so } \|R_F(v)\| \to 0 \text{ as } v \to 0.$$

So $\lim\limits_{v\to 0} F(p+v) = F(p) + 0 + 0 = F(p)$. ☐

Beginning of April 9, 2021

Proof of Chain Rule. Let R_F be a function of $v \in \mathbb{R}^n$ (the sublinear error term by when approximating F, i.e., $R_F(v) = F(p+v) - F(p) - Av)$ for $v, p \in U$. Likewise, let R_G be the error term when approximating G, i.e., $R_G(w) = G(q+w) - G(q) - Bw$ for $w, q \in V$.
Now we let $R_{G\circ F}$ be the error term of $G \circ F$ by $R_{G\circ F}(v) := G(F(p+v)) - G(F(p)) - BAv$. We want to show that $\lim\limits_{v\to 0} R_{G\circ F}(v)/\|v\| = 0$ (which then by definition means the derivative of differentiability means $G \circ F$ is

differentiable at p. In the formula for $R_{G\circ F}(v)$ we plug on the approximations for $F(p+v)$:

$$\begin{aligned} R_{G\circ F}(v) &= G(\underbrace{F(p)}_{q} + \underbrace{Av + R_F(v)}_{w}) - G(q) - BAv \\ &= G(q) + B(Av + R_F(v)) + R_G(Av + R_F(v)) - G(q) - BAv \\ &= BR_F(v) + R_G(Av + R_F(v)). \end{aligned}$$

Then, by triangle inequality,

$$\frac{\|R_{G\circ F}(v)\|}{\|v\|} \leqslant \frac{\|BF_F(v)\|}{\|v\|} + \frac{\|R_G(Av + R_F(v))\|}{\|v\|}.$$

The first term $\to 0$ because it is further bounded by $\|B\|_{\text{op}}\dfrac{\|R_F(v)\|}{\|v\|}$ where the operator norm of B is finite and the quotient $\to 0$ as $v \to 0$ by differentiability of R.

We now show that the second term also $\to 0$. We can multiply both the numerator and detonator by $\|Av+R_F(v)\|$. (If $Av + R_F(v) = 0$ then the numerator is simply $R_G(0) = G(q) - G(q) = B \cdot 0 = 0$ and the fraction still evaluates to 0, and now we assume $Av + R_F(v) \neq 0$ so that multiplication on both sides makes sense.) Then

$$(\|R_G\frac{Av + R_F(v)\|}{\|v\|} = \frac{\|R_G(Av + R_F(v))}{\|Av + R_F(v)\|} \cdot \frac{\|Av + R_F(v)\|}{\|v\|}.$$

Note that, as $v \to 0$, $Av \to 0$ since multiplication by A is continuous. Also, $R_F(v) = F(p+v) - F(p) - Av$ is differentiable (and thus continuous) and also $\to 0$. Therefore, as $Av + R_F(v) \to 0$, the first term $\to 0$ by sublinearity of R_G. Now it suffices to show that $\dfrac{\|Av + R_F(v)\|}{\|v\|}$ is bounded as $v \to 0$. Indeed,

$$\frac{\|Av + R_F(v)\|}{\|v\|} \leqslant \frac{\|Av\|}{\|v\|} + \frac{\|R_F(v)\|}{\|v\|} \leqslant \|A\|_{\text{op}} + \frac{\|R_F(v)\|}{\|v\|}$$

where the second fraction $\to 0$ as $v \to 0$, once again by the sublinearity of R_F. This concludes the proof. □

Recall that one way to define T_PM (tangent space; M *smooth manifold*, $p \in M$) is by defining it as the set of equivalence classes of paths in M through p (these paths can be expressed as functions $\gamma : (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$ for all these paths). Equivalence relation is given by $\gamma_1 \sim \gamma_2$ if $\gamma_1'(0) = \gamma_2'(0)$ "in some chart" (derivative because this is the tangent space).

Important question to ask: how does $(\mathrm{D}F)_p$ act on vectors $v = \gamma'(0)$ for $\gamma : (-\epsilon, \epsilon) \to U$ where $r(0) = p, p \in U$, and $U \in \mathbb{R}^n$ open?

Proposition 5.2.9

Let $t_0 \in (a, b)$ and let $\gamma : (a, b) \to U$ be a differentiable function (curve) with $\gamma(t_0) = p$ ($p \in U$ and $U \subset \mathbb{R}^n$ open). Let $F : U \to \mathbb{R}^m$ be differentiable at p. Then

$$(\mathrm{D}F)_p(\gamma'(t_0)) = (F \circ \gamma)'(t_0).$$

Proof is clear by chain rule:

$$(\mathrm{D}(F \circ \gamma))_{t_0} = (\mathrm{D}F)_{\gamma(t_0)} (\mathrm{D}\gamma)_{t_0}.$$

□

Remark. If we are trying to define $(\mathrm{D}F)_p$ as a map $T_pM \to T_qN$ for a map of smooth manifolds $F : M \to N$, and suppose T_PM is the equivalence classes of curves γ in M through p and T_qN likewise, then how should $(\mathrm{D}F)_p$ act on some element of T_pM?

(1) Pick one representative in the equivalence class (some $\gamma : (a,b) \to M$).

(2) Compose γ with $F : F \circ \gamma : (a,b) \to N$.

(3) Take equivalence classes of these new functions (this is independent of choice of representative in (1)).

Corollary 5.2.10

If $U \subset \mathbb{R}^n$ is open, $p \in U$, and $F : U \to \mathbb{R}^m$ is differentiable at p, then

$$(\mathrm{D}F)_p(v) = \lim_{t \to 0} \frac{F(p+tv) - F(p)}{t}.$$

Proof. Proof: simply let $\gamma(t) := p + tv$ for $t \in (-\epsilon, \epsilon)$ for sufficiently small ϵ ensuring $\gamma : (\epsilon, \epsilon) \to U$. Then $\gamma(0) = p and \gamma'(0) = v$. Then

$$(\mathrm{D}F)_p(v) = (\mathrm{D}F)_p(\gamma'(0)) = (F \circ \gamma)'(0) = \lim_{t \to 0} \frac{(F \circ \gamma)(t) - (F \circ \gamma)(0)}{t} = \lim_{t \to 0} \frac{F(p+tv) - F(p)}{t}.$$

□

Remark. Above is the "directional derivative" formula. We can also generalize differentiability to functions between certain infinite-dimensional spaces. The approach by linear approximation is called **Frechét differentiability** and the one by directional derivatives is called **Gateaux differentiability**.

We can use the above corollary to compute the (standard-basis) matrix for $(\mathrm{D}F)_p$ in terms of partial derivatives.

Definition 5.2.11

Let $U \subset \mathbb{R}^n$ be open. Let $F : U \to \mathbb{R}^m$ and $p \in U$. The **partial derivative** $\dfrac{\partial F}{\partial x_i}(p)$ (for $1 \leqslant i \leqslant n$) is

$$\frac{\partial F}{\partial x_i}(p) = \lim_{t \to 0} \frac{F(p + te_i) - F(p)}{t},$$

if it exists. Here e_i is the i^{th} standard basis factor, i.e., the one defined by corresponding Kronecker deltas.

Remark. If F is differentiable at p then $\dfrac{\partial F}{\partial x_i}(p)$ exists for all $1 \leqslant i \leqslant n$: in particular $\dfrac{\partial F}{\partial x_i}(p) = (\mathrm{D}F)_p(e_i)$. This gives us a systematic way to deduce the matrix for $(\mathrm{D}F)_p$ in standard basis. Specifically, if $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear and $\{e_i\}_{i=1}^n$ are the standard basis vectors for $\mathbb{R}^n$, then for $v \in \mathbb{R}^n$ we have

$$T(v) = Av \text{ where } A = \begin{bmatrix} | & \cdots & | \\ T(e_1) & \cdots & T(e_n) \\ | & \cdots & | \end{bmatrix}.$$

Since $(\mathrm{D}F)_p : \mathbb{R}^n \to \mathbb{R}^m$ is also linear, for $v \in \mathbb{R}^n$ we have $(\mathrm{D}F)_p(v) = \mathcal{J}v$ where

$$\mathcal{J} = \begin{bmatrix} | & \cdots & | \\ (\mathrm{D}F)_p(e_1) & \cdots & (\mathrm{D}F)_p(e_n) \\ | & \cdots & | \end{bmatrix} = \begin{bmatrix} | & \cdots & | \\ \dfrac{\partial F}{\partial x_1}(p) & \cdots & \dfrac{\partial F}{\partial x_n}(p) \\ | & \cdots & | \end{bmatrix} = \begin{bmatrix} \dfrac{\partial F_1}{\partial x_1}(p) & \cdots & \dfrac{\partial F_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_m}{\partial x_1}(p) & \cdots & \dfrac{\partial F_m}{\partial x_n}(p) \end{bmatrix} \text{ where } F = \begin{bmatrix} F_1 \\ \vdots \\ F_m \end{bmatrix}.$$

Definition 5.2.12

The **Jacobian matrix** of F, $\mathcal{J}_{F,p}$, is defined as above, assuming all partials exist.

We have now shown that if F is differentiable at p then all partials at p exist and $(\mathrm{D}F)_p(v) = \mathcal{J}_{F,p}(v)$. Next time: we'll show that we can go backwards, i.e., if all partials exist and are continuous then F is differentiable with derivative being the Jacobian.

Beginning of April 12, 2021

Definition 5.2.13

Let $U \subset \mathbb{R}^n$ be open and let $F : U \to \mathbb{R}^m$ a function. We say F is of class C^r ($r \in [1, \infty]$) if all partials $\partial^r F/(\partial x_{j_1} \cdots \partial x_{i_r})$ (iterated partial derivatives) exist and are continuous on U.

By a previous proposition, this is true if and only if all component-wise partials $\partial^r F_i/(\dots)$ exist and are continuous on U. In particular, F is of class C^1 if and only if $\partial F_i/\partial x_j$ exists and is continuous on U for all i, j.

Theorem 5.2.14

Let $U \subset \mathbb{R}^n$ be open and let $F : U \to \mathbb{R}^m$ be a C^1 function. Then f is differentiable at all $p \in U$ (and $(\mathrm{D}F)_p$ is represented by the Jacobian as always).

Proof. Let $\mathcal{J}_{F,p}$ be the Jacobian of F at $p \in U$. Fix p (and also F). Let $\mathcal{J} := \mathcal{J}_{F,p}$. We want to show that $\mathcal{J}$ is a linear approximation of F near p.

Let $T(v) := \mathcal{J}v$ (clearly T is linear). Let $R(v) = R_{F,p,T}(v) = F(p+v) - F(p) - \mathcal{J}v$. We want to show it is sublinear, i.e., $R(v)/\|v\| \to 0$ as $v \to 0$.

For $1 \leqslant i \leqslant m$, let $R_i(v)$ be the i^{th} coordinate of $R(v)$. Then the i^{th} component $\mathcal{J}v$ only cares about the i^{th} row of

$\mathcal{J}$. Therefore it suffices to show that

$$R_i(v) = F_i(p+v) - F_i(p) - \left[\frac{\partial F_i}{\partial x_1}(p) \quad \cdots \quad \frac{\partial F_i}{\partial x_n}(p)\right] v,$$

when divided by $\|v\|$, tends to 0 as $v \to 0$, for all $1 \leqslant i \leqslant m$. Given $\epsilon > 0$, we choose $\delta > 0$ such that if $\|v\| < \delta$ then $p + v \in U$ (recall U is open) and

$$\left|\frac{\partial F_i}{\partial x_j}(p+v) - \frac{\partial F_i}{\partial x_j}(p)\right| < \frac{\epsilon}{n}. \qquad (\Delta)$$

(This is also possible since $\partial F_i/\partial x_j$ is continuous.) We claim that this δ proves the claim.
Indeed, pick $v = [v_1 \ \ldots \ v_n]^T$ be such that $\|v\| < \delta$. Then

$$\begin{aligned} R_i(v) &= F_i(p+v) - F_i(p) - \sum_{i=1}^{n} v_i \frac{\partial F_i}{\partial x_i}(p) \qquad \text{(literally a dot product)} \\ &= F_i(p + \sum_{i=1}^{n} v_i e_i) - F_i(p + \sum_{i=1}^{n-1} v_i e_i) - v_n \frac{\partial F_i}{\partial x_n}(p) + \\ &\quad F_i(p + \sum_{i=1}^{n-1} v_i e_i) - F_i(p + \sum_{i=1}^{n-2} v_i e_i) - v_{n-1} \frac{\partial F_i}{\partial x_{n-1}}(p) + \\ &\quad \cdots + F_i(p + v_1 e_1) - F_i(p) - v_1 \frac{\partial F_i}{\partial x_1}(p). \end{aligned}$$

where $v = v_1 e_1 + \cdots + v_n e_n$ denotes the standard basis expansion of $v \in \mathbb{R}^n$ $((e_i)_j = \delta_{i,j})$. We want to show that each line, when divided by $\|v\|$, is less than ϵ/n. Indeed, this can be done using the single-variable MVT. Notice that the function

$$g : t \mapsto F_i(p + \sum_{i=1}^{j-1} v_i e_i + t e_j)$$

is differentiable on $t \in [0, v_j]$ because the derivative $\partial F_i/\partial x_j$ exists on U and so

$$g'(t) = \frac{\partial F_i}{\partial x_j}(p + \sum_{i=1}^{j-1} v_i e_i + t e_j).$$

Then, MVT says $g(v_j) - g(0) = g'(\theta)(v_j - 0)$ for some $\theta \in [0, v_j]$, i.e.,

$$F_i(p + \sum_{i=1}^{j} v_i e_i) - F_i(p + \sum_{i=1}^{j-1} v_i e_i) = v_j \frac{\partial F_i}{\partial x_j}(p + \sum_{i=1}^{j-1} v_i e_i + \theta e_j).$$

By (Δ) and the fact that $|v_j|/\|v\| \leqslant 1$ (property of norm),

$$\left|\frac{\partial F_i}{\partial x_j}(p + \sum_{i=1}^{j-1} v_i e_i + \theta e_j) - \frac{\partial F_i}{\partial x_j}(p)\right| < \frac{\epsilon}{n} \text{ because } \|(v_1, \ldots, v_{j-1}, \theta, 0, \ldots)\| \leqslant \|(v_1, \ldots, v_j, \ldots, v_n)\|.$$

This proves the claim. □

Example 5.2.15. Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $F(x,y,z) = (e^z \cos y, e^{-z} \sin y, x^2 + y^2 + z^2)$. Then F is differentiable on all of $\mathbb{R}^3$ with Jacobian

$$\mathcal{J}_{F,(x,y,z)} = \begin{bmatrix} 0 & -e^z \sin y & e^z \cos y \\ 0 & e^{-z} \cos y & -e^{-z} \sin y \\ 2x & 2y & 2z \end{bmatrix}.$$

Each entry is a continuous function of $x, y.z$ and so F is C^1 (in fact C^∞).

Theorem 5.2.16: MVT for $\mathbb{R}^n \to \mathbb{R}$

Let $U \subset \mathbb{R}^n$ be open and let $F : U \to \mathbb{R}$ be differentiable. Let $p, q \in U$ be such that $p + t(q-p) \in U$ for all $t \in [0,1]$ (i.e., the line segment connecting p, q lies entirely in U). Then there exists $\theta \in [0,1]$ such that $f(q) - f(p) = (\mathrm{D}F)_{p+\theta(q-p)}(q-p)$.

The MVT does not apply to functions whose target space is of higher dimension. Consider $F : \mathbb{R} \to \mathbb{R}^2$ defined by $F(t) = (\cos t, \sin t)$. Clearly $F(2\pi) - F(0) = 0$, but does it equal to $(\mathrm{D}F)_t(2\pi - 0)$ for any t? The answer is clearly no, as the Jacobian $(\mathrm{D}F)_t(2\pi) = 2\pi[-\sin t \quad \cos t]^T$ which can never be 0.

Theorem 5.2.17: Mean Value Inequality, $n, m \geqslant 1$

Let $U \subset \mathbb{R}^n$ be open and $F : U \to \mathbb{R}^m$ differentiable. For $p, q \in U$ such that the segment $\{p + t(q-p) : t \in [0.1]\} \subset U$, we have

$$\|F(q) - F(p)\| \leqslant \Big(\sup_{t \in [0,1]} \|(\mathrm{D}F)_{p+t(q-p)}\|_{\text{op}} \Big) \|q-p\| \leqslant \sup_{x \in U} \|(\mathrm{D}F)_x\|_{\text{op}} \|q-p\|.$$

Beginning of April 14, 2021

Proof. Call $\sup_{t \in [0,1]} \|(\mathrm{D}F)_{p+t(q-p)}\|_{\text{op}}$ M for convenience. Instead of working with norms, we will make use of inner products. If we can show $\langle F(q) - F(p), u \rangle \leqslant M\|q-p\|$ for all unit vectors $u \in \mathbb{R}^m$ then we are done. Why? If we take

$$\tilde{u} := \frac{F(q) - F(p)}{\|F(q) - F(p)\|} \text{ if } F(p) \neq F(q)$$

then the claim follows (if $F(p) = F(q)$ then this is trivially true).
Let $u = [u_1 \; \ldots \; u_m]^T$ be a unit vector in $\mathbb{R}^m$ and let $v = q - p \in \mathbb{R}^n$. For $t \in [0,1]$, let

$$g(t) = \langle F(p + tv), u \rangle = u_1 F_1(p + tv) + \cdots + u_m F_m(p + tv).$$

Note that each F_i is a differentiable function of t with derivative $(\mathrm{D}F_i)_{p+tv}(v)$. Therefore g is also differentiable on $[0,1]$ with

$$g'(t) = u_1(\mathrm{D}F_1)_{p+tv}(v) + \cdots + u_m(\mathrm{D}F_m)_{p+tv}(v) = \langle (\mathrm{D}F)_{p+tv}(v), u \rangle.$$

(*One can also show this using Leibniz product rule*.) Therefore the 1-dimensional MVT applies to g: there exists $\theta \in [0,1]$ such that $g(1) - g(0) = g'(\theta)(1-0) = g'(\theta)$. Notice that

$$g(1) - g(0) = \langle F(p) + F(v), u \rangle - \langle F(p), u \rangle = \langle F(q), u \rangle - \langle F(p), u \rangle = \langle F(q) - F(p) - u \rangle$$

where

$$g'(\theta) = \langle (\mathrm{D}F)_{p+\theta v}(v), u \rangle \leqslant \|(\mathrm{D}F)_{p+\theta v}(v)\| \|u\| = \|(\mathrm{D}F)_{p+\theta v}(v)\| \leqslant \|(\mathrm{D}F)_{p+\theta}\|_{\text{op}} \|v\|$$

where the first $\leqslant$ uses Cauchy-Schwarz and the last term can be further bounded by $M\|q-p\|$. □

Recall how we needed extra caution when interchanging integrals and sums (uniform convergence is required).

Now we present an analogue.

Theorem 5.2.18: Differentiation under integral sign

For simplicity, assume we have a function of two variables $f : [a,b] \times (c,d) \to \mathbb{R}$ (or $\mathbb{R}^m$) be continuous. Assume that $\frac{\partial f}{\partial y}(x,y)$ exists and is continuous for all $(x,y) \in [a,b] \times (c,d)$. Let

$$F(y) = \int_a^b f(x,y)\,\mathrm{d}x.$$

Then F is differentiable on (c,d) with derivative

$$F'(y) = \frac{\partial}{\partial y}\int_a^b f(x,y)\,\mathrm{d}x = \int_a^b \frac{\partial f}{\partial y}(x,y)\,\mathrm{d}x,$$

i.e., in this case $\frac{\partial f}{\partial y}$ and $\int_a^b$ are interchangeable.

5.4 Implicit and Inverse Function Theorems

We start with the *implicit function theorem* and then use it to prove the *inverse function theorem*.

The idea for implicit function theorem: say we have $n+m$ variables $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ and we "impose m (nonlinear) constraints". We look at points $(x,y) := (x_1, \ldots, x_n, y_1, \ldots, y_m)$ such that $F(x,y) = z$ ($F : \mathbb{R}^{n+m} \to \mathbb{R}^m$) such that the m "constraint equations" hold, i.e.,

$$z = (z_1, \ldots, z_m) \qquad \text{and} \qquad F_1(x,y) = z_1, \quad \ldots, \quad F_m(x,y) = z_m.$$

The set of (x,y) such that these equations hold forms the level set of F at level z [think of it as $F^{-1}(z)$].

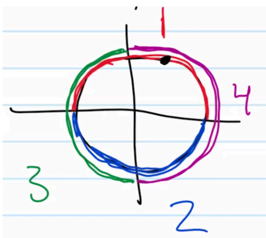
Clearly this is a very standard way to describe curves surfaces, etc., for example the circle $x^2 + y^2 = z$ in $\mathbb{R}^2$. For theory of smooth manifolds to be reasonable, we need the fact that "suitably nice" level sets are *smooth manifolds*, and this is given by the implicit function theorem.

Main idea: locally near (x_0, y_0), can we write $F^{-1}(z)$ as a graph of some function? (why "some"? See below.)

Example 5.4.1. Same old example: $n = m = 1, z = 1, F(x,y) = x^2 + y^2 = z$: clearly the condition fails globally: a circle is not a graph of a function as it fails the horizontal line test and the vertical line test (so it is not a graph of x as a function of y and not a graph of y as a function of x, either).

However, if we pick any point on the circle, for example $(\sqrt{2}/2, \sqrt{2}/2)$, there exists a neighborhood in which the horizontal line test is satisfied (so it's a graph of x as a function of y, in this case $x = \sqrt{1-y^2}$) and also a neighborhood in which the vertical line test is locally satisfied (so it can also be written as a graph of y as a function of x, in this case $y = \sqrt{1-x^2}$).

Even for "special" points like $(0,1)$ where no neighborhood satisfies the horizontal line test (thus we cannot locally express this level set as a graph of x as a function of y), there are neighborhoods that pass the vertical line test, e.g., take $y = \sqrt{1-x^2}$. Locally, the level set coincides with the graph of this function.



Note that we have 4 open subsets of the circle, each of which corresponds to the graph of some function:

(1) $y = \sqrt{1-x^2}$, taking x as a coordinate function,

(2) $y = -\sqrt{1-x^2}$, taking x as a coordinate function,

(3) $x = -\sqrt{1-y^2}$, taking y as a coordinate function, and

(4) $x = \sqrt{1-y^2}$, taking y as a coordinate function.

We can also perform "coordinate changes" between two patches: $t \mapsto \pm\sqrt{1-t^2}$ is smooth. The functions whose graphs recover the level set locally are smooth in this case. This defines a "smooth atlas" on the circle, making it a smooth manifold.

Beginning of April 16, 2021

Recall from last time that, for our specific example, level set of F can be locally expressed as the graph of some coordinates as functions of other coordinates.

To build our hypothesis, suppose $U \in \mathbb{R}^{m+n}$ open, $F : U \to \mathbb{R}^m$ of class C^k $(k \geqslant 1)$, $z_0 \in \mathbb{R}^m$. Then we have a level set $F^{-1}(z_0) \in U$ (pre-image, not inverse, and we will keep using such notation in this section). We also have a "linearized version" of the level set near $\ker(\mathrm{D}F)_{(x_0,y_0)}$ for any $(x_0, y_0) \in F^{-1}(z)$. Recall $(\mathrm{D}F)_{(x_0,y_0)}$ is a linear transformation from $\mathbb{R}^{m+n} \to \mathbb{R}^m$, so its kernel, $\ker(\mathrm{D}F)_{(x_0,y_0)} \subset \mathbb{R}^{m+n}$, heuristically gives the tangent space to the level set at (x_0, y_0), if we treat (x_0, y_0) as the origin of this space (all points in the kernel get mapped to the very same value, namely (x_0, y_0) because $(x_0 + y_0) + 0 = (x_0 + y_0)$).

Example 5.4.2. Again let $n = n = 1$ and consider $F(x, y) = x^2 + y^2$ and $z_0 = 1$. Thus we have the unit circle in $\mathbb{R}^2$ has the level set. Take $(x_0, y_0) := (1/\sqrt{2}, 1/\sqrt{2})$. Then

$$(\mathrm{D}F)_{(x_0,y_0)} = \begin{bmatrix} 2x_0 & 2y_0 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \sqrt{2} \end{bmatrix},$$

and so

$$\ker(\mathrm{D}F)_{(x_0,y_0)} = \{(u,v) \in \mathbb{R}^2 \mid (u,v)^T(\sqrt{2},\sqrt{2}) = 0\} = \operatorname{span}\{(-1,1)\}.$$

What does this give? The span containing $(x_0, y_)$ gives the tangent line to the circle at (x_0, y_0), indeed the tangent space! *Alternatively (and preferably) one can view the tangent space as $\ker(DF)_{(x_0,y_0)}$ with (x_0, y_0) being the origin of the space.*

Thinking abstractly, assume $\gamma(0) = 0$ for some $\gamma : (a,b) \to F^{-1}(\{z_0\})$. We can view γ as $\gamma : (a,b) \to \mathbb{R}^{m+n}$ and

$$(F \circ \gamma)'(0) = 0 \text{ since } F(\gamma(t)) = z_0 \text{ for all } t.$$

Recall the chain rule; this tells us $(\mathrm{D}F)_{(x_0,y_0)}(\gamma'(0)) = 0$, i.e., abstract tangent vectors are in $\ker(\mathrm{D}F)_{(x_0,y_0)}$.

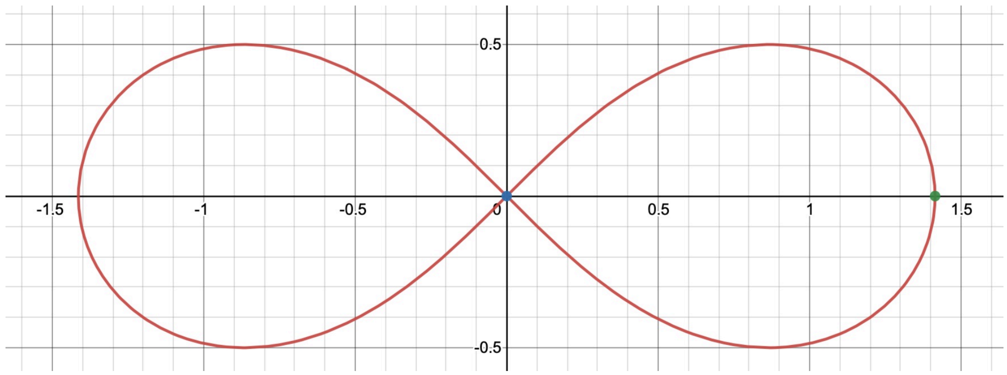
The linearization tells us that a vector in $\mathbb{R}^{m+n}$ being in $\ker(\mathrm{D}F)_{(x_0,y_0)}$ amounts to satisfying m linear[!] equations (where the original F gives nonlinear equations).

Best case scenario (no redundancies): each time an equation is added (each time a constraint is added), the dimension of the solutions decreases by 1. Then $\ker(\mathrm{D}F)_{(x_0,y_0)}$ is (precisely) n-dimensional (rather than $> n$). Then the *rank-nullity theorem* states that $\operatorname{rank}(\mathrm{D}F)_{(x_0,y_0)} = m + n - n = m$, i.e., the image of $(\mathrm{D}F)_{(x_0,y_0)}$ is m-dimensional.

Refresher: if rank$(\mathrm{D}F)_{(x_0,y_0}=m$, solving $(\mathrm{D}F)_{(x_0,y_0)}(v)=0$ amounts to solving the RREF version. While doing so, we are able to express some variables as a linear function of other variables. Thus we can write $\ker(\mathrm{D}F)_{(x_0,y_0)}$ as the graph of m of the coordinates (the pivot columns) as linear functions of the other n (the free variables).

This "maximal rank" assumption will be our hypothesis for the implicit function theorem, and we'll show that even the nonlinear version holds, but of course it's harder.

Example 5.4.3. Take again $n=m=1$. Consider $F:\mathbb{R}^2\to\mathbb{R}$ by $(x,y)\mapsto(x^2+y^2)^2-2(x^2-y^2)$. Take $z_0=0$, Then $F^{-1}(\{z_0\})=\{(0,0),(\pm\sqrt{2},0)\}$. See figure below. (This is called the *Lemniscate of Bernoulli*.)



Since $(\mathrm{D}F)_{(x,y)}=\begin{bmatrix}2(x^2+y^2)(2x)-4x & 2(x^2+y^2)(2y)+4y\end{bmatrix}$, a simple substituion with $(x,y)=(\sqrt{2},0)$ gives

$$(\mathrm{D}F)_{(\sqrt{2},0)}=\begin{bmatrix}4\sqrt{2} & 0\end{bmatrix},$$

which is of rank 1. It follows that $\ker(\mathrm{D}F)_{(\sqrt{2},0)}=\mathrm{span}\{(0,1)\}$ and indeed the tangent line is just the vertical line passing $(\sqrt{2},0)$.

However, what about $(0,0)$? Clearly $(\mathrm{D}F)_{(0,0)}=[0\quad 0]$. Its rank is 0 and the rank-nullity says $\ker(\mathrm{D}F)_{(0,0)}$ is of dimension 2 and thus the entire $\mathbb{R}^2$. This violates the "maximal rank" assumption and the implicit function theorem will not apply. This is what happens at a singular point (studied especially in algebraic geometry).

Theorem 5.4.4: Implicit function theorem

(Dini, 1876) Let $U\subset\mathbb{R}^{m+n}$ open, let $F:U\to\mathbb{R}^m$ be a function of class C^k $(1\leqslant k\leqslant\infty)$, and let $(x_0,y_0)\in U$, $z_0\in\mathbb{R}^m$ be such that $F(x_0,y_0)=z_0$ (i.e., $(x_0,y_0)\in F^{-1}(\{z_0\})$. We write the standard basis matrix of $(\mathrm{D}F)_{(x_0,y_0)}$ (an $m\times(m+n)$ matrix) as

$$\begin{bmatrix}A_{m\times n} \mid B_{m\times m}\end{bmatrix} \text{ and assume } B \text{ is invertible.}$$

(This is possible if we assume rank$(\mathrm{D}F)_{(x_0,y_0)}=m$ because we can simply apply RREF and put all pivot columns into B.)

Then, for all sufficiently $r>0$, there exists τ_0 such that, for all $\tau<\tau_0$, the following holds:

(1) $B(x_0,\tau)\times B(y_0,r)\in U$, and

(2) there exists a unique function $g:B(x_0,\tau)\to B(y_0,r)$ with

$$F^{-1}(\{z_0\})\cap(B(x_0,\tau)\times B(y_0,r))=\text{ graph of } g:=\{x,g(x)\mid x\in B(x_0,\tau)\}.$$

(3) Furthermore, g is of class C^k as well.

The proof of this theorem will be divided into several lemmas.

Beginning of April 19, 2021

Lemma 5.4.5

Let $U \subset \mathbb{R}^{n+m}$ be open and let $F : U \to \mathbb{R}^m$ be C^k $(1 \leqslant k \leqslant \infty)$ such that $(0,0) \in U$ ($\mathbb{R}^{n+m}$ viewed as $\mathbb{R}^n \times \mathbb{R}^m$, so first component in $\mathbb{R}^n$ and second in $\mathbb{R}^m$) and $(0,0) \in F^{-1}(\{0\})$ (preimage of $0 \in \mathbb{R}^m$). (*More generally, if we replace* $(0,0) \in U$ *and* $(0,0) \in F^{-1}(\{0\})$ *to get the i.f.t.*) Assume

$$(\mathrm{D}F)_{(0,0)} = \begin{bmatrix} A_{m\times n} & | & B_{m\times m} \end{bmatrix} \text{ where } B \text{ is invertible, i.e., } \operatorname{rank}(B) = m.$$

(*Again we can replace* $(0,0)$ *by more general coordinates, but this reduces the notational simplicity.*) Then for sufficiently small $r > 0$, there exists $\tau_0 > 0$ such that, for all $\tau \leqslant \tau_0$,

(1) $B(0,\tau) \times B(0,r) \subset U$,

(2) there exists a unique function $g : B(0,\tau) \to B(0,r)$ (from $\mathbb{R}^n$ to $\mathbb{R}^m$) with

$$F^{-1}(\{0\}) \cap (B(0,\tau) \times B(0,r)) = \text{ graph of } g := \{(x, g(x)) \mid x \in B(0,\tau)\}.$$

The function g is locally Lipschitz at 0, i.e., there exists L such that for $x \in B(0,\tau)$ small enough, we have

$$\|g(x) - g(0)\| \leqslant L\|x - 0\|.$$

(*Note that* $g(0) = 0$ *since* $F(0,0) = 0$.)

Proof. For $(x,y) \in U$, let $R(x,y) := F(x,y) - F(0,0) - [A \mid B][x \mid y]^T$, i.e., (since $F(0,0) = 0$)

$$F(x,y) = Ax + By + R(x,y).$$

Thus for $(x,y) \in U$, we have

$$(x,y) \in F^{-1}(\{0\}) \iff F(x,y) = 0 \iff Ax + By + R(x,y) = 0.$$

The RHS looks almost like we are trying to solve a function of y: the $\iff$ chain is further extended to

$$(x,y) \in F^{-1}(\{0\}) \iff By = -Ax - R(xy) \iff y = -B^{-1}(Ax + R(x,y)).$$

The last step is justified as B is assumed to be invertible. We would have been done if $R(x,y)$ depends solely on x. Unfortunately it is not the case, but we can view it as a fixed-point theorem. For a fixed x, we can view the RHS as a function of y. By the uniqueness of fixed-point there will be a unique y where the LHS=RHS.

Define $r_0 > 0$ such that $\overline{B(0,r_0)} \times \overline{B(0,r_0)} \subset U$. If $x \in B(0,r_0)$, we can define an operator $K_x : \overline{B(0,r_0)} \to \mathbb{R}^m$ by

$$K_x(y) := -B^{-1}(Ax + R(x,y)).$$

To invoke Banach contraction mapping theorem, we need to ensure K_x is a (strong) contraction and that the space on which K_x maps into itself is Banach.
First note that $R(x,y) = F(x,y) - Ax - By$ is C^k just like F is (as linear functions Ax, By are C^∞) and

$$(\mathrm{D}R)_{(0,0)} = \begin{bmatrix} A & \mid & B \end{bmatrix} - \begin{bmatrix} A & \mid & B \end{bmatrix} = 0 \text{ (zero matrix)}.$$

Let $\frac{\partial R}{\partial y}(x,y)$ be the "partial Jacobian of R" at (x,y) where we only take the y partials not the x partials. Note that $\frac{\partial F}{\partial y}(0,0) = B$ so $\frac{\partial R}{\partial y} = (0,0) = B - B = 0$. It follows that $\frac{\partial R}{\partial y}$ is a function $U \to \mathrm{M}_{m\times n}(\mathbb{R})$ where each entry is a partial derivative of R. Since R is of class C^k, these entries are in particular continuous. For convenience we pick the Euclidean norm on U. We know

$$\frac{\partial R}{\partial y} : (U, \|\cdot\|_E) \to (\mathrm{M}_{m\times n}(\mathbb{R}), \|\cdot\|_{\mathrm{op}})$$

is continuous. Thus, for $\epsilon := 1/(2\|B^{-1}\|_{\mathrm{op}})$, there exists $r > 0$ (WLOG, also $\leqslant r_0$) such that if $\|x\|, \|y\| \leqslant r$ then

$$\left\|\frac{\partial R}{\partial y}(x,y) - \frac{\partial R}{\partial y}(0,0)\right\|_{\mathrm{op}} = \left\|\frac{\partial R}{\partial y}(x,y)\right\|_{\mathrm{op}} < \frac{1}{2\|B^{-1}\|_{\mathrm{op}}}. \tag{Δ}$$

(Note $B^{-1} \neq 0$ so its norm is not zero and division makes sense.) On the other hand, by making r smaller if necessary, we can also ensure

$$\det\left[\frac{\partial F}{\partial y}(x,y)\right] \neq 0 \text{ for } \|x\|, \|y\| \leqslant r$$

because $\det\left[\frac{\partial F}{\partial y}(0,0)\right] = \det(B) \neq 0$ and (1) all entries of $\frac{\partial F}{\partial y}(x,y)$ are continuous and (2) the determinant itself is a continuous operator w.r.t. the entries.

Claim. If $\|x\|, \|y_1\|, \|y_2\| \leqslant r$, then $\|K_x(y_1) - K_x(y_2)\| \leqslant \|y_1 - y_2\|/2$ so K_x is a contraction. To see this,

$$\begin{aligned}
\|K_x(y_1) - K_x(y_2)\| &= \| - B^{-1}(Ax + R(x,y_1)) - (-B^{-1}(Ax + R(x,y_2))\| \\
&= \| - B^{-1}(R(x,y_1) - R(x,y_2))\| \\
&\leqslant \|B^{-1}\|_{\mathrm{op}}\|R(x,y_1) - R(x,y_2)\|.
\end{aligned}$$

The second term $\|R(x,y_1) - R(x,y_2)\|$ can be further bounded by $\|y_2 - y_1\|/(2\|B^{-1}\|_{\mathrm{op}})$ using mean value inequality (which we will continue next lecture).

Beginning of April 19, 2021

By the mean value inequality, for a fixed x, $R(x,\cdot)$ is a C^1 function of y. Thus it is differentiable with total derivative $\frac{\partial R}{\partial y}(x,y)$. Therefore

$$\|R(x,y_1) - R(x,y_2)\| \leqslant \sup_{t\in[0,1]} \left\|\frac{\partial R}{\partial y}(x, y_1 + t(y_2 - y_1))\right\| \cdot \|y_2 - y_1\| < \frac{\|y_2 - y_1\|}{2\|B^{-1}\|_{\mathrm{op}}}$$

by (Δ) (since $\|x\|, \|y_1\|, \|y_2\|$ are all < 1 and the ball is convex so $\|y_1 + t(y_2 - y_1)\| < r$ too). Therefore K_x is a contraction. `[End of proof of claim]`

To ensure K_x maps $\overline{B(0,r)}$ maps into itself, we will make x smaller if necessary. Since the function

$$x \mapsto K_x(0) = -B^{-1}(Ax + R(x,0))$$

is continuous, for $\epsilon = r/2$ there exists $\tau > 0$ (WLOG assume $\tau \leqslant r$) such that if $\|x\| < \tau$ then $\|K_x(0)\| < r/2$.

Claim. If $\|x\| < \tau$ and $\|y\| < r$ then $\|K_x(y)\| \leqslant r$. Indeed,

$$\begin{aligned}\|K_x(y)\| &= \|K_x(y) - K_x(0) + K_x(0)\| \\ &\leqslant \|K_x(y) - K_x(0)\| + \|K_x(0)\| \\ &\leqslant \frac{\|y\|}{2} + \|K_x(0)\| \leqslant \frac{r}{2} + \frac{r}{2} = r.\end{aligned}$$

[End of proof of claim]

Therefore if $\|x\| \leqslant \tau$ then K_x is a contraction on the complete $\overline{B(0,r)} \subset \mathbb{R}^m$. Therefore by Banach contraction mapping theorem, K_x has a unique fixed point, i.e., if $\|x\| \leqslant \tau$ then there exists a unique $y \in \overline{B(0,r)}$ such that $K_x(y) = y$, so

$$y = -B^{-1}(Ay + R(x,y)) \iff F(x,y) = 0.$$

Therefore given x, there exists a unique y such that $(x,y) \in \overline{B(0,\tau)} \times \overline{B(0,r)}$ satisfies $K_x(y) = y$. Now we define

$$g : \overline{B(0,\tau)} \to \overline{B(0,r)}$$

by taking $g(x)$ to be the unique fixed point y. It follows that the set of such (x,y)'s form the graph of g, and

$$F^{-1}(\{0\}) \cap (\overline{B(0,\tau)} \times \overline{B(0,r)}) = \text{ graph of } g.$$

Claim. If $L := \|B^{-1}\|_{\text{op}}\|A\|_{\text{op}}$, then $4L$ is a local Lipschitz constant for g at 0, i.e., for small enough x,

$$\|g(x) - g(0)\| \leqslant 4L\|x - 0\| = 4L\|x\|.$$

We know that $g(x)$ is (the) fixed point for K_x, so $g(x) = K_x(g(x))$. Therefore

$$\begin{aligned}\|g(x)\| = \|K_x(g(x))\| &= \|K_x(g(x)) - K_x(0) + K_x(0)\| \\ &\leqslant \|K_x(g(x)) - K_x(0)\| + \|K_x(0)\| \\ &= \|K_x(g(x)) - K_x(g(0))\| + \|K_x(0)\| \\ &\leqslant \frac{\|g(x) - g(0)\|}{2} + \|B^{-1}(Ax + R(x,0))\| \\ &\leqslant \frac{\|g(x)\|}{2} + \|B^{-1}\|_{\text{op}}\|Ax + R(x,0)\| \\ &\leqslant \frac{\|g(x)\|}{2} + \|B^{-1}\|_{\text{op}}\|A\|_{\text{op}}\|x\| + \|B^{-1}\|_{\text{op}}\|R(x,0)\|.\end{aligned}$$

Note that $\lim\limits_{(x,y)\to 0} \dfrac{R(x,y)}{\|(x,y)\|} = 0$ because F is differentiable. Therefore treating $y = 0$ we have $\lim\limits_{x\to 0} \dfrac{R(x,0)}{\|(x,0)\|} = 0$. Assuming $\|A\|_{\text{op}} = \epsilon > 0$ (for now we'll assume it's strictly positive), for small enough $x \leqslant \tau'$ (WLOG $\tau' \leqslant \tau$)

$$\frac{\|R(x,0)\|}{\|(x,0)\|} \leqslant \|A\|_{\text{op}} \implies \|R(x,0)\| \leqslant \|A\|_{\text{op}}\|x\|.$$

Then,

$$\|g(x)\| \leqslant \frac{\|g(x)\|}{2} + 2\|B^{-1}\|_{\text{op}}\|A\|_{\text{op}}\|x\|$$

so

$$\frac{1}{2}\|g(x)\| \leqslant 2\|B^{-1}\|_{\text{op}}\|A\|_{\text{op}}\|x\| \implies \|g(x)\| \leqslant 4L\|x\| \text{ where } L := \|B^{-1}\|_{\text{op}}\|A\|_{\text{op}}.$$

[End of proof of claim]

Local Lipschitz condition then implies that g is continuous at 0. Therefore there exists $\tau' \leqslant \tau$ such that $g(B(0,\tau')) \subset B(0,r)$. By construction, $F^{-1}(0) \cap (B(0,\tau) \times B(0,r))$ is still the graph of $g : B(0,\tau') \to B(0,r)$. This g is the function we have been looking for for so long. And clearly, if g_1, g_2 both have this property, then they must have the same graph (uniqueness of K_x's fixed points) and they must equal. □

Lemma 5.4.6

In the previous lemma, the unique function g is not just Lipschitz: it is differentiable at zero with total derivative given by $-B^{-1}A$ where $(\mathrm{D}F)_{(0,0)} = [A \mid B]$.

Proof. For convenience let τ be the τ' in the previous claim (the sufficiently small bound on x). Then for $x \in B(0,\tau)$, $g(x)$ is the unique fixed point of K_x. Thus

$$g(x) = K_x(g(x)) = -B^{-1}(Ax + R(x, g(x))).$$

It follows that

$$\begin{aligned}\|g(x) - g(0) - (-B^{-1}A)x\| &= \|g(x) - (-B^{-1}A)x\| \\ &= \| - B^{-1}(Ax + R(x,g(x))) + B^{-1}Ax\| \\ &= \|B^{-1}(R(x,g(x))\| \leqslant \|B^{-1}\|_{\mathrm{op}}\|R(x,g(x))\|,\end{aligned}$$

and so

$$\begin{aligned}\frac{\|g(x) - g(0) - (-B^{-1}A)x\|}{\|x\|} &\leqslant \|B^{-1}\|_{\mathrm{op}} \frac{\|R(x,g(x))\|}{\|x\|} \\ &= \|B^{-1}\|_{\mathrm{op}} \frac{\|R(x,g(x))\|}{\|(x,g(x))\|} \cdot \frac{\|(x,g(x))\|}{\|x\|}.\end{aligned}$$

The last equation since $x \neq 0$ implies $(x, g(x)) \neq 0$. Since g is continuous at 0, as $x \to 0$ we have $g(x) \to 0$ and so $(x, g(x)) \to 0 \in \mathbb{R}^{n+m}$. Since F is differentiable, the term $\|R(x,g(x))\|/\|(x,g(x))\| \to 0$. For the last term,

$$\begin{aligned}\frac{\|(x,g(x))\|}{\|x\|} &= \frac{\|(x,0) + (0,g(x))\|}{\|x\|} \\ &\leqslant \frac{\|x\| + \|g(x)\|}{\|x\|} \leqslant 1 + 4L\end{aligned}$$

for sufficiently small (but nonzero) x.

(In the case where $\|A\|_{\mathrm{op}} = 0$ (for example $F(x,y) = x^2 - y$ and so $(\mathrm{D}F)_{(x,y)} = [2x \quad 1]$ and $(\mathrm{D}F)_{(0,0)} = [0 \quad 1]$), pick an arbitrary $C > 0$ and use the fact that $\lim_{x \to 0} R(x,0)/\|(x,0)\| < \|A\|_{\mathrm{op}} + C$. Then the claim is analogous, with L being repalced by $L := \|B^{-1}\|_{\mathrm{op}}(\|A\|_{\mathrm{op}} + C)$.)

Therefore as $x \to 0$, this term $\to 0$ as well. Hence the entire thing $\to 0$, i.e., g is differentiable at 0 with derivative $-B^{-1}A$. □

Beginning of April 23, 2021

Lemma 5.4.7

Now we show that $g : B(0,\tau) \to B(0,r)$ is differentiable at all $x \in B(0,\tau)$ with derivative at x being

$$-\left(\frac{\partial F}{\partial y}(x, g(x))\right)^{-1}\left(\frac{\partial F}{\partial x}(x, g(x))\right) = -B^{-1}_{x,g(x)}A_{x,g(x)}.$$

Proof. Let $x_0 \in B(0,\tau)$ and let $y_0 = g(x_0)$ be arbitrarily chosen. Define

$$\tilde{F}(x,y) := F(x + x_0, y + x_0).$$

This is a function defined on $B(0,\tau') \times B(0,r')$ as long as $\tau < t - \|x_0\|$ and $r' < r - \|y_0\|$ (so $x + x_0$ is still in $B(0,\tau)$ and likewise for the other one). Notice that $\tilde{F}$ is of class C^k as is F itself. It follows clearly that

$$\frac{\partial \tilde{F}}{\partial y}(0,0) = \frac{\partial F}{\partial y}(x_0, y_0)$$

is invertible and we can simply apply the previous result to $\tilde{F}$: $B(0,\tau') \to B(0,r') \to \mathbb{R}^m$. Thus there exists a $\tau'' \leqslant \tau'$ and a function $\tilde{g} : B(0,\tau'') \to B(0,r')$ (*a weaker statement than we can actually get:* $\tilde{g}$ *to* $B(0,r'')$ *for some smaller* r'' *but we don't care*) such that $\tilde{F}(x,\tilde{g}(x)) = 0$ for all $x \in B(0,\tau'')$ and $\tilde{g}$ is differentiable with derivative

$$-\left(\frac{\partial \tilde{F}}{\partial y}(0,0)\right)^{-1}\left(\frac{\partial \tilde{F}}{\partial x}(0,0)\right).$$

Claim For small enough x with $\|x\| < \tau''$, we have $g(x) = \tilde{g}(x - x_0) + y_0$.
To see this, notice that $g(x)$ is the unique point in $B(0,r)$ with $F(x, g(x)) = 0$. On the other hand,

$$0 = \tilde{F}(x - x_0, \tilde{g}(x - x_0)) = F(x, \tilde{g}(x - x_0) + y_0)$$

so by uniqueness of fixed points, $g(x) = \tilde{g}(x - x_0) + y_0$. `[End of proof of claim]`
Therefore by chain rule, g is differentiable at x_0 (since $\tilde{g}$ is at 0) and

$$(\mathrm{D}g)_{x_0} = (\mathrm{D}\tilde{g})_0 = -\left(\frac{\partial \tilde{F}}{\partial y}(0,0)\right)^{-1}\left(\frac{\partial \tilde{F}}{\partial x}(0,0)\right) = -\left(\frac{\partial F}{\partial y}(x_0, g(x_0))\right)^{-1}\left(\frac{\partial F}{\partial x}(x_0, g(x_0))\right),$$

proving the differentiability of g on points other than the origin. □

Lemma 5.4.8

Our unique function g is of class C^1.

Proof. We want to show that each entry of $(\mathrm{D}g)_x$ is a continuous function of $x \in B(0,\tau)$. We have

$$(\mathrm{D}g)_x = -\left(\frac{\partial F}{\partial y}(x, g(x))\right)^{-1}\left(\frac{\partial F}{\partial y}(x, g(x)\right) = -B^{-1}_{x,g(x)}A_{x,g(x)}.$$

Notice that the entries of $A_{x,y}$ and $B_{x,y}$ are partial derivatives of components of F and are thus continuous. Since g is continuous on $B(0,\tau)$, entries of $A_{x,g(x)}, B_{x,g(x)}$ are composition of continuous functions and are therefore continuous. It remains to deal with $B^{-1}_{x,g(x)}$, but fortunately we have Cramer's rule which gives us an explicit way of computing inverses. Since $\det(B_{x,g(x)})$ is nonzero (B invertible),

$$B^{-1}_{x,g(x)} = \frac{1}{\det(B_{x,g(x)})} \cdot \text{cofactor}(B_{x,g(x)})^T$$

where the (i,j) entry of the cofactor matrix is $(-1)^{i+j}$ times the determinant of the original matrix (i.e. B) but without i^{th} row and j^{th} column. (Then we take transpose.) Since determinants are continuous, this entire thing is also continuous. Thus g is C^1. □

Lemma 5.4.9

One more final lemma before we prove the implicit function theorem: assuming F is C^k, then g is not only C^1 but also C^k. True for $1 \leqslant k \leqslant \infty$.

Proof. We use induction on k. The base case $k = 1$ is shown in the lemma above. Now for the inductive step, assume ($F \in C^k$ and in particular $F \in C^{k-1}$) implies ($g \in C^{k-1}$). To show g is of class C^k, it suffices to show that the entries of $(\mathrm{D}g)_x$ are C^{k-1} functions of x. (k^{th} partials of g is the same as $(k-1)^{\text{th}}$ partials of entries of $(\mathrm{D}g)_x$).

Recall that $(\mathrm{D}g)_x = -B^{-1}_{x,g(x)} A_{x,g(x)}$. Since F is C^k, entries of $(\mathrm{D}F)_{(x,y)}$ are C^{k-1} functions of (x,y), so entries of $A_{x,g(x)}, B_{x,g_(x)}$ of C^{k-1}. By Cramer's rule, $B^{-1}_{x,g(x)}$ is also C^{k-1} and thus $(\mathrm{D}g)_x$ is C^{k-1}. □

Proof of implicit function theorem. It should be obvious now. □

Implicit Function Theorem

Definition 5.4.10

If $U, V \subset \mathbb{R}^m$ are open, then $F : U \to V$ is a **C^k diffeomorphism** ($1 \leqslant k \leqslant \infty$) if:

(1) F is of class C^k, and

(2) F^{-1} is also C^k (this is not true in general; it might not even be differentiable).

Example 5.4.11. A non-example of diffeomorphism: $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$. Immediately we see f is of class C^∞ but the inverse $f^{-1}(y) = \sqrt[3]{y}$ is not even differentiable at origin. (*In this case f is a homeomorphism nonetheless, i.e, bicontinuous bijection.*)

Proposition 5.4.12

If $U \subset \mathbb{R}^m$, $V \subset \mathbb{R}^n$ are open and $F : U \to V$ is a diffeomorphism, then $m = n$.

Proof. We have $F^{-1} \circ F = \mathrm{id}_U$ so

$$(\mathrm{D}F^{-1})_{F(p)}(\mathrm{D}F)_p = I_{n \times n}$$

and likewise $F \circ F^{-1} = \mathrm{id}_V$. Of course this requires the invertibility of $(\mathrm{D}F)_p$ so this requires $(\mathrm{D}F)_p$ to be square!

Therefore $m = n$. □

Therefore, "nonlinear invertibility" implies "linear invertibility of linear approximations / total derivatives". But what about the converse?

Theorem 5.4.13: Inverse function theorem

Let $U_0 \subset \mathbb{R}^n$ be open and let $f : U_0 \to \mathbb{R}^m$ be a C^k function ($1 \leqslant k \leqslant \infty$). Let $p \in U_0$ such that $(\mathrm{D}f)_p$ is invertible (by above we already have $n = m$).
Claim: there exists an open subset $p \in U \subset U_0$ such that $f(p) \in V \subset \mathbb{R}^n = \mathbb{R}^m$ such that f is a C^k diffeomorphism from U to V.

Beginning of April 26, 2021

Proof. Main idea: set things up to apply the implicit function theorem. Note that $U_0 \times \mathbb{R}^n$ is an open subset of $\mathbb{R}^{n+n} = \mathbb{R}^{2n}$. Now we define $F : U_0 \times \mathbb{R}^n \to \mathbb{R}^n$ by $F(x, y) = f(x) - y$ (where $x \in U_0, y \in \mathbb{R}^n$). If we define $q := f(p)$ then immediately $F(p, q) = 0$, so $(p, q) \in F^{-1}(0)$. We also have

$$(\mathrm{D}F)_{(p,q)} = \left[\frac{\partial F}{\partial x}(p,q) \quad \frac{\partial F}{\partial y}(p,q)\right].$$

Notice that $\frac{\partial F}{\partial y}(p,q) = -I$ (for y, F a constant function minus y), and $\frac{\partial F}{\partial x}(p,q)$ is just $(\mathrm{D}f)_p$. So

$$(\mathrm{D}F)_{(p,q)} = \left[(\mathrm{D}f)_p \quad | \quad -I\right].$$

We know both blocks are invertible. Specifically, since $(\mathrm{D}f)_p$ is invertible, using these columns, we can apply the implicit function theorem to the first n coordinates and express the level set (of F at 0) as a graph <u>of x as a function of y</u>.
To put formally, applying the implicit function theorem to the <u>first n coordinates</u>, there exists an open neighborhood $U_p \times V_q$ of $(p, q) \in U_0 \times \mathbb{R}^n$ and a <u>unique</u> function $h : V_q \to U_q$ with

$$F^{-1}(0) \cap (U_p \times V_q) = \text{ graph of } h = \{(h(y), y) \mid y \in V_q\},$$

where h is of class C^k. It follows that $F(h(y), y) = 0$ for all $y \in V_q$, i.e.,

$$f(h(y)) - y = 0 \text{ for all } y \in V_q \implies f(h(y)) = y \implies f \circ h = \mathrm{id}_{V_q}.$$

Then $h(q) = p$ because $(p, q) \in F^{-1}(0) \cap (U_p \times V_q) = \text{graph }(h)$. Since $(\mathrm{D}f)_p$ and $(\mathrm{D}h)_q$ are <u>square</u>, chain rule and linear algebra gives

$$(\mathrm{D}f)_p \circ (\mathrm{D}h)_q = \mathrm{id} \implies (\mathrm{D}h)_q \circ (\mathrm{D}f)_p = \mathrm{id}.$$

(Thankfully any one-sided inverse of a square matrix is a two-sided inverse!) Thus $(\mathrm{D}h)_q$ is invertible.
We can repeat the process above and do the same thing with h instead of f. Define $H : U_p \times V_q \to \mathbb{R}^n$ by $H(x, y) = h(y) - x$. We again have $H(p, q) = 0$ so $(p, q) \in H^{-1}(0)$. Then,

$$(\mathrm{D}H)_{(p,q)} = \left[-I \mid (\mathrm{D}h)_q\right]$$

where we know both blocks are invertible. Using the implicit function theorem on the right, there exists open neighborhood $U_p' \times V_q' \subset U_p \times V_q$ and a unique function $g : U_p' \to V_q'$ with $H^{-1}(0) \cap (U_p' \times V_q') = \text{graph }(g)$. Following the same argument, $h \circ g = \mathrm{id}_{U_p'}$.

Claim. $g = f$ restricted on U'_p. Indeed, for $x \in U'_p$, $g(x) \in V'_q \subset V_q$. Using $f \circ h = \mathrm{id}_{V_q}, g(x) = f(h(g(x)))$, and since $h \circ g$ is also identity on U'_p, $g(x) = f(x)$, proving the subclaim. [END OF PROOF OF CLAIM]

However, since f may well be not linear, there is no guarantee that this we have a bisection — the inverse may only work for a smaller subset. The claim below addresses this issue and thus proves the main theorem.

Claim. Let $V''_q := h^{-1}(U'_p)$. Then $f(U'_p) \subset V''_q$.
Since h has domain V_q, we know $V''_q \subset V_q$. Note that since $h \circ g$ is $\mathrm{id}_{U'_p}$ and $g = f$ restircted to U'_p, we have $h \circ f$ resitrcted to U'_p is still the same identity map on U'_p. Then if $y \in V''_q$, $y = f(h(y)) = g(h(y))$ so $V''_q \subset V'_q$.
Therefore, if $x \in U'_p$ then $x = h(f(x))$, so $f(x) \in h^{-1}(U'_p) = V''_q$. [END OF PROOF OF CLAIM]

Now that we have obtained two bijective maps, $f : U'_p \to V''_q$ and $h : V''_q \to U'_p$, the composition each way gives the identities and this proves the inverse function theorem. □

5.5 A more abstract View on Differential Forms

Definition 5.5.1

If V is a vector space, let

$$T^*V = \bigoplus_{n=0}^{\infty} \underbrace{V \otimes \ldots \otimes V}_{n \text{ times}}.$$

This is a vector space and it has multiplication by $\otimes$. This defines a ring (even an $\mathbb{R}$-algebra) and is called the **tensor algebra** of V. (Think *noncommutative polynomials in the basis vectors of* V.) And we have a graded $\mathbb{R}$-algebra, where a piece of degree k is $V^{\otimes k}$.

Definition 5.5.2

The **exterior algebra** of V is

$$\bigwedge^* V = \frac{T^*V}{(x \otimes x \ \forall x \in V)}$$

where the denominator (quotient) denotes the two-sided ideal generated by elements of form $x \otimes x$. Even better: the quotient (graded ring / homogeneous ideal) is another graded ring. The *number of tensor factors* gives a well-defined grading on $\bigwedge^* V$:

$$\bigwedge^* V = \bigoplus_{k=0}^{\infty} \bigwedge^k V,$$

the k^{th} **exterior power** of V.

The natural map $V^k \to \wedge^k V$ defined by

$$v_1 \otimes \ldots . \otimes v_k \mapsto \frac{1}{k!}[v_1 \otimes \ldots \otimes v_k]$$

(equivalence class on RHS) restricts to an isomorphism

$$\{v \in V^{\otimes k} : v \text{ alternating bilinear}\} \xrightarrow{\cong} \bigwedge^k V,$$

with inverse

$$[v_1 \otimes \ldots \otimes v_k] \to \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma)\, v_{\sigma(1)} \otimes \ldots \otimes v_{\sigma(k)}.$$

This gives a highly useful approach to wedge products: we can naturally identify $v_1 \wedge \ldots \wedge v_k$ with the equivalence class $[v_1 \otimes \ldots \otimes v_k]$, a much cleaner approach than wedge products done in HW14.

It follows that if $\alpha \in \bigwedge^k V$ and $\beta \in \bigwedge^\ell V$, then $a \wedge \beta$ is in $\bigwedge^* V = \bigoplus_{n=0}^{\infty} \bigwedge^n V$.

Beginning of April 28, 2021 [last day of class!]

Vector Bundles

In general, let M be a smooth manifold second-countable Hausdorff topological space equipped with equivalence classes of smooth atlases := $\{(u_\alpha, \varphi_\alpha) : \alpha \in A\}$ where $U_\alpha \subset M$ and $\varphi_\alpha : U_\alpha \to V_\alpha \subset \mathbb{R}^n$ (open) is a homomorphism. In other words, $\varphi_\alpha^{-1} : V_\alpha \to U_\alpha$ is a local parametrization for M in the patch U_α. We further require $\bigcup_{\alpha \in A} U_\alpha = M$ and such that the "transition functions" (φ^{-1}) are smooth. We say two atlases are equivalent if their union is still a smooth atlas.

Definition 5.5.3

For $p \in M$, we've previously defined T_pM to be the set of equivalence classes of curves at p to M. We define T_p^*M to be the dual space of T_pM, **contangent space** to M at p.

Then we can define **vector bundles** over smooth manifold (*smoothly varying* collection of vector spaces [all of same dimension, called the **rank** of vector bundle] for each point of M). Similarly, the cotangent spaces T_p^*M organize into the **cotangent bundle** T^*M.

So far, we have linear algebra operations like $\oplus, \otimes, ^*, \bigwedge^*$, and so on. All of these can be extended and can be applied to bundles (we can take these operations "point by point" and this upgrades to a construction on vector bundles). For example, we can form $\bigwedge^k(T^*M)$, the k^{th} exterior power of the cotangent bundle of M.

Definition 5.5.4

Say $E \xrightarrow{\pi} M$ is a vector bundle. This means E is the disjoint union of all vector spaces over all points in M, and on top of that it has a smooth manifold itself. Here π sends a vector in (the vector space at p, a subset of E) to the point p, and we define $E_p := \pi^{-1}(\{p\})$, the vector space we have at point $p \in M$.

A **section** of the vector bundle is a map $f : M \to E$ such that $\pi \circ f = \text{id}$. In other words, $f(p) \in \pi^{-1}(\{p\}) = E_p$ for all $p \in M$. *Heuristically a section of a vector bundle is like a function with variable codomain: $f(p)$ is a point in the vector space that we assigned to point p.* For example a vector field on M is a section of TM.

Definition 5.5.5

A differential k-form on M is a section of $\bigwedge^k(T^*M)$. In particular a 1-form is a section of T^*M.

Pullbacks

Definition 5.5.6

Let M, N be smooth manifolds and $F : M \to N$ a smooth map. Let α be a k-form on N (at $q \in N$, we have $\alpha_q \in \bigwedge^k(T_q^* M) \cong \text{Alt}_k(T_q N, \mathbb{R})$).
Now we define the **pullback** $F^*\alpha$: for $p \in M$, we want an element of $\text{Alt}_k(T_p M, \mathbb{R})$, a alternating multilinear map taking k inputs in $T_p M$ and gives a real number. $(F^*\alpha)_p$ is defined to act on inputs $v_1, \ldots, v_k \in T_p M$ by taking

$$\alpha_{F(p)}((\mathrm{D}F)_p(v_1), \ldots, (\mathrm{D}F)_p(v_k)).$$

Facts. Some basic facts about pullbacks:

(1) Pullbacks + exterior derivatives: $F^*(\mathrm{d}\alpha) = \mathrm{d}(F^*\alpha)$.

(2) Pullbacks + wedge products: $F^*(\alpha \wedge \beta) = F^*(\alpha) \wedge F^*(\beta)$.

(3) (Exterior derivatives + wedge products: $\mathrm{d}(\alpha \wedge \beta) = (\mathrm{d}\alpha) \wedge \beta + (-1)^k \alpha \wedge (\mathrm{d}\beta)$ where α is a k-form.)

Integration of Differential Forms on Manifolds

Note that $\bigwedge^n(T^* M)$ for a n-manifold M is a rank 1 vector bundle (line bundle) on M.

Definition 5.5.7

We say M is **orientable** if there exists a nowhere vanishing (global) section ω of $\bigwedge^n(T^* M)$. If so, the **orientation** on M is defined to be the equivalence class of such sections *modulo multiplication*, $[\omega]$, by always positive functions. Non-orientable manifolds include the Mobius band, the Klein bottle, and so on.

Definition 5.5.8

If α is a **top-degree** (degree n) form on an oriented n-manifold M, then we can define the **integral** of α on M by *partition of unity*, that is, to write α as a sum of forms such that, for each term α_i of the sun, there exists some oriented coordinate patch (U_i, φ_i) such that α_i vanishes outside U.
Now to integrate α_i, we use pullback by φ_i^{-1} to get a compactly supported n-form on an open subset of $\mathbb{R}^n$.
Then it remains to sum over all i in the original sum decomposition of α.
This process is independent of choice of partitions.

For example, if α is a k-form on $\mathbb{R}^N$, we can pull back α by any $r : [0,1]^k \to \mathbb{R}^n$ (or more general subset). This leads to the notion of *integration of chains*.

Theorems about Integrations

Theorem 5.5.9

[Integrals + pullbacks] If M, N are oriented manifolds, $F : M \to N$ a smooth and *orientation-preserving* diffeomorphism, and α an n-form on N, then

$$\int_N \alpha = \int_M F^*(\alpha).$$

This is the differential form equivalence of change of variable formula.

Theorem 5.5.10: Stokes' Theorem

[Integrals + exterior derivatives] Let M be an oriented n-manifold with boundary ∂M (oriented too) and α a $(n-1)$-form. Then

$$\int_M \mathrm{d}\alpha = \int_{\partial M} \alpha.$$

Indeed, we can pull back α via the inclusion map $\partial M \hookrightarrow M$.

This generalizes the classical theorems.

For line integrals. For example we can define a 1-form $\mathrm{d}s$ on C by adding the requirement that in any local parametrization $r(t)$ for C, we have

$$r^*(\mathrm{d}s) = \|r'(t)\|\mathrm{d}t.$$

Then any 1-form on an oriented curve C is equal to $f\mathrm{d}s$ for some function on C. Then how do we integrate $f\mathrm{d}s$ on C using pullback?

$$\int_C f\ \mathrm{d}s = \int_a^b f(r(t))\|r'(t)\|\ \mathrm{d}t$$

where $[a, b]$ is the domain of r.

How about $C \in \mathbb{R}^n$ for some n and we have a 1-form $\alpha : p_1\mathrm{d}x_1 + \ldots + p_n\mathrm{d}x_n$ on $\mathbb{R}^n$? Similar! We can pull back α via inclusion map $C \hookrightarrow \mathbb{R}^n$ and get some 1-form on C: $(F \cdot T)\ \mathrm{d}s$ where

$$F = \begin{bmatrix} p_1 & \ldots & p_n \end{bmatrix}^T \text{ and } T(r(t)) = \frac{r'(t)}{\|r'(t)\|}.$$

This explains the line integrals of vector fields as curves.

For classical Green's theorem. We take Stokes' theorem for M on a 2-manifold in ambient $\mathbb{R}^2$. Assuming α a 1-form on M comes from 1-form on all of $\mathbb{R}^2$ with $\alpha = P\mathrm{d}x + Q\mathrm{d}y$, using properties of wedge product (cf. HW14) gives

$$\mathrm{d}\alpha = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}x \wedge \mathrm{d}y$$

and

$$\int_M \mathrm{d}\alpha = \oint_{\partial M} \alpha \implies \int_M \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\ \mathrm{d}x\ \mathrm{d}y = \oint_{\partial M} (F \cdot T)\ \mathrm{d}s \text{ where } F = \begin{bmatrix} P \\ Q \end{bmatrix}.$$

For surface analogue & classical Stokes' and divergence theorems. We define a 2-form $\mathrm{d}A$ on S such that, for any local parametrization $r(u, v)$, we set

$$r^*(\mathrm{d}A) = \left\|\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}\right\|\mathrm{d}u \wedge \mathrm{d}v.$$

(*This is well-defined.*) Just like $\mathrm{d}s$ above, here $\mathrm{d}A$ is nowhere vanishing and it represents orientation classes). Then the classical surface integral with respect to area, $\int_S f\, \mathrm{d}A$, is the same as integrals of arbitrary 2-forms on surfaces. Let $S \subset \mathbb{R}^3$. We have a 2-form on S that comes from a 2-form

$$\alpha : P\mathrm{d}x \wedge \mathrm{d}y + Q\mathrm{d}x + \mathrm{d}z + R\mathrm{d}y \wedge \mathrm{d}z \text{ on } \mathbb{R}^3.$$

We can pull back α via the inclusion map $S \hookrightarrow \mathbb{R}^3$ which is again some $f\mathrm{d}A$:

$$f = (F \cdot \hat{n})\mathrm{dA} \text{ where } F = \begin{bmatrix} R \\ -Q \\ P \end{bmatrix} \text{ and } \hat{n}(r(u,v)) = \frac{\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}}{\| \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \|}.$$

This gives the flux integrals and explains classical surface integrals of vector fields.

Stokes' theorem. Let a 1-form on S come from $\alpha = P\mathrm{d}x + Q\mathrm{d}y + R\mathrm{d}z$ on ambient $\mathbb{R}^3$. Applying wedge products on α (HW14),

$$\int_S \mathrm{d}\alpha = \oint_{\partial S} \alpha \implies \int_S ((\nabla \times F) \cdot \hat{n})\, \mathrm{d}A = \oint_{\partial S} (F \cdot T)\, \mathrm{d}s.$$

($\nabla \times F$ is just the curl of F.)

Divergence theorem: now we consider a 2-form on M coming from $\alpha = P\mathrm{d} \wedge \mathrm{d}y + Q\mathrm{d}x \wedge \mathrm{d}z + R\mathrm{d}y \wedge \mathrm{d}z$ on $\mathbb{R}^3$. It follows (again, from HW14) that

$$\mathrm{d}\alpha = \mathrm{div}(F)\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z$$

and thus

$$\int_M \mathrm{d}\alpha = \oint_{\partial M} \alpha \implies \iiint_M \mathrm{div}(F)\, \mathrm{d}x\mathrm{d}y\mathrm{d}z = \oiint_{\partial M} (F \cdot \hat{n})\mathrm{d}A.$$

226 flashback intensifies

End of Course