Analysis II Notes

Pramana

Spring 2023

Special functions, power series, Fourier series, approximation, contraction principle, characterizations of compactness in metric spaces, applications to differential equations. Differential calculus in normed spaces, including implicit and inverse function theorems. Course is essential for graduate work in mathematics.

Professor: Andreas Seeger.

Contents

1. Uniform convergence and series

1.1. Definitions and tests for uniform convergence

January 25, 2023

I will be using $\mathbb{F}$ in these notes to denote $\mathbb{R}$ or $\mathbb{C}$

Definition 1.1 (Uniform convergence of functions)

Given a sequence of functions $(f_n)_{n\in\mathbb{N}}$, $f_n : X \to \mathbb{R}$ or $\mathbb{C}$ (can be generalized to $f : X \to Y$, where Y is a *metric space*). (f_n) is said to be **uniformly convergent** with limit f if for every $\varepsilon > 0$, there is an $N = N(\varepsilon)$ (dependent on ε) such that

$$|f_n(x) - f(x)| < \varepsilon, \quad \text{for all } n \geq N(\varepsilon), \quad x \in X.$$

Definition 1.2 (Alternative definition)

$f_n \to f$ uniformly if $\sup_{x\in X} |f_n(x) - f(x)| \to 0$ as $n \to \infty$.

Remark 1.1. The key difference between uniform convergence and pointwise convergence of a function is that N depends on ε *and* x. For example, consider $f_n : [0,1] \to \mathbb{R} : x \mapsto x^n$. Suppose we compare it to $f(x) \equiv 0$. By Definition 1.2,

$$\sup_{x\in[0,1)} |f_n(x) - f(x)| = \sup_{x\in[0,1)} x^n = 1,$$

so we do not have uniform convergence, but $f_n \to f$ pointwise.

Definition 1.3 (Uniform convergence for function series)

The series $\sum_{k=1}^{\infty} f_k(x)$ **converges uniformly** if the partial sums $s_n(x) := \sum_{k=1}^{n} f_k(x)$ converge uniformly.

Definition 1.4 (Uniformly Cauchy)

f_n is **uniformly Cauchy** if for every ε, there exists $N = N(\varepsilon)$ s.t.

$$m, n > N \implies |f_m(x) - f_n(x)| < \varepsilon, x \in X.$$

Theorem 1.2 (Uniformly Cauchy implies uniformly convergent)

Any real-valued sequence which is uniformly Cauchy is uniformly convergent.

Proof. Note that if f_n is uniformly Cauchy, then the sequence of numbers $f_n(x)$ is Cauchy for all $x \in X$. Thus, for every x, $\lim_{n\to\infty} f_n(x)$ exists. Call it $f(x)$. There exists $N = N\left(\frac{\varepsilon}{2}\right)$ s.t.

$$|f_m(x) - f_n(x)| < \frac{\varepsilon}{2}, \quad \text{for } m, n > N\left(\frac{\varepsilon}{2}\right), \quad x \in X$$

Then

$$\lim_{n\to\infty} s\,|f_m(x) - f_n(x)| = |f_m(x) - f(x)| \leq \frac{\varepsilon}{2} < \varepsilon.$$ □

The Cauchy criterion can check for convergence without having to calculate the limit. Due to this useful property, we generalize it for uniform convergence as well.

Example 1.3 – We can show

$$\sum_{k=1}^{\infty} \frac{\sin kx}{k^2 + |\cos kx|}$$

is uniformly convergent on $\mathbb{R}$ by Theorem 1.2. Consider the difference of partial sums where $m < n$

$$\begin{aligned} |s_n(x) - s_m(x)| &= \left| \sum_{k=m+1}^{n} f_k(x) \right| \\ &\leq \sum_{k=m+1}^{n} |f_k(x)| \\ &\leq \sum_{k=m+1}^{n} \frac{1}{k^2} \end{aligned}$$

To bound this sum, we may take several approaches:

1. Note that
$$\frac{1}{k^2} \leq \frac{1}{k(k-1)} = \frac{1}{k-1} - \frac{1}{k}.$$
Thus,
$$\sum_{k=m+1}^{n} \frac{1}{k^2} \leq \sum_{k=m+1}^{n} \frac{1}{k(k-1)} = \frac{1}{m} - \frac{1}{n} \to 0.$$
This is an example of the *comparison test*.
2. Since
$$\begin{aligned} \sum_{k=m+1}^{n} \frac{1}{k^2} &\leq \sum_{k=m+1}^{n} \int_{k-1}^{k} \frac{1}{x^2}\, \mathrm{d}x \\ &= \int_{m}^{n} \frac{1}{x^2}\, \mathrm{d}x. \end{aligned}$$
We compute
$$\int_{m}^{n} \frac{1}{x^2}\, \mathrm{d}x = \frac{1}{m} - \frac{1}{n} \to 0.$$
This is the *integral test*.
3. We can set an upper/lower bound by considering "chunks" of length 2^n:
$$\frac{1}{4} \cdot 2^{-\ell} \leq \sum_{k=2^\ell+1}^{2^{\ell+1}} \frac{1}{k^2} \leq 2^{-\ell}.$$
This is a **dyadic decomposition**.

January 27, 2023

Proposition 1.4 (Weierstrass M-test)
You can prove uniform convergence by the following test. If there exists a nonnegative sequence $(a_k)_{k\in\mathbb{N}} \geq 0$ such that $\sum_{k=1}^{\infty} a_k$ converges, and a sequence of functions $f_k \colon X \to \mathbb{F}$ such that $|f_k(x)| \leq a_k$, then $\sum_{k=1}^{\infty} f_k(x)$ converges uniformly on X.

Proof. Let $\varepsilon > 0$ be arbitrary. Let $N(\varepsilon) = N < m < n$ and $s_n(x) = \sum_{k=1}^{n} f_k(x)$. Then

$$\begin{aligned}|s_n(x) - s_m(x)| &= \left|\sum_{k=m+1}^{n} f_k(x)\right| \\ &\le \sum_{k=m+1}^{n} |f_k(x)| \\ &\le \sum_{k=m+1}^{n} a_k.\end{aligned}$$

Since $\sum_k a_k$ converges, it is Cauchy, and we can find N such that the sum is less than ε. □

This uses the underlying fact that if (b_n) is a sequence s.t. $\lim_{n\to\infty} b_{2n}$ exists, and $\lim_{n\to\infty} b_{2n+1} - b_{2n} = 0$, then $\lim_{n\to\infty} b_n = \lim_{n\to\infty} b_{2n}$.

Example 1.5 – Does

$$\sum_{k=1}^{\infty} \frac{(-1)^k}{k} x^k$$

converge uniformly? Proposition 1.4 fails here since $\sum_n \frac{1}{n}$ diverges.

Let $k = 2\ell - 1$ for $\ell \in \mathbb{N}$. How big is $|f_{2\ell-1}(x) - f_{2\ell}(x)| = \frac{x^{2\ell-1}}{2\ell-1} - \frac{x^{2\ell}}{2\ell} =: g_\ell(x)$? We take $g'_\ell(x)$:

$$g'_\ell(x) = x^{2\ell-2} - x^{2\ell-1} = x^{2\ell-2}(1-x).$$

The maximum is attained at $x = 1$, so

$$|f_{2\ell-1}(x) - f_{2\ell}(x)| \le \frac{1}{2\ell-1} - \frac{1}{2\ell} = \frac{1}{(2\ell-1)(2\ell)} = O(\ell^{-2}).$$

With this comparison, we can show $s_{2n}(x)$ is Cauchy. Moreover, $s_{2n+1} - s_{2n} = \frac{(-1)^{2n+1}}{2n+1} \to 0$ as $n \to \infty$. Since odd and even are both Cauchy, the entire series is Cauchy, and Theorem 1.2 finishes.

January 30, 2023

Remark 1.6 (Review of integration by parts)**.** Is

$$a_n = \int_1^n \frac{\sin t}{\sqrt{t}}\, \mathrm{d}t.$$

Cauchy? The integral $\int_1^n \frac{1}{\sqrt{t}}\, \mathrm{d}t$ does *not* converge as $n \to \infty$, so we have to use a different method. Consider $a_n - a_m$ s.t. $m < n$. Then

$$a_n - a_m = \int_m^n \frac{\sin t}{\sqrt{t}}\, \mathrm{d}t.$$

By integration by parts with $u(t) = -\cos t, v(t) = 1/\sqrt{t}$:

$$\int_m^n \frac{\sin t}{\sqrt{t}}\, \mathrm{d}t = -\frac{\cos n}{\sqrt{n}} + \frac{\cos m}{\sqrt{m}} - \int_m^n \sin t \left(\frac{-1}{2}\right) t^{-3/2}\, \mathrm{d}t.$$

Now we bound the new integral by

$$\left|\int_m^n \sin t \left(\frac{-1}{2}\right) t^{-3/2}\, \mathrm{d}t\right| \le \int_m^n t^{-3/2}\, \mathrm{d}t \le \frac{2}{\sqrt{m}}.$$

1.2. Applications of uniform convergence

Theorem 1.7

Let $f_n\colon X \to \mathbb{F}$, where X is a metric space. Assume

1. f_n converges uniformly on X to f,
2. f_n is continuous at $a \in X$.

Then f is also continuous at a.

Proof. We can compare

$$|f(x) - f_N(x) + f_N(x) - f_N(a) + f_N(a) - f(a)| \le |f(x) - f_N(x)| + |f_N(x) - f_N(a)| + |f_N(a) - f(a)|$$

The first and last terms can be bounded by $\frac{\varepsilon}{3}$ by uniform convergence. Since f_n is continuous at a, $\exists \delta$ s.t. $|f_N(x) - f_N(a)| < \frac{\varepsilon}{3}$ for all x s.t. $d(x, a) < \delta$. □

Theorem 1.8

Suppose that $f\colon [a, b] \to \mathbb{F}$ are continuous functions and $f_n \to f$ uniformly on $[a, b]$. Then

$$\int_a^b f_n(x)\,\mathrm{d}x \to \int_a^b f(x)\,\mathrm{d}x, \qquad n \to \infty$$

Proof. We have

$$\left|\int_a^b f_n(x)\,\mathrm{d}x - \int_a^b f(x)\,\mathrm{d}x\right| \le \left|\int_a^b f_n(x) - f(x)\,\mathrm{d}x\right| \le \int_a^b |f_n(x) - f(x)|\,\mathrm{d}x.$$

$M_n = \sup_{x\in[a,b]} |f_n(x) - f(x)| \to 0$ since $f_n \to f$ uniformly. Thus, we can bound the integral

$$\int_a^b M_n\,\mathrm{d}x \le M_n(b - a).$$ □

Theorem 1.9

Suppose that $f_n\colon [a, b] \to \mathbb{R}$ is a sequence of functions such that f_n is differentiable and f_n' is continuous for all n. Assume that $f_n' \to g$ uniformly on $[a, b]$ and there exists x_0 s.t. $f_n(x_0)$ converges. Then $f_n \to f$ uniformly, where the limit function is differentiable and $f' = g$.

Proof sketch. Use

$$f_n(x) = f_n(x_0) + \int_{x_0}^x f_n'(t)\,\mathrm{d}t \to c + \int_{x_0}^x g(t)\,\mathrm{d}t.$$

The limit has derivative $g(x)$ by the fundamental theorem of calculus. □

1.3. Taylor's theorem

February 1, 2023

Definition 1.5 (C^k function)
A function f is of **class** C^k if $f, f', \dots, f^{(k)}$ are continuous on a given interval.

The simplest form of Taylor's theorem is as follows: On some interval I we consider a C^1-function and $a \in I$. Then

$$\begin{aligned} f(a+h) &= f(a) + \int_a^{a+h} f'(t)\,\mathrm{d}t \\ &= f(a) + h\int_0^1 f'(a+sh)\,\mathrm{d}s. \end{aligned}$$

This change is made by substituting $t = a + sh$. We can now say

$$\min_{0\le s\le 1} f'(a+sh) \le f'(a+sh) \le \max_{0\le s\le 1} f'(a+sh),$$

$$\min_{0\le s\le 1} f'(a+sh) \le \int_0^1 f'(a+sh)\,\mathrm{d}s \le \max_{0\le s\le 1} f'(a+sh).$$

Applying the IVT to the function $s \mapsto f'(a+sh)$ tells us that

$$\int_0^1 f'(a+sh)\,\mathrm{d}s = f'(\xi),$$

where ξ is between a and $a+h$, or $\xi = a + s^* h$ such that $s^* \in [0,1]$. Thus,

$$f(a+h) = f(a) + \underbrace{f'(\xi)h}_{\text{remainder}}.$$

By rearranging, we find that it is just the mean value theorem. Now suppose we try and extend it. Let $f \in C^2$ on I. Use integration by parts with $u(s) = 1-s$ and $v(s) = f'(a+sh)$:

$$\int_0^1 f'(a+sh)\,\mathrm{d}s = -\int_0^1 \frac{\mathrm{d}}{\mathrm{d}s}(1-s)f'(a+sh)\,\mathrm{d}s = f'(a) + \int_0^1 (1-s)f''(a+sh)h\,\mathrm{d}s.$$

Thus,

$$f(a+h) = f(a) + f'(a)h + h^2\int_0^1 (1-s)f''(a+sh)\,\mathrm{d}s.$$

Motivated by the fact that $\int_0^1 1 - s\,\mathrm{d}s = \frac{1}{2}$.

The last integral is not quite an average, but if we rewrite to

$$\frac{h^2}{2}\int_0^1 2(1-s)f''(a+sh)\,\mathrm{d}s =: E_1(a,h).$$

We can bound

$$|E_1(a,h)| \le \frac{h^2}{2}\max\left|f''(a+sh)\right|.$$

Then

$$f(a+h) = f(a) + f'(a) + f''(\xi)\frac{h^2}{2}.$$

Theorem 1.10 (Taylor's/Maclaurin's theorem)
Assume that $f \in C^{k+1}$ on I. Let a be in the interior of I. For $a, a+h \in I$, we have the identity

$$f(a+h) = \sum_{j=0}^{k} \frac{h^j}{j!} f^{(j)}(a) + E_k f(a,h), \tag{1.1}$$

where the error term $E_k f(a,h)$ is given by

$$E_k f(a,h) = \frac{h^{k+1}}{(k+1)!} \int_0^1 (k+1)(1-s)^k f^{(k+1)}(a+sh)\,\mathrm{d}s.$$

We may also write, as before,

$$E_k f(a,h) = \frac{h^{k+1}}{(k+1)!} f^{(k+1)}(\xi), \quad \xi \text{ between } a \text{ and } a+h.$$

Proof. We induct on k. Assume it holds for some $k \in \{0,1,2,\dots\}$ and that $f \in C^{k+2}$. Then we use integration by parts by noting $0 - \frac{\mathrm{d}}{\mathrm{d}s}(1-s)^{k+1} = (k+1)(1-s)^k$.

$$\begin{aligned}\int_0^1 (k+1)(1-s)^k f^{(k+1)}(a+sh)\,\mathrm{d}s &= -(1-s)^{k+1} f^{(k+1)}(a+sh)\Big|_{s=0}^{1} \\ &\quad + \int_0^1 (1-s)^{k+1} f^{(k+2)}(a+sh)h\,\mathrm{d}s \\ &= f^{(k+1)}(a) + \int_0^1 (1-s)^{k+1} f^{(k+2)}(a+sh)h\,\mathrm{d}s\end{aligned}$$

Thus

$$f(a+h) = \sum_{j=0}^{k} \frac{h^j}{j!} f^{(j)}(a) + \frac{f^{(k+1)}(a)}{(k+1)!} h^{k+1} + E_{k+1}(a,h),$$

where

$$E_{k+1}(a,h) = \frac{h^{k+1+1}}{(k+1)!(k+2)} \int_0^1 (1-s)^{k+1}(k+2) f^{(k+2)}(a+sh)\,\mathrm{d}s.$$

Since

$$\int_0^1 (1-s)^{k+1}(k+2)\,\mathrm{d}s = 1,$$

this is a "weighted average" and

$$E_{k+1}(a,h) = \frac{h^{k+1+1}}{(k+1)!(k+2)} \int_0^1 (1-s)^{k+1}(k+2) f^{(k+2)}(a+sh)\,\mathrm{d}s = \frac{h^{k+2}}{(k+2)!} f^{(k+2)}(\xi).$$

□

1.3.1. Applications of Taylor's formula

February 3, 2023

Example 1.11 (Recentering polynomials) – Let $P = \sum_{j=0}^{n} c_j x^j$ be a polynomial of degree n. Suppose we wanted to write P in the form

$$P(x) = \sum_{j=0}^{n} b_j (x-a)^j.$$

A long-winded way to solve this would be to write $x^j = ((x-a)+a)^j$ and expand with the binomial theorem, but this takes a very long time.

Instead, by Theorem 1.10,

$$P(x) = \sum_{k=0}^{n} \frac{P^{(k)}(a)}{k!}(x-a)^k.$$

Where the error term is 0.

Example 1.12 (Exponential function) – Consider $f(x) = e^x \in C^\infty$, which is the unique function such that $f'(x) = f(x), f'(0) = 1$. To define e, we call it the unique number a that makes the integral $\int_1^a \frac{1}{t}\,\mathrm{d}t = 1$. With this, we can define $\log(x) := \int_1^x \frac{1}{t}\,\mathrm{d}t$. Clearly, $\log\colon (0,\infty) \to \mathbb{R}$. Then let

$$\exp = \log^{-1}\colon \mathbb{R} \to (0,\infty).$$

There are two important properties of $\exp$:

$$\exp(x+y) = \exp(x)\exp(y), \qquad \exp'(x) = \exp(x).$$

By Theorem 1.10,

$$\exp(x) = \sum_{k=0}^{n} \frac{x^k}{k!} + \frac{x^{n+1}}{(n+1)!}e^{\xi}, \quad \xi \text{ between } 0 \text{ and } x.$$

For the remainder, let $|x| \le b$. Then

$$\left|\frac{x^{n+1}}{(n+1)!}\right| \le \frac{b^{n+1}}{(n+1)!} \xrightarrow{n\to\infty} 0.$$

Thus, the remainder goes to 0 *uniformly* on $[-b,b]$.

To show that $e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$, take $\log\left(1+\frac{1}{n}\right)^n = \frac{\log(1+1/n)}{1/n}$. Then taking

$$\lim_{x\to 0} \frac{\log(1+x) - \log 1}{x} = \log'(1) = 1.$$

Proposition 1.13

e is not a rational number.

Proof. We know

$$e = \sum_{k=0}^{n} \frac{1}{k!} + \frac{e^{\xi}}{(k+1)!}$$

is true for all n. Suppose that e is rational. Then Let $e = \frac{p}{q}$, $q \geq 4$. Then

$$\begin{aligned} e &= \sum_{k=0}^{q} \frac{1}{k!} + \frac{e^{\xi}}{(q+1)!} \\ \underbrace{\frac{p}{q} q!}_{\in \mathbb{N}} &= q! \left(\sum_{k=0}^{q} \frac{1}{k!} + \frac{e^{\xi}}{(q+1)!} \right) \\ &= \sum_{k=0}^{q} \underbrace{\frac{q!}{k!}}_{\in \mathbb{N}} + \underbrace{\frac{1}{q+1} e^{\xi}}_{<1, >0}. \end{aligned}$$

We cannot write an integer as the sum of an integer and a number between 0 and 1. Contradiction. □

Remark 1.14. To extend e^x to the complex numbers, we can write

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}, \quad z \in \mathbb{C}.$$

We can extend cos and sin similarly. If $z = ib$ is a pure imaginary number, then

$$e^z = \sum_{n=0}^{\infty} \frac{(ib)^n}{n!} = \left(1 - \frac{b^2}{2!} + \cdots\right) + i\left(b - \frac{b^3}{3!} + \cdots\right) = \cos b + i \sin b.$$

1.4. Power series

February 6, 2023

Definition 1.6 (Power series)

Power series are series of the form

$$\sum_{n=0}^{\infty} a_n z^n, \quad z \in \mathbb{C}.$$

Remark 1.15. In the complex numbers, we can write a number in terms of polar coordinates:

$$z_1 = r(\cos\alpha + i\sin\alpha), \quad z_2 = R(\cos\beta + i\sin\beta).$$

We then have the rule

$$z_1 z_2 = rR(\cos(\alpha+\beta) + i\sin(\alpha+\beta)).$$

Therefore, if $z = re^{i\alpha}$ (which is the same as before), then

$$z^n = r^n e^{i(n\alpha)}.$$

Power series results in this section generalize to complex numbers as well.

Theorem 1.16 (Radius of convergence)

There exists a unique $R \in [0, \infty]$ s.t. $\sum_{n=0}^{\infty} a_n z^n$ converges for all z with $|z| < R$. The series is divergent for all z with $|z| > R$. R can be computed:

$$R = \frac{1}{\limsup_{n\to\infty} |a_n|^{\frac{1}{n}}}.$$

It is a harder question for whether it converges/diverges for $|z| = R$

Proof. Let R be defined as above. To show divergence for $|z| > R$, there is an $\varepsilon > 0$ s.t. $|z| > R(1+\varepsilon) \implies \frac{|z|}{R} > 1+\varepsilon$. Thus,

$$\limsup_{n\to\infty} |a_n|^{\frac{1}{n}} |z| > 1+\varepsilon.$$

Recalling the definition of $\limsup$, there is a subsequence a_{k_n} s.t. $|a_{k_n}|^{1/k_n} |z| \geq 1+\varepsilon$ for all $n \in \mathbb{N}$. Thus,

$$|a_{k_n}| |z|^{k_n} \geq (1+\varepsilon)^{k_n} \to \infty.$$

So $\sum_{n=0}^{\infty} a_n z^n$ diverges.

To show convergence for $|z| < R$,

$$\limsup_{n\to\infty} |a_n|^{\frac{1}{n}} \cdot |z| < 1-\varepsilon.$$

Thus, there is an N s.t. $n > N \implies |a_n|^{\frac{1}{n}} |z| < 1 - \frac{\varepsilon}{2}$. So

$$|a_n| |z^n| < \left(1 - \frac{\varepsilon}{2}\right)^n.$$

Use a comparison to a geometric series to show that it converges. □

Remark 1.17. For the series $\sum_n a_n (z-z_0)^n$, we can just shift the function by $z - z_0 \mapsto w$ and apply the previous theorems.

Remark 1.18. The above proof gives *uniform* convergence for $|z| \leq r$ where $r < R$.

1.4.1. Computing sums with power series

February 8, 2023 Consider the two identities

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \log 2, \tag{1.2}$$

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \frac{\pi}{4}. \tag{1.3}$$

To get the sum $\sum_n a_n$, consider the power series $\sum_n a_n x^n$ evaluated at $x = 1$.

To find Equation 1.3, consider the power series

$$\begin{aligned}
\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1} &= \sum_{k=0}^{\infty} (-1)^k \int_0^x t^{2k} \, \mathrm{d}t \\
&= \int_0^x \sum_{k=0}^{\infty} (-1)^k t^{2k} \, \mathrm{d}t \\
&= \int_0^x \sum_{k=0}^{\infty} (-t^2)^k \, \mathrm{d}t \\
&= \int_0^x \frac{1}{1+t^2} \, \mathrm{d}t \\
&= \arctan x.
\end{aligned}$$

Letting $x = 1$, $\arctan 1 = \frac{\pi}{4}$. Note that we did not justify the steps in this calculation. In particular switching the integral and the sum was done without justification.

Theorem 1.19
Given $[a, b]$ such that $f_n \colon [a, b] \to \mathbb{F}$ (f_n continuous) such that the series

$$\sum_{n=0}^{\infty} f_n(x)$$

converges uniformly on $[a, b]$. Then the sum is a continuous function $s(t)$, and for $x_0, x \in [a, b]$,

$$\int_{x_0}^{x} s(t)\,\mathrm{d}t = \int_{x_0}^{x} \sum_{n=0}^{\infty} f_n(t)\,\mathrm{d}t = \sum_{n=0}^{\infty} \int_{x_0}^{x} f_n(t)\,\mathrm{d}t$$

Apply the analogous theorem for sequences of functions to the partial sums of the series to prove the above theorem.

Consider

$$s(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1}$$

If this series converges uniformly on $[-1, 1]$, then the sum is a continuous function.

In our case, to show that the series converges uniformly on $[-1, 1]$, we use a previous trick by bounding $\left|\frac{x^{2k+1}}{2k+1} - \frac{x^{2k+3}}{2k+3}\right| = O(k^{-2})$. Then uniform convergence is shown.

$$s(x) = \arctan x, \qquad -1 < x < 1,$$

and continuous on $[-1, 1]$. So

$$s(1) = \lim_{x \to 1^-} s(x) = \lim_{x \to 1^-} \arctan x = \frac{\pi}{4}.$$

For Equation 1.2, consider $-\sum_{n=1}^{\infty} \frac{x^n}{n}$ at $x = -1$.

1.4.2. Differentiation of power series

Consider $\sum_{n=0}^{\infty} a_n x^n$ with radius of convergence $R > 0$. We already proved that the sum converges uniformly on $[-r, r]$ for $r < R$. We already know we can differentiate the power series term-by-term.

Theorem 1.20 (Differentiating a power seriess)
Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$. Then f is differentiable and

$$f'(x) = \sum_{n=1}^{\infty} a_n n x^{n-1}$$

Proof. We need to show that the radius of convergence is the same (so any smaller radius has uniform convergence). Let

$$R = \frac{1}{\limsup |a_n|^{\frac{1}{n}}}$$

Let $|x| < R(1-\varepsilon) \implies \frac{|x|}{R} < 1 - \varepsilon$. Then

$$|a_n|^{1/n} |x| \le (1-\varepsilon) \implies |a_n| |x|^n \le (1-\varepsilon)^n.$$

For the derivative,

$$(n+1)^{1/(n+1)} |a_n|^{1/(n+1)} |x| \le (1-\varepsilon).$$

For large n,

$$|a_{n+1}| |x|^n \le \frac{(1-\varepsilon)^{n+1}}{|x|}.$$

So this power series converges as well. □

In fact, all formally differentiated series satisfy

$$f^{(k)}(x) = \sum_{n=k}^{\infty} a_n x^{n-k} n(n-1)\cdots(n-k+1).$$

Thus, the kth derivative exists and is continuous for all k.

Corollary 1.21
A (convergent) power series is its own Taylor series.

Corollary 1.22
If $\sum_{k=0}^{\infty} a_k x^k = \sum_{k=0}^{\infty} b_k x^k$ on $(-R, R)$, then $a_k = b_k$ for all k.

1.4.3. Abel's theorem

February 10, 2023

Look at power series $\sum_{k=0}^{\infty} a_k x^k$. To evaluate at 1, we may consider

$$\lim_{r\to 1} \sum_{k=0}^{\infty} a_k r^k.$$

If $\sum_n a_n r^n$ converges uniformly for $r \in [0,1]$, then we will show that the limit agrees with the actual sum.

Theorem 1.23 (Abel's theorem)
If $\sum_{k=0}^{\infty} a_k r^n$ converges at $r = 1$, then

$$\sum_{k=0}^{\infty} a_k = \lim_{r\to 1^-} \sum_{k=0}^{\infty} a_k r^k.$$

Proof. Write the partial sums as $s_n := \sum_{k=0}^{n} a_k$. We may write $a_n = s_n - s_{n-1}$ and $a_0 = s_0$. By assumption, $s = \lim_{n\to\infty} s_n$. Using the above notation, we can write a_k in terms of the partial sums:

$$\begin{aligned}\sum_{k=0}^{\infty} a_k r^k &= s_0 + \sum_{k=1}^{\infty}(s_k - s_{k-1})r^k \\ &= s_0 + (s_1 - s_0)r + (s_2 - s_1)r^2 + \cdots \\ &= s_0(1-r) + s_1(r - r^2) + s_2(r^2 - r^3) + \cdots \\ &= (1-r)\left(\sum_{k=0}^{\infty} s_n r^n\right) \\ &= (1-r)\left(\sum_{k=1}^{\infty}(s_n - s)r^n\right) + (1-r)\left(\sum_{k=1}^{\infty} s r^n\right).\end{aligned}$$

The second term is just s, so we will show the first term goes to 0.

Fix $\varepsilon > 0$. We find $N(\varepsilon)$ s.t. $|s_n - s| < \frac{\varepsilon}{C}$ for all $n > N$. For all $r \in [0,1)$,

$$\begin{aligned}\left|(1-r)\left(\sum_{k=N}^{\infty}(s_n - s)r^k\right)\right| &\leq \sum_{k=N}^{\infty} \frac{\varepsilon}{C} r^k (1-r) \\ &< \frac{\varepsilon}{C}.\end{aligned}$$

For the rest of the sum, consider what happens as $r \to 1$:

$$\underbrace{(1-r)}_{\to 0} \underbrace{\sum_{n=0}^{N-1}(s_n - s)r^n}_{\text{finite}}.$$

So there is an $r_0 < 1$ s.t.

$$\left|(1-r)\sum_{n=0}^{N-1}(s_n - s)r^n\right| < \frac{\varepsilon}{2}$$

for $r_0 < r < 1$. Combining these two sums and letting $C = 3$,

$$\left|(1-r)\left(\sum_{k=1}^{\infty}(s_n - s)r^n\right)\right| < \varepsilon.$$

□

1.4.4. Abel summation

In the first part of Theorem 1.23, we used a summation technique worth noting. Abel summation is also known as *summation by parts* because of its similarities to integration by parts.

Theorem 1.24 (Summation by parts)
Given (a_n) and (b_n), let $A_n := \sum_{k=0}^{n} a_k$. Then for $p < q$,

$$\sum_{n=p}^{q} a_n b_n = A_q b_q - A_{p-1} b_p + \sum_{n=p}^{q} A_n (b_n - b_{n+1}).$$

In the case $p = 0$, define $A_{-1} = 0$.

Proof.

$$\begin{aligned}\sum_{n=p}^{q} a_n b_n &= \sum_{n=p}^{q}(A_n - A_{n-1})b_n \\ &= \sum_{n=p}^{q} A_n b_n - \sum_{n=p}^{q} A_{n-1} b_n \\ &= \sum_{n=p}^{q} A_n b_n - \sum_{n=p-1}^{q-1} A_n b_{n+1} \\ &= A_q b_q + \sum_{n=p}^{q-1} A_n (b_n - b_{n-1}) - A_{p-1} b_p.\end{aligned}$$

□

Example 1.25 (Application of Abel summations) – For what values of α does $\sum_{k=1}^{\infty} \frac{\sin kx}{k^a}$ converge? Similarly, does $\sum_{k=1}^{\infty} \frac{\cos kx}{k^a}$ converge? We first note

We may create bounds by considering $\sin \frac{x}{2} \sum_{k=1}^{N} \sin kx$, and finding cancellation. However, this is not intuitive, so we won't use it.

- There are good bounds for $\sum_{k=1}^{N} \sin kx$. Write

$$\begin{aligned}\sum_{k=1}^{N} \sin kx &= \operatorname{Im}\left(\sum_{k=0}^{N} e^{ikx}\right) \\ &= \operatorname{Im}\left(\sum_{k=0}^{N} z^k\right)\Big|_{z=e^{ix}} \\ &= \operatorname{Im}\left(\frac{1-z^{N+1}}{1-z}\right) \\ &\le \operatorname{Im}\left(\frac{|1-z^{n+1}|}{|1-z|}\right) \\ &\le \frac{2}{|1-z|}.\end{aligned}$$

  $z = 1 \iff \cos x = 1, \sin x = 0$ only when x is an integer multiple of 2π. These values will not work.

- $$\begin{aligned}|b_k - b_{k+1}| &= \left|\frac{1}{(k+1)^a} - \frac{1}{k^a}\right| \\ &= \left|\frac{k^a - (k+1)^a}{k^a (k+1)^a}\right|.\end{aligned}$$

  To bound this, let $f(x) = x^a$. Then $f(k+1) - f(k) = f'(\xi)$ such that ξ is between k and $k+1$. Then $f'(\xi)$ is between ak^{a-1} and $a(k+1)^{a-1}$. Thus,

$$\begin{aligned}|b_k - b_{k+1}| &\le C\frac{k^{a-1}}{k^{2a}} \\ &\le Ck^{-a-1}\end{aligned}$$

$b_k \to 0$ as $k \to \infty$. $|A_q - b_q|$ and $|A_{p-1} b_q|$ both go to 0 as p and $q \to \infty$. We can show convergence by using Abel sums.

February 13, 2023

Remark 1.26. In the homework, we showed that

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k} - \log n$$

converges. The value that it converges to is called γ, the *Euler-Mascheroni* constant.

Problem 1.1 (Open problem)
Is γ irrational?

2. Metric spaces

2.1. Norms

If we have a norm, we induce a metric: $d(x, y) = \|x - y\|$. However, mectrics do not induce norms, because norms are only defined on *vector spaces*, and there are plenty of metric spaces that are not vector spaces.

Example 2.1 (Infinite-dimensional normed spaces) – Metrics behave differently in infinite-dimensional spaces:

- **Function metric spaces:** If X is a set, we let $B_{\mathbb{R}}(x)$ be the set of bounded real-valued functions. Define a norm $\|f\| := \sup_{x\in X} |f(x)|$. This norm induces the metric defined as $d(f, g) = \sup_{x\in X} |f(x) - g(x)|$.
- $\ell^k(\mathbb{N})$ **spaces:** The space ℓ^1 (of $\mathbb{N}$) is the set of functions on $\mathbb{N}$ a.k.a. sequences such that $\sum_{n=1}^{\infty} |a_n|$ converges. $\ell^\infty(\mathbb{N})$ is the set of series (a_n) such that $\sup_{n\in\mathbb{N}} |a_n|$ is finite.

2.2. Compactness

2.2.1. Balls

February 15, 2023

Given a metric space (X, d), and $Y \subseteq X$, Y is a metric space with the metric $d|_{Y\times Y}$.

Definition 2.1 (Open ball)
The **open ball** in X, centered at $y \in X$ is denoted

$$B_X(y, r) = \{x \in X \mid d(x, y) < r\}.$$

We note that with X and Y defined as before, $B_Y(y, r) = B_X(y, r) \cap Y$.

Theorem 2.2 (Condition for open set of $(Y, d|_{Y\times Y}) \subset (X, d)$)
Let $Y \subseteq X$, with the metric inherited from X. Then $U \subseteq Y$ is open in $Y \iff$ there is an open set (w.r.t. d_X metric) $O \subseteq X$ such that $O \cap Y = U$.

Proof. ($\impliedby$) Let O be open in X. Let $w \in O \cap Y$. Since w is an *interior point* of O, there is $B_X(w, r_w) \subseteq O$. Thus,

$$\underbrace{B_X(w, r_w) \cap Y}_{B_Y(w, r_w)} \subset O \cap Y.$$

So w is an interior point of $O \cap Y$ (w.r.t. Y's metric).

($\implies$) Assume that U is open in Y, so w is an interior point of U (w.r.t. Y's metric), and $B_Y(w, r_w) \subseteq U$. We define

We note $\bigcup_{w\in U} B_Y(w, r_w) = U$.

$$O := \bigcup_{w\in U} B_X(w, r_w),$$

which is an open set in X. To show this satisfies the conditions, we note

$$O \cap Y = \bigcup_{w\in U} (B_X(w, r_w) \cap Y) = \bigcup_{w\in U} B_Y(w, r_w) = U.$$

□

2.2.2. Notions of compactness

Definition

We may examine compactness in different contexts:

Definition 2.2 (Compact metric space)

A metric space (X, d) is **compact** if the following holds: whenever X can be written as the union of open sets: $\bigcup_{\alpha \in A} O_x$ there is a finite number of indices $\alpha_1, \ldots, \alpha_n$ whose corresponding Os union to X.

Definition 2.3 (Compact subset)

Given a metric space X and $K \subseteq X$, K is **compact** if every cover of K with open sets has a **finite subcover**.

These definitions are compatible with each other. This means that a compact subset K is a compact metric space with K as the metric. Thus, the context of X does not matter. We can prove this by using Theorem 2.2.

Remark 2.3 (Heine-Borel in $\mathbb{R}^n$ does not hold in all spaces)**.** The Heine-Borel theorem requires that the set be *bounded* and *closed*. In a normed space, a set Y is *bounded* if $\|y\| \leq C$ for all $y \in Y$. This is not "topologially invariant", so we may have issues past $\mathbb{R}^n$. Consider $X = \ell^\infty(\mathbb{N})$, where $\|f\|_\infty := \sup |f(n)|$. Define

$$e^k(n) := \begin{cases} 1 & \text{if } k = n, \\ 0 & \text{if } k \neq n. \end{cases}$$

Note that

$$d(e^k, e^\ell) = \left\|e^k - e^\ell\right\|_\infty = 1$$

for all k, ℓ. Let Y be the set of sequences e^k for all $k \in \mathbb{N}$. Consider all balls $B(e^k, 4^{-1})$. Any intersection is empty. Y is not compact because removing any ball does not cover X. However, the set is bounded by 1, and is closed because all points are isolated.

2.2.3. Three useful classifications of compactness

February 17, 2023

Definition 2.4 (Sequentially compact)

A metric space X is **sequentially compact** or has the *Bolzano-Weierstrass property* if every sequence in X has a convergent subsequence.

For notation, let $(a_n) \subseteq (b_n)$ mean (a_n) is a subsequence of (b_n).

Does there exist a sequence (d_n) which is a subsequence of $(a_n^{(0)})$ and for all but the first ℓ entries are part of $(a_n^{(\ell)})$ (for $(a_n^{(0)}) \supseteq (a_n^{(1)}) \supseteq (a_n^{(2)}) \supseteq (a_n^{(3)}) \supseteq \cdots$)? Yes, by letting $d_n = a_n^{(n)}$. This is called the *Cantor diagonal subsequence*. We will use this idea in the proof of Theorem 2.4

Definition 2.5 (Totally bounded)

A metric space X is **totally bounded** if for every $\varepsilon > 0$, X is a union of finitely many balls of radius ε.

Totally bounded is a stronger condition than bounded.

Theorem 2.4 (Equivalence of notions of compactness)
The following are equivalent for a given metric space X:

1. X is compact.
2. X is sequentially compact.
3. X is totally bounded and *complete*.

A metric space is **complete** if every Cauchy sequence converges.

Proof. (1) $\implies$ (2). Suppose X compact and there is a sequence (a_n) in X which does not have a convergent subsequence. Note that (a_n) does not have a term that repeats an infinite amount of times.

Consider A, the set of a_n's that are distinct. A does not have any *accumulation points* (limit points), and therefore A is closed. Thus, for every $x_\alpha \in A$, there exists $\varepsilon(x_\alpha) = \varepsilon > 0$ s.t.

$$B(x_\alpha, \varepsilon) \cap A = \{x_\alpha\}.$$

Moreover, $X \setminus A$ is open. Consider the collection of open sets, $X \setminus A$ (unless X is A), and all the $B(x_\alpha, \varepsilon)$'s. This is an open cover, but we cannot remove any of them, so there is no finite subcover.

(2) $\implies$ (3). Let (x_n) be Cauchy. By the definition of sequentially compact, it has a convergent subsequence, so it is convergent. Thus, X is complete.

Exercise. Show that a Cauchy sequence with a convergent subsequence is convergent.

Suppose that X is not totally bounded, then there exists an $\varepsilon > 0$ s.t. X is not a finite union of ε-balls. Pick a ball $B(p_1, \varepsilon)$, and there is p_2 so that $d(p_1, p_2) \geq \varepsilon$. Repeat this construction until we have $p_1, \ldots, p_\ell$ so that $d(p_i, p_j) \geq \varepsilon$ if $i \neq j$. The ε-balls at these points do not cover X, so there is $p_{\ell+1}$ that has the same properties. The sequence (p_n) has no subsequence that is Cauchy, so it fails to be sequentially compact.

(3) $\implies$ (2). Assume that X is totally bounded. We will prove that every sequence that has a convergence subsequence which is Cauchy.

Let (x_n) be an arbitrary sequence. Cover X with finitely many 1-balls. At least one of B_1 contains x_n for infinitely many n. Thus, there is a subsequence $(x_n^{(1)})$ so that all values lie in B_1.

There is a subsequence $(x_n^{(2)}) \subseteq (x_n^{(1)})$ which belongs to a ball $B_{1/2}$.

In general, for all ℓ, there is a subsequence $(x_n^{(\ell+1)}) \subseteq (x_n^{(\ell)})$, which belongs to B_ℓ, and $B_\ell \cap B_{\ell-1} \neq \varnothing$. Choose the diagonal sequence, $d_n := x_n^{(n)}$ is a subsequence of (x_n). For $n, m > \ell$,

$$d(d_n, d_m) \leq 2^{-\ell+1}.$$

(2), (3) $\implies$ (1). Assume that X is sequentially compact (equivalently totally bounded and complete) then every collection $(O_\alpha)_{\alpha \in A}$ with open sets has a finite subcover.

Claim 2.1. There exists an $\varepsilon > 0$ s.t. all balls of radius ε are contained in at least one of the O_α.

Proof. Suppose the claim is false. Then for all $\varepsilon_n = \frac{1}{2^n}$, we find a ball $B(p_n, \frac{1}{2^n})$ which is not contained in any of the O_α's. But (p_n), by assumption, has a convergent subsequence. Let that be $p_{\kappa(n)}$, and suppose $p_{\kappa(n)} \to p$. There is an index α_0 so that $p \in O_{\alpha_0}$. Thus, p is an interior point of $p \in O_{\alpha_0}$.

There is an ε_j such that

$$B(p, \varepsilon_j) \subseteq O_\alpha.$$

Choose a suitably large n such that

$$d(p_{\kappa(n)}, p) < \frac{\varepsilon_j}{2}, \qquad \frac{1}{2^{\kappa(n)}} < \frac{\varepsilon_j}{2}.$$

For every $y \in B(p_{\kappa(n)}, \frac{1}{2^{\kappa(n)}})$ has $d(y,p) < \varepsilon$. But the n$B(p_{\kappa(n)}, \frac{1}{2^{\kappa(n)}})$ are contained in O_α for the large n. ■

Since X is totally bounded, we can find a finite cover $B(p_1, \varepsilon), \ldots, B(p_N, \varepsilon)$ that are contained in finitely many O_α, showing compactness. □

2.3. Office hour tangent: O notation and similar

February 20, 2023

Definition 2.6 (O notation)

Given the sequences (a_n) and (b_n), we write

- $a_n = O(b_n)$ if there exists a $C > 0$ such that $|a_n| \le C\,|b_n|$,
- $a_n = o(b_n)$ if $\frac{a_n}{b_n} \to 0$ as $n \to \infty$,
- $a_n \sim b_n$ if $\frac{a_n}{b_n} \to L$ for some $L \in \mathbb{R}$ as $n \to \infty$.

2.4. Results from compactness

February 22, 2023

Theorem 2.5 (Properties of continuous functions on compact metric spaces)

Let X be a *compact* metric space and $f : X \to \mathbb{R}$ be continuous. Then

1. f is bounded,
2. f attains its maximum and minimum,
3. f is uniformly continuous.

For every $\varepsilon > 0$ there is a δ s.t. $|f(x) - f(y)| < \varepsilon$ if $f(x, y) < \delta(\varepsilon)$

Proof. (1) By continuity, for every $x \in X$, there is a ball $B(x, \varepsilon_x)$ s.t. for all w in the ball, $|f(w) - f(x)| < 1$. All the balls form an open cover, so we may choose a finite subcover:

$$B(x_1, \varepsilon_{x_1}), \ldots, B(x_N, \varepsilon_{x_N}).$$

Fix $w \in X$. $w \in B(x_i, \varepsilon_{x_i})$. Then

$$\begin{aligned} |f(w)| &= |f(x_i)| + |f(w) - f(x_i)| \\ &\le \max_j |f(x_j)| + 1. \end{aligned}$$

(2) Consider a sequence (x_n) s.t. $f(x_n) > \sup f - \frac{1}{n}$. By sequential compactness, there is a subsequence $(x_{\kappa(n)})$ which converges to x_0. Since f is continuous, $f(x_{\kappa(n)}) \to$

$f(x_0)$. Thus,

$$f(x_\kappa(n)) \geq \sup f - \frac{1}{\kappa(n)} \xrightarrow{n\to\infty} f(x_0) \geq \sup f \implies f(x_0) = \sup f.$$

Use a similar argument for $\inf$.

(3) Pick $\varepsilon > 0$. For every $x \in X$, there is a δ_x s.t. $|f(w) - f(x)| < \frac{\varepsilon}{2}$ if $d(w,x) < \delta_x$. The balls $B(x, \delta_x/2)$ cover X. By compactness, we choose a finite subcover $B(x_i, \delta_{x_i}/2)$, $i = 1, \dots, N$. Let $\delta := \min_i \frac{\delta_{x_i}}{2}$. Pick w_1, w_2 such that $d(w_1, w_2) < \delta$. Since this is a covering, $w_1 \in B(x_i, \delta_{x_i}/2)$ for some i. One can show

$$w_2 \in B(x_i, \delta_x/2 + \delta) \subseteq B(x_i, \delta_x).$$

Looking at f,

$$\begin{aligned} |f(w_1) - f(w_2)| &\leq |f(w_1) - f(x_i)| + |f(x_i) - f(w_2)| \\ &< \varepsilon. \end{aligned}$$

□

We will now greatly simplify the proof of the Heine-Borel theorem using the equivalent notions of compactness.

Theorem 2.6 (Heine-Borel theorem)
If $E \subseteq \mathbb{R}^n$ is bounded and closed, then it is compact.

Proof. Since $\mathbb{R}^n$ is complete, E (a closed subset) is as well. E bounded means $E \subseteq [-R, R]^n$. Choose $a < \frac{\varepsilon}{2\sqrt{n}}$. Consider the balls centered at

$$(k_1 a, \dots, k_n a), \quad |k_i a| < R + 1, \quad k_i \in \mathbb{Z}.$$

□

February 24, 2023

Example 2.7 (Examples of compactness where Heine-Borel does not apply) – Let c_0 be the vector space of real-valued sequences that converge to 0. Let $a := (a_n)_{n=1}^\infty$. Let

$$\|a\|_\infty = \sup_{n\in\mathbb{N}} |a_n|.$$

If $\left\|a^{(n)} - a\right\|_\infty \to 0$. This gives a notion of uniform convergence (if we think of sequences as functions from $\mathbb{N} \to \mathbb{R}$). Given $\varepsilon > 0$, find an $N = N(\varepsilon)$ such that

$$\left|a_j - a_j^{(n)}\right| \leq \frac{\varepsilon}{2}, \quad \forall j, n \geq N.$$

Similarly, we already know $|a_j^{(n)}| < \frac{\varepsilon}{2}$ for suitably large n.

$$\begin{aligned} |a_j| &\leq |a_j^{(n)}| + |a_j - a_j^{(n)}| \\ &< \varepsilon \end{aligned}$$

Thus c_0 is a closed subspace of ℓ^∞.

Lemma
A subset K of c_0 is totally bounded if (and only if) $\lim_{j\to\infty} \|(a_n)\| = 0$.

Definition 2.7 (Relatively compact)
A metric space X is **relatively compact** if the closure of X is compact.

Proposition 2.8
A subset $U \subseteq X$ of a complete metric space that is totally bounded is relatively compact.

The proof of the above statement follows from Theorem 2.4, since closed subsets of complete metric spaces are also complete, and the closure of a totally bounded set is totally bounded as well.

February 27, 2023

Lemma 2.9
Suppose $\mathscr{A} \subseteq X$ and for every $\varepsilon > 0$, there is a finite number of balls

$$B(\xi, \varepsilon), \quad \xi \in \mathcal{X} \subseteq X, \quad |\mathcal{X}| < \infty$$

such that

$$\mathscr{A} \subseteq \bigcup_{\xi \in \mathcal{X}} B(\xi, \varepsilon).$$

Then $\mathscr{A}$ is totally bounded.

Remark 2.10. The difference between this and the definition of totally bounded is that the ξ's may belong to X, which may be a larger space than $\mathscr{A}$. This lemma tells us that it doesn't matter if we consider centers in the larger metric space.

2.4.1. Application: Characterizing totally bounded function classes

Definition 2.8 (Pointwise and uniformly bounded)
A class $\mathscr{A}$ of functions defined on X is **pointwise bounded** if for every x there is an $M_x < \infty$ such that

$$\sup_{f \in \mathscr{A}} |f(x)| \leq M_x.$$

$\mathscr{A}$ is **uniformly bounded** if

$$\sup_{x \in X} \sup_{f \in \mathscr{A}} |f(x)| \leq M < \infty.$$

Definition 2.9
A class $\mathscr{A} \subseteq C(K)$ is **equicontinuous** if for every $\varepsilon > 0$, there exists δ s.t. $d(x, y) < \delta$ implies $|f(x) - f(y)| < \varepsilon$ for all $f \in \mathscr{A}$.

Theorem 2.11 (Arzelà-Ascoli)
Consider $C(K)$ on a compact set K equipped with the sup norm:

$$\|f\|_\infty := \sup_{x\in K} |f(x)|.$$

$\mathscr{A} \subseteq C(K)$ is totally bounded if and only if it is pointwise bounded and equicontinuous.

If $\mathscr{A} \subseteq C(K)$ is equicontinuous, then $\mathscr{A}$ is pointwise bounded $\iff$ it is uniformly bounded.

Proof. ($\implies$) Let Ξ be a collection of N balls $B(f_j, \frac{\varepsilon}{3})$, $1 \le j \le N$. $f_1, \ldots, f_N$ are continuous on K, a compact set. Thus, there is an M such that

$$\sup_{i=1,\ldots,N} |f_i(x)| \le M.$$

For arbitrary $f \in \mathscr{A}$, there is an index i such that

$$\|f - f_i\|_\infty \le \frac{\varepsilon}{3}.$$

Thus,

$$|f(x)| \le |f_i(x)| + |f(x) - f_i(x)| \le M + \frac{\varepsilon}{3}.$$

This implies $\mathscr{A}$ is uniformly bounded.

For each i there is a δ_i s.t. $d(x,y) < \delta_i \implies |f_i(x) - f_i(y)| < \frac{\varepsilon}{3}$. Let $\delta = \min_i \delta_i$. We can also find i such that

$$|f(t) - f_i(t)| < \frac{\varepsilon}{3}, \quad \forall t \in K,$$

by uniform continuity (continuity on a compact set). Thus,

$$\begin{aligned} |f(x) - f(y)| &\le |f(x) - f_i(x) + f_i(x) - f_i(y) + f_i(y) - f(y)| \\ &\le |f(x) - f_i(x)| + |f_i(x) - f_i(y)| + |f_i(y) - f(y)| \\ &< \varepsilon. \end{aligned}$$

March 1, 2023

($\impliedby$) Since K is compact and $\mathscr{A}$ is equicontinuous, choose a covering of δ-balls

$$B(x_1, \delta), \ldots, B(x_n, \delta),$$

and f changes by less than $\frac{\varepsilon}{3}$ in any of these balls. Ideally, we want to define a set of *constant* functions that give a covering, but overlaps may cause us to define a function twice on a point. To prevent this, let

$$E_1 = B(x_1, \delta), \quad E_2 = B(x_2, \delta) \setminus B(x_1, \delta), \ldots,$$

$$E_j := B(x_j, \delta) \setminus \left(\bigcup_{k=1}^{j-1} B(x_k, \delta) \right).$$

Each E_j has diameter $\le \frac{\delta}{2}$. Let $\mathscr{C}$ be the set of functions which are (1) constant on each E_i (2) takes values of the form $k\frac{\varepsilon}{3}$, where $\left|k\frac{\varepsilon}{2}\right| \le M, k \in \mathbb{Z}$. These conditions mean that $\mathscr{C}$ is finite.

To show this covers, consider $f \in \mathscr{A}$. Consider $f|_{E_i}$. For any point $x \in E_i$ has distance δ to some x_i. There is also some $k\frac{\varepsilon}{3}$ s.t. $\left|f(x_i) - k\frac{\varepsilon}{3}\right| < \frac{\varepsilon}{3}$. ☐

2.5. Banach's fixed point theorem

March 6, 2023 We begin with a motivating problem: Find $\mathbf{x} = (x_1, x_2)$ such that

$$10x_1 + \sin(x_1 + 2x_2) = 10000,$$

$$\cos(x_1 + x_2) + 5x_2 = 100.$$

Finding the solution to this would be hard, but showing that it has a solution may be easier.

Definition 2.10 (Contraction)
If (X, d) is a metric space, and $T\colon X \to X$ is called a **contraction** if there exists an $\alpha < 1$ such that

$$d(T(x), T(y)) \leq \alpha \cdot d(x, y)$$

Theorem 2.12 (Banach fixed point theorem)
If (X, d) is *complete*, then all contractions have a *unique* fixed point, which exists.

Proof. (Uniqueness) Suppose $x, y \in X$ s.t. $T(x) = x$, $T(y) = y$. Then

$$d(x, y) = d(T(x), T(y)) \leq \alpha \cdot d(x, y).$$

Thus, $d(x, y) = 0$.

(Existence) We use "the method of successive approximation". We construct a sequence. $x_0 \in X$. Define $x_n = T(x_{n-1})$. I claim this is Cauchy. We find

$$d(x_{n+1}, x_n) \leq \alpha^n \cdot d(x_1, x_0).$$

Let $m > n$. Then

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m-1}) + \cdots + d(x_{n+1}, x_n) \\ &\leq \alpha^n (1 + \alpha + \cdots + \alpha^{m-1-n}) \cdot d(x_1, x_0) \\ &\leq \frac{\alpha^n}{1 - \alpha} d(x_1, x_0). \end{aligned}$$

This shows (x_n) is Cauchy. Since X is complete, $x_n \to x \in X$. Also, $x_{n+1} = T(x_n) \to x$. Thus, since T is continuous, $T(x_n) \to T(x)$. □

Example 2.13 (Integral equation with a fixed point) – Find $f \in C[0,1]$ s.t.

$$f(x) = \int_0^1 K(x,t)f(t)\,\mathrm{d}t + g(x).$$

Assume that $K \in C[0,1]^2$ $|K(x,t)| \le \alpha < 1$. We may define an **operator** T s.t.

$$Tf(x) := \int_0^1 K(x,t)f(t)\,\mathrm{d}t + g(x).$$

Then

$$\begin{aligned} d(Tf_1, Tf_2) &= \sup_x |Tf_1(x) - Tf_2(x)| \\ &\le \int_0^1 |K(x,t)| \cdot |f_1(t) - f_2(t)|\,\mathrm{d}t \\ &\le \alpha \cdot d(f_1, f_2). \end{aligned}$$

2.6. Differential equations

2.6.1. Initial value problems

March 9, 2023 Give the system

$$\begin{aligned} y(x_0) &= y_0 \\ y'(x) &= F(x, y(x)) \end{aligned} \tag{2.1}$$

For a function $(x,y) \mapsto F(x,y)$ which is continuous near x_0, y_0. This is an **initial value problem**: can we find $x \mapsto y(x)$ near x_0, such that y solves the above conditions? How many solutions are there?

Example 2.14 (Multiple solutions example) – Consider

$$y'(x) = \sqrt{|y(x)|}, \quad y(0) = 0$$

There are multiple solutions to this:

- $y(x) \equiv 0$.
- $y(x) = \begin{cases} \frac{x^2}{4} & x > 0, \\ 0 & x = 0. \end{cases}$

Theorem 2.15 (Peano)
For an IVP of the form in Equation 2.1, if F is continuous near (x_0, y_0), then it has a solution.

We will not prove this theorem right now.

Theorem 2.16 (Picard-Lindelöf)
If in addition, for F continuous on

$$R := \{(x, y) : |x - x_0| \leq a, |y - y_0| \leq b\},$$

and

$$|F(x, y) - F(x, \widetilde{y})| \leq C|y - \widetilde{y}|,$$

such that

$$\sup_{x,y \in R} |F(x, y)| = M < \infty.$$

For $x \in [x_0 - \delta, x_0 + \delta]$, such that

We will prove that δ has a better bound later.

$$\delta < \min\left\{a, \frac{b}{M}, \frac{1}{C}\right\},$$

then the solution guaranteed by Theorem 2.15 is *unique*.

Lemma 2.17
For a continuous function $x \mapsto y(x)$ defined on $[x_0 - \delta, x_0 + \delta]$ with values in $[y_0 - b, y_0 + b]$, the following two conditions are equivalent:

1. y is of class C^1 on the interval, and satisfies our initial value problem.
2. y solves the *integral eqation*

$$y(x) = x_0 + \int_{x_0}^{x} F(t, y(t))\, \mathrm{d}t.$$

Proof. ($\Longrightarrow$)

$$y'(t) = F(t, y(t)), \qquad t \in [x_0 - \delta, x_0 + \delta].$$

Thus,

$$\int_{x_0}^{x} y'(t)\, \mathrm{d}t = \int_{x_0}^{x} F(t, y(t))\, \mathrm{d}t$$
$$y(x) - y(x_0) = \int_{x_0}^{x} F(t, y(t))\, \mathrm{d}t.$$

Use the substitution $y(x_0) = y_0$.
($\Longleftarrow$) $t \mapsto (t, y(t))$ is continuous. $(t, y) \mapsto F(t, y)$ is continuous. Thus, $t \mapsto F(t, y(t))$ is continuous. By the fundamental theorem of calculus,

$$x \mapsto \int_{x_0}^{x} F(t, y(t))\, \mathrm{d}t$$

is C^1 and the derivative is $F(x, y(x))$. This yields the IVP form. □

Proof of Picard-Lindelöf. Assume that F is defined and continuous on

$$R = \{(x, y) : x_0 - a \leq x \leq x_0 + a, y_0 - b \leq y \leq y_0 + b\}.$$

We want to apply Theorem 2.12. We define the operator

$$Ty(x) \coloneqq y_0 + \int_{x_0}^{x} F(t, y(t))\, \mathrm{d}t.$$

March 10, 2023

Thus, the integral equation is $y = Ty$. Our metric space will be the set of all continuous functions on $[x_0 - \delta, x_0 + \delta]$ which have values in $[y_0 - b, y_0 + b]$. Let $\mathcal{M}$ be this metric space, and equip it with the $\sup$ norm.

Claim 2.2. $T \colon \mathcal{M} \to \mathcal{M}$.

Let $y \in \mathcal{M}$. Consider

$$\begin{aligned} |Ty(x) - y_0| &= \left| \int_{x_0}^{x} F(t, y(t))\, \mathrm{d}t \right| \\ &\le M|x - x_0| \\ &\le M\delta. \end{aligned}$$

By requiring $\delta < \frac{b}{M}$, we are done. ■

Claim 2.3. $\mathcal{M}$ is complete.

Let

$$\mathcal{M}_x \coloneqq \{ f \in C[x_0 - \delta, x_0 + \delta] : |f(x) - y_0| \le \delta \}.$$

Each of these sets are clearly closed. Since

$$\mathcal{M} = \bigcap_{x \in [x_0 - \delta, x_0 + \delta]} \mathcal{M}_x,$$

$\mathcal{M}$ is closed, and therefore it is complete. ■

Claim 2.4. T is a contraction.

We will bound T to shown contraction.

$$\begin{aligned} d(Ty_1, Ty_2) &= \sup_{|x - x_0| < \delta} |Ty_1(x) - Ty_2(x)| \\ &= \sup_{|x - x_0| < \delta} \left| \int_{x_0}^{x} F(t, y_1(t)) - F(t, y_2(t))\, \mathrm{d}t \right| \end{aligned}$$

By the Lipschitz condition, we have

$$|F(t, y_1(t)) - F(t, y_2(t))| \le C|y_1(t) - y_2(t)| \le Cd(y_1, y_2).$$

Thus,

$$\begin{aligned} d(Ty_1, Ty_2) &\le \sup_{|x - x_0| < \delta} Cd(y_1, y_2) \\ &\le C\delta d(y_1, y_2). \end{aligned}$$

If we assume that $\delta < \frac{1}{C}$, then this is a contraction. □

Remark 2.18 (Small intervals where solution is guaranteed)**.** Consider the IVP

$$y'(x) = R^2 + y(x)^2, \qquad y(0) = 0, \qquad (x, y) \in [-1, 1] \times [-1, 1].$$

If we let R be large, then the interval will have to be less than $\frac{1}{R^2 + 1}$.

Let's show where it actually works. We may rewrite this as

$$1 = \frac{y'}{R^2 + y^2} = \frac{1}{R}\frac{y/R}{1 + (y/R)^2} = \frac{1}{R}\frac{d}{dx}\arctan\frac{y}{R}.$$

We find that

$$\arctan\frac{y(x)}{R} = Rx \implies y(x) = R\tan Rx.$$

This means that $Rx < \frac{\pi}{2}$. Clearly, if R gets large, the solution interval can get arbitrarily small.

Proposition 2.19

We can improve the bounds of δ in Theorem 2.16 to

$$\delta < \min\left\{a, \frac{b}{M}\right\}.$$

Proof. Let's choose a different metric. For a large constant L, let

$$d_L(f, g) := \sup_{|x-x_0|<\delta} |f(x) - g(x)|e^{-L|x-x_0|}.$$

Claim 2.5. $d_L(Ty_1, Ty_2) \leq \frac{C}{L} d_L(y_1, y_2)$.

$$\begin{aligned}
d_L(Ty_1, Ty_2) &= \sup_{|x-x_0|<\delta} |Ty_1(x) - Ty_2(x)|e^{-L|x-x_0|} \\
&= e^{-L|x-x_0|} \sup_{|x-x_0|<\delta} \left|\int_{x_0}^{x} F(t, y_1(t)) - F(t, y_2(t))\,\mathrm{d}t\right| \\
&\leq e^{-L|x-x_0|} \int_{x_0}^{x} C d_L(y_1, y_2) e^{L|t-x_0|}\,\mathrm{d}t \\
&= \frac{C}{L} d_L(y_1, y_2).
\end{aligned}$$

□

2.6.2. Successive approximation

March 20, 2023 The fixed point theorem tells us, given $y(t_0) = y_0$, we can create an interative solution to an integral equation by letting

$$y_{n+1}(t) = y_0 + \int_{t_0}^{t} F(s, y_n(s))\,\mathrm{d}s.$$

Proposition 2.20

Let a_* be the δ defined in Proposition 2.19. We can bound successive functions by

$$\max_{|t-t_0|\leq a_*} |y_n(t) - y_{n-1}(t)| \leq \frac{MC^{n-1}a_*^n}{n!}$$

Proof. Assume that $t \geq t_0$.

$$\begin{aligned} |y_1(t) - y_0(t)| &= \left| \int_{t_0}^{t} F(s, y_0) \, \mathrm{d}s \right| \\ &\leq M|t - t_0| \\ |y_2(t) - y_1(t)| &= \left| \int_{t_0}^{t} F(s, y_1(s)) - F(s, y_0(s)) \, \mathrm{d}s \right| \\ &\leq \int_{t_0}^{t} |F(s, y_1(s)) - F(s, y_0(s))| \, \mathrm{d}s \\ &\leq \int_{t_0}^{t} C|y_1(s) - y_0(s)| \, \mathrm{d}s \\ &\leq \int_{t_0}^{t} CM|s - t_0| \, \mathrm{d}s \\ &= CM\frac{|t - t_0|^2}{2}. \end{aligned}$$

We can continue with the same method by induction to show the bound. □

Proof of Peano's theorem. "Near" in the theorem statement has a formal definition. Let F be continuous on

$$R = [x_0 - a, x_0 + a] \times [y_0 - b, y_0 + b].$$

Moreover, let $M = \max_R |F|$ and $a_* := \min\left\{a, \frac{b}{M}\right\}$.

Let

$$y_n = \int_{t_0}^{t} F(s, y_{n-1}(s)) \, \mathrm{d}s = T[y_{n-1}].$$

I claim the set

$$\{T[y] : y \in C[t_0 - a_*, t_0 + a_*]\}$$

is compact. We use Arzela-Ascoli; we don't work with the entire interval, but a similar proof follows. Totally bounded is true. For equicontinuity, we need to use *Euler's (polygonal) method*: Given a partition $\mathscr{P} = t_0 < t_1 < \cdots < t_N = t_0 + a_*$. We may approximate y by

$$y_{\mathscr{P}}(t) := \begin{cases} y_0 + F(t_0, y_0)(t - t_0) & t \in [t_0, t_1] \\ y_{\mathscr{P}}(t_1) + F(t_1, y(t_1))(t - t_1) & t \in (t_1, t_2] \\ \vdots & \\ y_{\mathscr{P}}(t_k) + F(t_k, y(t_k))(t - t_k) & t \in (t_k, t_{k+1}] \\ \vdots & \\ y_{\mathscr{P}}(t_{N-1}) + F(t_{N-1}, y(t_{N_1}))(t - t_{N-1}) & t \in (t_{N-1}, t_N] \end{cases}$$

March 22, 2023

We can show by induction that this is well-defined on $[t_0 - a_*, t_0 + a_*]$. We also have

$$|y_{\mathscr{P}}(t) - y_{\mathscr{P}}(t')| \leq M|t - t'|.$$

The set

$$\{y_{\mathscr{P}} : \mathscr{P} \text{ partition}, |\text{slope}| \leq M\} \subseteq C[t_0, t_0 + a_*],$$

satisfies Arzela-Ascoli, so it is compact. Thus, this has a convergent subsequence that is convergent to some $y \in C[t_0, t_0 + a_*]$.

Given $\varepsilon > 0$, there exists a $\delta(\varepsilon)$ such that

$$|t - t'| + |x - x'| < 100\delta \implies |F(t, y) - F(t', y')| < \varepsilon.$$

Let the maximum width of the partition be $\leq \min\left\{\delta, \frac{\delta}{M}\right\}$. Define step function

$$g(t) = \begin{cases} F(t_k, y_{\mathscr{P}}(t_k)) & t \in [t_k, t_{k+1}] \end{cases}.$$

Then

$$y_{\mathscr{P}}(t) = y_0 + \int_{t_0}^{t} g_{\mathscr{P}}(s)\,\mathrm{d}s.$$

We now bound

$$\begin{aligned} \left| y_{\mathscr{P}}(t) - \left[y_0 + \int_{t_0}^{t} F(s, y_{\mathscr{P}}(s))\,\mathrm{d}s \right] \right| &= \left| y_0 + \int_{t_0}^{t} g_{\mathscr{P}}(s)\,\mathrm{d}s - \left[y_0 + \int_{t_0}^{t} F(s, y_{\mathscr{P}}(s))\,\mathrm{d}s \right] \right| \\ &= \left| \int_{t_0}^{t} g_{\mathscr{P}}(s) - F(s, y_{\mathscr{P}}(s))\,\mathrm{d}s \right| \\ &\leq (t - t_0)\varepsilon \leq a_*\varepsilon. \end{aligned}$$

□

3. Norms, operators, and derivatives

3.1. More norms

March 24, 2023 Given $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{F}^n$, the **$\ell^p$-norms** are defined as

$$\|\mathbf{x}\|_p \coloneqq \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}, \tag{3.1}$$

$$\|\mathbf{x}\|_\infty \coloneqq \max_{x=1}^n |x_i|. \tag{3.2}$$

Lemma 3.1 (Generalized AM-GM)
If $a, b \geq 0$ and $\theta \in (0,1)$,

$$a^{1-\theta} b^\theta \leq (1-\theta)a + \theta b. \tag{3.3}$$

Proof. This is true if a or b is 0, so assume $a, b > 0$. Diving by a on both sides we get

$$\left(\frac{b}{a}\right)^\theta \leq (1-\theta) + \theta \frac{b}{a}.$$

Instead, let's show for $x \geq 0$, $f(x) \coloneqq (1-\theta) + \theta x - x^\theta$ is always non-negative. Observe $f(1) = 0$.

$$f'(x) = \theta - \theta x^{\theta - 1} \begin{cases} < 0 & 0 < x < 1 \\ > 0 & x > 1 \end{cases}$$

This proves the inequality. □

Theorem 3.2 (Hölder's inequality)
Suppose $1 < p < \infty$. Given p, let the *conjugate exponent* q satisfy $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\left|\sum_{i=1}^n x_i y_i\right| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |x_i|^q\right)^{\frac{1}{q}}.$$

Proof. Consider the special case where $\|\mathbf{x}\|_p = 1, \|\mathbf{y}\|_q = 1$. All we need to show is

$$\left|\sum_{i=1}^n x_i y_i\right| \leq 1, \qquad \left|\sum_{i=1}^n x_i y_i\right| \leq \sum_{i=1}^n |x_i y_i|.$$

Without loss of generality, assume $x_i \geq 0$, $y_i \geq 0$. Let

$$x_i y_i = (x_i^p)^{\frac{1}{p}} (y_i^q)^{\frac{1}{q}}.$$

Then, since $\frac{1}{p} + \frac{1}{q} = 1$, by Equation 3.3 we have

$$(x_i^p)^{\frac{1}{p}} (y_i^q)^{\frac{1}{q}} \leq \frac{1}{p} x_i^p + \frac{1}{q} y_i^q.$$

Thus,

$$\sum_{i=1}^n |x_i y_i| \le \frac{1}{p}\sum_{i=1}^n |x_i|^p + \frac{1}{q}\sum_{i=1}^n |y_i|^q = \frac{1}{p} + \frac{1}{q} = 1.$$

For $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, consider normalizing the vectors:

$$\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_p} \right\|_p = \left\| \frac{\mathbf{y}}{\|\mathbf{y}\|_p} \right\|_p = 1.$$

So we get

$$\left| \sum_{i=1}^n \frac{x_i}{\|\mathbf{x}\|_p} \frac{y_i}{\|\mathbf{y}\|_p} \right| \le 1 \implies \left| \sum_{i=1}^n x_i y_i \right| \le \|\mathbf{x}\|_p \|\mathbf{y}\|_p .$$

□

Corollary 3.3 (Minkowski's inequality)
$\mathbf{x} \mapsto \|\mathbf{x}\|_p$ is a norm if and only if $p \ge 1$.

Proof. Positive definite and absolute homogeneity is clear. We can show that the norm fails to have the triangle inequality if $p < 1$ for $\mathbf{x} = (1, 0, \dots, 0)$ and $\mathbf{y} = (0, 1, \dots, 0)$, for instance. For the other direction, assume that $x_i, y_i \ge 0$.

$$\begin{aligned}
\|\mathbf{x}+\mathbf{y}\|_p^p &= \sum_{i=1}^n (x_i+y_i)^p = \sum_{i=1}^n (x_i+y_i)^{p-1+1} \\
&= \sum_{i=1}^n (x_i+y_i)^{p-1}(x_i+y_i) \\
&= \sum_{i=1}^n (x_i+y_i)^{p-1} x_i + \sum_{i=1}^n (x_i+y_i)^{p-1} y_i \\
&\le \underbrace{\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n (x_i+y_i)^{\overbrace{(p-1)q}^{p}}\right)^{\frac{1}{q}} + \left(\sum_{i=1}^n y_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n (x_i+y_i)^{(p-1)q}\right)^{\frac{1}{q}}}_{\text{Theorem 3.2}} \\
&= \|\mathbf{x}\|_p \|\mathbf{x}+\mathbf{y}\|_p^{\frac{p}{q}} + \|\mathbf{y}\|_p \|\mathbf{x}+\mathbf{y}\|_p^{\frac{p}{q}} \\
&= (\|x\|_p + \|y\|_p) \|\mathbf{x}+\mathbf{y}\|^{p-1} \\
\implies \|\mathbf{x}+\mathbf{y}\|_p &\le \|x\|_p + \|y\|_p .
\end{aligned}$$

□

3.2. Equivalent norms

March 27, 2023

Definition 3.1 (Equivalent norms)
Two norms $\|\cdot\|_a, \|\cdot\|_b$ over a vector space, V are said to be **equivalent** if there exists constants c, C such that

$$c\|v\|_a \le \|v\|_b \le C\|v\|_a$$

for all $v \in V$.

This is an equivalence relation $\sim$ on norms, since $\|v\|_a \sim \|v\|_a$, $\|v\|_a \sim \|v\|_b$ implies

$$\frac{1}{C}\|v\|_b \le \|v\|_a \le \frac{1}{c}\|v\|_b .$$

Suppose $\|v\|_a \sim \|v\|_b$ and $\|v\|_b \sim \|v\|_c$, such that the constants are m, M and ℓ, L, respectively. Then

$$m\ell \|v\|_c \leq \|v\|_a \leq ML \|v\|_c .$$

Example 3.4 (Equivalent norms) – In $\mathbb{R}^n$, we have

$$\max |x_i| \leq \sum_{i=1}^{n} |x_i| \leq n \max |x_i| \implies \|\cdot\|_\infty \sim \|\cdot\|_1 .$$

Theorem 3.5

On a finite dimensional vector space V (over $\mathbb{F}$), all norms are equivalent.

Proof. Let the basis of V be $\{v_1, \ldots, v_n\}$. Then every vector $\mathbf{v} \in V$ can be written as a *unique* linear combination of the basis vectors: $v = \sum_{i=1}^{n} x_i(v) v_i$. $x_i \colon V \to \mathbb{F}$ is a linear functional. Let

$$\mathbf{v} \mapsto \|\mathbf{v}\|_* := \max_{i=1}^{n} |x_i(\mathbf{v})|$$

be a norm on V. Let $v \mapsto \|v\|$ be any other norm. Then

$$\begin{aligned}
\|\mathbf{v}\| &= \left\| \sum_{i=1}^{n} x_i(\mathbf{v}) v_i \right\| \\
&\leq \sum_{i=1}^{n} \|x_i(\mathbf{v}) v_i\| \leq \sum_{i=1}^{n} |x_i(\mathbf{v})| \, \|v_i\| \\
&\leq \max_{i=1}^{n} |x_i(\mathbf{v})| \cdot \sum_{i=1}^{n} \|v_i\| .
\end{aligned}$$

So we have found a C such that $\|v\| \leq C \|\mathbf{v}\|_*$. Define a function F on $\mathbb{R}^n$ as

$$F(\mathbf{x}) = \left\| \sum_{i=1}^{n} x_i v_i \right\| .$$

Claim 3.1. *F is a continuous function on $\mathbb{R}^n$.*

$$\begin{aligned}
|F(\mathbf{x}) - F(\mathbf{y})| &= \left| \left\| \sum_{i=1}^{n} x_i v_i \right\| - \left\| \sum_{i=1}^{n} y_i v_i \right\| \right| \\
&\leq \left\| \sum_{i=1}^{n} x_i v_i - \sum_{i=1}^{n} y_i v_i \right\| \\
&= \left\| \sum_{i=1}^{n} (x_i - y_i) v_i \right\| \leq \max_i |x_i - y_i| \cdot \sum_{i=1}^{n} \|v_i\|
\end{aligned}$$ ■

Consider this on the set $K = \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^n, \max_{i=1}^{n} |x_i| = 1\} \subseteq \mathbb{R}^n$, which is closed and bounded, and is therefore compact. F attains its infimum on K, so

$$F(\mathbf{x}) \geq a, \quad (\forall \mathbf{x} \in K).$$

For arbitrary $v \neq 0$,

$$\|\mathbf{v}\| \leq \left\|\sum_{i=1}^{n} x_i(\mathbf{v}) v_i\right\| = \left\|\max_i |x_i(v)| \cdot \sum_{i=1}^{n} \frac{x_i(\mathbf{v})}{\max_i |x_i(\mathbf{v})|} v_i\right\|$$

$$= \|\mathbf{v}\|_* \underbrace{\left\|\sum_{i=1}^{n} \frac{x_i(\mathbf{v})}{\max_i |x_i(\mathbf{v})|} v_i\right\|}_{\geq a}.$$

□

3.2.1. Remarks on equivalent norms

March 29, 2023

Example 3.6 – Does the inequality

$$\max_{x\in[0,1]} |f(x)| \leq C \int_0^{\frac{1}{3}} |f(t)|\, \mathrm{d}t$$

hold for

- $f \in C[0,1]$? (No)
- All polynomials?
- All polynomials of degree $\leq 10^6$?

For the last question, we know the LHS and RHS are both norms on polynomials of a bounded finite degree, which, by Theorem 3.5, are equivalent. Thus, a constant does exist.

Proposition 3.7 (Inequality on ℓ^p norms)
If $p_1 \leq p_2$, $\|\mathbf{x}\|_{\ell^{p_2}} \leq \|\mathbf{x}\|_{\ell^{p_1}}$.

Proof. It helps to normalize:

$$\|\mathbf{x}\|_{\ell^{p_1}} = 1, \left(\sum_{i=1}^{n} |x_i|^{p_1}\right)^{\frac{1}{p_1}} = 1 \implies |x_i| \leq 1.$$

Note that

$$|x_i|^{p_2} \leq |x_i|^{p_1} \implies \left(\sum_{i=1}^{n} |x_i|^{p_2}\right)^{\frac{1}{p_2}} \leq 1.$$

For general x, consider

$$\left\|\frac{\mathbf{x}}{\|\mathbf{x}\|_{\ell^{p_1}}}\right\|_{\ell^{p_2}} \leq 1 \implies \|\mathbf{x}\|_{\ell^{p_2}} \leq \|\mathbf{x}\|_{\ell^{p_1}}.$$

□

Definition 3.2 ($\ell^p(\mathbb{N})$ norms on infinte series)
Let $(x_n) \in \ell^p(\mathbb{N}) \equiv \ell^p$, the set (vector space) of *p-summable sequences*, so that

$$\sum_{i=1}^{\infty} |x_i|^p$$

is finite. The norm is given by

$$\|x\|_{\ell^p} := \left(\sum_{i=1}^{\infty} |x_i|^p\right)^{\frac{1}{p}}.$$

3.3. Continuous function norms

Definition 3.3 (p-norms on continuous functions)
For any function $f \in C[a, b]$, define the **p-norm** as

$$\|f\|_p := \left(\int_a^b |f(x)|^p \, \mathrm{d}x\right)^{\frac{1}{p}}$$

Remark 3.8. To show this is a norm, we could try and prove a version of Hölder's inequality:

$$\int_a^b |f(x)g(x)| \, \mathrm{d}x \leq \left(\int_a^b |f(x)|^p \, \mathrm{d}x\right)^{\frac{1}{p}} \left(\int_a^b |g(x)|^q \, \mathrm{d}x\right)^{\frac{1}{q}}, \qquad \frac{1}{p} + \frac{1}{q} = 1.$$

Another idea is to apply our results for ℓ^p-norms to Riemann sums, and take the limit.

Definition 3.4 (Banach spaces)
Complete normed spaces are caled **Banach spaces**.

Example 3.9 (Banach spaces) –
1. $\ell^p(\mathbb{N})$ is a Banach space.
2. $C[a, b]$ is not complete, so it is not a Banach space.
3. To rectify this, we can create the space $L^p[a, b]$, which requires the *Lebesgue integral*.

3.4. Linear operators

Definition 3.5 (Linear operator)
Let V, W be vector spaces. $T : V \to W$ is a **linear operator** if

$$T(\alpha v + \beta w) = \alpha T(v) + \beta T(w).$$

What is the criterion for a linear operator to be continuous?

Theorem 3.10

For normed vector spaces V, W (both over $\mathbb{F}$), fix two norms on those space. For the linear operator $T: V \to W$, the following are equivalent:

1. T is continuous everywhere.
2. T is continuous at 0.
3. There is a constant C such that $\|T\mathbf{v}\|_W \le C\|\mathbf{v}\|_V$.

Proof. (1) $\implies$ (2) is trivial.

(2) $\implies$ (3). For $\varepsilon = 1$, there is a δ so that

$$\|\mathbf{v}\|_V < \delta \implies \|T\mathbf{v}\|_W < 1.$$

Observe that $\frac{\delta}{2} \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|_V}$ has norm $< \delta$. So

$$\left\| T\left(\frac{\delta}{2} \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|_V}\right)\right\|_W < 1.$$

But the LHS is the same as

$$\|T\mathbf{v}\|_W \le \frac{2}{\delta} \cdot \|\mathbf{v}\|_V.$$

(3) $\implies$ (1). For $\mathbf{v}_1, \mathbf{v}_2 \in V$,

$$\|T(\mathbf{v}_1) - T(\mathbf{v}_2)\|_W = \|T(\mathbf{v}_1 - \mathbf{v}_2)\| \le C\|\mathbf{v}_1 - \mathbf{v}_2\|.$$

This implies T is *Lipschitz* continuous. Thus, T is continuous everywhere. ☐

Example 3.11 – Let V be the set of real polynomials with norm $\|p\| = \max_{x\in[0,1]|p(t)|}$ and let $W = \mathbb{R}$. Let $Tp = p'(1)$. Is there a constant C so that $|p'(1)| \le C \max_{x\in[0,1]|p(t)|}$? Let $p_n(t) = t^n$. Then

$$|p_n'(1)| = n.$$

But this can get arbitrarily large.

March 31, 2023

Definition 3.6 (Bounded linear operator, operator norm)

The set of all bounded linear operators $T: V \to W$ is $\mathcal{L}(V, W)$. This is a vector space. A linear operator $T: V \to W$ so that

$$\|T\mathbf{v}\|_W \le C\|\mathbf{v}\|_V$$

is called a **bounded linear operator**. It gets this name because it maps bounded sets to bounded sets. The smallest such C is called the **operator norm**. An equivalent way to define this is

$$\|T\|_{\text{op}} := \sup_{\mathbf{v}=1} \|T\mathbf{v}\|_W.$$

Proposition 3.12

If V is finite dimensional, then the linear operator $T: V \to W$ where V, W are normed spaces is (always) continuous.

Proof. Let $v_1, \ldots, v_n$ be the basis. Write $\mathbf{v} = \sum_{i=1}^n x_i(\mathbf{v})v_i$. Then

$$\|T\mathbf{v}\|_W = \left\| T\left[\sum_{i=1}^n x_i(\mathbf{v})v_i\right]\right\| = \left\|\sum_{i=1}^n x_i(\mathbf{v})T(v_i)\right\| \le \sum_{i=1}^n |x_i(\mathbf{v})| \, \|T(v_i)\|_W .$$

Let $C = \sum_{i=1}^n \|Tv_i\|_W$. Then with the max norm, we can dominate this by

$$\le C \max |x_i(\mathbf{v})| \le CA \, \|\mathbf{v}\|_V ,$$

by Theorem 3.5. Theorem 3.10 finishes. □

3.4.1. Matrix norms

Definition 3.7 (Matrix norm)
Let $A \in \mathbb{R}^{n\times m}$, so that $A\colon \mathbb{R}^m \to \mathbb{R}^n$. Equip $\mathbb{R}^m$ and $\mathbb{R}^n$ with the ℓ^a and ℓ^b norm respectively ($a, b \in [1,\infty]$). The **matrix norm** is defined as

$$\|A\|_{a\to b} := \sup_{\|\mathbf{x}\|_a = 1} \|A\mathbf{x}\|_b .$$

Many of these matrix norms can be interesting, and can motivate discussions of other parts of linear algebra to solve.

$$\begin{aligned}
\|A\mathbf{x}\|_{2\to 2} &= \left(\sum_{i=1}^m \left|\sum_j a_{i,j}x_j\right|^2\right)^{\frac{1}{2}} \\
&\le \left(\sum_{i=1}^m \left|\left(\sum_{j=1}^n a_{i,j}^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n |x_j|^2\right)^{\frac{1}{2}}\right|^2\right)^{\frac{1}{2}} \\
&= \|\mathbf{x}\|_2 \underbrace{\left(\sum_{i=1}^m \sum_{j=1}^n |a_{i,j}|^2\right)^{\frac{1}{2}}}_{S} =: \|\mathbf{x}\|_2 \|A\|_{\text{HS}}
\end{aligned}$$

We call sum (S) the *Hilbert-Schmidt norm* of the matrix, which does indeed define a norm, but not the 2 norm we are looking for.

Indeed, suppose A has e-values $\lambda_1, \ldots, \lambda_n$. Then

$$\|A\|_{HS} = \left(\sum_{j=1}^n \lambda_j^2\right)^{\frac{1}{2}} ,$$

$$\|A\mathbf{x}\|_2 = \left(\sum_{i=1}^n |\lambda_j x_j|^2\right)^{\frac{1}{2}} \le \max_j |\lambda_j| \, \|\mathbf{x}\|_2 ,$$

which gives us a better bound than the HS norm.

Definition 3.8 (Orthogonal vectors)
x is **orthogonal** to y if $\mathbf{x}^T\mathbf{y} = \sum_i x_i y_i = 0$.

April 3, 2023

Remark 3.13. The *spectral theorem* tells us that every real symmetric matrix has real eigenvalues and has an orthonormal basis (orthogonal basis where all basis vectors have norm 1) of eigenvectors.

Let U be a matrix formed out of n orthogonal basis vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Then $U^T U = UU^T = I$. With this fact, we can prove the following proposition.

Proposition 3.14

For $A \in \mathbb{R}^{n\times n}$ so that A is symmetric,

$$\|A\|_{2\to 2} = \max_j |\lambda_j|,$$

where λ_j are the eigenvalues of A.

Proof. Create the matrix U defined before by the orthogonal basis guaranteed by the spectral theorem. Given a standard basis vector $\mathbf{e}_j$, $U\mathbf{e}_j = \mathbf{u}_j \iff \mathbf{e}_j = U^{-1}\mathbf{u}_j$. Thus, $AU\mathbf{e}_j = A\mathbf{u}_j$, and by the linearity of U, we can show that U preserves length. $U^{-1}AU = D$, where D is diagonal with eigenvalues along the diagonal. This implies $A = UDU^{-1}$ since U preserves length, we conclude that the maximum we can "stretch" a vector is $\max_j |\lambda_j|$.

Constructing a vector that achieves this bound requires picking an eigenvector that has the corresponding largest eigenvalue. □

Remark 3.15 ($A \in \mathbb{C}^{n\times n}$)**.** In the complex case, we replace the word symmetric with **Hermitian**, which means that A equals its *conjugate transpose*, $\overline{A^T}$.

Theorem 3.16

For any matrix A,

$$\|A\|_{2\to 2} = \sqrt{\max_j |\mu_j|},$$

where μ_j are the eigenvalues of $A^T A$.

Proof.

$$\begin{aligned}\|A\mathbf{x}\|_2^2 = \left\|(A\mathbf{x})^T(A\mathbf{x})\right\|_2 &= \left\|\mathbf{x}^T A^T A\mathbf{x}\right\|_2 \\ &= \|\mathbf{x}\|_2 \left\|A^T A\mathbf{x}\right\|_2 \\ &\le \|\mathbf{x}\|_2 \cdot \max_j |\mu_j| \cdot \|\mathbf{x}\|_2 .\end{aligned}$$

This implies $\|A\|_{2\to 2} \le \sqrt{\max_j |\mu_j|}$. To show that it reaches this bound, pick the eigenvector of $A^T A$ that has the maximum $|\mu_{j_0}|$. □

3.5. Derivatives

April 7, 2023

3.5.1. Fréchet derivatives

Definition 3.9 (Fréchet differentiability)

A function $f\colon V \to W$, where V and W are normed vector spaces, with $U \subseteq V$ which is open, is said to be **(Fréchet) differentiable at $x \in U$** is there exists a (bounded) linear operator $R\colon V \to W$ so that

$$\lim_{h\to 0} \frac{\|F(x+h) - F(x) - Th\|_W}{\|h\|_V} = 0. \tag{3.4}$$

If this is true, then we write $T = DF(x) = DF|_x$. f is **(Fréchet) differentiable** if all $x \in U \subseteq V$ are differentiable. When $V = \mathbb{R}^n$, we say that f is **totally differentiable** with **total derivative** $Df(x)$.

Recall that with *little-o notation*, $f(h) = o(g(h))$ if

$$\lim_{h\to 0} \frac{\|f(h)\|}{\|g(h)\|} = 0.$$

We can form an equivalent requirement for differentiability: $Df(x) = T$ for $x \in V$ if and only if

$$f(x+h) = f(x) + Th + o(\|h\|), \qquad \text{as } h \to 0. \tag{3.5}$$

We can replace $o(\|h\|)$ with any function $E(h)$ so that $\|E(h)\| \leq o(\|h\|)$.

3.5.2. Directional (Gâteaux) derivatives

Definition 3.10 (Directional (Gâteaux) derivative)

Let V, W be normed vector spaces. Suppose $U \subseteq V$ is open. Let $v \in V$ so that $v \neq 0$. If the limit

$$\lim_{t\to 0} \frac{f(x+tv) - f(x)}{t} \tag{3.6}$$

exists, we call it the **directional derivative** of f at x with direction v, and denote it $D_v f|_x = D_v f(x)$.

If directional derivatives exist for all $v \neq 0$, then f is **Gâteaux differentiable** at x.

Definition 3.11

If $V = \mathbb{R}^n$, consider the standard basis $e_j = (0, \dots, \underbrace{1}_{j\text{th index}}, \dots, 0)$ for $1 \leq j \leq n$. We denote the **jth partial derivative** of f at a as

$$\frac{\partial f(a)}{\partial x_j} := D_{e_j} f(a) = \lim_{t\to 0} \frac{f(a+e_j) - f(a)}{t}. \tag{3.7}$$

Assume that $f\colon V \to W$ is (Fréchet) differentiable at a. Let $h = ta$ in Equation 3.5 so that

$$D_v f(a) = \frac{f(a+tv) - f(a)}{t} = \frac{Df_a(tv)}{t} + \frac{o(\|tv\|)}{t} = Df_a(v) + o(1).$$

Taking $t \to 0$, we find $D_v f(a) = Df_a(v)$. Therefore, (Fréchet) differentiability implies directional differentiability in all directions.

Example 3.17 – The converse of this, that the directional deriative existing in all directions implies the derivative exists, is not true in general.

Consider

$$f\colon \mathbb{R}^2 \to \mathbb{R}\colon (x_1, x_2) \mapsto \begin{cases} \frac{x_1^3 + 2x_2^2}{x_1^2 + x_2^2} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Pick $\mathbf{u} = (u_1, u_2)$, and take the directional derivative of f at 0 in direction $\mathbf{u}$:

$$D_{\mathbf{u}} f(0) = \lim_{t \to 0} \frac{f(0 + t\mathbf{u}) - f(0)}{t} = \frac{1}{t} \cdot \frac{t^3 u_1^3 + 2t^3 u_2^3}{t^2 u_1^2 + t^2 u_2^2} = \frac{u_1^3 + u_2^3}{u_1^2 + u_2^2}.$$

Note that this is certainly not linear in $\mathbf{u}$. However, if f were differentiable, then it would be linear in the entries of $\mathbf{u}$:

$$Df(\mathbf{u}) = c_1 u_1 + c_2 u_2.$$

So the derivative does not exist.

3.5.3. Calculus in $\mathbb{R}^n$

Definition 3.12

April 10, 2023

Given $\Omega \subseteq \mathbb{R}^n$ which is open, $f\colon \Omega \to \mathbb{R}^m$ is of **class $C^1(\Omega)$** if $\frac{\partial f_i}{\partial x_j}$ exists and is continuous for all i, j.

Theorem 3.18

If $f \in C^1(\Omega)$, then f is Fréchet differentiable.

Proof for $n = 2$, $m = 1$. Assume $\mathbf{a} = (a_1, a_2)$. Then $\frac{\partial f}{\partial x_1}(\mathbf{a})$, and $\frac{\partial f}{\partial x_2}(\mathbf{a})$ are continuous at $\mathbf{a}$. We let

$$T(h_1, h_2) = \frac{\partial f}{\partial x_1}(a_1, a_2) h_1 + \frac{\partial f}{\partial x_2}(a_1, a_2) h_2$$

be our conjectured derivative. Consider

$$\begin{aligned} f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) - T(h_1, h_2) = {} & f(a_1 + h_1, a_2 + h_2) - f(a_1 + h_1, a_2) \\ & + f(a_1 + h_1, a_2) - f(a_1, a_2) \\ & - \frac{\partial f}{\partial x_1}(a_1, a_2) h_1 - \frac{\partial f}{\partial x_2}(a_1, a_2) h_2. \end{aligned}$$

We can bound

$$f(a_1 + h_1, a_2) - f(a_1, a_2) - \frac{\partial f}{x_2}(a_1, a_2) h_1 = o(h_1) = o(\|h\|).$$

For the rest of the equation, we can find use the fact that $\frac{\partial f}{\partial x_2}$ is continuous to find a δ so that

$$\left| \frac{\partial f}{\partial x_2}(x_1, x_2) - \frac{\partial f}{\partial x_2}(a_1, a_2) \right| < \varepsilon,$$

provided that $|x_1 - a_1|, |x_2 - a_2| < \delta$. We have (for some ξ between a_2 and $a_2 + h_2$),

$$\begin{aligned} f(a_1+h_1, a_2+h_2) - f(a_1+h_1, a_2) - \frac{\partial f}{\partial x_2}(a_1,a_2)h_2 &= \frac{\partial f}{\partial x_2}(a_1+h_1,\xi)h_2 - \frac{\partial f}{\partial x_2}(a_1,a_2)h_2 \\ &= \left(\frac{\partial f}{\partial x_2}(a_1,\xi) \pm \varepsilon\right) h_2 - \frac{\partial f}{\partial x_2}(a_1,a_2)h_2 \\ &= \pm\varepsilon \cdot h_2 \\ &= o(\|h\|). \end{aligned}$$

Combining these two, we find that T is the Fréchet derivative. □

Remark 3.19. While we didn't prove this for the general case, a similar proof follows for higher n, and we can seperate the function $f = (f_1, \dots, f_m)$ by each dimension m, making the proof the same as $m = 1$, repeated for each dimension.

3.5.4. Derivative rules

Theorem 3.20 (Chain rule)

Let X, Y, Z be normed vector spaces. Given open sets $\Omega_X \subseteq X$, $\Omega_Y \subseteq Y$, let $g\colon \Omega_X \to \Omega_Y$, $f\colon \Omega_Y \to Z$, and $a \in \Omega_X, g(a) \in \Omega_Y$. Assume f is differentiable at $g(a)$ and g is differentiable at a. Then $f \circ g$ is differentiable at a with

$$D(f \circ g)\big|_a = Df\big|_{g(a)} \circ Dg\big|_a.$$

Since all these derivatives are linear operators, the chained derivative is a linear operator from X to Z.

Proof. We know $g(a+h) - g(a) = Dg\big|_a(h) + o(\|h\|)$, and $f(g(a)+k) - f(g(a)) = Df\big|_{g(a)}(k) + o(\|h\|)$. To combine these to get our result, let $k = g(a+h) - g(a)$. Then

$$\begin{aligned} f(g(a+h)) - f(g(a)) &= Df\big|_{g(a)}[g(a+h) - g(a)] + o(\|g(a+h) - g(a)\|) \\ &= \underbrace{Df\big|_{g(a)}[Dg\big|_a(h)]}_{(1)} + \underbrace{Dg\big|_a[o(\|h\|)]}_{(2)} + \underbrace{o(\big\|Dg\big|_a(h) + o(\|h\|)\big\|)}_{(3)}. \end{aligned}$$

(1) is what we want to show is the derivative. Showing that (2) is $o(\|h\|)$ directly follows from the fact that $Dg\big|_a$ is a bounded linear operator. For (3), we write

$$\begin{aligned} \frac{o(\|Dg|_a(h) + o(\|h\|)\|)}{\|h\|} &= \frac{o(\|Dg|_a(h) + o(\|h\|)\|)}{\|Dg|_a(h) + o(\|h\|)\|} \cdot \frac{\|Dg|_a(h) + o(\|h\|)\|}{\|h\|} \\ &\le \frac{o(\|Dg|_a(h) + o(\|h\|)\|)}{\|Dg|_a(h) + o(\|h\|)\|} \cdot \left(\big\|Dg|_a\big\|_{op} + \frac{o(\|h\|)}{\|h\|}\right). \end{aligned}$$

The term in the parenthesis remains bounded, and converges to $\|Dg|_a\|_{op}$. The term

$$\big\|Dg\big|_a(h) + o(\|h\|)\big\| \le \big\|Dg\big|_a\big\|_{op} \cdot \|h\| + o(\|h\|).$$

As $h \to 0$, this term tends to 0, so by the definition of small-o notation, the other part tends to 0 as well. □

4. Topics in multivariable calculus

April 17, 2023

Multivariable calculus concerns continuous functions $f\colon \mathbb{R}^n \to \mathbb{R}^m$, where n and m are positive integers. By our analysis of derivatives in general normed vector spaces, we can develop a theory of multivariable calculus.

Definition 4.1 (Gradient)
The **gradient** of a function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ at $x \in \mathbb{R}^n$ is

$$\nabla f(x) := \begin{bmatrix} \partial_1 f & \partial_2 f & \cdots & \partial_m f \end{bmatrix}\Big|_x,$$

where ∂_i represents the partial derivative of f: $\partial_i f := \frac{\partial f}{\partial x_i}$.

4.1. Mean value theorem

Recall the standard mean value theorem: for $g\colon I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is an interval, and $a, b \in I$, there is a ξ between a and b so that $g(b) - g(a) = g'(\xi)(b-a)$.

If we want $\ell(\mathbf{a}, \mathbf{b})$ to exist *for all* $\mathbf{a}, \mathbf{b} \in \Omega$, then we need Ω to be *convex*.

Theorem 4.1 (Mean value theorem for real-valued functions)
Let $\Omega \subseteq \mathbb{R}^n$ be an open set. Suppose $\mathbf{a}, \mathbf{b} \in \Omega$ so that the line segement between $\mathbf{a}$ and $\mathbf{b}$, $\ell(\mathbf{a}, \mathbf{b}) := \{\mathbf{a} + t(\mathbf{b} - \mathbf{a}) \mid 0 \le t \le 1\}$ is contained in Ω. Suppose $f\colon \Omega \to \mathbb{R}$ is differentiable on $\ell(\mathbf{a}, \mathbf{b})$. Then there exists a vector $\boldsymbol{\xi} \in \ell(\mathbf{a}, \mathbf{b})$ so that

$$f(\mathbf{b}) - f(\mathbf{a}) = Df\big|_{\boldsymbol{\xi}}[\mathbf{b} - \mathbf{a}].$$

Proof. We apply a change of varibles to $g(t) := f(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$. We may now apply the MVT on $\mathbb{R} \to \mathbb{R}$ to g. So

$$g(1) - g(0) = g'(\tau)(1 - 0), \qquad \text{for some } \tau \in (0, 1).$$

Let $U(t) = \mathbf{a} + t(\mathbf{b} - \mathbf{a})$. Then $g(t) = f(U(t))$, so by the chain rule,

$$g'(\tau) = Dg\big|_\tau = Df\big|_{U(\tau)} DU\big|_\tau = Df\big|_{\mathbf{a} + \tau(\mathbf{b} - \mathbf{a})}[\mathbf{b} - \mathbf{a}].$$ □

We naturally want to ask if there is a mean value theorem on $\mathbb{R}^m$-valued functions. The only reasonable way for us to do this would be to apply the mean value theorem to each coordinate. However, we can quickly find counterexamples to this.

Example 4.2 (MVT does not hold for $\mathbb{R}^n \to \mathbb{R}^m$) – Define $f_1(t) = t^2$, $f_2(t) = t^3$, and let $f\colon \mathbb{R} \to \mathbb{R}^2$ be defined as $f(x) = (f_1(x), f_2(x)) = (x^2, x^3)$. Consider applying the mean value theorem to each entry. Then

$$1 = f_1(1) - f_1(0) = f'(t_1)(1 - 0) \implies t_1 = \frac{1}{2},$$

$$1 = f_2(1) - f_2(0) = f'(t_2)(1 - 0) \implies t_2 = \sqrt{\frac{1}{3}}.$$

These would be our only options of t, but they do not agree, so a generalized MVT will not work for $\mathbb{R}^m$-valued functions.

Remark 4.3. This also means that we cannot generalize MVT to complex valued functions, since $\mathbb{C}$ behaves somewhat like $\mathbb{R}^2$ (formal proof is not given here).

Despite this, we can achieve a bound similar to the mean value theorem.

Theorem 4.4 (Mean value theorem on $\mathbb{R}^m$ for integrals)
Let $\Omega \subseteq \mathbb{R}^n$ be open, and convex, and let $f \in C^1(\Omega)$. Then for all $\mathbf{a}, \mathbf{b} \in \Omega$,

$$f(\mathbf{b}) - f(\mathbf{a}) = \int_0^1 \nabla f(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))(\mathbf{b} - \mathbf{a})\, \mathrm{d}t.$$

Proof. Apply the formula in 1 dimension to $g(t) = f(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$. Since $f(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}$, we have

$$\begin{bmatrix} f_1(\mathbf{b}) - f_1(\mathbf{a}) \\ \vdots \\ f_m(\mathbf{b}) - f_m(\mathbf{a}) \end{bmatrix} = \begin{bmatrix} \int_0^1 \nabla f_1(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))(\mathbf{b} - \mathbf{a})\, \mathrm{d}t \\ \vdots \\ \int_0^1 \nabla f_m(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))(\mathbf{b} - \mathbf{a})\, \mathrm{d}t \end{bmatrix} =: \int_0^1 Df(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) \cdot (\mathbf{b} - \mathbf{a})\mathrm{d}t$$

□

We note that $Df(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$ is an $m \times n$ matrix, and is therefore also an operator, which acts on the $n \times 1$ vector $(\mathbf{b} - \mathbf{a})$. With that noted, we can prove the following theorem.

A proof exists for all norms, but using the 2-norm is convenient for us.

Theorem 4.5 (Mean value inequality)
Let $\mathbb{R}^n$ and $\mathbb{R}^m$ use the Euclidean norm Let the norm on matrices be the operator norm. Then for $f \colon \Omega \to \mathbb{R}^m$ that is differentiable on $\ell(\mathbf{a}, \mathbf{b}) \in \Omega$, we have

$$\|f(\mathbf{b}) - f(\mathbf{a})\| \le \sup_{\boldsymbol{\xi} \in \ell(\mathbf{a}, \mathbf{b})} \left\| Df|_{\boldsymbol{\xi}} \right\|_{op} \|\mathbf{b} - \mathbf{a}\|.$$

Proof. Since we are working in the 2-norm, we have

$$\|f(\mathbf{b}) - f(\mathbf{a})\|^2 = \sum_{i=1}^{m} (f_i(\mathbf{b}) - f_i(\mathbf{a}))(f_i(\mathbf{b}) - f_i(\mathbf{a})).$$

Define $g(t) = \sum_{i=1}^{m}(f_i(\mathbf{b}) - f_i(\mathbf{a}))f_i(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$. Then

$$\begin{aligned}
\|f(\mathbf{b}) - f(\mathbf{a})\|^2 &= |g(1) - g(0)| \\
&= |g'(\tau)(1 - 0)| \\
&= \left|\sum_{i=1}^{m}(f_i(\mathbf{b}) - f_i(\mathbf{a}))\nabla f_i\big|_{\mathbf{a}+\tau(\mathbf{b}-\mathbf{a})} \cdot (\mathbf{b} - \mathbf{a})\right| \\
&\le \|f(\mathbf{b}) - f(\mathbf{a})\| \left(\sum_{i=1}^{m}\left|\nabla f_i\big|_{\mathbf{a}+\tau(\mathbf{b}-\mathbf{a})}(\mathbf{b} - \mathbf{a})\right|^2\right)^{\frac{1}{2}} \qquad \text{(Cauchy-Schwarz)} \\
&= \|f(\mathbf{b}) - f(\mathbf{a})\| \left\|Df\big|_{\mathbf{a}+\tau(\mathbf{b}-\mathbf{a})}(\mathbf{b} - \mathbf{a})\right\| \\
&\le \|f(\mathbf{b}) - f(\mathbf{a})\| \cdot \sup_{\boldsymbol{\xi}\in\ell(\mathbf{a},\mathbf{b})} \left\|Df\big|_{\boldsymbol{\xi}}\right\|_{op} \|\mathbf{b} - \mathbf{a}\|
\end{aligned}$$

By cancelling $\|f(\mathbf{b}) - f(\mathbf{a})\|$ on both sides, we get the inequality. □

4.2. Inverse function theorem

April 19, 2023 Recall the inverse derivative result from analysis I: Let $f\colon [x_1, x_2] \to \mathbb{R}$ be differentiable on (x_1, x_2) so that $f'(x) > 0$ for all x (i.e. f is strictly increasing). $f\colon [x_1, x_2] \to [y_1, y_2]$ is bijective with an inverse $f^{-1}\colon [y_1, y_2] \to [x_1, x_2]$. Then we have

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}.$$

Proposition 4.6 (Local inverse function theorem $\mathbb{R} \to \mathbb{R}$)
Consider $f\colon [x_1, x_2] \to \mathbb{R}$. Suppose f' exists near $a \in (x_1, x_2)$ and $f'(a) > 0$ for all such values in that interval. Further, suppose that f' is continuous at a. Then there exists an interval $U, V \subseteq \mathbb{R}$ so that $f\colon U \to V$ is invertible.

Example 4.7 (Global vs. local invertibility) – Let $f(x_1, x_2) = (x_1 \cos x_2, x_1 \sin x_2)$. This function is not globally invertible, as $f(x_1, x_2) = f(x_1, x_2 + 2\pi)$. However, if we restrict ourselves to $(x_1, x_2) \in (a, b) \times (c, d)$, where $0 < d - c < 2\pi$, $0 < a < b$, then we have invertibility.

Definition 4.2
Let $\Omega \subseteq \mathbb{R}^n$ be open, and let $f\colon \Omega \to \mathbb{R}^n$. For $\mathbf{a} \in \Omega$, we say that f is **locally invertible** if there exists an open set U containing $\mathbf{a}$, and an open set V containing $f(\mathbf{a})$ so that $f\colon U \to V$ is invertible, with inverse denoted $f|_U^{-1}$, so that

$$f^{-1}(f(\mathbf{x})) = \mathbf{x}, \quad \forall \mathbf{x} \in U, \qquad f(f^{-1}(\mathbf{y})) = \mathbf{y}, \quad \forall \mathbf{y} \in V.$$

Recall that $Df|_{\mathbf{a}}$ is a matrix in $\mathbb{R}^{n\times n}$.

Theorem 4.8 (Inverse function theorem)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and let $\mathbf{a} \in \Omega$. Suppose $f\colon \Omega \to \mathbb{R}^n$ is a function so that Df exists on Ω, and $Df|_{\mathbf{a}}$ is invertible. Further, suppose that $\mathbf{x} \mapsto Df|_{\mathbf{x}}$ is continuous. Then there exists an open U containing $\mathbf{a}$ and an open set V containing $f(\mathbf{a})$ so that $f|_U$ is invertible, $f^{-1}\colon V \to U$ is differentiable, and

$$Df^{-1}\big|_{\mathbf{y}} = \left(Df\big|_{\mathbf{x}}\right)^{-1},$$

where $\mathbf{y} = f(\mathbf{x})$.

4.3. Implicit function theorem

April 21, 2023

Some functions are implicit, in that they are not solvable purely in terms of x or y. For example, the function of a circle, $x^2 + y^2 = c$, cannot be turned into a well-defined graph in terms of y (or x).

Example 4.9 – While the above is true, we can try and solve parts of the function. Given $x^2 + y^2 - c = 0$, we want to find $g(x)$ so that $x^2 + g(x)^2 - c = 0$. Suppose we look "near" $x = 0$ and $y > 0$. Then we can write $y = \sqrt{c - x^2} =: g(x)$, which will give a function for our circle in terms of x near 0.

In general, suppose we are given a function f from some open set in $\mathbb{R}^{n+m}$ to $\mathbb{R}^m$. We may split the input variables into two vectors, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{y} \in \mathbb{R}^m$. We start with writing f as $f(\mathbf{x}, \mathbf{y})$. Then we want to find $g(\mathbf{x})$ so that

$$f(\mathbf{x}, g(\mathbf{x})) = 0.$$

This is the setup for the following theorem.

Theorem 4.10 (Implicit function theorem)

Given an open set $\Omega \subseteq \mathbb{R}^{n+m}$, let $f\colon \Omega \to \mathbb{R}^m$. Consider $(\mathbf{x}_0, \mathbf{y}_0) \in \mathcal{Z}$, where $\mathcal{Z} := \{(\mathbf{x}, \mathbf{y}) : f(\mathbf{x}, \mathbf{y}) = 0\}$ is the zero set of f. Suppose that $f \in C^1(\Omega, \mathbb{R}^m)$ and $D_y f|_{(\mathbf{x}_0,\mathbf{y}_0)}$ is invertible. Then there is an open U containing $\mathbf{x}_0$ and an open V containing $\mathbf{y}_0$ so that there is a differentiable function $g\colon U \to \mathbb{R}^m$ such that

$$\mathcal{Z} \cap (U \times V) = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in U, \mathbf{y} \in g(\mathbf{x})\}.$$

This means that $f(\mathbf{x}, g(\mathbf{x}))$ is part of the zero set, as described before.

We also have

$$Dg\big|_{\mathbf{x}} = -(D_y f)^{-1}\big|_{(\mathbf{x}, g(\mathbf{x}))} \circ D_x f\big|_{(\mathbf{x}, g(\mathbf{x}))}.$$

Proof. Define $F\colon (\mathbf{x}, \mathbf{y}) \mapsto (\mathbf{x}, f(\mathbf{x}, \mathbf{y}))$. To prove this, we would like to show that F is locally invertible, and find the inverse $F^{-1}(\mathbf{x}, \mathbf{0})$. To do this, we want to apply Theorem 4.8. We first need to show that DF is invertible. Indeed, we can show that

$$DF\big|_{(\mathbf{x}_0,\mathbf{y}_0)} = \left[\begin{array}{c|c} I_{n\times n} & \mathbf{0} \\ \hline D_x f\big|_{(\mathbf{x}_0,\mathbf{y}_0)} & D_y f\big|_{(\mathbf{x}_0,\mathbf{y}_0)} \end{array}\right].$$

Thus,

$$\det(DF\big|_{(\mathbf{x}_0,\mathbf{y}_0)}) = \det(D_y f\big|_{(\mathbf{x}_0,\mathbf{y}_0)}) \neq 0,$$

since we have assumed $D_y f|_{(\mathbf{x}_0,\mathbf{y}_0)}$ is invertible. $F(\mathbf{x}_0, \mathbf{y}_0) = (\mathbf{x}_0, \mathbf{0})$ by assumption.

F has an inverse $G = F^{-1}$ near $(\mathbf{x}_0, \mathbf{0})$. Consider $F^{\text{im}}(U, V)$ of $(\mathbf{x}_0, \mathbf{0})$. Define G as

$$G(\mathbf{x}, \mathbf{y}) := (G_1(\mathbf{x}, \mathbf{y}), G_2(\mathbf{x}, \mathbf{y})) = (\mathbf{x}, G_2(\mathbf{x}, \mathbf{y})).$$

By setting $g(x) := G_2(\mathbf{x}, \mathbf{0})$,we have $f(G(\mathbf{x}, \mathbf{0})) = f(\mathbf{x}, g(\mathbf{x})) = \mathbf{0}$.

To get our formula, we now know for $\mathbf{x}$ near $\mathbf{x}_0$, $f(\mathbf{x}, g(\mathbf{x})) = \mathbf{0}$. Let $w(\mathbf{x}) = (\mathbf{x}, g(\mathbf{x}))$, so $(f \circ w)(\mathbf{x}) \equiv 0$. Then by the chain rule,

$$\begin{aligned} D(f \circ w)\big|_{(\mathbf{x}, g(\mathbf{x}))} &= Df\big|_{w(\mathbf{x})} \circ Dw\big|_{\mathbf{x}} \\ &= \begin{bmatrix} D_{\mathbf{x}}f|_{(\mathbf{x}, g(\mathbf{x}))} & D_{\mathbf{y}}f|_{(\mathbf{x}, g(\mathbf{x}))} \end{bmatrix} \begin{bmatrix} I_{n\times m} \\ Dg|_{\mathbf{x}} \end{bmatrix} \\ &= D_{\mathbf{x}}f + D_{\mathbf{y}}f\big|_{\mathbf{x}, g(\mathbf{x})} \circ Dg\big|_{\mathbf{x}} = 0, \end{aligned}$$

which is equivalent to our formula. □

Remark 4.11. If $A \in \mathrm{GL}_n(\mathbb{R})$ (the set of real, invertible $n \times n$ matrices), then $A + H \in \mathrm{GL}_n(\mathbb{R})$ for sufficiently small H. To show this, consider first $I + H$. I claim this is invertible if $\|H\| < 1$. We have

$$(I + H)^{-1} = I - H + H^2 - H^3 + \cdots, \qquad (I - H)^{-1} = I + H + H^2 + H^3 + \cdots.$$

With our assumption, we can write

$$(I + H + \cdots + H^m)(I - H) = I - H^{m+1},$$

We use $\|H^m\| \leq \|H\|^m$ for this.

and a similar formula for $I + H$. Since $\left\|H^{m+1}\right\|$ converges to 0, we are done. To extend this to general A, we can write

$$A + H = A(I + A^{-1}H).$$

Since $\left\|A^{-1}H\right\| \leq \left\|A^{-1}\right\| \|H\|$, we can get invertibility of $A + H$ when $\|H\| < \frac{1}{\|A^{-1}\|}$.

4.4. Proof of inverse function theorem

April 27, 2023

Take a moment to recall the conditions for Theorem 4.8. We will prove that (1) an inverse f^{-1} exists, (2) the set $V = f^{\text{im}}(U)$ is open, and (3) $Df^{-1}\big|_{\mathbf{y}} = (Df|_{\mathbf{x}})^{-1}$.

Proof of inverse function theorem. We will turn $f(x) = y$ into a fixed point problem and apply Banach's fixed point theorem. Let $x = \varphi_y(x) := x + (Df|_a)^{-1}(y - f(x))$. Thus, finding a fixed point of $\varphi_y(x)$ is the same as finding x so that $f(x) = y$. For every y,

$$\begin{aligned} D\varphi_y &= I - (Df\big|_a)^{-1}Df\big|_x \\ &= (Df\big|_a)^{-1}(Df\big|_a - Df\big|_x) \\ \implies \|D\varphi_y\| &\leq \left\|(Df\big|_a)^{-1}\right\| \cdot \left\|Df\big|_a - Df\big|_x\right\|. \end{aligned}$$

Choose U so that for $x \in U$, $\|Df|_a - Df|_x\| < \frac{1}{2\|(Df|_a)^{-1}\|}$. We now apply Theorem 4.5. For $x, \tilde{x} \in U$,

$$\|\varphi_y(x) - \varphi_y(\tilde{x})\| \leq \sup_{\xi \in \ell(x, \tilde{x})} \left\|D\varphi_y\big|_\xi\right\| \cdot \|x - \tilde{x}\| \leq \frac{1}{2}\|x - \tilde{x}\|.$$

Thus, φ_y has a unique fixed point x for all y. Define

$$f : U \to f^{\text{im}}(U) =: V.$$

Next we show that V is an open set containing $f(a)$. For every $y \in V$, we show there is an open neighborhood $B(y, \varepsilon) \subseteq V$. Fix $y_0 \in V$. Since $f|_U$ is injective, there is only one x_0 so that $f(x_0) = y_0$. Consider $r > 0$ so that $B(x_0, r) \subseteq U$. This is a Banach space. We want to show that for $\|y - y_0\| < \varepsilon := \frac{r}{2\|(Df|_a)^{-1}\|}$, the map φ_y is a contraction on $\overline{B}(x_0, r)$ (the closed ball). This implies $y \in f^{\text{im}}(B(x_0, r)) \subseteq V$.

$$\begin{aligned}\|\varphi_y(x) - x_0\| &= \|\varphi_y(x) - \varphi_y(x_0) + \varphi_y(x_0) - x_0\| \\ &\leq \underbrace{\|\varphi_y(x) - \varphi_y(x_0)\|}_{(1)} + \underbrace{\|\varphi_y(x_0) - x_0\|}_{(2)}.\end{aligned}$$

For (1), we can bound

$$\|\varphi_y(x) - \varphi_y(x_0)\| \leq \frac{1}{2}\|x - x_0\| \leq \frac{r}{2}.$$

For (2),

$$\begin{aligned}\|\varphi_y(x_0) - x_0\| &= \left\|(Df|_a)^{-1}(y - f(x_0))\right\| \\ &\leq \left\|(Df|_a)^{-1}\right\|_{op} \cdot \|y - \underbrace{f(x_0)}_{y_0}\| \\ &\leq \varepsilon \\ &= \frac{r}{2}.\end{aligned}$$

We now use the fact that there is a funcion $g \colon V \to U$, where $g(y) = f^{-1}(y)$. Let $x = g(y)$, and $k := f(x + h) - f(x)$. Consider

$$\left\|g(y + k) - g(y) - (Df|_x)^{-1}[k]\right\|.$$

We want to show that this is $o(\|k\|)$. Write

$$\begin{aligned}\|h\| &= \left\|h - (Df|_a)^{-1}[k] + (Df|_a)^{-1}[k]\right\| \\ &\leq \left\|h - (Df|_a)^{-1}[k]\right\| + \left\|(Df|_a)^{-1}[k]\right\| \\ &\leq \left\|h - (Df|_a)^{-1}[k]\right\| + \left\|(Df|_a)^{-1}\right\|_{op}\|k\|.\end{aligned}$$

To bound the first term,

$$\begin{aligned}h - (Df|_a)^{-1}[k] &= h + (\nabla f|_a)^{-1}(f(x) - f(x + h)) \\ &= x + h - x + (Df|_a)^{-1}[f(x) - y + y - f(x + h)] \\ &= \varphi_y(x + h) - \varphi_y(x).\end{aligned}$$

The norm of the last line is bounded by $\frac{1}{2}\|h\|$. Thus,

$$\frac{1}{2}\|h\| \leq \left\|(Df|_a)^{-1}\right\|_{op}\|k\|.$$

To prove the total derivative of the inverse function, we have

$$\begin{aligned}\left\|\frac{g(y+k)-g(y)-(Df|_x)^{-1}[k]}{\|k\|}\right\| &= \left\|\frac{g(y+k)-g(y)-(Df|_x)^{-1}[f(x+h)-f(x)]}{\|k\|}\right\| \\ &= \left\|\frac{g(y+k)-g(y)-(Df|_x)^{-1}[f(x+h)-f(x)]}{\|h\|}\right\| \cdot \frac{\|h\|}{\|k\|} \\ &\le C \cdot \frac{\left\|(Df|_x)^{-1}\right\|_{op} \cdot \|f(x+h)-f(x)-Df|_x[h]\|}{\|h\|} \to 0\square\end{aligned}$$

4.5. Higher-order derivatives

April 26, 2023 Write $\frac{\partial f}{\partial x_i} =: \partial_i f$. For higher order, we denote the **mixed derivative** as $\partial_i \partial_j f := \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f$. We want to say something about the relationship between $\partial_i \partial_j f$ and $\partial_j \partial_i f$. Define the *finite difference operator* as $\Delta_h f(x) := f(x+h) - f(x)$. Then we can see

$$\partial_i f(a) = \lim_{h_i \to 0} \frac{\Delta_{h_i e_i}(f(x))}{h_i},$$

and

$$\partial_j \partial_i f(a) = \lim_{h_j \to 0} \lim_{h_i \to 0} \frac{\Delta_{h_j e_j}(\Delta_{h_i e_i} f(x))}{h_j h_i}.$$

For any vector $\mathbf{h}$, $\mathbf{k}$, we have $\Delta_{\mathbf{k}}(\Delta_{\mathbf{h}} f) = \Delta_{\mathbf{h}}(\Delta_{\mathbf{k}} f)$. However, we may see different effects in the limit.

Theorem 4.12 (Schwarz's theorem)
Let $U \subseteq \mathbb{R}^n$ so that $\partial_i f, \partial_j f$, and $\partial_i \partial_j f$ exist in U and $\partial_i \partial_j f$ is continuous at $a \in U$. Then $\partial_j \partial_i f$ exists at a and $\partial_i \partial_j f(a) = \partial_j \partial_i f(a)$.

Example 4.13 (Conservative vector field) – Let $U \subseteq \mathbb{R}^n$, and $F : U \to \mathbb{R}^n$. Is there a function g so that

$$\frac{\partial}{\partial x_i} g = F?$$

If we assume this is true, then

$$\partial_i \partial_j g = \partial_j \partial_i g \implies \partial_i F_j = \partial_j F_i.$$

It turns out that this is necessarily true if $F = \nabla g$, and sufficient if U is a convex set. In that case,

$$g(\mathbf{x}) = \int_0^1 (\mathbf{x} - \mathbf{a})^T F_j(\mathbf{a} + s(\mathbf{x} - \mathbf{a}))\, \mathrm{d}s,$$

for an arbitrary line γ (parametrized from 0 to 1),

$$g(\mathbf{x}) = \int_0^1 \gamma'(t) F(\gamma(s))\, \mathrm{d}s.$$

5. Approximation Theory

5.1. Weierstrass approximation theorem

April 28, 2023

Theorem 5.1 (Weierstrass approximation theorem)
For every continuous function f on $[a, b]$, there exists a sequence of polynomials that converge uniformly to f.

For a proof of this, see my Analysis I notes. The main tool for this is to use *Bernstein polynomials*:

$$B_n f(t) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} t^k(1-t)^{n-k}.$$

We can prove that for $[0, 1]$,

$$\lim_{n\to\infty} \sup_{t\in[0,1]} (B_n f(t) - f(t)) = 0,$$

which is the result we want.

Keep in mind that an equivalent way to write the Weierstrass approximation theorem is "$\mathbb{R}[x]$, the set of all polynomials with real-valued coefficients, is *dense* in $C[a, b]$ (with respect to the $\sup$ metric)". This wording will show up in the more general Stone-Weierstrass theorem.

5.2. The Stone-Weierstrass theorem

Theorem 5.2 (Stone-Weierstrass)
Let K be a compact metric space, and $\mathcal{A} \subseteq C(K)$. Assume that $\mathcal{A}$:

1. Is a **self-adjoint algebra**. That is, given $f, g \in \mathcal{A}$ and $c \in \mathbb{C}$,

$$f + g, f \cdot g, cf, \overline{f} \in \mathcal{A}.$$

2. **Seperate points**. For $x, y \in K$ so that $x \neq y$, there exists $f \in \mathcal{A}$ so that $f(x) \neq f(y)$.

3. **Vanishes nowhere**. For all $x \in K$, there exists $f \in \mathcal{A}$ so that $f(x) \neq 0$.

Then $\mathcal{A}$ is dense in $C(K)$.

Proof of Stone-Weierstrass in $\mathbb{R}$. When we assume that all functions maps to $\mathbb{R}$, then we can remove the assumption that $\overline{f} \in \mathcal{A}$, and let $c \in \mathbb{R}$.

Claim 5.1. Assume that $\mathcal{A}$ is an algebra satisfying the assumptions of Theorem 5.2. Let $x_1, \ldots, x_n$ be distinct points in K, and let $c_1, \ldots, c_n$ be arbitrary scalars. Then there exists $f \in \mathcal{A}$ so that $f(x_i) = c_i$.

Proof. For every pair (i, j), $i \neq j$. There exists $g_{i,j}$ so that $g_{i,j}(x_i) \neq g_{i,j}(x_j)$. For every j there exists h_j so that $h_j(x_j) \neq 0$. Let

$$u_i(x) := \left[\prod_{i \neq j} (g_{i,j}(x) - g_{i,j}(x_j))\right] \cdot h_i(x),$$

$$u_i(x_k) = \begin{cases} 0 & \text{if } k \neq i, \\ \neq 0 & \text{if } k = i. \end{cases}$$

Then taking a linear combination of u_i gives us the desired function f. ■

This proves Stone-Weierstrass in the case where K is finite.

If $\mathcal{A}$ is an algebra, then $\overline{\mathcal{A}}$ is an algebra, since $f_n + g_n \to f + g$, $f_n g_n \to fg$, $cf_n \to cf$.

Claim 5.2. If $f \in \overline{\mathcal{A}}$, then $|f| \in \mathcal{A}$.

Proof. Use Theorem 5.1 for the function $g(y) = |y|$, and put $y = f(x)$. Given f, since K is compact, we have $-M \leq f(x) \leq M$. By Theorem 5.1, there is a sqeuence (q_n) that converges uniformly to $|y|$. Since $q_n(0) \to 0$, define $p_n(y) := q_n(y) - q_n(0)$. Then

$$\sup_{x \in K} |p_n(f(x)) - |f(x)|| \to 0, \qquad n \to \infty.$$ ■

Claim 5.3. If $f, g \in \overline{\mathcal{A}}$, then the functions $x \mapsto \max\{f(x), g(x)\}$ and $x \mapsto \min\{f(x), g(x)\}$ are both in $\mathcal{A}$.

Proof. We can write

$$\max\{f, g\} = \frac{f+g}{2} + \frac{|f-g|}{2}, \qquad \min\{f, g\} = \frac{f+g}{2} - \frac{|f-g|}{2}.$$ ■

Claim 5.4. Fix a function $f \in C(K)$. For every $y \in K$, there is a function $g_y \in \overline{\mathcal{A}}$ so that $g_y(y) = f(y)$, and $g_y(x) < f(x) + \varepsilon$ for all $x \in K$.

Proof. Fix y. Choose a function $h_{x,y} \in \overline{\mathcal{A}}$ so that $h_{x,y}(y) = f(y)$ and $h_{x,y}(x) = f(x)$. There exists a neighborhood U_x so that $h_{x,y}(t) < f(x) + \varepsilon$ for $t \in U_x$. Choose an open cover $(U_x)_{x \in K}$ of K. There is a finite subcover $\{U_{x_i}\}_{i=1,\ldots,n}$ since K is compact. Therefore, we have $h_{x_i,y}(t) < f(t) + \varepsilon$ for $t \in U_{x_i}$. Then

$$g_y := \min_{1 \leq i \leq n} h_{x_i,y} \implies g_y(t) < f(t) + \varepsilon, \quad \forall t \in K.$$ ■

With our function g_y, we can find an open ball V_y so that $g_y(t) > f(t) - \varepsilon$ for all $t \in V_y$. Let $(V_y)_{y \in K}$ be an open cover of K, and choose a finite subcover $(V_{y_j})_{j=1,\ldots,m}$. Therefore,

$$\phi := \max_{1 \leq j \leq m} g_{y_j} \implies f(t) - \varepsilon < \phi(t).$$

Since $\phi(t) < f(t) + \varepsilon$ as well,

$$|f(t) - \phi(t)| < \varepsilon, \quad \forall t \in K,$$

therefore, $f \in \overline{\mathcal{A}}$. □