Part IB — Analysis and Topology

Based on lectures by Dr P. Russell

Michaelmas 2022

Contents

I Generalizing continuity and convergence 3

1 Three Examples of Convergence 3
1.1 Convergence in $\mathbb{R}$ . . . 3
1.2 Convergence in $\mathbb{R}^2$ . . . 3
1.3 Convergence of Functions . . . 5
1.4 Uniform Continuity . . . 15

2 Metric Spaces 19
2.1 Definitions and Examples . . . 19
2.2 Completeness . . . 28
2.3 Sequential Compactness . . . 34
2.4 The Topology of Metric Spaces . . . 36

3 Topological Spaces 44
3.1 Definitions and Examples . . . 44
3.2 Sequences and Hausdorff spaces . . . 47
3.3 Subspaces . . . 50
3.4 Compactness . . . 51
3.5 Products . . . 54
3.6 Quotients . . . 57
3.6.1 Basics on equivalence relations and quotients . . . 59
3.7 Connectedness . . . 60

II Generalizing differentiation 68

4 Differentiation 68
4.1 The First Derivative . . . 68

4.2 The Second Derivative 80
4.3 Ordinary Differential Equations 84
4.4 The Inverse Function Theorem 86

Part I
Generalizing continuity and convergence

§1 Three Examples of Convergence

§1.1 Convergence in $\mathbb{R}$

Let (x_n) be a sequence in $\mathbb{R}$ and $x \in \mathbb{R}$. We say (x_n) *converges* to x and write $x_n \to x$ if

$$\forall\, \varepsilon > 0 \quad \exists\, N \quad \forall\, n \geq N \quad |x_n - x| < \varepsilon.$$

Useful fact: $\forall\, a, b \in \mathbb{R}$ $|a+b| \leq |a| + |b|$ (Triangle Inequality).

Bolzano-Weierstrass Theorem (BWT) A bounded sequence in $\mathbb{R}$ must have a convergent subsequence (Proof by interval bisection).

Recall: A sequence (x_n) in $\mathbb{R}$ is Cauchy if

$$\forall\, \varepsilon > 0 \quad \exists\, N \quad \forall\, m, n \geq N \quad |x_m - x_n| < \varepsilon.$$

Exercise 1.1 (Easy)**.** Show convergent $\implies$ Cauchy.

General Principle of Convergence (GPC) Any Cauchy sequence in $\mathbb{R}$ converges.

> *Outline.* If (x_n) Cauchy then (x_n) bounded so by BWT has a convergent subsequence, say $x_{n_j} \to x$. But as (x_n) Cauchy, $x_n \to x$. □

§1.2 Convergence in $\mathbb{R}^2$

Remark 1. This all works in $\mathbb{R}^n$

Let (z_n) be a sequence in $\mathbb{R}^2$ and $z \in \mathbb{R}^2$. What should $z_n \to z$ mean?

In $\mathbb{R}$: "As n gets large, z_n gets arbitrarily close to z."

What does 'close' mean in $\mathbb{R}^2$?

In $\mathbb{R}$: a, b close if $|a-b|$ small. In $\mathbb{R}^2$: Replace $|\cdot|$ by $\|\cdot\|$

Recall: If $z = (x, y)$ then $\|z\| = \sqrt{x^2 + y^2}$.

Triangle Inequality If $a, b \in \mathbb{R}^2$ then $\|a + b\| \leq \|a\| + \|b\|$.

Definition 1.1

Let (z_n) be a sequence in $\mathbb{R}^2$ and $z \in \mathbb{R}^2$. We say (z_n) **converges** to z and write $z_n \to z$ if $\forall\, \varepsilon > 0\; \exists\, N\; \forall\, n \geq N\;\; \|z_n - z\| < \varepsilon$.

Equivalently, $z_n \to z$ iff $\|z_n - z\| \to 0$ (convergence in $\mathbb{R}$).

Example 1.1

Let $(z_n), (w_n)$ be sequences in $\mathbb{R}^2$ with $z_n \to z, w_n \to w$. Then $z_n + w_n \to z + w$.

Proof.

$$\begin{aligned}\|(z_n + w_n) - (z + w)\| &\leq \|z_n - z\| + \|w_n - w\| \\ &\to 0 + 0 = 0 \text{ (by results from IA).}\end{aligned}$$

□

In fact, given convergence in $\mathbb{R}$, convergence in $\mathbb{R}^2$ is easy:

Proposition 1.1

Let (z_n) be a sequence in $\mathbb{R}^2$ and let $z \in \mathbb{R}^2$. Write $z_n = (x_n, y_n)$ and $z = (x, y)$. Then $z_n \to z$ iff $x_n \to x$ and $y_n \to y$.

Proof. ($\implies$): $|x_n - x|, |y_n - y| \leq \|z_n - z\|$. So if $\|z_n - z\| \to 0$ then $|x_n - x| \to 0$ and $|y_n - y| \to 0$.

($\impliedby$): If $|x_n - x| \to 0$ and $|y_n - y| \to 0$ then $\|z_n - z\| = \sqrt{(x_n - x)^2 + (y_n - y)^2} \to 0$ by results in $\mathbb{R}$. □

Definition 1.2 (Bounded Sequence)

A sequence (z_n) in $\mathbb{R}^2$ is **bounded** if $\exists\, M \in \mathbb{R}$ s.t. $\forall\, n\; \|z_n\| \leq M$.

Theorem 1.2 (BWT in $\mathbb{R}^2$)

A bounded sequence in $\mathbb{R}^2$ must have a convergent subsequence.

Theorem 1.3 (GPC for $\mathbb{R}^2$)

Any Cauchy sequence in $\mathbb{R}^2$ converges.

Proof. Let (z_n) be a Cauchy sequence in $\mathbb{R}^2$. Write $z_n = (x_n, y_n)$. For all m, n, $|x_m - x_n| \leq \|z_m - z_n\|$ so (x_n) is a Cauchy sequence in $\mathbb{R}$, so converges by GPC. Similarly, (y_n) converges in $\mathbb{R}$. So by 1.1, (z_n) converges. □

<u>Thought for the day</u> What about continuity? Let $f : \mathbb{R}^2 \to \mathbb{R}$. What does it mean for f to be continuous? (Simple modification of defn for $\mathbb{R} \to \mathbb{R}$).

What can we do with it?

Big theorem in IA: If $f : \mathbb{R} \to \mathbb{R}$ is a continuous function on a closed bounded interval then f is bounded and attains its bounds.

Is there a similar theorem for $\mathbb{R}^2 \to \mathbb{R}$. What do we replace 'closed bounded interval' by? We proved the theorem using BWT. Why did it work? Why did we need a closed bounded interval to make it work? What can we do in $\mathbb{R}^2$?

§1.3 Convergence of Functions

Let $X \subset \mathbb{R}$[1], let $f_n : X \to \mathbb{R}$ $(n \geq 1)$ and let $f : X \to \mathbb{R}$. What does it mean for f_n to converge to f.

Obvious idea:

Definition 1.3 (Pointwise convergence)
Say (f_n) **converges pointwise** to f and write $f_n \to f$ pointwise if $\forall\, x \in X$ $f_n(x) \to f(x)$ as $n \to \infty$.

Pros

- Simple
- Easy to check
- Defined in terms of convergence in $\mathbb{R}$

Cons

- Doesn't preserve 'nice' properties.
- 'Doesn't feel right'.

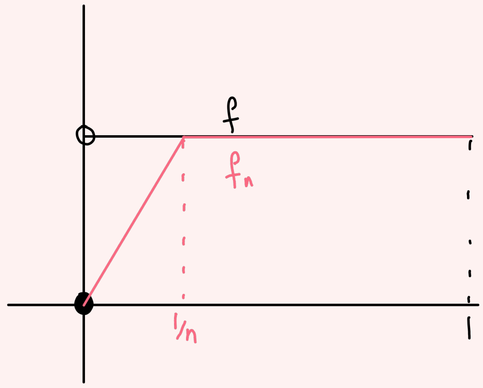
In all three examples, have $X = [0, 1]$, $f_n \to f$ pointwise.

Example 1.2 (Every f_n continuous but f not)

[1] Mostly can think of $X = \mathbb{R}$ or some interval

$$f_n(x) = \begin{cases} nx & x \le \frac{1}{n} \\ 1 & x \ge \frac{1}{n} \end{cases}$$

$$f(x) = \begin{cases} 0 & x = 0 \\ 1 & x > 0 \end{cases}$$



Clearly f_n continuous for all n but f not. If $x = 0$, $\forall\ n\ f_n(0) = 0 = f(0)$. If $x > 0$, for sufficiently large n $f_n(x) = 1 = f(x)$ so $f_n(x) \to f(x)$.

Example 1.3 (Every f_n integrable but f not)

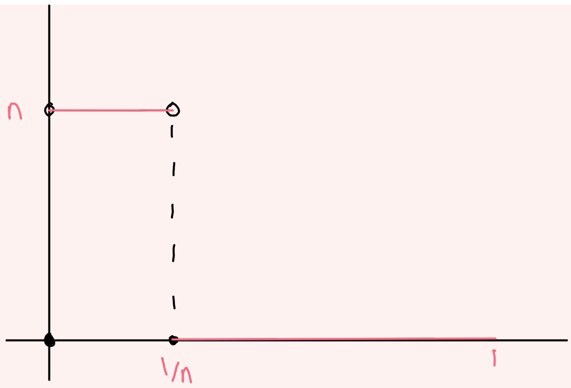
$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}.$$

This is a non integrable[a] function so now we want to find f_n such that they converge pointwise to this. Enumerate the rationals in $[0, 1]$ as $q_1, q_2, \ldots$ For $n \ge 1$, set $f_n(x) = \mathbb{1}_{q_1,\ldots,q_n}$. f_n integrable as it is nonzero at finitely many points.

[a] N.B. As in IA 'integrable' means 'Riemann integrable'

Example 1.4 (Every f_n and f integrable but $\int_0^1 f_n \not\to \int_0^1 f$)

Let $f(x) = 0$ for all x, so $\int_0^1 f = 0$. Define f_n s.t. $\int_0^1 f_n = 1$ for all n.



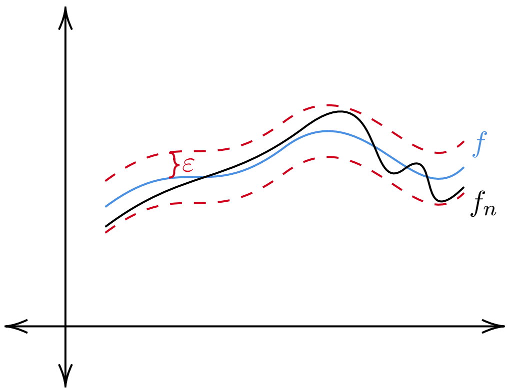
$$f_n(x) = \begin{cases} n & 0 < x < \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}.$$

Better definition:

Definition 1.4 (Uniform convergence)

Let $X \subset \mathbb{R}$, $f_n : X \to \mathbb{R}$ $(n \geq 1)$, $f : X \to \mathbb{R}$. We say (f_n) **converges uniformly** to f and write $f_n \to f$ uniformly if $\forall\, \varepsilon > 0\, \exists\, N\, \forall\, x \in X\, \forall\, n \geq N\ |f_n(x) - f(x)| < \varepsilon$.

cf $f_n \to f$ pointwise: $\forall\, \varepsilon > 0\, \forall\, x \in X\, \exists\, N\, \forall\, n \geq N\ |f_n(x) - f(x)| < \varepsilon$. (We have swapped the $\forall\, x \in x$ and $\exists\, N$). Pointwise convergence allows for N to be a function of x whilst uniform convergence requires N to work for all x even the worst case. In particular, $f_n \to f$ uniformly $\implies$ $f_n \to f$ pointwise.



Equivalently, $f_n \to f$ uniformly if for sufficiently large n $f_n - f$ is bounded and $\sup_{x \in X} |f_n - f| \to 0$.

Theorem 1.4 (A uniform limit of cts functions is cts)

Let $X \subset \mathbb{R}$, let $f_n : X \to \mathbb{R}$ be continuous ($n \geq 1$) and let $f_n \to f : X \to \mathbb{R}$ uniformly. Then f is cts.

Proof. Let $x \in X$. Let $\varepsilon > 0$. As $f_n \to f$ uniformly, we can find N s.t. $\forall\, n \geq N\ \forall\, y \in X\ |f_n(y) - f(y)| < \varepsilon$. In particular, $\forall\, y \in X\ |f_N(y) - f(y)| < \varepsilon$. As f_N is cts, we can find $\delta > 0$ s.t. $\forall\, y \in X,\ |y - x| < \delta \implies |f_N(y) - f_N(x)| < \varepsilon$. Now let $y \in X$ with $|y - x| < \delta$. Then

$$\begin{aligned} |f(y) - f(x)| &\leq |f(y) - f_N(y)| + |f_N(y) - f_N(x)| + |f_N(x) - f(x)|^{a} \\ &< \varepsilon + \varepsilon + \varepsilon = 3\varepsilon. \end{aligned}$$

Hence f is cts. □

[a] The core of this proof is this inequality.

Remark 2. This is often called a '3ε proof' (or an $\frac{\varepsilon}{3}$ proof).

Theorem 1.5

Let $f_n : [a,b] \to \mathbb{R}$ ($n \geq 1$) be integrable and let $f_n \to f : [a,b] \to \mathbb{R}$ uniformly. Then f is integrable and $\int_a^b f_n \to \int_a^b f$ as $n \to \infty$.

Proof. As $f_n \to f$ uniformly, we can pick n suff. large s.t. $f_n - f$ is bounded. Also f_n is bounded (as integrable). So by triangle inequality, $f = (f - f_n) + f_n$ is bounded. Let $\varepsilon > 0$. As $f_n \to f$ uniformly there is some N s.t. $\forall\, n \geq N\ \forall\, x \in [a,b]$ we have $|f_n(x) - f(x)| < \varepsilon$.
In particular, $\forall\, x \in [a,b]\ |f_N(x) - f(x)| < \varepsilon$.

By Riemann's criterion, there is some dissection $\mathcal{D}$ of $[a,b]$ for which $S(f_n, \mathcal{D}) - s(f_n, \mathcal{D}) < \varepsilon$. Let $\mathcal{D} = \{x_0, x_1, x_2, \ldots, x_k\}$ where $a = x_0 < x_1 < \cdots < x_k = b$. Now

$$\begin{aligned} S(f, \mathcal{D}) &= \sum_{i=1}^{k} (x_i - x_{i-1}) \sup_{x \in [x_{i-1}, x_i]} f(x) \\ &\leq \sum_{i=1}^{k} (x_i - x_{i-1}) \sup_{x \in [x_{i-1}, x_i]} (f_N(x) + \varepsilon) \\ &= \sum_{i=1}^{k} (x_i - x_{i-1}) \left(\left(\sup_{x \in [x_{i-1}, x_i]} f_N(x) \right) + \varepsilon \right) \\ &= \sum_{i=1}^{k} (x_i - x_{i-1}) \sup_{x \in [x_{i-1}, x_i]} f_N(x) + \sum_{i=1}^{k} (x_i - x_{i-1}) \varepsilon \end{aligned}$$

$$= S(f_N, \mathcal{D}) + (b-a)\varepsilon.$$

That is $S(f, \mathcal{D}) \leq S(f_N, \mathcal{D}) + (b-a)\varepsilon$. Similarly $s(f, \mathcal{D}) \geq s(f_N, \mathcal{D}) - (b-a)\varepsilon$. Hence

$$\begin{aligned} S(f,\mathcal{D}) - s(f,\mathcal{D}) &\leq S(f_N,\mathcal{D}) - s(f_N,\mathcal{D}) + 2(b-a)\varepsilon \\ &< (2(b-a)+1)\varepsilon \end{aligned}$$

But $2(b-a)+1$ is a constant so $(2(b-a)+1)\varepsilon$ can be made arbitrarily small. Hence by Riemann's criterion, f is integrable over $[a, b]$.

Now, for any n suff. large that $f_n - f$ is bounded,

$$\begin{aligned} \left| \int_a^b f_n - \int_a^b f \right| &= \left| \int_a^b (f_n - f) \right| \\ &\leq \int_a^b |f_n - f| \\ &\leq (b-a) \sup_{x \in [a,b]} |f_n - f| \\ &\to 0 \text{ as } n \to \infty \text{ since } f_n \to f \text{ uniformly.[a]} \end{aligned}$$

□

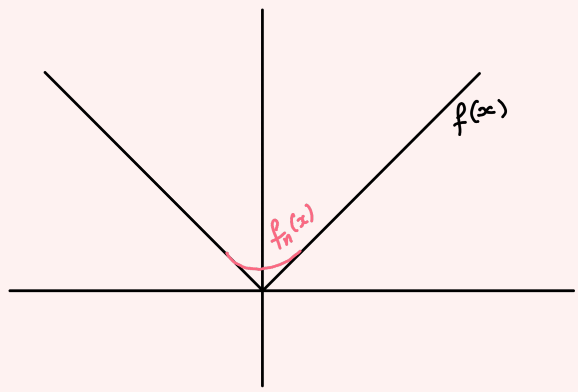
[a] Note we said that $f_n \to f$ uniformly if $\sup |f_n - f| \to 0$.

What about differentiation? Here even uniform convergence isn't enough.

Example 1.5

$f_n : (-1, 1) \to \mathbb{R}$, each f_n differentiable, $f_n \to f$ uniformly, f not diff.

Let $f(x) = |x|$ which is not differentiable at 0.



$$f_n = \begin{cases} |x| & |x| \geq \frac{1}{n} \\ ax^2 + bx + c & |x| < \frac{1}{n} \end{cases}.$$

We need $a(\frac{1}{n})^2 + \frac{b}{n} + c = \frac{1}{n}$ for continuity. Thus $b = 0$ and $c = \frac{1}{n} - \frac{a}{n^2}$.

Also need $2a\frac{1}{n} + b = 1$ and $2a(-\frac{1}{n}) = -1$ for differentiability so take $a = \frac{n}{2}$, $c = \frac{1}{n} - \frac{1}{2n} = \frac{1}{2n}$.

If $|x| \geq \frac{1}{n}$ then $|f_n(x) - f(x)| = 0$. If $|x| < \frac{1}{n}$:

$$\begin{aligned} |f_n(x) - f(x)| &= \left|\frac{n}{2}x^2 + \frac{1}{2n} - |x|\right| \\ &\leq \frac{n}{2}x^2 + \frac{1}{2n} + |x| \\ &\leq \frac{n}{2}(\frac{1}{n})^2 + \frac{1}{2n} + \frac{1}{n} \\ &= \frac{1}{2n} + \frac{1}{2n} + \frac{1}{n} \\ &= \frac{2}{n} \end{aligned}$$

So $\sup_{x \in (-1,1)} |f_n(x) - f(x)| \leq \frac{2}{n} \to 0$ as $n \to \infty$. So $f_n \to f$ uniformly.

If fact we need uniform convergence of the derivatives.

Theorem 1.6

Let $f_n : (u, v) \to \mathbb{R}$ $(n \geq 1)$ with $f_n \to f : (u, v) \to \mathbb{R}$ pointwise. Suppose further each f_n is continuously differentiable and that $f_n' \to g : (u, v) \to \mathbb{R}$ uniformly. Then f is differentiable with $f' = g$.

Proof. Fix $a \in (u, v)$. Let $x \in (u, v)$, by FTC we have each f_n' is integrable over $[a, x]$ and $\int_a^x f_n' = f_n(x) - f_n(a)$. But $f_n' \to g$ uniformly so by theorem 1.5 g is integrable over $[a, x]$ and $\int_a^x g = \lim_{n\to\infty} \int_a^x f_n' = f(x) - f(a)$. So we have shown that for all $x \in (u, v)$

$$f(x) = f(a) + \int_a^x g.$$

By theorem 1.4, g is cts so by FTC, f is differentiable with $f' = g$. □

Remark 3. It would have sufficed to assume $f_n(x) \to f(x)$ for a single value of x rather than $f_n \to f$ pointwise.

GPC?

Definition 1.5 (Uniform Cauchy)

Let $X \subset \mathbb{R}$ and let $f_n : X \to \mathbb{R}$ for each $n \geq 1$. We say (f_n) is **uniformly Cauchy** if $\forall\, \varepsilon > 0\, \exists\, N\, \forall\, m, n \geq N\, \forall\, x \in X\, |f_m(x) - f_n(x)| < \varepsilon$

Exercise 1.7. Show uniformly convergence $\implies$ uniformly Cauchy.

Theorem 1.8 (General principle of Uniform Convergence (GPUC))

Let (f_n) be a uniformly Cauchy sequence of functions $X \to \mathbb{R}$ $(X \subset \mathbb{R})$. Then (f_n) is uniformly convergent.

Proof. Let $x \in X$. Let $\varepsilon > 0$. Then $\exists\, N\, \forall\, m, n \geq N\, \forall\, y \in X\, |f_m(y) - f_n(y)| < \varepsilon$. In particular, $\forall\, m, n \geq N\, |f_m(x) - f_n(x)| < \varepsilon$. So $(f_n(x))$ is a Cauchy sequence in $\mathbb{R}$ so by GPC it converges, say $f_n(x) \to f(x)$ as $n \to \infty$.

We have now constructed $f : X \to \mathbb{R}$ s.t. $f_n \to f$ pointwise.
Let $\varepsilon > 0$. Then we can find a N s.t. $\forall\, m, n \geq N\, \forall\, y \in X\, |f_m(y) - f_n(y)| < \varepsilon$. Fix $y \in X$, keep $m \geq N$ fixed and let $n \to \infty$: $|f_m(y) - f(y)| \leq \varepsilon$. So we have shown that $\forall\, m \geq N,\ |f_m(y) - f(y)| < \varepsilon$.

But y was arbitrary so $\forall\ x \in X\ \forall\ m \geq N\ |f_m(x) - f(x)| \leq \varepsilon$. That is $f_n \to f$ uniformly. ☐

BW?

Definition 1.6 (Pointwise bounded)

Let $X \subset \mathbb{R}$ and let $f_n : X \to \mathbb{R}$ for each $n \geq 1$. We say (f_n) is **pointwise bounded** if $\forall\, x\, \exists\, M\, \forall\, n\, |f_n(x)| \leq M$.

Definition 1.7 (Uniformly bounded)

Let $X \subset \mathbb{R}$ and let $f_n : X \to \mathbb{R}$ for each $n \geq 1$. We say (f_n) is **uniformly bounded** if $\exists\, M\, \forall\, x\, \forall\, n\, |f_n(x)| \leq M$.[a]

[a] Again we have just swapped $\forall\, x\, \exists\, M$ as in convergence.

What would uniform BW say? 'If (f_n) is a uniformly bounded sequence of functions that it has a uniformly convergent subsequence'. But this is not true.

Example 1.6 (Counterexample of BW)

$$f_n : \mathbb{R} \to \mathbb{R}$$
$$f_n(x) = \begin{cases} 1 & x = n \\ 0 & x \neq n. \end{cases}$$

Obviously (f_n) uniformly bounded (by 1). However, if $m \neq n$ then $f_m(m) = 1$ and $f_n(m) = 0$ so $|f_m(m) - f_n(m)| = 1$ so no subsequence can be uniformly Cauchy so no subsequence can be uniformly convergent.

Application to power series Recall that if $\sum a_n x^n$ is a real power series with r.o.c $R > 0$ then we can differentiate/ integrate it term-by-term within $(-R, R)$.

Definition 1.8

Let $f_n : X \to \mathbb{R}$ $(X \subset \mathbb{R})$ for each $n \geq 0$. We say the series $\sum_{n=0}^{\infty} f_n$ **uniformly converges** if the sequence of partial sums (F_n) does, where $F_n = \sum_{m=0}^{n} f_m$.

We can apply theorems 1.4 to 1.6 to get e.g. if conditions hold with f_n cts diff and uniform convergence then $\sum f_n$ has derivative $\sum f_n'$.

Hope Prove $\sum a_n x^n$ converges uniformly on $(-R, R)$ then hit it with earlier theorems.

Not quite true:

Example 1.7

$\sum_{n=0}^{\infty} x^n$ r.o.c 1. This does not converge uniformly on $(-1, 1)$. Let $f(x) = \sum_{n=0}^{\infty} x^n$ and $F_n(x) = \sum_{m=0}^{n} x^m$. Note $f(x) = \frac{1}{1-x} \to \infty$ as $x \to 1$. However, $\forall\ x \in (-1, 1)$ $|F_n(x)| \leq n + 1$.

Fix any n. We can find a point $x \in (-1, 1)$ where $f(x) \geq n+2$ and so $|f(x) - F_n(x)| \geq 1$. So we don't have uniform convergence.

Back-up plan: It does work if we look at a smaller interval.

New plan: show if $0 < r < R$ then we do have uniform convergence on $(-r, r)$.
Given $x \in (-R, R)$ there's some r with $|x| < r < R$: use uniform convergence on $(-r, r)$ to check everything is nice at x. 'Local uniform convergence of power series'.

Aside

In $\mathbb{R}$ $x_n \to 0$ if

1. $\forall\, \varepsilon > 0\, \exists\, N\, \forall\, n \geq N\ |x_n| < \varepsilon$.
2. Equivalently: $\forall\, \varepsilon > 0\, \exists\, N\, \forall\, n \geq N\ |x_n| \leq \varepsilon$.

Proof. i $\implies$ ii: obvious
ii $\implies$ ii: Let $\varepsilon > 0$. Pick N s.t. $\forall\, n \geq N\ |x_n| \leq \frac{1}{2}\varepsilon$. Then $\forall\, n \geq N\ |x_n| < \varepsilon$. ☐

Also: $f_n, f : X \to \mathbb{R}$, $f_n \to f$ uniformly.

1. $\forall\, \varepsilon > 0\, \exists\, N\, \forall\, x \in X\, \forall\, n \geq N\ |f_n(x) - f(x)| < \varepsilon$.
2. For n suff large $f_n - f$ is bounded and $\forall\, \varepsilon > 0\, \exists\, N\, \forall\, n \geq N\ \sup_{x\in X} |f_n(x) - f(x)| < \varepsilon$.

Proof. ii $\implies$ i: obvious
i $\implies$ ii: if i holds then $\sup_{x\in X} |f_n(x) - f(x)| \leq \varepsilon$. But OK by same argument as previously. ☐

Lemma 1.1
Let $\sum a_n x^n$ be a real power series with r.o.c $R > 0$. Let $0 < r < R$. Then $\sum a_n x^n$ converges uniformly on $(-r, r)$.

Proof. Define $f, f_n : (-r, r) \to \mathbb{R}$ by $f(x) = \sum_{n=0}^{\infty} a_n x^n$ and $f_m(x) = \sum_{n=0}^{m} a_n x^n$. Recall that $\sum a_n x^n$ converges absolutely for all x with $|x| < R$.

Let $x \in (-r, r)$. Then f

$$\begin{aligned} |f(x) - f_m(x)| &= \left| \sum_{n=m+1}^{\infty} a_n x^n \right| \\ &\leq \sum_{n=m+1}^{\infty} |a_n||x|^n \\ &\leq \sum_{n=m+1}^{\infty} |a_n| r^n \end{aligned}$$

which converges by absolute convergence at r. Hence if m suff large, $f - f_m$ is bounded and

$$\sup_{x\in(-r,r)} |f(x) - f_m(x)| \leq \sum_{n=m+1}^{\infty} |a_n| r^n \to 0$$

as $m \to \infty$ by absolute convergence of r. ☐

Theorem 1.9

Let $\sum a_n x^n$ be a real power series with r.o.c $R > 0$. Define $f : (-R, R) \to \mathbb{R}$ by $f(x) = \sum_{n=0}^{\infty} a_n x^n$. Then

1. f is continuous;
2. for any $x \in (-R, R)$ f is integrable over $[0, x]$ with

$$\int_0^x f = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}.$$

Proof. Let $x \in (-R, R)$. Pick r s.t. $|x| < r < R$. By lemma 1.1, $\sum a_n y^n$ converges uniformly on $(-r, r)$. But the partial sum functions $y \mapsto \sum_{n=0}^{m} a_n y^n$ $(m \geq 0)$ are all cts functions on $(-r, r)$ (as they are polynomials). Hence by theorem 1.4, $f\,|_{(-r,r)}$[a] is cts. Hence f is cts at x, but x was arbitrary so f is a cts fcn on $(-R, R)$.

Moreover, $[0, x] \subset (-r, r)$ so we also have $\sum a_n y^n$ converges uniformly on $[0, x]$. Each partial sum function on $[0, x]$ is a poly so can be integrated with $\int_0^x \sum_{n=0}^{m} a_n y^n \, dy = \sum_{n=0}^{m} \int_0^x a_n y^n \, dy = \sum_{n=0}^{m} \frac{a_n}{n+1} x^{n+1}$. Hence by theorem 1.5, f is integrable over $[0, x]$ with

$$\begin{aligned} \int_0^x f &= \lim_{m\to\infty} \int_0^x \sum_{n=0}^{m} a_n y^n \, dy \\ &= \lim_{m\to\infty} \sum_{n=0}^{m} \frac{a_n}{n+1} x^{n+1} \\ &= \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}. \end{aligned}$$

□

[a] f restricted to domain $(-r, r)$

For differentiation, need technical lemma:

Lemma 1.2

Let $\sum a_n x^n$ be a real power series with r.o.c $R > 0$. Then the power series $\sum_{n\geq 1} n a_n x^{n-1}$ has r.o.c at least R.

Proof. Let $x \in \mathbb{R}$ with $0 < x < R$. Pick w with $x < w < R$. Then $\sum a_n w^n$ is absolutely convergent, so $a_n w^n \to 0$ (terms of a convergent series) so $\exists\ M$ s.t. $\forall\ n$, $|a_n w^n| \leq M$.

For each n,

$$|na_nx^{n-1}| = |a_nw^n|\left|\frac{x}{w}\right|^n \frac{1}{|x|}n.$$

Fix n. Let $\alpha = |\frac{x}{w}| < 1$. Let $c = \frac{M}{|x|}$, a constant. Then $|na_nx^{n-1}| \le cn\alpha^n$. By comparison test, ETS (enough to show) $\sum n\alpha^n$ converges.
Note $\left|\frac{(n+1)\alpha^{n+1}}{n\alpha^n}\right| = (1+\frac{1}{n})\alpha \to \alpha < 1$ as $n \to \infty$ so done by ratio test. □

Theorem 1.10

Let $\sum a_nx^n$ be a real power series with r.o.c. $R > 0$. Let $f : (-R, R) \to \mathbb{R}$ be defined by $f(x) = \sum_{n=0}^{\infty} a_nx^n$. Then f is differentiable and $\forall\, x \in (-R, R)$ $f'(x) = \sum_{n=1}^{\infty} na_nx^{n-1}$.

Proof. Let $x \in (-R, R)$. Pick r with $|x| < r < R$. Then $\sum a_ny^n$ converges uniformly on $(-r, r)$. Moreover, the power series $\sum_{n\ge1} na_ny^{n-1}$ has r.o.c at least R and so also converges uniformly on $(-r, r)$.

The partial sum functions $f_m(y) = \sum_{n=0}^{m} a_ny^n$ are polys so differentiable with $f'_m(y) = \sum_{n=1}^{m} na_ny^{n-1}$. We now have f'_m converging uniformly on $(-r, r)$ to the function $g(y) = \sum_{n=1}^{\infty} na_ny^{n-1}$.

Hence by theorem 1.6, $f\mid_{(-r,r)}$ is differentiable and $\forall\, y \in (-r, r)$ $f'(y) = g(y)$.

In particular, f is differentiable at x with $f'(x) = g(x)$. Hence f is a differentiable function on $(-R, R)$ with derivative g as desired. □

§1.4 Uniform Continuity

Let $X \subset \mathbb{R}$. Let $f : X \to \mathbb{R}$. (May as well think of $X = \mathbb{R}$ or $X = (a, b)$).

Definition 1.9 (Continuous function)

f is **continuous** if

$$\forall\, \varepsilon > 0\, \forall\, x \in X\, \exists\, \delta > 0\, \forall\, y \in X\ |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon.$$

Definition 1.10 (Unifomly Continuous function)

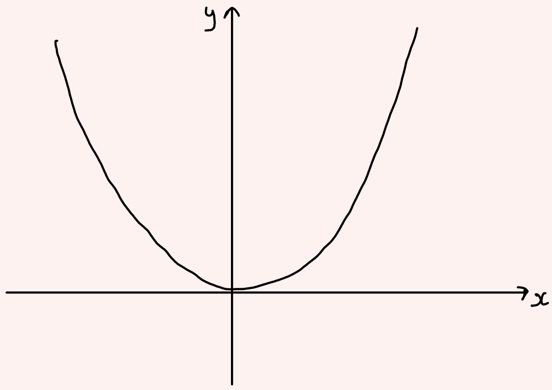
f is **uniformly continuous** if

$$\forall\, \varepsilon > 0\, \exists\, \delta > 0\, \forall\, x \in X\, \forall\, y \in X\ |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon.$$

Remark 4. Clearly if f is uniformly cts then f is cts. We would suspect that f cts doesn't imply f uniformly cts.

Example 1.8

A function $f : \mathbb{R} \to \mathbb{R}$ that is cts but not uniformly cts.



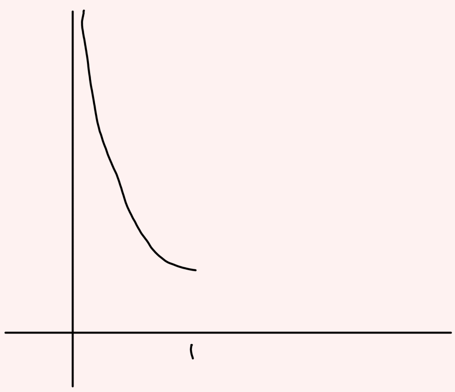
We want some function that looks like this, a continuous function which gets steeper as we go to infinity. So $f(x) = x^2$ ought to work. We know f is cts (as it's a poly). Suppose $\delta > 0$. Then

$$\begin{aligned} f(x+\delta) - f(x) &= (x+\delta)^2 - x^2 \\ &= 2\delta x + \delta^2 \to \infty \text{ as } x \to \infty. \end{aligned}$$

So in particular, $\forall\, \delta > 0\, \exists\, x, y \in \mathbb{R}$ s.t. $|x-y| < \delta$ but $|f(x)-f(y)| \geq 1$. So conditions for uniform cty fails for $\varepsilon = 1$. So f not uniform cty.

Example 1.9

Make domain bounded. We can still fail, e.g. $f : (0,1) \to \mathbb{R}$ cts but not uniform cts.



Let $f(x) = \frac{1}{x}$, clearly cts. Proof that its not uniform continuity is left as an exercise to the reader.

Theorem 1.11

A continuous real-valued function on a closed bounded interval is uniformly continuous.

Proof. Let $f : [a, b] \to \mathbb{R}$ and suppose f is cts but not uniformly cts. Then we can find $\varepsilon > 0$ st. $\forall\, \delta > 0\ \exists\, x, y \in [a, b]$ with $|x - y| < \delta$ but $|f(x) - f(y)| \geq \varepsilon$.

In particular, taking $\delta = \frac{1}{n}$ we can find sequences $(x_n), (y_n) \in [a, b]$ with, for each n, $|x_n - y_n| < \frac{1}{n}$ but $|f(x_n) - f(y_n)| \geq \varepsilon$. The sequence (x_n) is bounded so by BW[a] it has a convergent subsequence $x_{n_j} \to x$. And $[a, b]$ is a closed interval so $x \in [a, b]$. Then $x_{n_j} - y_{n_j} \to 0$ so $y_{n_j} \to x$.

But f is cts at x so $\exists\ \delta > 0$ s.t. $\forall\ y \in [a, b]\ |y - x| < \delta \implies |f(y) - f(x)| < \frac{\varepsilon}{2}$. Take such a δ. As $x_{n_j} \to x$ we can find J_1 s.t. $j \geq J_1 \implies |x_{n_j} - x| < \delta$. Similarly we can find J_2 s.t. $j \geq J_2 \implies |y_{n_j} - x| < \delta$. Now let $j = \max(J_1, J_2)$ then $|x_{n_j} - x|, |y_{n_j} - x| < \delta$ so we have $|f(x_{n_j}) - f(x)|, |f(y_{n_j}) - f(x)| < \varepsilon/2$. Then $|f(x_{n_j}) - f(y_{n_j})| \leq |f(x_{n_j}) - f(x)| + |f(y_{n_j}) - f(x)| < \varepsilon$ ↯. □

[a] Bolzano Weierstrass

Corollary 1.1

A continuous real-valued function on a closed bounded interval is bounded.

Proof. Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and so uniformly continuous by theorem 1.11. Then we can find $\delta > 0$ s.t. $\forall\ x, y \in [a, b]\ |x - y| < \delta \implies |f(x) - f(y)| < 1$.

Let $M = \lceil \frac{b-a}{\delta} \rceil$. Let $x \in [a, b]$. We can find $a = x_0 \leq x_1 \leq \cdots \leq x_M = x$ with $|x_i - x_{i-1}| < \delta$ for each i. Hence

$$\begin{aligned} |f(x)| &= \left| f(a) + \sum_{i=1}^{M} f(x_i) - f(x_{i-1}) \right| \\ &\leq |f(a)| + \sum_{i=1}^{M} |f(x_i) - f(x_{i-1})| \\ &< |f(a)| + \sum_{i=1}^{M} 1 \\ &= |f(a)| + M. \end{aligned}$$

□

Remark 5. Referring back to example 1.9, starting at $x = 1$ and going towards $x = 0$ we can that δ gets smaller and smaller s.t. you require an infinite number of steps to get 0. So $M = \infty$ essentially.

Corollary 1.2

A continuous real-valued function on a closed bounded interval is integrable.

Proof. Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and so uniformly continuous by theorem 1.11. Let $\varepsilon > 0$. Then we can find $\delta > 0$ s.t. $\forall\ x, y \in [a, b]\ |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon$. Let $\mathcal{D} = \{x_0 < x_1 < \cdots < x_n\}$ be a dissection s.t. for each i we have $x_i - x_{i-1} < \delta$.

Let $i \in \{1, \ldots, n\}$. Then for any $u, v \in [x_{i-1}, x_i]$ we have $|u - v| < \delta$ so $|f(u) - f(v)| < \varepsilon$. Hence

$$\sup_{x \in [x_{i-1}, x_i]} f(x) - \inf_{x \in [x_{i-1}, x_i]} f(x) \leq \varepsilon.$$

Hence:

$$\begin{aligned} S(f, \mathcal{D}) - s(f, \mathcal{D}) &= \sum_{i=1}^{n} (x_i - x_{i-1}) \left(\sup_{x \in [x_{i-1}, x_i]} f(x) - \inf_{x \in [x_{i-1}, x_i]} f(x) \right) \\ &\leq \sum_{i=1}^{n} (x_i - x_{i-1}) \varepsilon \\ &= \varepsilon \sum_{i=1}^{n} (x_i - x_{i-1}) \\ &= \varepsilon (b - a). \end{aligned}$$

But $\varepsilon(b - a)$ can be made arbitrarily small by taking ε small. So by Riemann's criterion f is integrable over $[a, b]$. □

§2 Metric Spaces

§2.1 Definitions and Examples

Question

Can we think about convergence in a more general setting? Convergence seemed similar in our 3 settings.

What do we really need?

Answer

We need a notion of distance.

In $\mathbb{R}$: distance x to y is $|x - y|$.
In $\mathbb{R}^2$: its $\|x - y\|$.
For functions: distance f to g is $\sup_{x \in X} |f(x) - g(x)|$ (where this exists, i.e. if $f - g$ bounded).

The triangle inequality was often important (see the proof of uniqueness of limits).

Definition 2.1 (Metric)

A **metric** d is a function $d : X^2 \to \mathbb{R}$ satisfying:

- $d(x, y) \geq 0$ for all $x, y \in X$ with equality iff $x = y$;
- $d(x, y) = d(y, x)$ for all $x, y \in X$.
- $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

Definition 2.2 (Metric Space)

A **metric space** is a set X endowed with a metric d.

We could also define a metric space as an ordered pair (X, d). If it is obvious what d is, we sometimes write 'The metric space X ...'.

Example 2.1

$X = \mathbb{R}$, $d(x, y) = |x - y|$ 'The usual metric on $\mathbb{R}$'.

Example 2.2

$X = \mathbb{R}^n$ with the Euclidean metric, $d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$.

Uniform convergence of functions doesn't quite work: we want $d(f, g) = \sup |f - g|$ but this might not exist if $f - g$ is unbounded. However, we can do something with appropriate sets of functions.

Example 2.3

Let $Y \subset \mathbb{R}$. Take $X = B(Y) = \{f : Y \to \mathbb{R} \mid f \text{ bounded}\}$ with the uniform metric $d(f, g) = \sup_{x \in Y} |f - g|$.

Checking triangle inequality:

Proof. Let $f, g, h \in B(Y)$. Let $x \in Y$. Then

$$\begin{aligned} |f(x) - h(x)| &\leq |f(x) - g(x)| + |g(x) - h(x)| \\ &\leq d(f, g) + d(g, h) \end{aligned}$$

Taking sup over all $x \in Y$

$$d(f, h) \leq d(f, g) + d(g, h).$$

□

Definition 2.3 (Subspace)

Suppose (X, d) a metric space and $Y \subset X$. Then $d\ |_{Y^2}$ is a metric on Y. We say Y with this metric is a **subspace** of X.

Example 2.4

Subspaces of $\mathbb{R}$: any of $\mathbb{Q}, \mathbb{Z}, \mathbb{N}, [0, 1], \ldots$ with the usual metric $d(x, y) = |x - y|$.

Example 2.5

Recall that a cts function on a closed bounded interval is bounded. Define $C([a, b]) = \{f : [a, b] \to \mathbb{R} \mid f \text{ cts}\}$. This is a subspace of $B([a, b])$, example 2.3. That is $C([a, b])$ is a metric space with the uniform metric $\mathcal{L}(f, g) = \sup_{x \in [a,b]} |f(x) - g(x)|$

Example 2.6

The empty metric space $X = \varnothing$ with the empty metric.

Could maybe define different metrics on the same set:

Example 2.7

The ℓ_1 metric on $\mathbb{R}^n$: $d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$.

Example 2.8

The ℓ_∞ metric on $\mathbb{R}^n$: $d(x, y) = \max_i |x_i - y_i|$.[a]

[a]Proof of triangle inequality similar to example 2.3

Example 2.9

On $C([a, b])$ we can define the L_1 metric: $d(f, g) = \int_a^b |f - g|$.

Example 2.10

$X = \mathbb{C}$ with

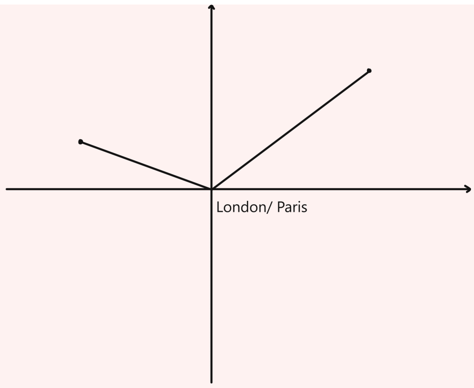
$$d(z, w) = \begin{cases} 0 & z = w \\ |z| + |w| & z \neq w. \end{cases}.$$

First two conditions of a metric hold obviously, for triangle inequality we need $d(u, w) \leq d(u, v) + d(v, w)$.

1. If $u = w$, LHS = 0 ✓
2. If $u = v$ or $v = w$ then LHS = RHS ✓
3. If u, v, w distinct:

$$LHS = |u| + |w|$$
$$RHS = |u| + |w| + 2|v| \checkmark$$

This metric is often called the British Rail metric or SNCF metric, you can think of it as for distinct points you have to travel through the origin.



Example 2.11 (Discrete metric)
Let X be any set. Define a metric d on X by

$$d(x,y) = \begin{cases} 0 & x = y \\ 1 & x \neq y. \end{cases}$$

Easy to check this works. This is called the <u>discrete metric</u> on X.

Example 2.12 (p-adic metric)
Let $\mathbb{X} = \mathbb{Z}$. Let p be a prime. The **p-adic metric** on $\mathbb{Z}$ is the metric d defined by:

$$d(x,y) = \begin{cases} 0 & x = y \\ p^{-a} & \text{if } x \neq y \text{ and } x - y = p^a m \text{ with } p \nmid m. \end{cases}$$

'Two numbers are close if difference is divisible by a large power of p'.

Only thing we need to check is triangle inequality

Proof. STP: $d(x,z) \leq d(x,y) + d(y,z)$

1. If $x = z$, LHS = 0 ✓
2. If $x = y$ or $y = z$ then LHS = RHS ✓

So easy if any two of x, y, z the same so assume x, y, z all distinct. Let $x - y = p^a m$ and $y - z = p^b n$ where $p \nmid m, p \nmid n$ and wlog $a \leq b$. So $d(x,y) = p^{-a}$ and $d(y,z) = p^{-b}$.

Now:

$$\begin{aligned} x - z &= (x - y) = (y - z) \\ &= p^a m + p^b n \\ &= p^a \underbrace{\left(m + p^{b-a} n\right)}_{\text{integer}} \text{ as } a \leq b. \end{aligned}$$

So $p^a \mid x - z$ so $d(x,z) \leq p^{-a}$. But $d(x,y) + d(y,z) \geq d(x,y) = p^{-a}$. □

[a] p^a is the largest a s.t. $p^a \mid x - y$

Definition 2.4 (Convergence)

Let (X,d) be a metric space, let (x_n) be a sequence in X and let $x \in X$. We say (x_n) **converges** to x and write '$x_n \to x$' or '$x_n \to x$ as $n \to \infty$' if

$$\forall\, \varepsilon > 0\, \exists\, N\, \forall\, n \geq N\ d(x_n, x) < \varepsilon.$$

Equivalently $x_n \to x$ iff $d(x_n, x) \to 0$ in $\mathbb{R}$.

Proposition 2.1

Limits are unique. That is, if (X,d) is a metric space, (x_n) a sequence in X, $x, y \in X$ with $x_n \to x$ and $x_n \to y$ then $x = y$.

Proof. For each n,

$$\begin{aligned} d(x,y) &\leq d(x,x_n) + d(x_n,y) \text{ by triangle inequality} \\ &\leq d(x_n,x) + d(x_n,y) \text{ by symmetry} \\ &\to 0 + 0 = 0 \text{ as } d(x_n,x), d(x_n,y) \to 0 \end{aligned}$$

So $d(x,y) \to 0$ as $n \to \infty$. But $d(x,y)$ is constant so $d(x,y) = 0$ so $x = y$. □

Remark 6. This justifies talking about the limit of a convergent sequence in a metric space, and writing $x = \lim_{n\to\infty} x_n$ if $x_n \to x$.

Remark 7 (Remarks on definition of convergence in a metric space).

1. Constant sequences obviously converge. More over, eventually constant sequences converge.
2. Suppose (X, d) is a metric space and Y is a subspace of X. Suppose (x_n) is a sequence in Y which converges in Y to x. Then also (x_n) converges in X to x.

   However, converse is false: e.g. in $\mathbb{R}$ with the usual metric then $\frac{1}{n} \to 0$ as $n \to \infty$. Consider the subspace $\mathbb{R} \setminus \{0\}$. Then $(\frac{1}{n})$ is a sequence in $\mathbb{R} \setminus \{0\}$ but it doesn't converge in $\mathbb{R} \setminus \{0\}$. (Why? Suppose $\frac{1}{n} \to x$ in $\mathbb{R} \setminus \{0\}$. Then also $\frac{1}{n} \to x$ in $\mathbb{R}$. But $\frac{1}{n} \to 0$ in $\mathbb{R}$ so by uniqueness of limits $x = 0$. But $x \in \mathbb{R} \setminus \{0\}$ and $0 \notin \mathbb{R} \setminus \{0\}$ $\lightning$.)

Example 2.13

Let d be the Euclidean metric on $\mathbb{R}^n$. Exactly as in $\mathbb{R}^2$, we have $x_n \to x$ iff the sequence converges in each coordinate in the usual way in $\mathbb{R}$.

What about other metrics on $\mathbb{R}^n$? E.g. let d_∞ be the uniform metric: $d_\infty(x, y) = \max_i |x_i - y_i|$. Which sequences converge in $(\mathbb{R}^n, d_\infty)$? $d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \le \sqrt{\sum_{i=1}^n d_\infty(x, y)^2}$ so $d(x, y) \le \sqrt{n} d_\infty(x, y)$. But also $d_\infty(x, y) \le d(x, y)$ as one of the terms in $d(x, y)$ is d_∞^2.

Now suppose (x_n) is a sequence in $\mathbb{R}^n$. Then $d(x_n, x) \to 0 \iff d(x_n, x) \to 0$. So exactly same sequences converge in $(\mathbb{R}^n, d)$ and $(\mathbb{R}^n, d_\infty)$

What about ℓ_1 metric d_1? $d_1(x, y) = \sum_{i=1}^n |x_i - y_i|$. Similarly, $d_\infty(x, y) \le d_1(x, y) \le n d_\infty(x, y)$. So again, exactly the same sequences converge in $(\mathbb{R}^n, d_1)$.

Example 2.14

Let $X = C([0, 1]) = \{f : [0, 1] \to \mathbb{R} \mid f \text{ continuous}\}$. Let d_∞ be the uniform metric on X: $d_\infty(f, g) = \sup_{x \in [0,1]} |f(x) - g(x)|$.

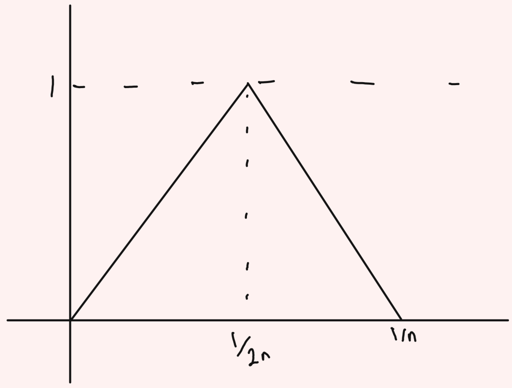
$$\begin{aligned} f_n \to f \text{ in } (X, d_\infty) &\iff d_\infty(f_n, f) \to 0 \\ &\iff \sup_{x \in [0,1]} |f_n(x) - f(x)| \to 0 \\ &\iff f_n \to f \text{ uniformly.} \end{aligned}$$

We also have L_1-metric d_1 on X: $d_1(f, g) = \int_0^1 |f - g|$. Now $d_1(f, g) = \int_0^1 |f - g| \le \int_0^1 d_\infty(f, g) = d_\infty(f, g)$. So similarly to previous example,

$$f_n \to f \text{ in } (X, d_\infty) \implies f_n \to f \text{ in } (X, d_1).$$

But converse does not hold, i.e. we can find a sequence (f_n) in X s.t. $f_n \to 0$ in

d_1-metric but f_n doesn't converge in d_∞-metric, i.e. $\int_0^1 |f_n| \to 0$ as $n \to \infty$ but (f_n) does not converge uniformly.



$$f_n(x) = \begin{cases} 2nx & x \le \frac{1}{2n} \\ 2n(\frac{1}{n} - x) & \frac{1}{2n} < x \le \frac{1}{n} \\ 0 & x > \frac{1}{n}. \end{cases}$$

Then $d_1(f_n, 0) = \frac{1}{2} \times \frac{1}{n} \times 1 = \frac{1}{2n} \to 0$. So in (X, d_1) we have $f_n \to 0$. But f_n does not converge uniformly: indeed, $f_n \to 0$ pointwise; if we have uniform convergence then uniform limit is the same as pointwise limit; but $\forall\ n\ f_n(\frac{1}{2n}) = 1$ so $f_n \not\to 0$ uniformly.

Example 2.15

Let (X, d) be a discrete metric space; $d(x, y) = \begin{cases} 0 & x = y \\ 1 & x \ne y \end{cases}$. When do we have $x_n \to x$ if (X, d)?

Suppose $x_n \to x$, i.e. $\forall\ \varepsilon > 0\ \exists\ N\ \forall\ n \ge N\ d(x_n, x) < \varepsilon$. Setting $\varepsilon = 1$ in this, we can find N s.t. $\forall\ n \ge N\ d(x_n, x) < 1$, i.e. $\forall\ n \ge N\ d(x_n, x) = 0$ i.e. $\forall\ n \ge N\ x_n = x$. Thus (x_n) is eventually constant.
But we know in any metric space, eventually constant sequences converge.

So in this space, (x_n) converges iff (x_n) eventually constant.

Definition 2.5 (Continuity)

Let (X, d) and (Y, e) be metric spaces and let $f : X \to Y$.

1. Let $a \in X$ and $b \in Y$. We say $f(x) \to b$ as $x \to a$ if $\forall\ \varepsilon > 0\ \exists\ \delta > 0\ \forall\ x \in X\ 0 <$

$d(x,a) < \delta \implies e(f(x), b) < \varepsilon$.

2. Let $a \in X$. We say f is **continuous** at a if $f(x) \to f(a)$ as $x \to a$. That is: $\forall\, \varepsilon > 0\, \exists\, \delta > 0\, \forall\, x \in X\ d(x,a) < \delta \implies e(f(x), f(a)) < \varepsilon$.

3. If $\forall\, a \in X$ f is continuous at a we say f is a **continuous** function or simply f is **continuous**.

4. We say f is **uniformly continuous** if $\forall\, \varepsilon > 0\, \exists\, \delta > 0\, \forall\, x, y \in X\ d(x,y) < \delta \implies e(f(x), f(y)) < \varepsilon$

5. Suppose $W \subset X$. We say f is **continuous on W** (respectively **uniformly continuous on W**) if the function $f\,|_W$ is continuous (resp. uniformly continuous), as a function from $W \to Y$ where we are now thinking of W as a subspace of X.

Remark 8. 1. Don't have a nice rephrasing of item 1 in terms of similar concepts in the reals. We would want to write '$e(f(x), b) \to 0$ as $d(x,a) \to 0$'. But this is meaningless, we haven't defined such a concept in the reals.

2. Item 1 says nothing about what happens at the point a itself. E.g. let $f : \mathbb{R} \to \mathbb{R}, f(x) = \begin{cases} 1 & x = 0 \\ 0 & x \neq 0 \end{cases}$. Then $f(x) \to 0$ as $x \to 0$ (but $f(0) \neq 0$ so f is not continuous at 0).
If we have f cts then $d(x,a) = 0 \implies x = a \implies f(x) = f(a) \implies e(f(x), f(a)) = 0$. So we can drop the '$0 <$' from definition of continuity.

3. We can rewrite item 5: f is continuous on W iff $f\,|_W$ is a continuous function $f\,|_W$: $W \to Y$ thinking of W as a subspace of X. That is: $\forall\, a \in W\, \forall\, \varepsilon > 0\, \exists\, \delta > 0\, \forall\, x \in X\ d(x,a) < \delta \implies e(f(x), f(a)) < \varepsilon$. In particular, note the subtlety that this <u>only</u> mentions points of W. So under this definition, e.g. $f : \mathbb{R} \to \mathbb{R}, f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$ then $f\,|_{[0,1]}$ is cts. But f is not cts at points $0, 1$.

Proposition 2.2

Let $(X,d), (Y,e)$ be metric spaces, $f : X \to Y$ and $a \in X$. Then f is continuous at a iff whenever (x_n) is a sequence in X with $x_n \to a$ then $f(x_n) \to f(a)$.

Proof. ($\implies$): Suppose f is cts at a. Let (x_n) be a sequence in X with $x_n \to a$. Let $\varepsilon > 0$. As f cts at a we can find $\delta > 0$ s.t. $\forall\, x \in X$ s.t. $d(x,a) < \delta \implies e(f(x), f(a)) < \varepsilon$. As $x_n \to x$ we can find N s.t. $n \geq N \implies d(x_n, a) < \delta$. Let $n \geq N$ then $d(x_n, a) < \delta$ so $e(f(x_n), f(a)) < \varepsilon$. Hence $f(x_n) \to f(a)$.

($\impliedby$): Suppose f is not cts at a. Then there is some $\varepsilon > 0$ s.t. $\forall\, \delta > 0\, \exists\, x \in X$ with

$d(x, a) < \delta$ but $e(f(x), f(a)) \geq \varepsilon$. Now take $\delta = \frac{1}{n}$ we obtain a sequence (x_n) with, for each n $d(x_n, a) < \frac{1}{n}$ but $e(f(x_n), f(a)) > \varepsilon$. Hence $x_n \to a$ but $f(x_n) \not\to f(a)$. □

Proposition 2.3

Let $(W, c), (X, d), (Y, e)$ be metric spaces, left $f : W \to X$, let $g : X \to Y$ and let $a \in W$. Suppose f is cts at a and g is cts at $f(a)$. Then $g \circ f$ is cts at a.

Proof. Let (x_n) be a sequence in W with $x_n \to a$. Then by proposition 2.2, $f(x_n) \to f(a)$ and so also $g(f(x_n)) \to g(f(a))$. So by proposition 2.2 $g \circ f$ cts at a. □

Example 2.16

In $\mathbb{R} \to \mathbb{R}$ with the usual metric, this is the same definition as when we defined continuity directly for $\mathbb{R}$ only. So we already have lots of cts fcns $\mathbb{R} \to \mathbb{R}$: polynomials, sin, e^x, ...

Example 2.17

Constant functions are continuous. Also if X is any metric space and $f : X \to X$ by $f(x) = x$ for all $x \in X$ (the indentity function) then that is continuous.

Example 2.18 (Projection Maps)

Consider $\mathbb{R}^n$ with the usual metric and $\mathbb{R}$ with the usual metric. The **projection maps** $\pi_i : \mathbb{R}^n \to \mathbb{R}$ given by $\pi_i(x) = x_i$ are continuous.

(Why? We've seen convergence in $\mathbb{R}^n$ of sequences is the same as convergence in each coordinate. Let's denote a sequence in $\mathbb{R}^n$ by $(x^{(m)})_{m \geq 1}$. So e.g. $x_5^{(3)}$ is the 5th coord of the 3rd term. We know $x^{(m)} \to x$ iff for each i $x_i^{(m)} \to x_i$, i.e. for each i $\pi_i(x^{(m)}) \to \pi_i(x)$. Then we can use proposition 2.2)

Similarly, suppose $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$ Let $f : \mathbb{R} \to \mathbb{R}^n$ defined by $f(x) = (f_1(x), \dots, f_n(x))$. Then f is cts at a point iff all of $f_1, \dots, f_n$ are. Using these facts example 2.16 and proposition 2.3, we have many cts fcns $\mathbb{R}^n \to \mathbb{R}^m$. E.g. $f : \mathbb{R}^3 \to \mathbb{R}^2, f(x, y, z) = (e^{-x} \sin y, 2x \cos z)$ is cts. (Why? write $w = (x, y, z) \in \mathbb{R}^3$, we have $f_1(w) = e^{-\pi_1(w)} \sin \pi_2(w)$ and $f_2(w) = 2\pi_1(w) \cos \pi_3(w)$. So f_1, f_2 cts so f cts.)

Example 2.19

Recall that if we have the Euclidean metric, the l_1 or l_∞ metric on $\mathbb{R}^n$ then the convergent sequences are the same. So by proposition 2.2, the ctf fcns $X \to \mathbb{R}^n$ or from $\mathbb{R}^n \to Y$ are the same with each of these three metrics.

Example 2.20

Let (X, d) be the discrete metric space, example 2.11, and let (Y, e) be any metric space. Which functions $f : X \to Y$ are cts? Suppose $a \in X$ and (x_n) a sequence in X with $x_n \to a$. Then (x_n) is eventually constant, i.e. for sufficiently large n $x_n = a$ and so $f(x_n) = f(a)$. So $f(x_n) \to f(a)$.

Hence every function on a discrete metric space is cts.

§2.2 Completeness

Question

In section 1 we saw a version of GPC held in each of the three examples we considered. Does GPC hold in a general metric space?

Definition 2.6 (Cauchy Sequences)

Let (X, d) be a metric space and let (x_n) be a sequence in X. We say (x_n) is **Cauchy** if $\forall\, \varepsilon > 0\, \exists\, N\, \forall\, m, n \geq N\ d(x_m, x_n) < \varepsilon$.

Theorem 2.1

(x_n) convergent $\implies$ (x_n) Cauchy.

Proof. Left as an exercise. □

But converse is not true in general.

Example 2.21

Let $X = \mathbb{R} \setminus \{0\}$ with the usual metric and $x_n = \frac{1}{n}$. We say previously that (x_n) does not converge.
Note that X is a subspace of $\mathbb{R}$. In $\mathbb{R}$ (x_n) is convergent $(x_n \to 0)$ so (x_n) is Cauchy in $\mathbb{R}$ so (x_n) is Cauchy in X.

Example 2.22

$\mathbb{Q}$ with the usual metric. Let x_n be $\sqrt{2}$ to n decimal places. This converges in $\mathbb{R}$ so is Cauchy in $\mathbb{Q}$ but clearly doesn't converge in $\mathbb{Q}$.

Definition 2.7 (Completeness)

Let (X, d) be a metric space. We say X is **complete** if every Cauchy sequence in X converges.

Example 2.23

Example 2.21 says $\mathbb{R} \setminus \{0\}$ with the usual metric is not complete.
Similarly $\mathbb{Q}$ with usual metric is not complete.

Example 2.24

GPC says $\mathbb{R}$ with the usual metric is complete.

Example 2.25

GPC for $\mathbb{R}^n$ says $\mathbb{R}^n$ with Euclidean metric is complete.

Example 2.26

GPUC, theorem 1.8, (almost) says if $X \subset \mathbb{R}$ and $B(X) = \{f : X \to \mathbb{R} \mid f \text{ is bounded}\}$ with the uniform norm then $B(X)$ is complete.

Proof. Let (f_n) be a Cauchy sequence in $B(X)$. Then (f_n) is uniformly Cauchy so by GPUC is uniformly convergent. That is $f_n \to f$ uniformly for some $f : X \to \mathbb{R}$. As $f_n \to f$ uniformly we know $f_n - f$ is bounded for n suff. large. Take such an n, then $f_n - f$ and f_n are bounded so $f = f_n - (f_n - f)$ is bounded. That is, $f \in B(X)$. Finally, $f_n \to f$ uniformly and $d(f_n, f) \to 0$, i.e. $f_n \to f$ in $(B(X), d)$. □

Remark 9. In many ways, this is typical of a proof that a given space (X, d) is complete:

1. Take (x_n) Cauchy in X;
2. Construct/ find a putative limit object x where it seems (x_n) converges to x in some sense;
3. Show $x \in X$,
4. Show $x_n \to x$ in metric space (X, d) i.e. that $d(x_n, x) \to 0$.

This is often tricky/ fiddly/ annoying/ repetitive/ boring. But we need to take care as for example, it's tempting to talk about $d(x_n, x)$ while doing (ii) or (iii); but makes no sense to write '$d(x_n, x)$' until we have completed (iii) as d is only defined on X^2 (if $x \notin X$ then can't use d).

Example 2.27

If $[a, b]$ is a closed interval then $C([a, b])$ with uniform norm d is complete.

Proof. (i): Let (f_n) be a Cauchy sequence in $C([a, b])$.
(ii): We know $C([a, b])$ is a subspace of $B([a, b])$ with uniform metric. We know $B([a, b])$ is complete by example 2.26 and (f_n) is a Cauchy sequence in $B([a, b])$ so in $B([a, b])$, $f_n \to f$ for some f.
(iii) Each f_n is cts and $f_n \to f$ uniformly so f is cts, i.e. $f \in C([a, b])$.
(iv) Finally, each $f_n \in C([a, b])$, $f \in C([a, b])$ and $f_n \to f$ uniformly so $d(f_n, f) \to 0$. □

This generalises:

Definition 2.8 (Closed Metric Space)

Let (X, d) be a metric space and $Y \subset X$. We say Y is **closed** if whenever (x_n) a sequence in Y with $x_n \to x \in X$ then $x \in Y$.

Proposition 2.4

A closed subset of a complete metric space is complete.

Remark 10. This does make sense: if $Y \subset X$ then Y is itself a metric space or a subspace of X so we can say e.g. 'Y is complete' to mean the metric space Y (as a subspace of X) is complete.

We could do exactly the same with any other properties of metric spaces we define.

Proof. Let (X, d) be a metric space and $Y \subset X$ with X complete and Y closed. (i): Let (x_n) be a Cauchy sequence in Y.
(ii): Now (x_n) is a Cauchy sequence in X so by completeness $x_n \to x$ in X for some $x \in X$.
(iii) $Y \subset X$ is closed so $x \in Y$.
(iv) Finally we now have each $x_n \in Y, x \in Y$ and $x_n \to x$ in X, so $d(x_n, x) \to 0$ so $x_n \to x$ in Y. □

Example 2.28

Define $\ell_1 = \{(x_n)_{n\geq 1} \in \mathbb{R}^{\mathbb{N}} \mid \sum_{n=1}^{\infty} |x_n| \text{ converges}\}$. Define a metric d on ℓ_1 by $d((x_n),(y_n)) = \sum_{n=1}^{\infty} |x_n - y_n|$.

Note we have $\sum |x_n|, \sum |y_n|$ converge as we are in ℓ_1. For each n $|x_n - y_n| \leq |x_n| + |y_n|$ so by comparison test $\sum |x_n - y_n|$ converges. So d is well-defined. Easy to check d is a metric on ℓ_1. Then (ℓ_1, d) is complete.

Proof. (i): Let $(x^{(n)})_{n\geq 1}$ be a Cauchy sequence in ℓ_1, so for each n $(x_i^{(n)})_{i\geq 1}$ is a sequence in $\mathbb{R}$ with $\sum_{i=1}^{\infty} |x_i^{(n)}|$ convergent.
(ii) For each i, $(x_i^{(n)})_{n\geq 1}$ is a Cauchy sequence in $\mathbb{R}$, since if $y, z \in \ell_1$ then $|y_i - z_i| \leq d(y,z)$. But $\mathbb{R}$ is complete, so for each i we can find $x_i \in \mathbb{R}$ s.t. $x_i^{(n)} \to x_i$ as $n \to \infty$. Let $x = (x_1, x_2, \dots) \in \mathbb{R}^{\mathbb{N}}$.
(iii) We next show $x \in \ell_1$, i.e. that $\sum_{i=1}^{\infty} |x_i|$ converges.

Given $y \in \ell_1$, define $\sigma(y) = \sum_{i=1}^{\infty} |y_i|$, i.e. $\sigma(y) = d(y,z)$ where z is the constant zero sequence.
We now have, for any m, n

$$\begin{aligned}\sigma(x^{(m)}) &= d(x^{(m)}, z) \\ &\leq d(x^{(m)}, x^{(n)}) + d(x^{(n)}, z) \\ &= d(x^{(m)}, x^{(n)}) + \sigma(x^{(n)})\end{aligned}$$

So $\sigma(x^{(m)}) - \sigma(x^{(n)}) \leq d(x^{(m)}, x^{(n)})$. Similarly, for any m, n $\sigma(x^{(n)}) - \sigma(x^{(m)}) \leq d(x^{(m)}, x^{(n)})$ and so $|\sigma(x^{(m)}) - \sigma(x^{(n)})| \leq d(x^{(m)}, x^{(n)})$. Hence $(\sigma(x^{(m)}))_{m\geq 1}$ is a Cauchy sequence in $\mathbb{R}$, and so by GPC converges, say $\sigma(x^{(m)}) \to K$ as $m \to \infty$.

Claim 2.1

For any $I \in \mathbb{N}$, $\sum_{i=1}^{I} |x_i| \leq K + 2$.

Proof. As $\sigma(x^{(n)}) \to K$ as $n \to \infty$ we can find N_1 s.t. $n \geq N_1 \implies \sum_{i=1}^{\infty} |x_i^{(n)}| \leq K + 1$. Also, $n \geq N_1 \implies \sum_{i=1}^{I} |x_i^{(n)}| \leq K + 1$ (as each term non-negative).

Next, for each $i \in \{1, 2, \dots, I\}$ we have $x_i^{(n)} \to x_i$ as $n \to \infty$. So we can find N_2 s.t. $n \geq N_2 \implies \forall\, i \in \{1, \dots, I\}\ |x_i^{(n)} - x_i| < I^{-1}$.

Now let $n = \max(N_1, N_2)$ then $\sum_{i=1}^{I} |x_i| \leq \sum_{i=1}^{I} |x_i^{(n)}| + \sum_{i=1}^{I} |x_i^{(n)} - x_i| \leq K + 1 + I(I^{-1}) = K + 2$. □

Now the partial sums of $\sum |x_i|$ are increasing and bounded above so $\sum |x_i|$ converges. That is $x \in \ell_1$.

(iv) Finally, need to check $x^{(n)} \to x$ as $n \to \infty$ in ℓ_1, i.e. that $d(x^{(n)}, x) \to 0$ as $n \to \infty$.
We have, for all n, I:

$$\begin{aligned} d(x^{(n)}, x) &= \sum_{i=1}^{\infty} |x_i^{(n)} - x_i| \\ &\le \sum_{i=1}^{I} |x_i^{(n)} - x_i| + \sum_{i=I+1}^{\infty} |x_i^{(n)}| + \sum_{i=I+1}^{\infty} |x_i|. \end{aligned}$$

Let $\varepsilon > 0$. We know $\sum |x_i|$ convergent (as $x \in \ell_1$) so we can pick I_1 s.t. $\sum_{i=I_1+1}^{\infty} |x_i| < \varepsilon$.
As $(x^{(n)})$ is Cauchy, we can find N_1 s.t. $m, n \ge N_1 \implies d(x^{(m)}, x^{(n)}) < \varepsilon$. As $\sum_i |x_i^{(N_1)}|$ converges, we can find I_2 s.t. $\sum_{i=I_2+1}^{\infty} |x_i^{(N_1)}| < \varepsilon$. Then

$$\begin{aligned} n \ge N_1 \implies \sum_{i=I_2+1}^{\infty} |x_i^{(n)}| &\le \sum_{i=I_2+1}^{\infty} |x_i^{(N_1)}| + \sum_{i=I_2+1}^{\infty} |x_i^{(n)} - x_i^{(N_1)}| \\ &< \varepsilon + d(x^{(n)}, x^{(N_1)}) \\ &< 2\varepsilon. \end{aligned}$$

Let $I = \max(I_1, I_2)$. For each $i = 1, 2, \dots, I$ we have $|x_i^{(n)} - x_i| \to 0$ as $n \to \infty$, so $\sum_{i=1}^{I} |x_i^{(n)} - x_i| \to 0$ as $n \to \infty$. Hence we can find N_2 s.t $n \ge N_2 \implies \sum_{i=1}^{I} |x_i^{(n)} - x_i| < \varepsilon$. Let $N = \max(N_1, N_2)$ and let $n \ge N$. Then

$$\begin{aligned} d(x^{(n)}, x) &\le \sum_{i=1}^{I} |x_i^{(n)} - x_i| + \sum_{i=I+1}^{\infty} |x_i^{(n)}| + \sum_{i=I+1}^{\infty} |x_i| \\ &\le \sum_{i=1}^{I} |x_i^{(n)} - x_i| + \sum_{i=I_2+1}^{\infty} |x_i^{(n)}| + \sum_{i=I_1+1}^{\infty} |x_i| \\ &< \varepsilon + 2\varepsilon + \varepsilon = 4\varepsilon \end{aligned}$$

Hence $d(x^{(n)}, x) \to 0$ as $n \to \infty$, i.e. $x^{(n)} \to x$ in ℓ_1.
Hence ℓ_1 is complete. □

Now we will move on to main theorem of completeness.

Definition 2.9 (Contraction mapping)

Let (X, d) be a metric space and $f : X \to X$. We say f is a **contraction** if $\exists\, \lambda \in [0, 1)$ s.t. $\forall\, x, y \in X$ $d(f(x), f(y)) \le \lambda d(x, y)$.

Theorem 2.2 (The Contraction Mapping Theorem)

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

Proof. Let $\lambda \in [0, 1)$ satisfy $\forall\, x, y \in X\ d(f(x), f(y)) \le \lambda d(x, y)$.
Let $x_0 \in X$. Recursively define $x_n = f(x_{n-1})$ for $n \ge 1$. Let $\Delta = d(x_0, x_1)$. Then, by induction $d(x_n, x_{n+1}) \le \lambda^n \Delta$ for all n.
Now suppose $N \le m < n$. Then

$$\begin{aligned} d(x_m, x_n) &\le \sum_{i=m}^{n-1} d(x_i, x_{i+1}) \\ &\le \sum_{i=m}^{n-1} \lambda^i \Delta \\ &\le \sum_{i=N}^{\infty} \lambda^i \Delta \\ &= \frac{\lambda^N \Delta}{1-\lambda} \to 0 \text{ as } N \to \infty. \end{aligned}$$

So $\forall\, \varepsilon > 0\, \exists\, N\, \forall\, m, n \ge N\ d(x_m, x_n) < \varepsilon$ (i.e. we take N s.t. $\frac{\lambda^N \Delta}{1-\lambda} < \varepsilon$). Thus (x_n) is Cauchy, so by completeness converges, say $x_n \to x \in X$. But also $x_n = f(x_{n-1}) \to f(x)$ as f continuous[a]. So by uniqueness of limits, $f(x) = x$.

Suppose also $f(y) = y$ for some $y \in X$. Then $d(x, y) = d(f(x), f(y)) \le \lambda d(x, y)$ with $\lambda < 1$. So $d(x, y) = 0$ so $x = y$. □

[a] follows immediately from definition of contraction mapping (e.g. let $\delta = \varepsilon$ in definition of continuity).

Remark 11.

1. Why is f cts? We have, for all $x, y \in X\ d(f(x), f(y)) \le d(x, y)$. So $\forall\, \varepsilon > 0$, $d(x, y) < \varepsilon \implies d(f(x), f(y)) < \varepsilon$. (Indeed, this shows f is uniformly continuous.)
2. We have proved more than claimed. Not only does f have a unique fixed point, but starting from any point of the space and repeatedly apply f then the resulting sequence converges to the fixed point. In fact, the speed of convergence is exponential.

Example 2.29 (Application)

Suppose we want to numerically approximate the solution to $\cos x = x$. Any root must lie in $[-1, 1]$. Consider metric space $X = [-1, 1]$ with usual metric. X is a closed subset of a complete space $\mathbb{R}$ so is complete. Obviously X is non-empty.

Think of $\cos : [-1, 1] \to [-1, 1]$. Suppose $x, y \in [-1, 1]$.

$$\begin{aligned} |\cos x - \cos y| &= |x - y||\cos' z| \text{ for some } z \in [-1, 1] \text{ by MVT} \\ &|x - y||-\sin' z| \\ &\le |x - y| \sin 1 \end{aligned}$$

But $0 \le \sin 1 < 1$ so $\cos$ is a contraction of $[-1, 1]$. So by The Contraction Mapping Theorem, $\cos$ has a unique fixed point in $[-1, 1]$. That is $\cos x = x$ has a unique solution.

How do we find it numerically? Use remark 2, we will have rapid convergence to the root.

We will see two major applications of CMT (The Contraction Mapping Theorem) later.

§2.3 Sequential Compactness

Recall BW for $\mathbb{R}^n$ says a bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Definition 2.10 (Bounded)

Let (X, d) be a metric space. We say X is **bounded** if

$$\exists\, M \in \mathbb{R}\, \forall\, x, y \in X\ d(x, y) \le M.$$

Remark 12. Easy to check by triangle inequality that X bounded $\iff$ ($X = \varnothing$ or $\exists\, M \in \mathbb{R}$, $x \in X$ s.t. $\forall\ y \in X\ d(x, y) \le M$). So definition agrees with earlier definition for subsets of $\mathbb{R}^n$.

Definition 2.11 (Closed subspace)

Let (X, d) be a metric space and $Y \subset X$. We say Y is **closed** in X if whenever (x_n) is a sequence in Y with, in X, $x_n \to x \in X$ then $x \in Y$.

Definition 2.12 (Sequentially Compact)

A metric space is **sequentially compact** if every sequence has a convergent subsequence.

BW for $\mathbb{R}^n$ is essentially the following:

Theorem 2.3

Let $X \subset \mathbb{R}^n$ with the Euclidean metric. Then X is sequentially compact iff X is closed and bounded.

Proof. ($\Longleftarrow$) Suppose X is closed and bounded. Let (x_n) be a sequence in X. Then (x_n) is a bounded sequence in $\mathbb{R}^n$ so by BW, in $\mathbb{R}^n$, $x_{n_j} \to x$ for some $x \in \mathbb{R}^n$ and some subsequence (x_{n_j}) of (x_n).
As X is closed, $x \in X$. Hence the subsequence (x_{n_j}) converges in X. So X is sequentially compact.

($\Longrightarrow$) Suppose X is not closed. Then we can find a sequence (x_n) in X s.t. in $\mathbb{R}^n$ $x_n \to x \in \mathbb{R}^n$ with $x \notin X$. Now any subsequence $(x_{n_j}) \to x$ in $\mathbb{R}^n$. But $x \notin X$ so by uniqueness of limits (x_{n_j}) does not converge in X. So X is not sequentially compact.

Suppose instead X is not bounded. Then we can find a sequence (x_n) in X with $\forall\ n\ \|x_n\| \geq n$, i.e. $\|x_n\| \to \infty$ as $n \to \infty$. Suppose we have a subsequence $x_{n_j} \to x \in X$. Then $\left\|x_{n_j}\right\| \to \|x\|$ but $\left\|x_{n_j}\right\| \to \infty$ ↯. So, again, X is not sequentially compact. □

Remark 13. Does this hold in a general metric space? Obviously not: e.g. in $\mathbb{R} \setminus \{0\}$ with the usual metric, the set $[-1, 0) \cup (0, 1]$ is closed and bounded but the sequence $(\frac{1}{n})_{n\geq 1}$ has no convergent subsequence.

Problem: Space is not complete. Maybe complete and bounded $\Longrightarrow$ sequentially compact?

Even this doesn't work. Recall example from section 1: Let $X = \{f \in B(\mathbb{R}) : \sup_{x\in\mathbb{R}} |f(x)| \leq 1\}$ with uniform metric. Then X is complete (closed subset of a complete space $B(\mathbb{R})$) and bounded (if $f, g \in X$ then $d(f, g) \leq 2$). But consider $f_n(x) = \begin{cases} 1 & x = n \\ 0 & x \neq n \end{cases}$. Then f_n is a sequence in X but $\forall\ m, n\ m \neq n \implies d(f_m, f_n) = 1$. So (f_n) cannot have a convergent subsequence.

Problem: X is 'too big'.

We need a stronger concept of boundedness.

Definition 2.13 (Totally Bounded)

Let (X, d) be a metric space. We say X is **totally bounded** if $\forall\ \delta > 0$ we can find a finite subset $A \subset X$ s.t. $\forall\ x \in X\ \exists\ a \in A\ d(x, a) < \delta$.

Theorem 2.4

A metric space is sequentially compact iff it is complete and totally bounded.

Proof. ($\Longleftarrow$): Suppose the metric space (X, d) is complete and totally bounded. Let $(x_n)_{n\geq 1}$ be a sequence in X.
As X is totally bounded, we can find finite $A_1 \subset X$ s.t. $\forall\, x \in X\, \exists\, a \in A_1\, d(x, a) < 1$. In particular, there is an infinite set $N_1 \subset \mathbb{N}$ and a point $a_1 \in A_1$ s.t. $\forall\ n \in N_1\ d(x_n, a_1) < 1$. Hence $\forall\, m, n \in N_1\ d(x_m, x_n) < 2$.
Similarly, we can find finite $A_2 \subset X$ s.t. $\forall\ x \in X\ \exists\ a \in A_2\ d(x, a) < \frac{1}{2}$. In particular, there is an infinite $N_2 \subset N_1$ s.t. $\forall\ n \in N_2\ d(x_n, a_2) < \frac{1}{2}$ and thus $\forall\, m, n \in N_2\ d(x_m, x_n) < 1$.
Keep going. We get a sequence $N_1 \supset N_2 \supset N_3 \ldots$ of infinite subsets of $\mathbb{N}$ s.t. $\forall\, i\, \forall\, m, n \in N_i \implies d(x_m, x_n) < \frac{2}{i}$.
Now pick $n_1 \in N_1$. Then pick $n_2 \in N_2$ with $n_2 > n_1$. Then pick $n_i \in N_i$ with $n_i > n_{i-1}$.
We obtain a subsequence (x_{n_j}) of (x_n) s.t. $\forall\ j\ x_{n_j} \in N_j$. Thus if $i \leq j$ then $x_{n_i}, x_{n_j} \in N_i$ and so $d(x_{n_i}, x_{n_j}) < \frac{2}{i}$. Hence (x_{n_j}) is a Cauchy sequence and hence by completeness, converges. Thus X is sequentially compact.

($\implies$) Suppose X is not complete. Then X has a Cauchy sequence (x_n) which doesn't converge. Suppose we have a convergent subsequence, say $x_{n_j} \to x$. Then $x_n \to x$ (left as an exercise, same as in gpc for $\mathbb{R}$) $\lightning$.

Suppose instead X not totally bounded. Then there is some $\delta > 0$ s.t. whenever $A \subset X$ is finite $\exists\ x \in X\ \forall\ a \in A\ d(x, a) < \delta$. So pick $x_1 \in X$, pick $x_2 \in X$ s.t. $d(x_1, x_2) \geq \delta$, pick $x_3 \in X$ s.t. $d(x_1, x_3) \geq \delta$ and $d(x_2, x_3) \geq \delta$ ...We get a sequence $(x_n) \in X$ s.t. $\forall\ i, j\ i \neq j \implies d(x_i, x_j) \geq \delta$. Hence (x_n) has no convergent subsequence. $\square$

Exercise 2.5. A cts fcn on a sequentially compact metric space is uniformly compact. If the fcn is real-valued then it's bounded and attains its bounds.

§2.4 The Topology of Metric Spaces

Theme of section 2: to generalise convergence/ continuity, all we need is a distance.

But: e.g. in $\mathbb{R}^n$ we have three very different concepts of distance given by the Euclidean, ℓ_1 and ℓ_∞ metrics. But all give same concept of convergence and continuity.

Definition 2.14 (Homeomorphism)

Let (X, d) and (Y, e) be metric spaces. Let $f : X \to Y$. We say f is a **homeomorph-**

ism and that X, Y are **homeomorphic** if f is a cts bijection with a cts inverse.

Remark 14. Homeomorphism is an equivalence 'relation' (it satisfies symmetry, transitivity and reflexivity but not an actual relation).

Example 2.30

If $x, y \in \mathbb{R}^n$: $d_\infty(x,y) \leq d_1(x,y) \leq n d_\infty(x,y)$. ($d_1, d_\infty$ ℓ_1 and ℓ_∞ metrics respectively). So identity map $\mathbb{R}^n \to \mathbb{R}^n$ is continuous as map $(\mathbb{R}^n, d_1) \to (\mathbb{R}^n, d_\infty)$ and as a map $(\mathbb{R}^n, d_\infty) \to (\mathbb{R}^n, d_1)$. So it's a homeomorphism.
Similarly, $\mathbb{R}^n$ with the Euclidean metric is homeomorphic to both of these spaces.

Example 2.31

Same argument would show: If $(X,d), (Y,e)$ metric spaces and $f : X \to Y$ is a bijection satisfying

1. $\exists\, A\, \forall\, x, y \in X\ e(f(x), f(y)) \leq A d(x,y)$
2. $\exists\, B\, \forall\, x, y \in X\ d(x,y) \leq B e(f(x), f(y))$ then f, f^{-1} are cts so X, Y homeomorphic.

Example 2.32

Define $f : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ by $f(x) = \tan x$. Then f is a homeomorphism (usual metric in each case). But there is no constant A s.t. $\forall\ x, y \in (-\frac{\pi}{2}, \frac{\pi}{2})$ $|\tan x - \tan y| \leq A|x-y|$.

Proposition 2.5

Let $(V,b), (W,c), (X,d), (Y,e)$ be metric spaces and $f : X \to V$, $g : Y \to W$ be homeomorphisms.

1. In X, $x_n \to x$ iff in V $f(x_n) \to f(x)$;
2. A function $h : X \to Y$ is cts at $a \in X$ iff $g \circ h \circ f^{-1}$ is cts at $f(a) \in V$.

Proof. (i) $x_n \to x \implies f(x_n) \to f(x)$ as f cts; $f(x_n) \to f(x) \implies x_n = f^{-1}(f(x_n)) \to f^{-1}(f(x)) = x$ as f^{-1} cts.

(ii) h cts $\implies g \circ h \circ f^{-1}$ cts (composition of cts fcns). And $g \circ h \circ f^{-1}$ cts $\implies h = g^{-1} \circ (g \circ h \circ f^{-1}) \circ f$ is cts (similarly). □

We now have examples of metric spaces that look very different but behave identically w.r.t convergence/ continuity.

Thought: Could we dispense with distance altogether?

Another way to think about continuity.

Definition 2.15 (Open ball)

Let (X, d) be a metric space, let $a \in X$ and let $\varepsilon > 0$. The **open ball of radius ε about a** is the set $B_\varepsilon(a) = \{x \in X : d(x, a) < \varepsilon\}$.

Remark 15. Suppose $f : X \to Y$, $a \in X$. d metric on X, e metric on Y.

$$\begin{aligned} f \text{ cts at } a &\iff \forall\, \varepsilon > 0\, \exists\, \delta > 0\; d(x, a) < \delta \implies d(f(x), f(a)) < \varepsilon \\ &\iff \forall\, \varepsilon > 0\; \delta > 0\; x \in B_\delta(a) \implies f(x) \in B_\varepsilon(f(a)) \\ &\iff \forall\, \varepsilon > 0\, \exists\, \delta > 0\; f(B_\delta(a)) \subset B_\varepsilon(f(a)) \\ &\iff \forall\, \varepsilon > 0\, \exists\, \delta > 0\; B_\delta(a) \subset f^{-1}(B_\varepsilon(f(a))). \end{aligned}$$

So we have redefined continuity in terms of open balls. But open balls have radii so still mentioning distance.

Definition 2.16 (Open, Neighbourhood)

Let X be a metric space. A subset $G \subset X$ is **open** if $\forall\, x \in G\, \exists\, \varepsilon > 0\; B_\varepsilon(x) < G$. A subset $N \subset X$ is a **neighbourhood** (nbd) of a point $a \in X$ if there exists an open set $G \subset X$ s.t. $a \in G \subset N$.

Remark 16.

1. Intuition: A set is open if for each point in the set it contains all points nearby as well. A set is a neighbourhood of a if it contains all points near a.

2. The open ball $B_\varepsilon(a)$ is open. Why? If $x \in B_\varepsilon(a)$ then $d(x, a) = \delta < \varepsilon$, say so by Δ inequality $B_{\varepsilon-\delta}(x) \subset B_\varepsilon(a)$.

3. If $\mathcal{N}$ is an open set and $a \in \mathcal{N}$ then certainly $\mathcal{N}$ is a nbd (neighbourhood) of a: $a \in \underbrace{\mathcal{N}}_{\text{open}} \subset \mathcal{N}$.
   However, a nbd of a need not be open. E.g. in $\mathbb{R}$ with the usual metric then $[-1, 1]$ is a nbd of 0: $0 \in \underbrace{(0, 1)}_{\text{open}} \subset [-1, 1]$. And $[-1, 1] \cup \{396\}$ is a nbd of 0.

4. $\mathcal{N}$ is a nbd of a iff $\exists\, \varepsilon > 0$ s.t. $B_\varepsilon(a) \subset N$.

   Proof. ($\impliedby$): $a \in \underbrace{B_\varepsilon(a)}_{\text{open}} \subset \mathcal{N}$.

   ($\implies$): $a \in \underbrace{G}_{\text{open}} \subset \mathcal{N}$ and $\exists\, \varepsilon > 0$ s.t. $B_\varepsilon(a) \subset G$. □

5. A set G is open iff it's a nbd of each of its points.

Proposition 2.6

Let $(X, d), (Y, e)$ be metric spaces and let $f : X \to Y$.

1. f is cts at $a \in X$ iff whenever $\mathcal{N} \subset Y$ is a nbd of $f(a)$ we have $f^{-1}(\mathcal{N}) \subset X$ a nbd of a;
2. f is a cts function iff whenever $G \subset Y$ is open we have $f^{-1}(G) \subset X$ open.

Proof.

1. ($\implies$): Suppose f cts at $a \in X$. Let $\mathcal{N}$ be a nbd of $f(a)$. Then $\exists\ \varepsilon > 0$ s.t. $B_\varepsilon(f(a)) \subset \mathcal{N}$. But f cts at a so $\exists\ \delta > 0$ s.t. $B_\delta(a) \subset f^{-1}(B_\varepsilon(f(a))) \subset f^{-1}(\mathcal{N})$. So $f^{-1}(\mathcal{N})$ is a nbd of a.

   ($\impliedby$): Suppose $f^{-1}(\mathcal{N})$ is a nbd of a for every nbd $\mathcal{N}$ of $f(a)$. Let $\varepsilon > 0$. In particular, $B_\varepsilon(f(a))$ is a nbd of $f(a)$ so $f^{-1}(B_\varepsilon(f(a)))$ is a nbd of a so $\exists\ \delta > 0$ s.t. $B_\delta(a) \subset f^{-1}(B_\varepsilon(f(a)))$. So f is cts at a.

2. ($\implies$): Suppose f is a cts function. Let $G \subset Y$ be open. Let $a \in f^{-1}(G)$. Then $f(a) \in G$ and G open so G is a nbd of $f(a)$. Moreover, f is cts at a so by (i) we have $f^{-1}(G)$ a nbd of a. Hence $\exists\ \delta > 0$ s.t. $B_\delta(a) \subset f^{-1}(a)$. So $f^{-1}(a)$ is open.

   ($\impliedby$): Suppose $f^{-1}(G)$ open whenever G is open in Y. Let $a \in X$. Let $\mathcal{N} \subset Y$ be a nbd of $f(a)$. Then $\exists\ G \subset Y$ open s.t. $f(a) \in G \subset \mathcal{N}$. By assumption, $f^{-1}(G) \subset X$ open. Now $a \in f^{-1}(G) \subset f^{-1}(\mathcal{N})$ with $f^{-1}(G)$ open so $f^{-1}(\mathcal{N})$ is nbd of a. So by (i), f is cts at a. So f is a cts function.

□

Remark 17.
1. This says that we can define continuity entirely in terms of open sets without mentioning the metric.
2. We saw previously that homeomorphisms preserve convergence and continuity. Proposition 2.6 (ii) say homeomorphisms also preserve open sets: precisely, if $f : X \to Y$ is a homeomorphism then $G \subset X$ is open iff $f(G) \subset Y$ is open. (Why? $G = f^{-1}(f(G))$ with f cts and $f(G) = (f^{-1})^{-1}(G)$ with f^{-1} cts.)

What else is preserved by homeomorphisms?
Suppose $f : X \to Y$ is a homeomorphism and X is sequentially compact. Let (y_n) be a sequence in Y. Then $(f^{-1}(y_n))$ is a sequence in X and so has a convergent subsequence. $f^{-1}(y_{n_j}) \to x \in X$, say. But convergence of sequences is preserved by homeomorphisms. Hence $(y_{n_j}) = f(f^{-1}(y_{n_j})) \to f(x) \in Y$. So Y is sequentially compact.

Thus if X, Y homeomorphic spaces, X sequentially compact $\iff$ Y sequentially compact.

'Sequential compactness is a topological property[2]'. If X, Y homeomorphic and one has the property then so does the other.

What about completeness? Not so good.

Example 2.33

We saw $(0,1)$ and $\mathbb{R}$ with the usual metric in each case are homeomorphic. But $\mathbb{R}$ is complete and $(0,1)$ is not. So completeness is not a topological property.

What went wrong? Property of being a Cauchy sequence is not preserved by homeomorphisms.

Remark 18. Suppose (x_n) is a sequence in a metric space X and $x \in X$. Then

$$\begin{aligned} x_n \to x &\iff \forall\, \varepsilon > 0 \,\exists\, N \,\forall\, n \geq N \; d(x_n, x) < \varepsilon \\ &\iff \forall\, \varepsilon > 0 \,\exists\, N \,\forall\, n \geq N \; x_n \in B_\varepsilon(x) \\ &\iff \text{for all nbds } \mathcal{N} \text{ of } x \;\exists\, N \,\forall\, n \geq N \; x_n \in \mathcal{N}. \end{aligned}$$

This defines convergence solely in terms of nbds. Can't do something similar for Cauchy sequences using nbds/ open sets.

We have just seen that seq. compactness is a topological property. We can define seq. compactness just in terms of nbds/ open sets: Seq. compact is defined in terms of convergence of sequences which can be described using nbds. But is there a 'nicer' way to do this, can we describe just in terms of open sets?

Definition 2.17 (Open Cover)

Let X be a metric space. An **open cover** of X is a collection $\mathcal{C}$ of open subsets of X s.t. $X = \bigcup_{G \in \mathcal{C}} G$.

Definition 2.18 (Subcover)

A **subcover** of $\mathcal{C}$ is an open cover $\mathcal{B}$ of X with $\mathcal{B} \subset \mathcal{C}$.

Definition 2.19 (Compact)

We say X is **compact** if every open cover of X has a finite subcover.

[2] This means it is preserved by homeomorphisms

Example 2.34 (The Heine-Borel Theorem)

$[0,1]$ with the usual metric is compact.

Proof. Let $\mathcal{C}$ be an open cover of $[0,1]$.
Let $A = \{x \in [0,1] : \exists\, \mathcal{B} \subset \mathcal{C}$ finite with $[0,x] \subset \bigcup_{G \in \mathcal{B}} G\}$. We know $\exists\, G \in \mathcal{C}$ with $0 \in G$. So $0 \in A$ so $A \neq \varnothing$. Clearly A bounded above by 1. So A has supremum, say $\sigma = \sup A$.

As G open, $\exists\, \varepsilon > 0$ s.t. $[0,\varepsilon) = B_\varepsilon(0) \subset G$. So $\frac{\varepsilon}{2} \in A$ so $\sigma > 0$.

Suppose $\sigma < 1$. We can find $H \in \mathcal{C}$ with $\sigma \in H$. As $\sigma = \sup A$ we can find $x \in A$ with $x \in H$. So we have $\mathcal{B} \subset \mathcal{C}$ finite with $[0,x] \subset \bigcup_{G \in \mathcal{B}} G$. But $\exists\, \varepsilon > 0$ s.t. $(\sigma - \varepsilon, \sigma + \varepsilon) = B_\varepsilon(\sigma) \subset H$. So $[0, \sigma + \frac{\varepsilon}{2}] \subset \bigcup_{G \in \mathcal{B} \cup \{H\}} G$. So $\sigma + \frac{\varepsilon}{2} \in A$ ↯.

Hence $\sigma = 1$. We can find $K \in \mathcal{C}$ s.t. $1 \in K$. As K open, we can find $\varepsilon > 0$ s.t. $(1-\varepsilon, 1] = B_\varepsilon(1) \subset K$. As $1 = \sup A$ we can find $x \in A \cap (1-\varepsilon, 1]$. That says we have finite $\mathcal{B} \subset \mathcal{C}$ with $[0,x] \subset \bigcup_{G \in \mathcal{B}} G$.

Then $\mathcal{B} \cup \{K\}$ is an open cover of $[0,1]$ and so a subcover of $\mathcal{C}$. So $[0,1]$ is compact. □

Theorem 2.6

Let X be a metric space. Then TFAE:

1. X is compact;
2. X is sequentially compact;
3. X is complete and totally bounded;
4. X is a subspace of $\mathbb{R}^n$ with the Euclidean metric. $X \subset \mathbb{R}^n$ is closed and bounded.

Proof. We have already shown (2) $\iff$ (3) ($\iff$ (4) if appropriate) in previous section 2.3.
So only remains to prove (1) $\iff$ (2).

($\implies$): Suppose X is not sequentially compact. Then there is some sequence (x_n) in X with no convergent subsequence. Hence for every point $a \in X$ we can find a nbd of a and hence an open set G_a containing a but containing (x_n) for only finitely many values of n. (If not, pick an a for which this is not true; then take n s.t. $x_{n_1} \in B_1(a)$, then $n_2 > n_1$ s.t. $x_{n_2} \in B_{\frac{1}{2}}(a)$, then $x_{n_3} \in B_{\frac{1}{3}}(a)$ with $n_3 > n_2$, ..., giving $x_{n_j} \to a$ ↯).

Now let $\mathcal{C} = \{G_a : a \in X\}$. This is an open cover of X. But if $\mathcal{D} \subset \mathcal{C}$ is finite then $\bigcup_{G \in \mathcal{D}} G$ contains x_n for only finitely many n, so $\bigcup_{G \in \mathcal{D}} \neq X$. So $\mathcal{C}$ has no finite subcover. Hence X is not compact.

($\Longleftarrow$): Suppose X is sequentially compact.
Let $\mathcal{C}$ be an open cover of X.

Claim 2.2
$\exists\, \delta > 0\, \forall\, a \in X\, \exists\, G \in \mathcal{C}\; B_\delta(a) \subset G$.

Proof. Suppose not. Then $\forall\, \delta > 0\, \exists\, a \in X\, \forall\, G \in \mathcal{C}\; B_\delta(a) \not\subset G$. Taking $\delta = \frac{1}{n}$ for each $n \in \mathbb{N}$ we obtain a sequence (x_n) in X s.t. for each n, $\forall\, G \in \mathcal{C}\; B_{\frac{1}{n}}(x_n) \not\subset G$. By sequential compactness, we can find a convergent subsequence $x_{n_j} \to a \in X$. Pick $G \in \mathcal{C}$ s.t. $a \in G$. As G open, we can pick $\varepsilon > 0$ s.t. $B_\varepsilon(a) \subset G$. Pick j sufficiently large s.t. $x_{n_j} \in B_{\frac{\varepsilon}{2}}(a)$ and also $\frac{1}{n_j} < \frac{\varepsilon}{2}$. Then $B_{\frac{1}{n_j}}(x_{n_j}) \subset B_\varepsilon(a) \subset G$. □

Now take δ as in the claim. As X is sequentially compact, it is totally bounded so we can find a finite set $A \subset X$ s.t. $\forall\, x \in X\, \exists\, a \in A\; d(x,a) < \delta$. That is $\forall\, x \in X\, \exists\, a \in A\; x \in B_\delta(a)$. That is, $X = \bigcup_{a \in A} B_\delta(a)$. By choice of δ, for each $a \in A$ we can pick $G_a \in \mathcal{C}$ s.t. $B_\delta(a) \subset G_a$. So $\{G_a : a \in A\}$ is a finite subcover. So X is compact. □

Finally, two important properties of open sets. First: relationship between open/closed.

Proposition 2.7
Let X be a metric space and $G \subset X$. Then G is open iff $F = X \setminus G$ is closed.

Proof. ($\Longrightarrow$) Suppose F not closed. Then there is a sequence (x_n) in F with $x_n \to x \notin F$ so $x \in G$. Suppose $\mathcal{N}$ is a nbd of x. Then $\exists\, N$ s.t. $\forall\, n \geq N\; x_n \in \mathcal{N}$. But $\forall\, n\; x_n \notin G$. So $\mathcal{N} \neq G$. So G is not a nbd of $x \in G$. So G is not open.

($\Longleftarrow$): Suppose G is not open. Then there is some $x \in G$ s.t. $\forall\, \varepsilon > 0\; B_\varepsilon(x) \not\subset G$. That is $B_\varepsilon(x) \cap F \neq \varnothing$. So for $n \in \mathbb{N}$ we can pick $x_n \in B_{\frac{1}{n}}(x) \cap F$. Then (x_n) is a sequence in F with $x_n \to x \in G$. So F is not closed. □

Secondly: If X is a metric space, can we say something about the structure of the collection of all open subsets of X?

Proposition 2.8
Let X be a metric space and let $\tau = \{G \subset X : G \text{ open}\}$. Then

1. $\varnothing \in \tau$ and $X \in \tau$;

2. if $\sigma \subset \tau$ then $\bigcup_{G\in\sigma} G \in \tau$, 'any union of open sets is open';
3. If $G_1, G_2, \ldots, G_n \in \tau$ then $\bigcap_{i=1}^n G_i \in \tau$, 'a finite intersection of open sets is open'.

Remark 19. We do need finiteness condition in (3). E.g. $\forall\, n \in \mathcal{N}$, $(-\frac{1}{n}, \frac{1}{n})$ is open in $\mathbb{R}$ with the usual metric. But $\bigcap_{n=1}^\infty (-\frac{1}{n}, \frac{1}{n}) = \{0\}$ which is not open.

Proof.

1. Obvious
2. Suppose $\sigma \subset \tau$. Let $H = \bigcup_{G\in\sigma} G$. Suppose $a \in H$. Then $a \in G$ for some $G \in \sigma$. So G is a nbd of a (as G open). So H is a nbd of a (as $G \subset H$). Hence H is open, i.e. $H \in \tau$.
3. Suppose $G_1, \ldots, G_n \in \tau$ and let $J = \bigcap_{i=1}^n G_i$. Suppose $a \in J$. For each i, $a \in G_i$ and G_i open so $\exists\ \delta_i \geq 0$ s.t. $B_{\delta_i}(a) \subset G_i$. Let $\delta = \min\{\delta_1, \ldots, \delta_n\}$[a]. Then $\delta > 0$ and $B_\delta(a) = \bigcap_{i=1}^n B_{\delta_i}(a) \subset \bigcap_{i=1}^n G_i = J$. So J is open, i.e. $J \in \tau$.

□

[a] This is where we need finiteness

§3 Topological Spaces

'Do continuity entirely in terms of open sets without mentioning distance'.

Metric space: set with a distance.

Topological space: set with a collection of open subsets.

§3.1 Definitions and Examples

Definition 3.1 (Topological Space)

A **topological space** is a set X endowed with a **topology** τ that is a subset $\tau \subset \mathcal{P}(X)$ satisfying

1. $\varnothing \in \tau$ and $X \in \tau$;
2. If $\sigma \subset \tau$ then $\bigcup_{G \in \sigma} G \in \tau$;
3. If $G_1, \ldots, G_n \in \tau$ then $\bigcap_{i=1}^{n} G_i \in \tau$.

Remark 20. Could replace (3) by $G, H \in \tau \implies G \cap H \in \tau$. Equivalent to (3) by induction.

Notation. Sometimes write '(X, τ) is a topological space'. If obvious what the topology is, we might just write 'X is a topological space'.

Example 3.1

Let (X, d) be a metric space. Let $\tau = \{G \subset X : G \text{ open}\}$. Then by proposition 2.8, τ is a topology on X.

We say τ is the topology, induced by the metric d.

We want to define open/ closed/ cts etc for topological spaces. As metric spaces are top. spaces try to make sure it's 'backward compatible', so to say. The new defns shouldn't contradict the old metric space ones. So in making new defns, we'll be guided by section 2.4.

Definition 3.2 (Open)

Let (X, τ) be a topological space. We say that $G \subset X$ is **open** if $G \in \tau$

Definition 3.3 (Closed)

Let (X, τ) be a topological space. We say F is **closed** if $X \setminus F \in \tau$.

Definition 3.4 (Neighbourhood)

Let (X, τ) be a topological space. We say $\mathcal{N} \subset X$ is a **neighbourhood** of $a \in X$ if $\exists G \subset X$ open with $a \in G \subset \mathcal{N}$.

Definition 3.5 (Continuity)

Let (X, τ) be a topological space. Let (Y, σ) be another topological space, and let $f : X \to Y$. We say f is **continuous** if whenever $G \subset Y$ is open, then $f^{-1}(G) \subset X$ is open. In other words, f is continuous if $\forall\, G \in \sigma, f^{-1}(G) \subset \tau$.

We say that f is **continuous at** $a \in X$ if whenever $\mathcal{N} \subset Y$ is a neighbourhood of $f(a)$ then $f^{-1}(\mathcal{N}) \subset X$ is a neighbourhood of a.

Definition 3.6 (Homeomorphisms)

Let (X, τ) be a topological space. We say that f is a **homeomorphism** and X, Y are **homeomorphic** if f is a bijection and both f, f^{-1} are continuous.

Definition 3.7 (Topological)

Let (X, τ) be a topological space. A property is **topological** if it is preserved by homeomorphisms; that is to say, if X, Y are homeomorphic then X has the property iff Y does.

Remark 21. 1. if τ is induced by a metric then this is all consistent with the metric space definitions of these concepts.

2. Given our definition, G open iff $G \in \tau$, we often don't need to explicitly name the topology. E.g. 'Let $X = \mathbb{R}$ with the usual topology and $G \subset X$ be open...'. Other times more convenient to specify τ, write '$G \in \tau$' etc.

3. Homeomorphism is an equivalence 'relation'.

4. If $a \in G$ and G open then G is a neighbourhood of a, however neighbourhoods need not be open in general. A set $G \subset X$ is open iff G is a neighbourhood of each of its points.

Proposition 3.1

Let X, Y be topological spaces and $f : X \to Y$. Then f is continuous if and only if for all $a \in X$, f is continuous at a.

Proof. ($\implies$): Suppose f continuous and let $a \in X$. Let $\mathcal{N} \subset Y$ be a neighbourhood of $f(a)$. Then there is an open set $G \subset Y$ with $a \in G \subset \mathcal{N}$. As f is continuous, $f^{-1}(G) \subset X$ is open. Now $a \in f^{-1}(G) \subset f^{-1}(\mathcal{N})$ with $f^{-1}(G)$ open. Thus f is continuous at a.

($\Longleftarrow$): Suppose for all $a \in X$ we have f continuous at a. Let $G \subset Y$ be open. Let $a \in f^{-1}(G)$. Then $f(a) \in G$, but G is open so G is a neighbourhood of $f(a)$. Now f is continuous at a so $f^{-1}(G)$ is a neighbourhood of a in X. But a was arbitrary so $f^{-1}(G)$ is a neighbourhood of each of its points. That is $f^{-1}(G)$ is open. Hence f is continuous. □

Proposition 3.2

Let $(X, \tau), (Y, \sigma), (Z, \rho)$ be topological spaces, let $f : X \to Y$ and $g : Y \to Z$ be continuous. Then $g \circ f : X \to Z$ is continuous.

Proof. Let $G \in \rho$. As g is continuous, $g^{-1}(G) \in \sigma$. As f continuous, $f^{-1}(g^{-1}(G)) \in \tau$. That is, $(g \circ f)^{-1}(G) \in \tau$. So $g \circ f$ continuous. □

Example 3.2 (The discrete topology)

Let X be any set and $\tau = \mathcal{P}(X)$, where 'every set is open'.
However, this is not a new example; it is induced by the discrete metric

$$d(x, y) = \begin{cases} 1 & x \neq y \\ 0 & x = y \end{cases}.$$

Now in (X, d), for any $x \in X$ then $\{x\} = B_1(x)$ is open and so if $G \subset X$ then $G = \bigcup_{x \in G} \{x\}$ is open.

Example 3.3 (The indiscrete topology)

Let X be any set and $\tau = \{\varnothing, X\}$. Here, 'only open sets are $\varnothing$ and whole space'. This is something we have not seen before: τ cannot be induced by a metric as long as $|X| \geq 2$. Indeed, suppose $|X| \geq 2$ and that τ is induced by a metric d. Let $x, y \in X$ with $x \neq y$, so let $d(x, y) = \delta > 0$. Then $B_\delta(x)$ is open with $x \in B_\delta(x)$ and $y \notin B_\delta(x)$ $\lightning$.

Example 3.4 (The cofinite topology)

Let X be any infinite set and let $\tau = \{G \subset X : X \setminus G \text{ is finite}\} \cup \{\varnothing\}$. This is not induced by a metric.

Proof. 1. $\varnothing \in \tau$ and $X \setminus X = \varnothing$ is finite so $X \in \tau$.

2. Let $\sigma \subset \tau$. If σ is empty or only contains $\varnothing$ then $\bigcup_{G\in\sigma} G = \varnothing \in \tau$. Otherwise, pick $H \in \sigma$ with $H \neq \varnothing$. Then $X \setminus H$ is finite so $(X \setminus \bigcup_{G\in\sigma} G) = \bigcap_{G\in\sigma}(X \setminus G) \subset X \setminus H$ is finite. So $\bigcup_{G\in\sigma} G \in \tau$.

3. Let $G, H \in \tau$. If $G = \varnothing$ or $H = \varnothing$ then $G \cap H = \varnothing \in \tau$. Otherwise $X \setminus G, X \setminus H$ are finite then $(X \setminus (G \cap H)) = (X \setminus G) \cup (X \setminus H)$ is finite. So $G \cap H \in \tau$.

So the cofinite topology is indeed a topology.

Furthermore, it is not induced by a metric d. Observe first that if G, H open and non-empty then $G \cap H \neq \varnothing$[a]. Now suppose $x, y \in X$ with $x \neq y$. Then $d(x, y) = \delta > 0$ so $B_\delta/2(x), B_\delta/2(y)$ are non-empty disjoint open sets. So d doesn't induce τ as open sets can't be disjoint in this topology. □

[a] As they are both infinite sets with finite complements

Example 3.5 (The cocountable topology)

Let X be any uncountable set and let $\tau = \{G \subset X : X \setminus G \text{ countable}\} \cup \{\varnothing\}$. Then, very similarly to previous example, we can show that this is a topology that is not induced by any metric.

§3.2 Sequences and Hausdorff spaces

Definition 3.8 (Convergence)

Let X be a topological space, let (x_n) be a sequence in X and let $x \in X$.
We say (x_n) **converges** to x and write $x_n \to x$ if whenever $\mathcal{N} \subset X$ is a neighbourhood of x then $\exists\, N \;\forall\, n \geq N$ we have $x_n \in \mathcal{N}$.

Example 3.6 (Cocountable Topology)

Let X be an uncountable set with the cocountable topology. What sequences converge in X?

The only convergent sequences in this space are eventually constant.

Proof. Suppose $x_n \to x$. Then let $\mathcal{N} = (X \setminus \{x_n : n \in \mathbb{N}\}) \cup \{x\}$. Then $\mathcal{N}$ open and $x \in \mathcal{N}$ so $\mathcal{N}$ is a neighbourhood of x. Thus $\exists\, N$ such that $\forall\, n \geq N, x_n \in \mathcal{N}$. So $\exists\, N, \forall\, n \geq N$ we have $x_n = x$.

Conversely, it is obvious if $\exists\, N$ such that $\forall\, n \geq N, x_n = x$ then $x_n \to x$.

So the only convergent sequences in this space are eventually constant. □

Example 3.7 (Indiscrete Topology)

Let $X = \{1, 2, 3\}$ with the indiscrete topology. Let $x_n = i \in X$ with $i \equiv n \mod 3$. So sequence is 1,2,3,1,2,3,. . . .

Claim 3.1

$x_n \to 2$

Proof. Let $\mathcal{N}$ be a neighbourhood of 2. Then $\exists\, G$ open such that $2 \in G \subset \mathcal{N}$. But only open sets are $\varnothing, \{1, 2, 3\}$ so $G = \{1, 2, 3\}$. So $\mathcal{N} = \{1, 2, 3\}$ so $\forall\, n, x_n \in \mathcal{N}$. □

Similarly $x_n \to 1$ and $x_n \to 3$, and so we arrive at the big revelation for topological spaces: **LIMITS OF CONVERGENT SEQUENCES NEED NOT BE UNIQUE.** Thus we can't write $\lim_{n\to\infty} x_n$ anymore.

Remark 22. The above proof shows that in any indiscrete space, every sequences converges to every point of the space.

Definition 3.9 (Hausdorff)

A topological space X is **Hausdorff** if whenever $x, y \in X$ with $x \neq y$ then there are disjoint open $G, H \subset X$ with $x \in G$ and $y \in H$.

Example 3.8

1. Metric spaces are Hausdorff. Indeed, if (X, d) metric and $x, y \in X$, $x \neq y$, let $\delta = d(x, y) > 0$ and take $G = B_{\delta/2}(x)$ and $H = B_{\delta/2}(y)$.

2. Indiscrete spaces are not Hausdorff (assuming $|X| \geq 2$).

3. The cofinite topology is not Hausdorff. Let X be an infinite set with the cofinite topology and let $x, y \in X$ with $x \neq y$. Let $G, H \subset X$ be open with $x \in G$, $y \in H$. Clearly $G, H \neq \varnothing$ so $X \setminus G, X \setminus H$ finite and so $X \setminus (G \cap H) = (X \setminus G) \cup (X \setminus H)$ is finite. In particular $G \cap H \neq \varnothing$.

Similarly, the cocountable topology is not Hausdorff.

Proposition 3.3

Limits of convergent sequences in Hausdorff spaces are unique.

Proof. Let X be Hausdorff, let $a, b \in X$, and let (x_n) be a sequence in X with $x_n \to a$ and $x_n \to b$.
Suppose $a \neq b$. Take open G, H with $a \in G$, $b \in H$ and $G \cap H = \varnothing$. Now G is a neighbourhood of a so there is some N_1 such that $\forall n \geq N_1$ we have $x_n \in G$. Similarly there is some N_2 such that $\forall n \geq N_2$, $x_n \in H$. Taking $n = \max\{N_1, N_2\}$, we get $x_n \in G \cap H = \varnothing$ ↯.
Hence $a = b$. □

Proposition 3.4 (Relationship to continuity)

Let X, Y be topological spaces and let $f : X \to Y$ be continuous at $a \in X$. Let (x_n) be a sequence in X with $x_n \to a$. Then $f(x_n) \to f(a)$.

Proof. Let $\mathcal{N} \subset Y$ be a neighbourhood of $f(a)$. As f continuous we know $f^{-1}(\mathcal{N})$ is a neighbourhood of a. As $x_n \to a$ we can find N such that $\forall\, n \geq N$, $x_n \in f^{-1}(\mathcal{N})$. Then $\forall\, n \geq N$, $f(x_n) \in \mathcal{N}$. So $f(x_n) \to f(a)$. □

Example 3.9 (Converse is not true in general!)

Let $X = Y = \mathbb{R}$, X with cocountable topology, Y with usual topology and let $f : X \to Y$ be the identity function.

Suppose $x_n \to 0$ in X. Then for sufficiently large n, $x_n = 0$ and so for sufficiently large n, $f(x_n) = x_n = 0 = f(0)$ so $f(x_n) \to f(0)$ in Y.

However, $(-1, 1) \subset Y$ is open and $0 \in (-1, 1)$ so $(-1, 1)$ is a neighbourhood of $0 \in Y$. But $f^{-1}((-1, 1)) = (-1, 1) \in X$ is not a neighbourhood[a] of 0 in X. So f is not continuous at 0.

[a] For $(-1, 1)$ to be a neighbourhood then it must contain some open sets containing zero but the only non empty open sets have finite complement. So cannot be a neighbourhood.

Example 3.10 (Converse is still not true even after imposing condition that spaces are Hausdorff.)

Take example as above but replace topology on X by

$$\sigma = \{G \subset \mathbb{R} : (X \setminus G) \text{ countable or } 0 \notin G\}.$$

Check that this is a topology, and this is Hausdorff: suppose $x, y \in X$ with $x \neq y$. If $x, y \neq 0$ then $\{x\}, \{y\} \in \sigma$. While if $x = 0$ say, then $\mathbb{R} \setminus \{y\}, \{y\} \in \sigma$.

Now, neighbourhoods of 0 in σ are exactly same as in the cocountable topology. So exactly as before, $x_n \to 0$ in $X \implies x_n \to 0$ in Y but f is not continuous at 0.

Remark 23. In a metric space, the topology is completely determined by convergence of sequences. Not true for a general topology space. Hence we'll tend to concentrate more on continuity than convergence of sequences.

§3.3 Subspaces

Definition 3.10 (Subspace Topology)
Let (X, τ) be a topological space and let $Y \subset X$. The **subspace topology** on Y is

$$\sigma = \{G \cap Y : G \in \tau\}.$$

Easy to check that this is a topology.

Let's check that this is indeed backward compatible with our definition of metric spaces.

Proposition 3.5
Let (X, d) be a metric space with topology τ be induced by d. Let Y be a subspace of the metric space X. Then Y has the subspace topology.

Proof. Let σ be the topology on Y induced by the metric $d|_{Y^2}$.

First suppose $G \in \tau$; we want to check that $G \cap Y$ is open in Y. Let $y \in G \cap Y$. As $y \in G$ and G is open in X we can find $\delta > 0$ such that $\forall x \in X$, we have $d(x, y) < \delta \implies x \in G$. Then $\forall x \in Y$, we have $d(x, y) < \delta \implies x \in G \cap Y$. Thus $G \cap Y$ is a neighbourhood of y. Since y is arbitrary $G \cap Y \in \sigma$.

Conversely, suppose $H \in \sigma$. For each $y \in H$ we can find $\delta_y > 0$ such that $\forall x \in Y$, $d(x, y) < \delta_y \implies x \in H$. Now consider the open balls

$$B_{\delta_y}(y) = \{x \in X : d(x, y) < \delta_y\}.$$

Each $B_{\delta_y}(y)$ is open, so for each $y \in H$, $y \in B_{\delta_y}(y)$ and $B_{\delta_y}(y) \cap Y \subset H$. Now let

$G = \bigcup_{y \in H} B_{\delta_y}(y)$. Then G is open and $G \cap Y = H$. That is, we've found $G \in \tau$ such that $G \cap Y = H$. □

Proposition 3.6

A subspace of a Hausdorff space is Hausdorff.

Proof. Let (X, τ) be Hausdorff, $Y \subset X$, and σ be the subspace topology on Y. Let $x, y \in Y$ with $x \neq y$. As X is Hausdorff we can find $G, H \in \tau$ with $x \in G, y \in H, G \cap H = \varnothing$. Then $G \cap Y, H \cap Y \in \sigma$ with $x \in G \cap Y$, $y \in H \cap Y$ and $(G \cap Y) \cap (H \cap Y) = \varnothing$. □

§3.4 Compactness

Definition 3.11 (Open Cover)

Let (X, τ) be a topological space. An **open cover** of X is a subset $\mathcal{C} \subset \tau$ such that $X = \bigcup_{G \in \mathcal{C}} G$.

Definition 3.12 (Subcover)

A **subcover** of $\mathcal{C}$ is a subset $\mathcal{D} \subset \mathcal{C}$ which is itself an open cover.

Definition 3.13 (Compactness)

We say that X is **compact** if every open cover of X has a finite subcover.

Definition 3.14 (Sequentially Compact)

We say that X is **sequentially compact** if every sequence in X has a convergent subsequence.

Exercise 3.1. Show that a continuous real-valued function on a sequentially compact topological space is bounded and attains its bounds.

Remark 24. This was the traditional wording we have been using; here and elsewhere, if no topology is specified $\mathbb{R}$ is generally assumed to have the usual topology, and proof is similar to the metric space case.

We've seen that for a metric space, compact and sequentially compact are equivalent; but **This is not true for general topological space!** In fact, $\exists$ compact space that is not sequentially compact, and $\exists$ sequentially compact space that is not compact. However both examples are beyond the scope of this course.

Observe that compactness and sequential compactness are both topological properties. Given that we don't want to think too much about sequences in a general topological space, we'll be concentrating primarily on compactness rather than sequential compactness.

Remark 25. If X is a topological space and $K \subset X$ we might want to say 'K is compact'. Clearly meaningful since K is a topological space with subspace topology. Thinking further: Let τ be the topology on X. Then K is *compact* iff whenever $\mathcal{C} \subset \tau$ with $K = \bigcup_{G \in \mathcal{C}} G \cap K$ then there is a finite $\mathcal{D} \subset \mathcal{C}$ such that $K = \bigcup_{G \in \mathcal{D}} G \cap K$.

Equivalently, K is compact iff whenever $\mathcal{C} \subset \tau$ with $K \subset \bigcup_{G \in \mathcal{C}} G$ then there is a finite $\mathcal{D} \subset \mathcal{C}$ with $K \subset \bigcup_{G \in \mathcal{D}} G$. Thus we sometimes refer to $\mathcal{C}$ as being an open cover of K (in X).

We shall now see some examples of compact spaces.

Example 3.11

$[0, 1]$ with the usual topology is compact, proven in Section 2.4 'Heine-Borel Theorem', using the 'creeping along proof'. More generally, $S \subset \mathbb{R}^n$ is compact iff S is closed and bounded.

Example 3.12

A metric space is compact iff it is complete and totally bounded.

Example 3.13

Suppose X is a discrete topological space. Then $\{\{x\} : x \in X\}$ is an open cover It does not have any subcovers apart from itself, so X is compact iff X is finite. (Note: any finite space if compact, since there are only finitely many subsets.)

Example 3.14

Let X be indiscrete. Then the only open covers of X are $\{\varnothing, X\}$ and $\{X\}$, both of which are finite. So X is compact.

Recall that sequential compactness and compactness are generally different in topological spaces. Hence, we need to find another way to prove properties involving "Boltzano-Weierstrass" differently.

Theorem 3.2

A continuous real-valued function on a compact topological space is bounded and attains its bounds.

Proof. Let X be compact and $f : X \to \mathbb{R}$ be continuous. Let $G_n = f^{-1}((-n, n)), (n \in \mathbb{N})$. Then $\{G_n : n \in \mathbb{N}\}$ is an open cover of X, so as X is compact we have a finite subcover $\{G_{n_1}, \ldots, G_{n_k}\}$. Now $\forall x \in G_{n_i}$, $|f(x)| < n_i$, and hence $\forall x \in X, |f(x)| < \max_{1 \leq i \leq k} n_i$, thus f is bounded.

Let $\sigma = \sup_{x \in X} f(x)$, and suppose σ is not attained by f. Now define $g : X \to \mathbb{R}$ by $g(x) = \frac{1}{\sigma - f(x)}$ which is well-defined and continuous. Hence by previous part, g is bounded. But by definition of sup given $\varepsilon > 0$ we can find x such that $\sigma - f(x) < \varepsilon$ where $g(x) > \frac{1}{\varepsilon}$ ↯.
Similarly, $\inf_{x \in X} f(x)$ is attained. □

Remark 26. Think of compactness as a 'smallness' condition, or we don't want a topological space to be 'too big'; in a sense, this is the next best thing to finiteness.

For instance, a real-valued function on a finite space is bounded (obvious). If we have a continuous function on a compact space - how can we show boundedness? We can use compactness to show space is not 'too big', i.e. we can cover it with finitely many sets on each of which f is bounded (then proof follows clearly).

More generally:

Theorem 3.3

A continuous image of a compact space is compact.

Proof. Let $f : X \to Y$ be continuous and X compact. Let $K = f(X) \subset Y$. Let $\mathcal{C}$ be an open cover of K in Y. Then $\{f^{-1}(G) : G \in \mathcal{C}\}$ is an open cover of X so by compactness, there is a finite $\mathcal{D} \subset \mathcal{C}$ such that $\{f^{-1}(G) : G \in \mathcal{D}\}$ is an open cover of X. Then $\mathcal{D}$ is an open cover of K in Y, so K is compact. □

Remark 27. This together with the fact that compact subsets of $\mathbb{R}$ are closed and bounded gives an alternative proof of theorem 3.2.

Lemma 3.1

1. A closed subset of a compact space is compact.
2. A compact subset of a Hausdorff space is closed.

Proof.

1. Let X be a compact topological space and let $F \subset X$ be closed. Let $\mathcal{C}$ be an open cover of F in X. Thus $X \setminus F$ is open so let $\mathcal{C}' = \mathcal{C} \cup \{X \setminus F\}$ then $\mathcal{C}'$ is an open cover of X. Since X is compact $\mathcal{C}'$ has a finite subcover $\mathcal{D}'$. Let $\mathcal{D} = \mathcal{D}' \setminus \{X \setminus F\}$ if $X \setminus F \in \mathcal{D}'$, and $\mathcal{D} = \mathcal{D}'$ otherwise. Then $\mathcal{D}$ is a finite subcover of $\mathcal{C}$, so F is compact.

2. Let X be a Hausdorff space and let $K \subset X$ be compact. We want to show K is closed, i.e. $X \setminus K$ is open, i.e. $X \setminus K$ is a neighbourhood of each of its points.

   Let $y \in X \setminus K$. We want to find some open set containing y, that is contained entirely within $X \setminus K$. Given $x \in K$, $x \neq y$ so as X Hausdorff, we can find disjoint open sets $U_x, V_x \subset X$ with $x \in U_x$ and $y \in V_x$. Then $\{U_x : x \in X\}$ is an open cover of K in X so it has a finite subcover $\{U_{x_1}, \dots, U_{x_n}\}$.
   Let $U = \bigcup_{i=1}^n U_{x_i}$ and $V = \bigcap_{i=1}^n V_{x_i}$. We have U, V open and $K \subset U$, $y \in V$ and $U \cap V = \varnothing$.
   In particular, we have found an open set V such that $y \in V \subset X \setminus K$. So $X \setminus K$ is a neighbourhood of each of its points and it's open, thus K is closed.

□

Theorem 3.4

A continuous bijection from a compact space to a Hausdorff space is a homeomorphism.

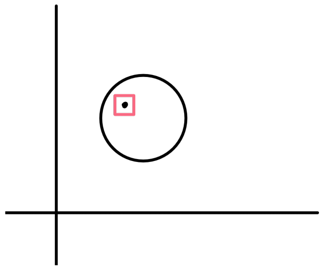
Proof. Let $f : X \to Y$ be a continuous bijection, X compact, Y Hausdorff, and our aim is to show $f^{-1} : Y \to X$ is continuous.

Let $G \subset X$ be open. Then $X \setminus G$ is closed, so by lemma 3.1 (a), $X \setminus G$ is compact. By theorem 3.3, $f(X \setminus G)$ is compact and so by lemma 3.1 (b), $f(X \setminus G)$ is closed. That is, $Y \setminus f(G)$ is closed, i.e. $f(G)$ is open. But f is a bijection so $(f^{-1})^{-1}(G) = f(G)$ is open. Thus f^{-1} is continuous. □

§3.5 Products

Suppose we have $\mathbb{R}$ with the usual topology. We would like to think about $\mathbb{R} \times \mathbb{R}$ as $\mathbb{R}^2$ with the Euclidean topology.
In general, if $(X, \tau), (Y, \sigma)$ are topological spaces, what sensible topology can we put on $X \times Y$?

In general, $\tau \times \sigma$ is not going to be a topology. E.g. $\mathbb{R} \times \mathbb{R}$, $\mathbb{R}^2$, an open ball in $\mathbb{R}^2$ is not in $\tau \times \sigma$. However, each point in the ball is some set in $\tau \times \sigma$ confined in the ball (i.e. for each point in the ball, there is some open square inside the ball that contains the point). Hence the open ball is a union of some sets in $\tau \times \sigma$.



Thus, in general, we would want our product topology from τ, σ to be the collection of all union of sets in $\tau \times \sigma$.

Definition 3.15 (π-system)

A **π-system** on a set X is a non-empty subset $\pi \subset \mathcal{P}(X)$ such that $A, B \in \pi \implies A \cap B \in \pi$.

Proposition 3.7

Let Π be a π-system on a set X. Then

$$\tau = \left\{ \bigcup_{A \in \Sigma} A : \Sigma \subset \Pi \right\} \cup \{\varnothing, X\}$$

is a topology on X.

Proof. Clearly, $\varnothing, X \in \tau$ and it's closed under arbitrary unions.
Now suppose $G, H \in \tau$. If $G = \varnothing, X$ or $H = \varnothing, X$ then $G \cap H \in \tau$ trivially.
Otherwise $G = \bigcup_{A \in \Phi} A$, $H = \bigcup_{B \in \Theta} B$ for some $\Phi, \Theta \subset \Pi$. Then $G \cap H = \bigcup_{A \in \Phi, B \in \Theta} (A \cap B) = \bigcup_{C \in \Sigma} C$, where $\Sigma = \{A \cap B : A \in \Phi, B \in \Theta\} \subset \Pi$. Hence $G \cap H \in \tau$. □

Definition 3.16 (Generated Topology)

We call this τ the topology **generated by** Π.

Proposition 3.8

Let $(X, \tau), (Y, \sigma)$ be topological spaces. Then $\tau \times \sigma$ is a π-system on $X \times Y$.

Proof. $\varnothing = \varnothing \times \varnothing \in \tau \times \sigma$ so $\tau \times \sigma \neq \varnothing$. Now suppose $A, B \in \tau \times \sigma$. Then $A = G \times H$, $B = K \times L$ for some $G, K \in \tau$ and some $H, L, \in \sigma$. So $A \cap B = (G \cap K) \times (H \cap L) \in$

$\tau \times \sigma$. □

Definition 3.17 (Product Topology)

Let $(X, \tau), (Y, \sigma)$ be topological spaces. The **product topology** on $X \times Y$ is the topology generated by the π-system $\tau \times \sigma$.

Exercise. If $X = Y = \mathbb{R}$, $\tau = \sigma$ is the usual topology, show that the product topology on $\mathbb{R}^2$ is the Euclidean topology. (Sheet 3)

Theorem 3.5

1. A product of Hausdorff spaces is Hausdorff.
2. A product of compact spaces is compact.

Proof. Let $(X, \tau), (Y, \sigma)$ be topological spaces and let ρ be the product topology on $X \times Y$.

1. Suppose X, Y are Hausdorff. Let $(x, y), (z, w) \in X \times Y$ with $(x, y) \neq (z, w)$, wlog $x \neq z$. As X is Hausdorff we can find $G, H \in \tau$ with $G \cap H = \varnothing$, $x \in G$, $z \in H$. Then $G \times Y, H \times Y \in \rho$ with $(G \times Y) \cap (H \times Y) = \varnothing$ and $(x, y) \in G \times Y$, $(z, w) \in H \times Y$. Thus $X \times Y$ is Hausdorff.

2. Suppose $X \times Y$ compact. Let $\mathcal{C} \subset \rho$ be an open cover of $X \times Y$. Fix $x \in X$. For each $y \in Y$, there is some $G_y \in \mathcal{C}$ such that $(x, y) \in G_y$. Hence we can find $U_y \in \tau$ and $V_y \in \sigma$ such that $(x, y) \in U_y \times V_y \subset G_y$. In particular, we have $x \in U_y$ and $y \in V_y$. Thus $\{V_y : y \in Y\} \subset \sigma$ is an open cover of Y. So, as Y is compact, it has a finite subcover $\{V_{y_1}, \ldots, V_{y_n}\}$, say.

   Let $W = \bigcap_{i=1}^n U_{y_i}$. Then W is open in X and $x \in W$. Moreover, $W \times Y \subset \bigcup_{i=1}^n (U_{y_i} \times V_{y_i}) \subset \bigcup_{i=1}^n G_{y_i}$. Now do this for each $x \in X$ to obtain $W_x = W, n_x = n$ and $G_{y_i}^{(x)} = G_{y_i}$ for $(1 \leq i \leq n_x)$ as above. Then $\{W_x : x \in X\} \subset \tau$ is an open cover of X. So, as X is compact, it has a finite subcover, $\{W_{x_1}, \ldots, W_{x_m}\}$, say.

   Now $X = \bigcup_{j=1}^m W_{x_j}$ and for each j, $W_{x_j} \times Y \subset \bigcup_{i=1}^{n_{x_j}} G_{y_i}^{(x_j)}$. Thus

   $$\{G_{y_i}^{(x_j)} : 1 \leq j \leq m, 1 \leq i \leq n_{x_j}\}$$

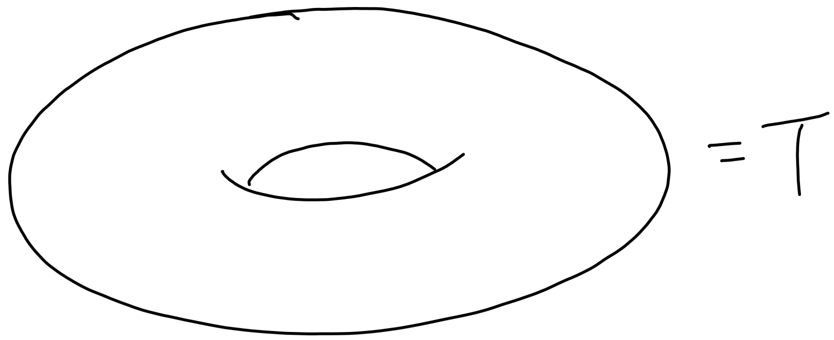
   is an open cover of $X \times Y$ and hence a finite subcover of $\mathcal{C}$. Thus $X \times Y$ is compact.

□

Remark 28. The second proof is a typical proof when working with products so it is useful to understand it and copy the idea.

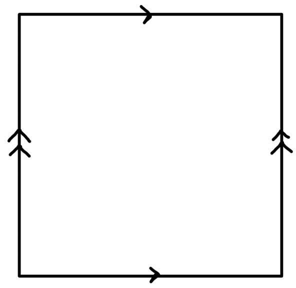
§3.6 Quotients

Consider the surface of a torus (donut) in $\mathbb{R}^3$. We might be interested in for example,



continuous functions $T \to T$. Analysis is likely to be unpleasant since the equation defining T might be complicated. However, given that we care about continuity and convergence, if we replace T by a space homeomorphic to T then we're happy, particularly if the new space is analytically easier to work with.

For example take the closed unit square $[0,1] \times [0,1]$ with Euclidean subspace topology. Glue $(x,0)$ to $(x,1)$ for each x, and glue $(0,y)$ to $(1,y)$ for each y; this seems to give us



T. More specifically, we defined an equivalence relation on $[0,1] \times [0,1]$, say by $\sim$, with equivalence classes:

$$\begin{aligned} \{(x,y)\} &\quad 0 < x, y < 1 \\ \{(x,0),(x,1)\} &\quad 0 < x < 1 \\ \{(0,y),(1,y)\} &\quad 0 < y < 1 \\ \{(0,0),(0,1),(1,0),(1,1)\}. & \end{aligned}$$

Essentially we could define $T = [0,1]^2/\sim$, the set of equivalence classes. This definition seems annoying - a better one could perhaps be defining an equivalence relation $\sim$ on $\mathbb{R}^2$, where every unit square is "mapped" to the $[0, 1]$ square, i.e. $(x, y) \sim (z, w) \Leftrightarrow x - z \in \mathbb{Z}$ and $y - w \in \mathbb{Z}$. Again, hopefully we can define $T = R^2/\sim$.

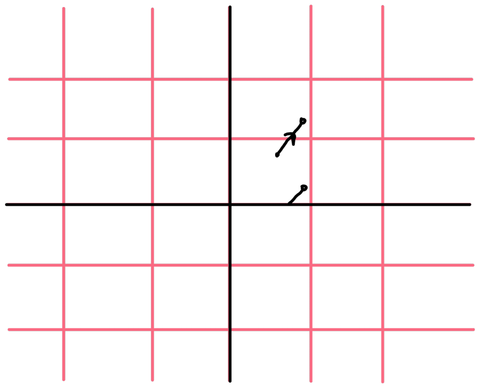


Figure 1: The squares 'wrap' over.

But what is the topology?

Definition 3.18 (Quotient Topology)
Let (X, τ) be a topological space and let $\sim$ be an equivalence relation on X. Let $q : X \to X/\sim$ be the quotient map, i.e. $\forall x \in X,\ q(x) = [x]_\sim$ (equivalence class of x). The **quotient topology** on $X/\sim$ is

$$\rho = \{G \subset X/\sim : q^{-1}(G) \in \tau\}$$

Remark 29.
1. ρ is indeed a topology using $q^{-1}(\bigcup G) = \bigcup q^{-1}(G)$ and $q^{-1}(G \cap H) = q^{-1}(G) \cap q^{-1}(H)$.
2. ρ is the largest topology on $X/\sim$ making the quotient map q continuous.

Example 3.15
Take $\mathbb{R}$ with the usual topology and $X \sim Y$ iff $x - y \in \mathbb{Z}$. Then $\mathbb{R}/\sim$ 'is' S^1, the unit circle (a subspace of $\mathbb{R}^2$ with the Euclidean topology), where 'is' means 'is homeomorphic to'. We will prove this later.

Example 3.16
Suppose as above, but now $x \sim y$ iff $x - y \in \mathbb{Q}$. What is the quotient topology on $\mathbb{R}/\sim$? Suppose $G \subset \mathbb{R}/\sim$ is open and $G \neq \varnothing$. Then $q^{-1}(G) \subset \mathbb{R}$ is open and non-empty so contains some interval $(a, b) \subset q^{-1}(G)$ with $a \neq b$. Now take any $x \in \mathbb{R}$, then $\exists y \in (a, b)$ with $x - y \in \mathbb{Q}$ i.e. $x \sim y$. Then $q(x) = [x]_\sim = [y]_\sim \in G$. Hence $G = \mathbb{R}/\sim$.

So quotient topology on $\mathbb{R}/\sim$ is the indiscrete topology.

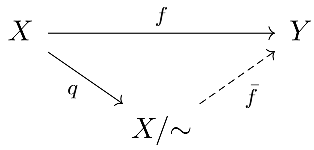
Example 3.17

Quotients of metrizable spaces need not be metrizable, and the same holds if 'metrizable' is replaced with 'Hausdorff'.

§3.6.1 Basics on equivalence relations and quotients

Suppose X is a set and $\sim$ is an equivalence relation on X. We have $X/\sim = \{[x]_\sim : x \in X\}$ and have quotient map $q : X \to X/\sim$, $x \mapsto [x]_\sim$. Clearly q is surjective.

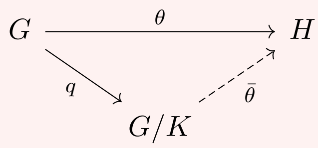
Suppose now Y is also a set and $f : X \to Y$. Assume f **respects** $\sim$, i.e. $x \sim y \implies f(x) = f(y)$.



Then there is a unique function $\bar{f} : X/\sim \to Y$ such that $f = \bar{f} \circ q$. Indeed, we must have $\forall\, x \in X$ $\bar{f}([x]_\sim) = \bar{f}(q(x)) = f(x)$. As f respects $\sim$, this is well defined: $[x]_\sim = [y]_\sim \implies x \sim y \implies f(x) = f(y)$.

Example 3.18

Suppose G, H are groups and $\theta : G \to H$ is a homomorphism. Let $K = \ker\theta$, and define $\sim$ on G by $g \sim h \Leftrightarrow g^{-1}h \in K$. Then $G/K = G/\sim$ and consider $\bar{\theta} : G/K \to H$.



We can check that $\bar{\theta}$ is a homomorphism and injective thus isomorphism onto $\theta(G)$. This is the first isomorphism theorem.

Proposition 3.9

Let (X, τ) be a topological space and $\sim$ be an equivalence relation on X. Let ρ be the quotient topology on $X/\sim$. Suppose $f : X \to Y$ is a continuous function respecting $\sim$, where (Y, σ) is a topological space. Then there is a unique continuous function $\bar{f} : X/\sim \to Y$ such that $f = \bar{f} \circ q$, where $q : X \to X/\sim$ is a quotient map.

Proof. Define $\bar{f} : X/\sim \to Y$ by $\bar{f}([x]_\sim) = f(x)$.
This is well-defined, $[x]_\sim = [y]_\sim \implies x \sim Y \implies f(x) = f(y)$. Clearly $\bar{f} \circ q = f$.
Let $G \in \sigma$. Then $q^{-1}(\bar{f}^{-1}(G)) = (\bar{f} \circ q)^{-1}(G) = f^{-1}(G) \in \tau$ as f is continuous. So by definition of quotient topology, $\bar{f}^{-1}(G) \in \rho$. Hence $\bar{f}$ is continuous.

Finally, if $f = h \circ q$ for some $h : X/\sim \to Y$ then $\forall x \in X$, $h([x]_\sim) = h(q(x)) = f(x) = \bar{f}([x]_\sim)$. So $h = \bar{f}$. □

Remark 30. This is what makes quotients useful. E.g. recall torus $T = \mathbb{R}^2/\sim$ for an appropriate relation $\sim$. Hopefully T is homeomorphic to a genuine torus as a surface in $\mathbb{R}^3$, e.g. T is nasty while R^2 is nice. Hence we want to work 'upstairs' in R^2 rather than 'downstairs' in T. For examples, if you want to think about a continuous function on T, we can instead think about an appropriate continuous function on $\mathbb{R}^2$ respecting $\sim$.

Example 3.19

Recall we had $\mathbb{R}$ with the usual topology, $x \sim y$ iff $x - y \in \mathbb{Z}$ and $S^1 = \{x \in \mathbb{R}^2 : ||x|| = 1\}$ with subspace topology inherited from Euclidean topology on $\mathbb{R}^2$.
We claimed $\mathbb{R}/\sim$ is homeomorphic to S^1. Define $f : \mathbb{R} \to S^1$ by $f(x) = (\sin 2\pi x, \cos 2\pi x)$. Clearly f is a continuous surjection and it respects $\sim$. By proposition 3.9 there is a unique continuous $\bar{f} : \mathbb{R}/\sim \to S^1$ with $\bar{f} \circ q = f$.

Clearly $\bar{f}$ is a continuous bijection (for injectivity, note each $x \in S^1$ is $f(a)$ for unique $a \in [0, 1)$ and each $b \in \mathbb{R}$ has $b \sim a$ for a unique $a \in [0, 1)$).

Now $\mathbb{R}/\sim = q([0, 1])$ is a continuous image of a compact set and so is compact. And S^1 is Hausdorff as its a metric space. We also know that any continuous bijection from a compact space to a Hausdorff space is a homeomorphism.

Thus we are done, as $\bar{f}$ is a homeomorphism from $\mathbb{R}/\sim$ to S^1.

§3.7 Connectedness

Recall the *Intermediate Value Theorem* (IVT): if $f : [a, b] \to \mathbb{R}$ is continuous and wlog $f(a) < f(b)$ then $[f(a), f(b)] \subset f([a, b])$.
Moreover, if $c, d \in f([a, b])$ with $c < d$ then $[c, d] \subset f([a, b])$.

However, this doesn't work more generally, e.g. if we replace $[a, b]$ by $[-1, 0) \cup (0, 1] = X$. Define $f : X \to \mathbb{R}$ by

$$\begin{cases} 1 & x < 0 \\ 0 & x > 0. \end{cases}$$

Then f is continuous on X, $0 \in f(X)$, $1 \in f(X)$, but e.g. $1/2 \notin f(X)$ so $[0, 1] \not\subset f(X)$. Indeed, this is because $[-1, 0) \cup (0, 1]$ is 'disconnected'.

Definition 3.19 (Connectedness)

A topological space X is **disconnected** if there exist disjoint, non-empty open sets U, V with $X = U \cup V$. We say X is **connected** if X is not disconnected.

Remark 31. 1. Recall that $U \subset X$ is closed iff its complement, $X \setminus U$ is open. If X is disconnected then $X = U \cup V$ so $U = X \setminus V$ and $V = X \setminus U$, so U, V closed. Thus, X is disconnected iff there exist disjoint, non-empty closed sets U, V with $X = U \cup V$.

   Further, if there exist an open and closed set of X, say A, then $A \cup (X \setminus A) = X$. So X is connected iff the only subsets of X that are both open and closed are $\varnothing, X$.

   X is connected iff whenever $U, V \subset X$ are open and disjoint with $X = U \cup V$ then $U = \varnothing$ or $V = \varnothing$. Again, we could replace 'open' by 'closed'.

   If X is disconnected, we say the sets U, V in the definition **disconnect** X.

2. Connectedness is a topological property.

3. if $S \subset X$, where X is a topological space, what does our definition of connectedness say when applied to S? Of course, as usual S has the subspace topology from X so is a topological space in its own right.

   S is disconnected $\iff$ there exist open sets $U, V \subset X$ such that $(S \cap U) \cap (S \cap V) = S \cap U \cap V = \varnothing$, $S \cap U, S \cap V \neq \varnothing$ and $(S \cap U) \cup (S \cap V) = S$ i.e. $S \subset U \cup V$. Again, we say U, V disconnect S.

   Warning 3.1

   It is not necessary to have $U \cap V = \varnothing$, e.g. in $\mathbb{N}$ with cofinite topology, the set $\{1, 2\}$ is disconnected in $\mathbb{N}$ by the open sets $\mathbb{N} \setminus \{1\}, \mathbb{N} \setminus \{2\}$. We have $(\mathbb{N} \setminus \{1\}) \cap (\mathbb{N} \setminus \{2\}) \cap \{1, 2\} = \varnothing$ but $(\mathbb{N} \setminus \{1\}) \cap (\mathbb{N} \setminus \{2\}) \neq \varnothing$. Indeed, if $U, V \subset \mathbb{N}$ are open and non-empty then $U \cap V \neq \varnothing$ (if U, V open in cofinite topology then they have finite complement so their intersection has finite complement.).

   Thus S is connected iff whenever $U, V \subset X$ are open, $S \cap U \cap V = \varnothing$ and $S \subset U \cup V$ then either $S \cap U = \varnothing$ or $S \cap V = \varnothing$.

   Finally, as in remark 1, could replace 'open' with 'closed' in these reformulations of the definition.

Question

Which subsets of $\mathbb{R}$ are connected?

Definition 3.20 (Interval)

A subset $I \subset \mathbb{R}$ is an **interval** if whenever $a < b < c$ with $a, c \in I$ then $b \in I$.

Proposition 3.10

Let $I \subset \mathbb{R}$ with the usual topology. Then I is connected iff I is an interval.

Proof. ($\implies$): Suppose I is not an interval. Then we can find $a < b < c$ with $a, c \in I$ but $b \notin I$. Then $(-\infty, b)$ and (b, ∞) disconnect I in $\mathbb{R}$.

($\impliedby$): Suppose I is an interval. Work in the subspace topology on I. Let $S \subset I$ be open, closed and non-empty. Let $a \in S$.
Suppose we have $b \in I \setminus S$. Wlog $b > a$. Let $c = \sup([a, b] \cap S)$. Then we can find a sequence (x_n) in S with $x_n \to c \in I$. But S is closed in I so $c \in S$. In particular, $c \neq b$. so $c < b$.

But also S is open in I so $\exists \delta > 0$ such that $(c - \delta, c + \delta) \subset S$, wlog $\delta < b - c$. Then $c + \frac{\delta}{2} \in S \cap [a, b]$ ↯. Thus in fact, $S = I$ and hence I is connected. ☐

Theorem 3.6 (Equivalent definition of connectedness)

Let X be a topological space. Then X is connected iff every continuous $f : X \to \mathbb{Z}$ (with the usual topology) is constant.

Proof. ($\implies$): Suppose X is connected and $f : X \to \mathbb{Z}$ continuous. For any $n \in \mathbb{Z}$, $\{n\} \subset \mathbb{Z}$ is open and closed so $f^{-1}(\{n\}) \subset X$ is open and closed, so $f^{-1}(\{n\}) = \varnothing$ or $f^{-1}(\{n\}) = X$[a]. Thus f is constant.

($\impliedby$): Suppose U, V disconnect X. Define $f : X \to \mathbb{Z}$ by $f(x) = \begin{cases} 0 & x \in U \\ 1 & x \in V \end{cases}$. Then for any $A \subset \mathbb{Z}$, $f^{-1}(A) = \varnothing, X, U$ or V[b] and so $f^{-1}(A)$ is open. So f is continuous and non-constant. ☐

[a] Connectedness implies only open and closed subsets of X are $\varnothing, X$.
[b] We get $\varnothing$ if $0, 1 \notin A$, X if $\{0, 1\} \subset A$, ...

Remark 32. 1. Theorem 3.6 together with the IVT can provide an alternative proof of proposition 3.10.

2. Theorem 3.6 remains true with same proof if $\mathbb{Z}$ is replaced by any discrete topological space Y with $|Y| \geq 2$ (need two points to construct $f(x)$ in ($\impliedby$)).

Proposition 3.11

A continuous image of a connected space is connected.

Proof. Let X be a connected topological space, let Y be a topological space and let $f : X \to Y$ be continuous. Suppose $U, V \subset Y$ are open with $f(X) \subset U \cup V$ and $U \cap V \cap f(X) = \varnothing$.

As f is continuous, $f^{-1}(U), f^{-1}(V) \subset X$ are open. Also $X = f^{-1}(U) \cup f^{-1}(V)$ and $f^{-1}(U) \cap f^{-1}(V) = \varnothing$. As X is connected, wlog $f^{-1}(U) = \varnothing$. Then $U \cap f(X) = \varnothing$. So $f(X)$ is connected. □

Proposition 3.12

A product of connected spaces is connected.

Proof. Let (X, τ) and (Y, σ) be connected topological spaces, and let ρ be the product topology on $X \times Y$.

Suppose $U, V \in \rho$ with $U \cap V = \varnothing$ and $U \cup V = X \times Y$. We want to show $U = \varnothing$, $V = X \times Y$ or $U = X \times Y$, $V = \varnothing$.

Fix $x \in X$. Then $\{x\} \times Y$ is homeomorphic to Y (Sheet 3). In particular, $\{x\} \times Y$ is connected. Then $\{x\} \times Y \subset U$ or $\{x\} \times Y \subset V$ as otherwise U, V would disconnect $\{x\} \times Y$ in $X \times Y$.

Let $A = \{x \in X : \{x\} \times Y \subset U\}$ and $B = \{x \in X : \{x\} \times Y \subset V\}$. Clearly $A \cap B = \varnothing$ (as $U \cap V = \varnothing$) and we've proved that $X = A \cup B$. Now we want to show that A, B are open to use connectedness in X.

Suppose $x \in A$. Then $\{x\} \times Y \subset U$. Then (assuming $Y \neq \varnothing$, which we do wlog) pick any $y \in Y$. Then $(x, y) \in U$. U is open so can find $T \in \tau$, $S \in \sigma$ such that $(x, y) \in T \times S \subset U$. In particular, for all $w \in T$ then $(w, y) \in U$ and so $\{w\} \times Y \subset U$[a], i.e. $w \in A$. We now have $T \in \tau$ with $x \in T \subset A$ so A is a neighbourhood of x in X. Hence A is open.

Similarly B is open. Since X is connected, so $A = \varnothing$ giving $U = \varnothing$ or $B = \varnothing$ giving $V = \varnothing$. Hence $X \times Y$ is connected. □

[a] If $\{w\} \times \{y\} \in U$ and $\{w\} \times \{a\} \in V$ for some $a \in Y$ this contradicts $\{x\} \times Y \subset U$ or $\{x\} \times Y \subset V$.

Example 3.20 (What can we do with spaces that aren't connected?)

Recall $[-1, 0) \cup (0, 1]$ is not connected. But it <u>is</u> a disjoint union of connected sets. Moreover, any proper superset of $[-1, 0)$ or $(0, 1]$ in $[-1, 0) \cup (0, 1]$ is disconnected.

We see here that the space is not connected, but we can take two maximal subsets of the space that are connected that gives us the entire space.

Definition 3.21 (Connected Component)

Let X be a topological space. A **connected component** of X is a maximal connected subset A of X: that is to say, A is connected but if $A \subset B \subset X$ with B connected then $A = B$.

Theorem 3.7

The connected components of a topological space X form a partition of X.

Proof. Define $\sim$ on X by $x \sim y$ iff $\exists A \subset X$ connected with $x, y \in A$. Clearly $\sim$ is reflexive ($\{x\}$ connected) and symmetric. Let's check that it is transitive.

Suppose $x, y, z \in X$ with $x \sim y$ and $y \sim z$. Then $\exists A, B \subset X$ connected with $x, y \in A$ and $y, z \in B$. Now $x, z \in A \cup B$. Suppose U, V disconnect $A \cup B$ in X. Wlog $y \in U$. Pick $w \in V \cap (A \cup B)$. Wlog $w \in A$. But also $y \in A$ so U, V disconnect A $\lightning$. So $A \cup B$ is connected, so $x \sim z$, thus $\sim$ is indeed an equivalence relation.

Now suppose S is an equivalence class of $\sim$. We want to show that S is connected. Suppose U, V disconnect S. Then we can find $x \in U \cap S$, $y \in V \cap S$ and $U \cap V \cap S = \varnothing$. Then $x \sim y$ so there is a connected $A \subset X$ with $x, y \in A$. For all $z \in A$, $x, z \in A$ are connected so $x \sim z$ so $z \in S$. So $A \subset S$ and so $U \cap V \cap A = \varnothing$ which gives U, V disconnects A $\lightning$. Hence S is connected.

Now let's show S is a connected component. Suppose $S \subset T \subset X$ with T connected. Let $x \in S$. Then for all $y \in T$, $x, y \in T$ with T connected so $x \sim y$. Thus $T \subset S$, so $S = T$ and S is a connected component.

Finally let R be a connected component and let $x, y \in R$. Then, as R is connected, $x \sim y$. So R is contained in some equivalence class Q of $\sim$. But $R \subset Q$ with Q connected so $R = Q$. Hence equivalence classes of $\sim$ are precisely the connected components. $\square$

Another concept of connectedness:

Remark 33. This is what tells us connected components exist.

Definition 3.22 (Paths)

A **path** from x to y in a topological space X is a continuous function $\varphi : [0, 1] \to X$ with $\varphi(0) = x$, $\varphi(1) = y$.

Definition 3.23 (Path-Connected)

X is **path-connected** if for all $x, y \in X$ there is a path from x to y.

Proposition 3.13

A path-connected space X is connected.

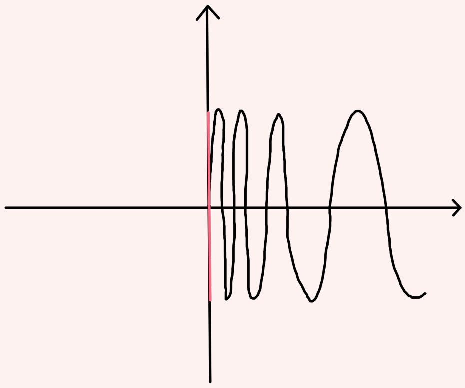
Proof. Suppose U, V disconnect X. Pick $a \in U, b \in V$. Let φ be a path in X from a to b. Then U, V disconnect $\varphi([0,1])$ $\lightning$. □

Warning 3.2

The converse is not true!

Example 3.21

Consider



$$A = \{(0, y) : -1 \leq y \leq 1\}$$
$$B = \left\{\left(x, \sin\frac{1}{x}\right) : 0 < x \leq 1\right\}$$

and $X = A \cup B \subset \mathbb{R}^2$. Let's show that X is connected but not path-connected.

Claim 3.2

X connected.

Proof. Clearly A, B themselves are path-connected so connected. Suppose U, V disconnect X in $\mathbb{R}^2$. Then wlog $A \subset U$, $B \subset V$. So $(0,0) \in A \subset U$. U is open so there exists $\delta > 0$ such that $B_\delta((0,0)) \subset U$. Pick n such that $\frac{1}{2n\pi} < \delta$. Then $(\frac{1}{2n\pi}, 0) \in U \cap B$ ↯. □

Claim 3.3

X not path-connected.

Proof. Suppose φ is a path from $(0,0)$ to $(1, \sin 1)$ in X. Let $\sigma = \sup\{t \in [0,1] : \varphi_1(t) = 0\}$. Let $y = \varphi_2(\sigma)$. Then, as φ is continuous, $\varphi(\sigma) = (0, y)$.

Choose $\delta > 0$ such that $|\sigma - t| < \delta \implies \|\varphi(\sigma) - \varphi(t)\| < 1$. Wlog $\delta < 1 - \sigma$. By definition of σ, $\varphi_1(\sigma + \frac{\delta}{2}) = x > 0$. Choose $w \in (0, x)$ such that $\left|\sin \frac{1}{w} - y\right| \geq 1$. Then by IVT, there is some $t \in \left(\sigma, \sigma + \frac{\delta}{2}\right)$ such that $\varphi_1(t) = w$. Then $|\sigma - t| < \delta$ but $\|\varphi(\sigma) - \varphi(t)\| \geq |\varphi_2(\sigma) - \varphi_2(t)| = \left|\sin \frac{1}{w} - y\right| \geq 1$ ↯. □

Proposition 3.14

An open, connected subset of Euclidean space is path-connected.

Proof. Let $X \subset \mathbb{R}^n$ be open and connected. If $X = \varnothing$ we are done. So assume $X \neq \varnothing$.

Fix $a \in X$ and let

$$U = \{x \in X : \exists \text{ path in } X \text{ from } a \text{ to } x\}.$$

Note that $U \neq \varnothing$ as $a \in U$, since there is a constant path from a to a.

Claim 3.4

U open in X.

Proof. Suppose $b \in U$. Since X is open we can pick $\delta > 0$ such that $B_\delta(b) \subset X$. Let φ be a path from a to b in X and let $x \in B_\delta(b)$. Then θ is a path in X from a to x where

$$\theta(t) = \begin{cases} \varphi(2t) & 0 \leq t \leq \frac{1}{2} \\ b + 2(t - \frac{1}{2})(x - b) & \frac{1}{2} \leq t \leq 1 \end{cases}.$$

□

Claim 3.5

U closed in X, i.e. $X \setminus U$ is open in X.

Proof. Let $b \in X \setminus U$. Choose $\delta > 0$ such that $B_\delta(b) \subset X$. Suppose $x \in B_\delta(b) \cap U$. Let φ be a path in X from a to x. Then

$$t \mapsto \begin{cases} \varphi(2t) & 0 \le t \le \frac{1}{2} \\ b + 2(t - \frac{1}{2})(b - x) & \frac{1}{2} \le t \le 1 \end{cases}$$

is a path from a to b in X ↯. Thus $B_\delta(b) \subset X \setminus U$. □

Hence, as X is connected, $U = X$. But the point a was arbitrary. Thus X is path-connected. □

Remark 34. Recall that we don't always specify the topology when defining a topological space; we should always assume it's the standard one. In particular, $\mathbb{R}$ comes with the usual topology, and $\mathbb{R}^n$ comes with the Euclidean topology. $X \subset \mathbb{R}, \mathbb{R}^n$ comes with the subspace topology from their standard topologies.

Products/Quotients come with the product/quotient topology respectively.

Part II
Generalizing differentiation

§4 Differentiation

§4.1 The First Derivative

Recall that $f : \mathbb{R} \to \mathbb{R}$ is *differentiable* at $a \in R$ with *derivative* A if

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

We write $f'(a) = A$.

Question
How can we generalise to $f : \mathbb{R}^n \to \mathbb{R}^n$?

Obviously if $n = 1$ exact same definition works.
Problem: If $n \geq 2$, then dividing by $h \in \mathbb{R}^n$ makes no sense.

Definition 4.1 (ith Partial Derivative)
If $f : \mathbb{R}^n \to \mathbb{R}^n$, the **$i$th partial derivative** of f at $a \in \mathbb{R}^n$ is

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

where this limit exists, where $e_1, \ldots, e_n$ is standard basis of $\mathbb{R}^n$.

However, with this definition all partial derivatives existing at a point doesn't mean that the function is differentiable at that point!

Example 4.1
Consider $f : \mathbb{R}^2 \to \mathbb{R}$ with

$$f(x, y) = \begin{cases} 0 & x = 0 \text{ or } y = 0 \\ 1 & \text{otherwise.} \end{cases}$$

Both partial derivatives exist at $(0, 0)$, but f is not continuous there.

Let's try think of a better definition. Return to $f : \mathbb{R} \to \mathbb{R}$, where

$$\begin{aligned} f'(a) = A &\Leftrightarrow \frac{f(a+h)-f(a)}{h} \to A \text{ as } h \to 0 \\ &\Leftrightarrow \frac{f(a+h)-f(a)}{h} = A + \varepsilon(h) \text{ where } \varepsilon(h) \to 0 \text{ as } h \to 0 \\ &\Leftrightarrow f(a+h) = f(a) + Ah + \underbrace{\varepsilon(h)h}_{o(h)} \text{ where } \varepsilon(h) \to 0 \text{ as } h \to 0. \end{aligned}$$

In words, 'small changes in a produce approximately linear changes in $f(a)$'.

Definition 4.2 (Differentiable)
Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $a \in \mathbb{R}^n$. We say that f is **differentiable** at a if there is a linear map $\alpha = \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with

$$f(a+h) = f(a) + \alpha(h) + \varepsilon(h)\|h\| \tag{$\star$}$$

where $\varepsilon(h) \to 0$ as $h \to 0$.

Proposition 4.1
Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$, $a \in \mathbb{R}^n$, $\alpha, \beta \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and

$$\begin{aligned} f(a+h) &= f(a) + \alpha(h) + \varepsilon(h)\|h\| \\ f(a+h) &= f(a) + \beta(h) + \eta(h)\|h\| \end{aligned}$$

with $\varepsilon(h), \eta(h) \to 0$ as $h \to 0$.
Then $\alpha = \beta$.

Definition 4.3 (Derivative)
After proving proposition 4.1, we know that the α in ($\star$) is unique. We say α is the **derivative** of f at a, and write $Df|_a = \alpha$. So, if f is differentiable at a, then

$$f(a+h) = f(a) + Df|_a(h) + \varepsilon(h)\|h\|$$

where $\varepsilon(h) \to 0$ as $h \to 0$.

Remark 35. If $f : \mathbb{R} \to \mathbb{R}^m$, then $Df|_a = f'(a)h$.

Proof of proposition 4.1. Let $h \in \mathbb{R}^n$, $h \neq 0$. Then

$$\alpha(h) - \beta(h) = (\eta(h) - \varepsilon(h))\|h\|$$

Then for $\lambda \in \mathbb{R}, \lambda \neq 0$,

$$\begin{aligned}\alpha(h) - \beta(h) &= \frac{\alpha(\lambda h) - \beta(\lambda h)}{\lambda} \\ &= \frac{(\eta(\lambda h) - \varepsilon(\lambda h))\|\lambda h\|}{\lambda} \\ \|\alpha(h) - \beta(h)\| &= \|\eta(\lambda h) - \varepsilon(\lambda h)\|\|h\| \to 0\end{aligned}$$

as $\lambda \to 0$. Hence $\alpha(h) = \beta(h)$. Hence $\alpha = \beta$. □

Remark 36. 1. To consider differentiability of f at a, it only matters what happens on some neighbourhood of a. So definition works if instead of $f : \mathbb{R}^n \to \mathbb{R}^m$, we have $f : \mathcal{N} \to \mathbb{R}^m$ where $\mathcal{N} \subset \mathbb{R}^n$ is a neighbourhood of a, or, in particular, if $f : B_\delta(a) \to \mathbb{R}^m$ where $\delta > 0$. (Imagine f defined as anything on the rest of $\mathbb{R}^n$ and it makes no difference.)

2. We can define the ℓ_1 and ℓ_∞ norms on $\mathbb{R}^n$ by

$$\begin{aligned}\|x\|_1 &= d_1(0, x) = \sum_{i=1}^{n} |x_i|, \text{ and} \\ \|x\|_\infty &= d_\infty(0, x) = \max_i |x_i|.\end{aligned}$$

Note $\|x\|_1 \geq 0$ with equality iff $x = 0$, $\|\lambda x\|_1 = |\lambda|\|x\|_1$; $\|x + y\|_1 \leq \|x\|_1 + \|y\|_1$.

Similarly for $\|\cdot\|_\infty$. We've seen that for all $x \in \mathbb{R}^n$,

$$\begin{aligned}\|x\|_\infty &\leq \|x\| \leq \sqrt{n}\|x\|_\infty, \text{ and} \\ \|x\|_\infty &\leq \|x\|_1 \leq n\|x\|_\infty\end{aligned}$$

So we can replace $\|\cdot\|$ in the definition of derivative by $\|.\|_1$ or $\|.\|_\infty$ and defintion doesn't change. Sometimes this is useful for computation.

Consider the vector space $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ of linear maps $\mathbb{R}^n \to \mathbb{R}^m$. We have $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{mn}$ with isomorphism (by writing map as matrix w.r.t standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$). So we could think about Euclidean norm of a linear map. But this seems a bit unnatural.

Definition 4.4 (Operator Norm)

The **operator norm** on $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is defined by

$$\underbrace{\|\alpha\|}_{\text{operator norm}} = \sup\{\|\alpha x\| : \|x\| = 1\}.$$

The $\|\alpha x\|$ in sup is the Euclidean norm as $\alpha x \in \mathbb{R}^m$.

Proposition 4.2

Let $\|\cdot\|$ be the operator norm on $V = \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Let $\alpha, \beta \in V$. Then:

1. $\|\alpha\| \geq 0$ with equality iff $\alpha = 0$;
2. $\forall \lambda \in \mathbb{R}, \|\lambda\alpha\| = |\lambda|\|\alpha\|$;
3. $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$;
4. $\forall x \in \mathbb{R}^n, \|\alpha x\| \leq \|\alpha\|\|x\|$;
5. $\|\alpha\beta\| \leq \|\alpha\|\|\beta\|$; ($\alpha\beta$ refers to composition in the case $m = n$.)
6. if $\|\cdot\|'$ is the Euclidean norm on $V \cong \mathbb{R}^{mn}$ with the standard isomorphism then there are constants (indep. of α) c, d (depending on n, m) such that

$$c\|\alpha\| \leq \|\alpha\|' \leq d\|\alpha\|.$$

Remark 37. A linear map from $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous and $\{x \in \mathbb{R}^n : \|x\| = 1\}$ is compact so operator norm is well-defined.

Notation. The standard is that $\|\cdot\|$ refers to the operator norm if applied to a linear map and Euclidean norm if applied to point of $\mathbb{R}^n$, unless otherwise stated.

Proof. (i) - (iii) is an exercise.

(iv): Let $x \in \mathbb{R}^n$. If $x = 0$ that's fine. Otherwise, $\alpha(x) = \|x\|\alpha\left(\frac{x}{\|x\|}\right)$ with $\left\|\frac{x}{\|x\|}\right\| = 1$. So $\|\alpha(x)\| \leq \|x\|\|\alpha\|$.

(v): Let $x \in \mathbb{R}^n$ with $\|x\| = 1$. Then $\|\alpha\beta x\| \leq \|\alpha\|\|\beta x\| \leq \|\alpha\|\|\beta\|\|x\| = \|\alpha\|\|\beta\|$ by taking (iv) twice. So $\|\alpha\beta\| \leq \|\alpha\|\|\beta\|$.

(vi): Let $x \in \mathbb{R}^n$ with $\|x\| = 1$. $\|\alpha x\| \leq \sqrt{m}\max_{1\leq i\leq m} |(\alpha x)_i|$. Let A be the matrix of α w.r.t the standard bases $e_1, \ldots, e_n$ of $\mathbb{R}^n$ and $f_1, \ldots, f_m$ of $\mathbb{R}^m$. Then

$$\begin{aligned}
\|\alpha x\| &\leq \sqrt{m} \max_{1\leq i\leq m} \left|\sum_{j=1}^n A_{ij}x_j\right| \\
&\leq \sqrt{m} \max_{1\leq i\leq m} \sum_{j=1}^n |A_{ij}||x_j| \\
&\leq \sqrt{m} \max_{1\leq i\leq m} \sum_{j=1}^n \|\alpha\|'^{a} \\
&= n\sqrt{m}\|\alpha\|'.
\end{aligned}$$

Hence $\|\alpha\| \leq n\sqrt{m}\|\alpha\|'$.

On the other hand, pick i, j maximize $|A_{ij}|$. Then $\|\alpha e_j\| \geq \|(\alpha e_j)_i\| = |A_{ij}|$. But

$$\begin{aligned} \|\alpha\|' &\leq \sqrt{mn}|A_{ij}| \\ &\leq \sqrt{mn}\|\alpha e_j\| \\ &\leq \sqrt{mn}\|\alpha\|. \end{aligned}$$

This proves (vi) with $d = \sqrt{mn}$ and $c = \frac{1}{n\sqrt{m}}$. □

[a] $|A_{ij}| \leq \|\alpha\|'$ as $\|\cdot\|'$ is the Euclidean norm. Also, $\|x\| = 1$ so $|x_j| \leq 1$.

Proposition 4.3

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ which is differentiable at $a \in \mathbb{R}^n$. Then f is continuous at a.

Proof. Write $f(a+h) = f(a) + Df|_a(h) + \varepsilon(h)\|h\|$ where both $\varepsilon(h), \|h\| \to 0$ as $h \to 0$. Also $Df|_a$ is linear so continuous so $Df|_a(h) \to Df|_a(0) = 0$ as $h \to 0$. Then $f(a+h) \to f(a)$ as $h \to 0$. □

Proposition 4.4

Let $f, g : \mathbb{R}^n \to \mathbb{R}^m$ and $\lambda : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $a \in \mathbb{R}^n$. Then $f + g, \lambda f$ are differentiable at a with

$$D(f+g)|_a = Df|_a + Dg|_a$$

and

$$D(\lambda f)|_a(h) = \lambda(a)Df|_a(h) + D\lambda|_a(h)f(a).$$

Proof. We have

$$\begin{aligned} f(a+h) &= f(a) + Df|_a(h) + \varepsilon(h)\|h\|, \\ g(a+h) &= g(a) + Dg|_a(h) + \eta(h)\|h\|, \\ \lambda(a+h) &= \lambda(a) + D\lambda|_a(h) + \zeta(h)\|h\|, \end{aligned}$$

where $\varepsilon(h), \eta(h), \zeta(h) \to 0$ as $h \to 0$. Now, $(f+g)(a+h) = (f+g)(a) + (Df|_a + Dg|_a)(h) + (\varepsilon(h) + \eta(h))\|h\|$ where $Df|_a + Dg|_a$ is linear and $\varepsilon(h) + \eta(h) \to 0$ as $h \to 0$.

Also,

$$\begin{aligned} (\lambda f)(a+h) = (\lambda f)(a) \\ &+ \lambda(a)Df|_a(h) + D\lambda|_a(h)f(a) \\ &+ \xi(h)\|h\| \end{aligned}$$

where $h \mapsto \lambda(a)Df|_a(h) + D\lambda|_a(h)f(a)$ is a linear map, and

$$\begin{aligned}\xi(h) = \zeta(h)f(a) + Df|_a(h)D\lambda|_a(h)\frac{1}{\|h\|} \\ + Df|_a(h)\zeta(h) + \lambda(a)\varepsilon(h) \\ + D\lambda|_a(h)\varepsilon(h) + \varepsilon(h)\zeta(h)\|h\| \to 0\end{aligned}$$

as $h \to 0$ since $\varepsilon(h), \zeta(h) \to 0$ as $h \to 0$, $\|h\| \to 0$ as $h \to 0$.

$Df|_a(h)$, $D\lambda|_a(h)$ are linear so continuous, so $Df|_a(h), D\lambda|_a(h) \to 0$ as $h \to 0$, and

$$\begin{aligned}\left\|Df|_a(h)D\lambda|_a(h)\frac{1}{\|h\|}\right\| &\leq \|Df|_a(h)\|\|D\lambda|_a(h)\|\frac{1}{\|h\|} \\ &\leq \|Df|_a\|\|h\|\|D\lambda|_a\|\|h\|\frac{1}{\|h\|} \\ &= \|Df|_a\|\|D\lambda|_a\|\|h\| \to 0\end{aligned}$$

as $h \to 0$, as required. □

Partial derivatives can still be useful for computation.

Proposition 4.5

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $a \in \mathbb{R}^n$. Write $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$, where for each i, $f_i : \mathbb{R}^n \to \mathbb{R}$. Then

1. f is differentiable at a iff each f_i is differentiable at a, in which case $Df|_a = \begin{pmatrix} Df_1|_a \\ \vdots \\ Df_m|_a \end{pmatrix}$; and
2. if f is differentiable at a and A is the matrix of $Df|_a$ in terms of the standard bases then $A_{ij} = D_j f_i(a)$.

Proof. 1. ($\implies$): Write $f(a+h) = f(a) + Df|_a(h) + \varepsilon(h)\|h\|$ where $\varepsilon(h) \to 0$ as $h \to 0$. Then

$$f_i(a+h) = f_i(a) + (Df|_a)_i(h) + \varepsilon_i(h)\|h\|$$

where $(Df|_a)_i : \mathbb{R}^n \to \mathbb{R}$ is linear[a] and $|\varepsilon_i(h)| \leq \|\varepsilon(h)\| \to 0$ as $h \to 0$.

($\impliedby$): For each i, write $f_i(a+h) = f_i(a) + (Df|_a)_i(h) + \varepsilon_i(h)\|h\|$. Then

$$f(a+h) = f(a) + \alpha(h) + \varepsilon(h)\|h\|$$

where $\alpha = \begin{pmatrix} Df_1|_a \\ \vdots \\ Df_m|_a \end{pmatrix} : \mathbb{R}^n \to \mathbb{R}^m$ is linear and $\|\varepsilon(h)\| = \left\| \begin{pmatrix} \varepsilon_1(h) \\ \vdots \\ \varepsilon_m(h) \end{pmatrix} \right\| = \sqrt{\sum_{i=1}^m \varepsilon_i(h)^2} \to 0$ as $h \to 0$.

2. Write $f(a+h) = f(a) + Df|_a(h) + \varepsilon(h)\|h\|$ where $\varepsilon(h) \to 0$ as $h \to 0$. Let $e_1, \dots, e_n$ be the standard basis in $\mathbb{R}^n$. Then

$$\begin{aligned} \frac{f(a+ke_j) - f(a)}{k} &= \frac{Df|_a(ke_j) + \varepsilon(ke_j)\|ke_j\|}{k} \\ &= Df|_a(e_j) \pm^{b} \varepsilon(ke_j) \\ &\to Df|_a(e_j) \end{aligned}$$

as $k \to 0$. So all partial derivatives of f exist at a and $D_j f(a) = Df|_a(e_j)$.

□

[a] As it is linear in each coordinate
[b] We don't know the sign of k

Definition 4.5 (Jacobian Matrix)

The matrix A in (b) is called the **Jacobian matrix** of f at a.

Theorem 4.1 (The Chain Rule)

Let $f : \mathbb{R}^p \to \mathbb{R}^n$ be differentiable at $a \in \mathbb{R}^p$ and let $g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $f(a) \in \mathbb{R}^n$. Then $g \circ f$ is differentiable at a with $D(g \circ f)|_a = Dg|_{f(a)} \circ Df|_a$.

Remark 38. Intuitively why this is true: If f is approximately linear near a and g is approximately linear near $f(a)$ then $g \circ f$ is approximately linear near a and the linear approximation to get near a is the chain rule result. Proof unfortunately is messy to make sure error terms behave.

Proof. Write

$$\begin{aligned} f(a+h) &= f(a) + \alpha(h) + \varepsilon(h)\|h\| \\ g(f(a)+k) &= g(f(a)) + \beta(k) + \eta(k)\|k\| \end{aligned}$$

where $\alpha = Df|_a, \beta = Dg|_{f(a)}$ are linear, $\varepsilon(h) \to 0$ as $h \to 0$, $\eta(k) \to 0$ as $k \to 0$. Now

$$g(f(a+h)) = g(f(a) + \underbrace{\alpha(h) + \varepsilon(h)\|h\|}_{k})$$

$$
\begin{aligned}
&= g(f(a)) + \beta(\alpha(h) + \varepsilon(h)\|h\|) \\
&\quad + \eta(\alpha(h) + \varepsilon(h)\|h\|)\Big\|\alpha(h) + \varepsilon(h)\|h\|\Big\| \\
&= g(f(a)) + \underbrace{\beta(\alpha(h))}_{\text{linear}} + \underbrace{\zeta(h)\|h\|}_{\text{small}}
\end{aligned}
$$

where

$$
\zeta(h) = \beta(\varepsilon(h)) + \eta(\alpha(h) + \varepsilon(h)\|h\|)\left\|\frac{\alpha(h)}{\|h\|} + \varepsilon(h)\right\|.
$$

Now $\varepsilon(h) \to 0$ as $h \to 0$ and β linear so continuous so $\beta(\varepsilon(h)) \to \beta(0) = 0$ as $h \to 0$. Next, α linear so continuous and so $\alpha(h) \to \alpha(0) = 0$ as $h \to 0$. And $\varepsilon(h)\|h\| \to 0 \times 0 = 0$ as $h \to 0$. So $\alpha(h) + \varepsilon(h)\|h\| \to 0$ as $h \to 0$.
Wlog $\eta(0) = 0$ so η continuous at 0. Then $\eta(\alpha(h) + \varepsilon(h)\|h\|) \to 0$ as $h \to 0$.

Finally,

$$
\begin{aligned}
\left\|\frac{\alpha(h)}{\|h\|} + \varepsilon(h)\right\| &\le \frac{\|\alpha(h)\|}{\|h\|} + \|\varepsilon(h)\| \\
&\le \frac{\|\alpha\|\|h\|}{\|h\|} + \|\varepsilon(h)\| \\
&= \|\alpha\| + \|\varepsilon(h)\| \to \|\alpha\|
\end{aligned}
$$

as $h \to 0$. Hence $\zeta(h) \to 0$ as $h \to 0$. □

Example 4.2

Suppose f is constant. Then $f(a+h) = f(a) + 0 + 0\|h\|$ so f is everywhere differentiable with derivative the zero map.

Example 4.3

Suppose f is linear. Then $f(a+h) = f(a) + f(h) + 0\|h\|$ so f is everywhere differentiable with $Df|_a = f$ for all a.

Example 4.4

Suppose $f : \mathbb{R} \to \mathbb{R}^m$. As remarked earlier, for $a \in \mathbb{R}$, f is differentiable in old sense at a iff it is differentiable in new sense, in which case $Df|_a(h) = hf'(a)$.

Example 4.5

Using the above together with chain rule, we obtain many differentiable functions. E.g. $f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} e^{x+y} \\ \cos(xy) \end{pmatrix}$ is differentiable. We can show this by considering the projection maps $\pi_1, \pi_2 : \mathbb{R}^2 \to \mathbb{R}$, $\pi_1 \begin{pmatrix} x \\ y \end{pmatrix} = x$, $\pi_2 \begin{pmatrix} x \\ y \end{pmatrix} = y$ are linear so differentiable. So by chain rule,

$$f_1(z) = e^{\pi_1(z)+\pi_2(z)}, f_2(z) = \cos(\pi_1(z)\pi_2(z))$$

are differentiable. So by proposition 4.5(1), f is differentiable.

What is derivative of f at $z = \begin{pmatrix} x \\ y \end{pmatrix}$? It is a linear map $\mathbb{R}^2 \to \mathbb{R}^2$. By proposition 4.5(2), the matrix of derivatives is given by the partial derivatives:

$$Df|_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)} \text{ has matrix } \begin{pmatrix} e^{x+y} & e^{x+y} \\ -y\sin(xy) & -x\sin(xy) \end{pmatrix}.$$

Example 4.6

Let $\mathcal{M}_n$ be the vector space of $n \times n$ real matrices. So $\mathcal{M}_n \cong \mathbb{R}^{n^2}$ so we can consider differentiability of $f : \mathcal{M}_n \to \mathcal{M}_n$.
Recall definition is still the same if we replace Euclidean norm by operator norm, so write $\|\cdot\|$ for operator norm on $\mathcal{M}_n$. Define $f : \mathcal{M}_n \to \mathcal{M}_n$ by $f(A) = A^2$. Then

$$\begin{aligned} f(A+H) &= (A+H)^2 \\ &= \underbrace{A^2}_{f(A)} + \underbrace{AH + HA}_{\text{linear}} + \underbrace{H^2}_{\text{higher order}} \end{aligned}$$

where $\left\|\frac{H^2}{\|H\|}\right\| \le \frac{\|H\|^2}{\|H\|} = \|H\| \to 0$ as $H \to 0$. So f is everywhere differentiable and $Df|_A(H) = AH + HA$.

Example 4.7

We have $\det : \mathcal{M}_n \to \mathbb{R}$. We have

$$\begin{aligned} \det(I+H) &= \begin{vmatrix} 1+H_{11} & & H'_{ij}s \\ & \ddots & \\ H'_{ij}s & & 1+H_{nn} \end{vmatrix} \\ &= 1 + \operatorname{tr}(H) + \text{terms involving two or more } H_{ij} \text{ multiplied together.} \end{aligned}$$

Note $\left|\frac{H_{ij}H_{kl}}{\|H\|_2}\right|$ [a] $\le |H_{kl}| \to 0$ as $H \to 0$. So det differentiable at I with $D\det|_I(H) = \operatorname{tr}(H)$.
Suppose $A \in \mathcal{M}_n$ invertible. Then

$$\begin{aligned}\det(A+H) &= \det(A)\det\left(I + A^{-1}H\right)\\ &= \det A(1 + \operatorname{tr}\left(A^{-1}H\right) + \varepsilon(A^{-1}H)\left\|A^{-1}H\right\|)\\ &= \det A + (\det A)(\operatorname{tr} A^{-1}H) + (\det A)\varepsilon(A^{-1}H)\left\|A^{-1}H\right\|\end{aligned}$$

where $\varepsilon(K) \to 0$ as $K \to 0$, and $\left|\frac{(\det A)\varepsilon(A^{-1}H)\|A^{-1}H\|}{\|H\|}\right| \le \left|(\det A)\varepsilon(A^{-1}H)\|A^{-1}\|\right| \to 0$ as $H \to 0$.
So det is differentiable at A with $D\det|_A(H) = (\operatorname{tr} A^{-1}H)(\det A)$.

[a] This is the Euclidean norm

Recall: if $f : \mathbb{R} \to \mathbb{R}$ differentiable with zero derivative everywhere then f is constant. This followed from the Mean Value Theorem. Let's look at what the MVT is like in higher dimensions.

Theorem 4.2 (Mean Value Inequality)
Let $f : \mathbb{R}^n \to \mathbb{R}^m$. Suppose f is differentiable on an open set $X \subset \mathbb{R}^n$ with $a, b \in X$. Suppose further $[a, b] = \{a + t(b - a) : 0 \le t \le 1\} \subset X$.
Then $\|f(b) - f(a)\| \le \|b - a\| \sup_{z\in(a,b)} \|Df|_z\|$ where $(a, b) = [a, b]\backslash\{a, b\}$.

Proof. Define $\varphi : [0, 1] \to \mathbb{R}$ by $\varphi = f(a + t(b - a)) \cdot (f(b) - f(a))$. Then $\varphi = \alpha \circ f \circ \beta$ where $\beta : [0, 1] \to \mathbb{R}^n, \beta(t) = a + t(b - a)$ and $\alpha : \mathbb{R}^m \to \mathbb{R}, \alpha(x) = x \cdot (f(b) - f(a))$. Clearly φ is continuous on $[0, 1]$.

Now α is a linear map so is everywhere differentiable with $D\alpha|_x = \alpha$. Next, $\beta([0, 1]) \subset X$ and f is differentiable on X. Finally, if $t \in (0, 1)$ then β differentiable at t with $\beta'(t) = b - a$, i.e. $D\beta|_t(h) = h(b - a)$.

Hence by the chain rule, if $t \in (0, 1)$ then φ is differentiable at t and

$$\begin{aligned}D\varphi|_t(h) &= D\alpha|_{f(\beta(t))}\left(Df|_{\beta(t)}(D\beta|_t(h))\right)\\ &= \alpha\left(Df|_{a+t(b-a)}(h(b-a))\right)\\ &= (f(b) - f(a)) \cdot \left(hDf|_{a+t(b-a)}(b-a)\right)\\ &= h\left((f(b) - f(a)) \cdot Df|_{a+t(b-a)}(b-a)\right).\end{aligned}$$

That is, $\varphi'(t) = (f(b) - f(a)) \cdot Df|_{a+t(b-a)}(b-a)$. So, by the MVT,

$$\begin{aligned}
\|f(b) - f(a)\|^2 &= (f(b) - f(a)) \cdot f(b) - (f(b) - f(a)) \cdot f(a) \\
&= \varphi(1) - \varphi(0) \\
&= \varphi'(t) \text{ for some } t \in (0,1) \\
&= (f(b) - f(a)) \cdot Df|_{a+t(b-a)}(b-a) \\
&\leq \|f(b) - f(a)\| \left\| Df|_{a+t(b-a)}(b-a) \right\| \text{ by Cauchy-Schwarz} \\
&\leq \|f(b) - f(a)\| \left\| Df|_{a+t(b-a)} \right\| \|b-a\|.
\end{aligned}$$

Hence $\|f(b) - f(a)\| \leq \|b-a\| \left\| Df|_{a+t(b-a)} \right\|$. □

Corollary 4.1

Let $X \subset \mathbb{R}^n$ be open and connected, and let $f : X \to \mathbb{R}^m$ be differentiable with $Df|_x$ the zero map for all $x \in X$. Then f is constant on X.

Proof. By MVI, f is 'locally' constant'. For each $x \in X$, there is some $\delta > 0$ such that $B_\delta(x) \subset X$ and so f is constant on $B_\delta(x)$. (Since $B_\delta(x)$ is convex so contains line segments joining each pair of points.)

Note that as X is open, if $U \subset X$ then U open in X iff U open in $\mathbb{R}^n$.
If $X = \varnothing$ we are done, so suppose not. Fix $a \in X$, and let

$$U = \{x \in X : f(x) = f(a)\}.$$

$U \neq \varnothing$: $a \in U$.

U is open: if $b \in U$ then there is some $\delta > 0$ such that $B_\delta(b) \subset X$ and f constant on $B_\delta(b)$ so $B_\delta(b) \subset U$.

U is closed in X: if $b \in X \backslash U$ then there is some $\delta > 0$ such that $B_\delta(b) \subset X$ and f constant on $B_\delta(b)$ so $B_\delta(b) \subset X \backslash U$. So $X \backslash U$ is open in $\mathbb{R}^n$, so open in X. Hence U is closed in X.

But X is connected, so only clopen set is X $\therefore U = X$. □

We've seen if f is differentiable at a then partial derivatives all exist at a and the matrix of $Df|_a$ is given by the partial derivatives. But, on the other hand, we can have all partial derivatives existing at a but f not differentiable at a. However, there is a partial converse to this.

Theorem 4.3

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and let $a \in \mathbb{R}^n$. Suppose there is some neighbourhood of a such that the partial derivatives $D_i f$ for $1 \le i \le n$ all exist on this neighbourhood and are continuous at a. Then f is differentiable at a.

How can we prove this?

For simplicity, let's try to prove this for $n = 2, m = 1$. So $f : \mathbb{R}^2 \to \mathbb{R}$. Write $a = (x, y)$. We want to think about $f(x + h, y + k)$ for small h, k. Now, by definition of partial derivatives,

$$f(x+h, y+k) = f(x+h, y) + kD_2 f(x+h, y) + o(k)$$

and

$$f(x+h, y) = f(x, y) + hD_1 f(x, y) + o(h).$$

Hence

$$\begin{aligned} f(x+h, y+k) &= f(x,y) + hD_1 f(x,y) + kD_2 f(x+h, y) + o(h) + o(k) \\ &= f(x,y) + hD_1 f(x,y) + k(D_2 f(x,y) + o(1)) + o(h) + o(k) \\ &= f(x,y) + \underbrace{hD_1 f(x,y) + kD_2 f(x,y)}_{\text{linear in } (h,k)} + \underbrace{o(h) + o(k)}_{o((h,k))}. \end{aligned}$$

Unfortunately, this proof does not work. The red-highlighted $o(k)$ is actually also dependent on h. Call it $\eta(h, k)$. We need $\frac{\eta(h,k)}{k} \to 0$ as $(h, k) \to (0, 0)$. But we only know for each h, $\frac{\eta(h,k)}{k} \to$ as $k \to 0$. This is weaker.

In fact, what we need is the Mean Value Theorem.

Proof. For simplicity, $n = 2, m = 1$. $a = (x, y)$. Take (h, k) small. Then, by MVT,

$$f(x+h, y+k) - f(x+h, y) = kD_2 f(x+h, y+\theta_{h,k} k)$$

for some $\theta_{h,k} \in (0, 1)$. By MVT,

$$f(x+h, y) - f(x, y) = hD_1 f(x + \varphi_h h, y)$$

for some $\varphi_h \in (0, 1)$. Hence

$$f(x+h, y+k) - f(x, y) = kD_2 f(x+h, y+\theta_{h,k} k) + hD_1 f(x+\varphi_h h, y).$$

As $(h, k) \to (0, 0)$, we have $(x + h, y + \theta_{h,k} k) \to (x, y)$ and $(x + \varphi_h h, y) \to (x, y)$ so by continuity of D_1, D_2 at (x, y), we have $D_2 f(x + h, y + \theta_{h,k} k) \to D_2 f(x, y)$ and $D_1 f(x + \varphi_h h, y) \to D_1 f(x, y)$.

Write $D_2 f(x+h, y+\theta_{h,k} k) = D_2 f(x, y) + \eta(h, k)$ and $D_1 f(x+\varphi_h h, y) = D_1 f(x, y) + \zeta(h, k)$, where $\eta(h, k), \zeta(h, k) \to 0$ as $(h, k) \to (0, 0)$.
Then $f(x+h, y+k) = f(x, y) + hD_1 f(x, y) + kD_2 f(x, y) + h\zeta(h, k) + k\eta(h, k)$. Now

$(h,k) \mapsto hD_1f(x,y) + kD_2f(x,y)$ is linear, and

$$\left|\frac{h\zeta(h,k) + k\eta(h,k)}{\sqrt{h^2+k^2}}\right| \le |\zeta(h,k)| + |\eta(h,k)| \to 0$$

as $(h,k) \to (0,0)$. So f is differentiable at $a = (x,y)$. □

Remark 39. 1. Same proof basically does $f : \mathbb{R}^n \to \mathbb{R}$ for general n with more involved notation. Then $f : \mathbb{R}^n \to \mathbb{R}^m$ by looking at each $f_i : \mathbb{R}^n \to \mathbb{R}$, $(1 \le i \le m)$.

2. If you try to prove anything similar without invoking the MVT at any point, it's likely that the proof is probably wrong.

§4.2 The Second Derivative

We'll start with a result on partial derivatives: '$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$'.

Theorem 4.4 (Symmetry of mixed partial derivatives)
Let $f : \mathbb{R}^n \to \mathbb{R}^m$, $a \in \mathbb{R}^n$ and $\varepsilon > 0$. Suppose D_iD_jf and D_jD_if exist on $B_\varepsilon(a)$ and are continuous at a. Then $D_iD_jf(a) = D_jD_if(a)$.

Proof. Wlog $m = 1, n = 2$[a]$, a = (x,y), i = 1, j = 2$.
Let

$$\Delta_h = f(x+h, y+h) - f(x, y+h) - f(x+h, y) + f(x,y) = g(y+h) - g(y)$$

where $g(t) = f(x+h,t) - f(x,t)$.
Let $0 < |h| < \sqrt{\varepsilon}$. Then

$$\begin{aligned}\Delta_h &= hg'(y+\theta_h h) \quad \text{some } \theta_h \in (0,1) \text{ by MVT} \\ &= h(D_2f(x+h, y+\theta_h h) - D_2f(x, y+\theta_h h)) \\ &= h^2 D_1D_2f(x+\varphi_h h, y+\theta_h h) \quad \text{some } \varphi_h \in (0,1) \text{ by MVT.}\end{aligned}$$

Similarly, $\Delta_h = h^2 D_2D_1f(x+\zeta_h h, y+\xi_h h)$[b] for some $\zeta_h, \xi_h \in (0,1)$.
Hence $D_1D_2f(x+\varphi_h h, y+\theta_h h) = D_2D_1f(x+\zeta_h h, y+\xi_h h)$. So let $h \to 0$ and use continuity of D_1D_2f, D_2D_1f at (x,y),

$$D_1D_2f(x,y) = D_2D_1f(x,y)$$

□

[a] We can consider f to be a vector of functions, f_i, so wlog $m = 1$. Also we can ignore the variables that we are not differentiating wrt by treating them as constants so wlog $n = 2$.

[b] Let $g(t) = f(t, y+h) - f(t, y)$.

Question

What is the second derivative really?

Definition 4.6 (Twice-Differentiable)

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be everywhere differentiable.
For each $x \in \mathbb{R}^n, Df_x \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Define $F : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{nm}$ by $F(x) = Df|_x$.

If F is differentiable at $a \in \mathbb{R}^n$ then we say that f is **twice-differentiable** at a and the **second derivative** of f at a is $D^2 f|_a = DF|_a$.

What is $D^2 f|_a$?

$$D^2 f|_a \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)) \cong Bi|(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m)$$

So $D^2 f|_a$ is a bilinear map from $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$.

If f is twice differentiable at a, this says

$$Df|_{a+h} = Df|_a + D^2 f|_a(h) + o(h)$$

where everything in the expression is a linear map, i.e.

$$Df|_{a+h}(k) = Df|_a(k) + \underbrace{D^2 f|_a(h,k)}_{\text{bilinear in } h,k} + \underbrace{o_k(h)}_{\text{for each fixed } k, \text{ this is } o(h)}$$

Example 4.8

Let $f : \mathcal{M}_n \to \mathcal{M}_n$, $f(A) = A^3$.
Then

$$f(A+K) = A^3 + \underbrace{A^2K + AKA + KA^2}_{\text{linear in } K} + \text{terms involving } K^2$$

so f is everywhere differentiable with $Df|_A(K) = A^2K + AKA + KA^2$.
Then

$$\begin{aligned} Df|_{A+H}(K) &= (A+H)^2K + (A+H)K(A+H) + K(A+H)^2 \\ &= \underbrace{A^2K + AKA + KA^2}_{Df|_A(K)} + \underbrace{AHK + HAK + AKH + HKA + KAH + KHA}_{\text{Bilinear}} \end{aligned}$$

$$+ \underbrace{H^2K + HKH + KH^2}_{o_K(H)}.$$

So f is twice differentiable at A and $D^2f|_A(H,K) = AHK + HAK + AKH + HKA + KAH + KHA$.

Remark 40. For definition to work, it's enough to have f differentiable on some neighbourhood of a.

Question

How does $D^2f|_a$ relate to $D_iD_jf(a)$?

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice-differentiable at $a \in \mathbb{R}^n$. Then, with $e_1, \ldots, e_n$ standard basis

$$\begin{aligned} \frac{D_jf(a+he_i) - D_jf(a)}{h} &= \frac{D^2f|_a(he_i, e_j) + o(h)}{h} \\ &= D^2f|_a(e_i, e_j) + o(1) \\ &\to D^2f|_a(e_i, e_j). \end{aligned}$$

So $D_iD_jf(a) = D^2f|_a(e_i, e_j)$. So if H is the $n \times n$ matrix respecting the bilinear form $D^2f|_a$, we have $H_{ij} = D_iD_jf(a)$.

Definition 4.7 (Hessian)

The **Hessian** matrix, H, of f is $H_{ij} = D_iD_jf(a)$.

If $f : \mathbb{R}^n \to \mathbb{R}^m$, we could do this for each $f_i : \mathbb{R}^n \to \mathbb{R}(i = 1, \ldots, m)$, or could think about matrices whose entries are elements of $\mathbb{R}^m$.

Definition 4.8 (Continuously Differentiable)

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $a \in \mathbb{R}^n$. We say f is **continuously differentiable** at a if $Df|_x$ exists for all x in some ball $B_\delta(a)$ $(\delta > 0)$ and the function $x \mapsto Df|_x$ is continuous at a.

Remark 41. If f is twice continuously differentiable at a then theorem 4.4 tells us that H is a symmetric matrix.
Hence, under this condition, $D^2f|_a$ is a symmetric bilinear form.

Let's use this to find stationary points.

Definition 4.9 (Local Maximum)

Let $f : \mathbb{R}^n \to \mathbb{R}$, $a \in \mathbb{R}^n$. We say a is a **local maximum** (resp. *minimum*) for f if there is some $\delta > 0$ such that for all $x \in B_\delta(a)$ we have $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$).

Proposition 4.6

Let $f : \mathbb{R}^n \to \mathbb{R}$ and let a be a local max/min for f. Suppose f is differentiable at a. Then $Df|_a$ is the zero map.

Proof. Let $u \in \mathbb{R}^n$. For $\lambda \neq 0$ in $\mathbb{R}$,

$$\frac{f(a+\lambda u) - f(a)}{\lambda} = \frac{Df|_a(\lambda u) + o(\lambda)}{\lambda} \to Df|_a(a)$$

as $\lambda \to 0$. Assume wlog a is a maximum (otherwise consider $-f$). Then $\frac{f(a+\lambda u)-f(a)}{\lambda}$ is ≥ 0 if $\lambda < 0$, and ≤ 0 if $\lambda > 0$. Hence $Df|_a(u) = 0$. □

Note the converse does not hold: e.g. $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$, $a = 0$.

Lemma 4.1 (Second-order Taylor Theorem)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice-differentiable at $a \in \mathbb{R}^n$. Then

$$f(a+h) = f(a) + Df|_a(h) + \frac{1}{2}D^2 f|_a(h,h) + o\Big(\|h\|^2\Big).$$

Proof. Define $g : [0,1] :\to \mathbb{R}$ by $g(t) = f(a+th) - f(a) - tDf|_a(h) - \frac{t^2}{2}Df|_a(h,h)$. Clearly g is continuous on $[0,1]$, $g(0) = 0$ and g is differentiable on $(0,1)$ with $g'(t) = Df|_{a+th}(h) - Df|_a(h) - tD^2 f|_a(h,h)$.

By MVT, $\exists\, t \in (0,1)$ such that $g(1) - g(0) = g'(t)$. Hence

$$\begin{aligned}
\frac{\left|f(a+h) - f(a) - Df|_a(h) - \frac{1}{2}D^2 f|_a(h,h)\right|}{\|h\|^2} &= \frac{\left|Df|_{a+th}(h) - Df|_a(h) - tD^2 f|_a(h,h)\right|}{\|h\|^2} \\
&= \frac{\left|D^2 f|_a(th,h) + o(\|h\|)^2 - tD^2 f|_a(h,h)\right|}{\|h\|^2} \\
&= \frac{|o(\|h\|)^2|}{\|h\|^2} \text{ by bilinearity} \\
&\to 0 \text{ as } h \to 0.
\end{aligned}$$

□

Theorem 4.5

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $a \in \mathbb{R}^n$. Suppose f is twice continuously differentiable at a (so, in particular, $D^2 f|_a$ is a symmetric bilinear form) and $Df|_a = 0$. Then

1. if $D^2 f|_a$ is positive definite, then a is a local minimum, and
2. if $D^2 f|_a$ is negative definite, then a is a local maximum,

Proof. Suppose wlog $D^2 f|_a$ is positive definite (otherwise consider $-f$). Then wrt some orthonormal basis $D^2 f|_a$ has diagonal matrix with strictly positive elements on leading diagonal.

Hence $\forall\ x \in \mathbb{R}^n$, $D^2 f|_a(x,x) \geq \mu \|x\|^2$ where $\mu > 0$ is the minimum eigenvalue of $D^2 f|_a$. By lemma 4.1,

$$\begin{aligned} \frac{f(a+h) - f(a)}{\|h\|^2} &= \frac{1}{2} \frac{D^2 f|_a(h,h)}{\|h\|^2} + o(1) \\ &\geq \frac{1}{2}\mu + o(1) \to \frac{1}{2}\mu \text{ as } h \to 0 \end{aligned}$$

but $\frac{1}{2}\mu > 0$ so for h sufficiently small, $\frac{f(a+h)-f(a)}{\|h\|^2} > 0$ and thus $f(a+h) - f(a) > 0$. Hence a is a local minimum for f. □

§4.3 Ordinary Differential Equations

Lemma 4.2

Let $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$ with A compact and B closed. Let $X = C(A,B) = \{f : A \to B : A \text{ continuous}\}$ with uniform metric $d(f,g) = \sup_{x \in A} \|f(x) - g(x)\|$. Then X is a complete metric space.

Proof. As A is compact, d is well-defined. Let (f_n) be a Cauchy sequence in X. Then (f_n) is uniformly Cauchy so uniformly convergent by GPUC on each coordinate. So $f_n \to f$ uniformly for some $f : A \to \mathbb{R}^m$. Uniform limit of continuous functions is continuous so f is continuous. And $\forall x \in A$, $f_n(x) \to f(x)$ so, as B is closed, $f(x) \in B$. Hence $f \in X$ and $d(f_n, f) \to 0$ as required. □

Often we want to solve an ODE but can't find a closed-form solution. Two ways we have tried so far: Numerical methods, and phase-plane portraits. However, these methods have no meaning if the ODE has no solution. So we would like a general result telling us that for ODE under appropriate conditions, they have unique solutions.

A typical ODE is: $\frac{dy}{dx} = \varphi(x,y)$ subject to $y = y_0$ when $x = x_0$. If we think about $\mathbb{R} \to \mathbb{R}^n$, we want to solve the initial value problem $f : \mathbb{R} \to \mathbb{R}^n$, and $f'(t) = \varphi(t, f(t))$, and initial condition $f(t_0) = y_0$.

Definition 4.10 (Closed Ball)
In $\mathbb{R}^n$, if $a \in \mathbb{R}^n$ and $\delta > 0$, the **closed ball of radius δ about a** is

$$\overline{B_\delta(a)} = \{x \in \mathbb{R}^n : \|x - a\| < \delta\}.$$

Theorem 4.6 (Lindelöf-Picard)
Let $a, b \in \mathbb{R}$, $(a < b)$, $y_0 \in \mathbb{R}^n$, $\delta > 0$ and $t_0 \in (a, b)$.
Let $\varphi : [a, b] \times \overline{B_\delta(y_0)} \to \mathbb{R}^n$ be continuous, and suppose there is some $K > 0$ such that

$$\forall\, t \in [a, b],\ \forall\, y, z \in \overline{B_\delta(y_0)},\ \text{we have } \|\varphi(t, y) - \varphi(t, z)\| \leq K\|y - z\|.$$

Then there is some $\varepsilon > 0$ such that $[t_0-\varepsilon, t_0+\varepsilon] \subset [a, b]$ and the initial value problem

$$f'(t) = \varphi(t, f(t)) \text{ with } f(t_0) = y_0 \tag{$\star$}$$

has a unique solution on $[t_0 - \varepsilon, t_0 + \varepsilon]$.

Proof. As φ is a continuous function on a compact set we can find M such that

$$\forall t \in [a, b], \forall y \in \overline{B_\delta(y_0)},\ \text{we have } \|\varphi(t, y)\| \leq M.$$

Take $\varepsilon > 0$ such that $[t_0-\varepsilon, t_0+\varepsilon] \subset [a, b]$. Let $X = C\Big([t_0 - \varepsilon, t_0 + \varepsilon], \overline{B_\delta(y_0)}\Big)$. Then by lemma 4.2, X is complete with uniform metric d. And obviously $X \neq \varphi$[a].

For $g \in X$, define $T_g : [t_0 - \varepsilon, t_0 + \varepsilon] \to \mathbb{R}^n$ by

$$Tg(t) = y_0 + \int_{t_0}^{t} \varphi(x, g(x))dx.$$

Note that by FTC, $T_f = f$ iff f is a solution of $(\star)$.

Now, if $g \in X$ and $t \in [t_0 - \varepsilon, t_0 + \varepsilon]$,

$$\begin{aligned}
\|T_g(t) - y_0\| &= \left\|\int_{t_0}^{t} \varphi(x, g(x))dx\right\| \\
&\leq \int_{t_0}^{t} \|\varphi(x, g(x))\|dx \\
&\leq M\varepsilon.
\end{aligned}$$

Also, if $g, h \in X$ and $t \in [t_0 - \varepsilon, t_0 + \varepsilon]$ then

$$\begin{aligned} \|T_g(t) - T_h(t)\| &= \left\| \int_{t_0}^{t} \varphi(x, g(x)) - \varphi(x, h(x)) dx \right\| \\ &\leq \int_{t_0}^{t} \|\varphi(x, g(x)) - \varphi(x, h(x))\| dx \\ &\leq \int_{t_0}^{t} K \|g(x) - h(x)\| dx \\ &\leq K\varepsilon d(g, h) \end{aligned}$$

i.e. $d(T_g, T_h) \leq K\varepsilon(g, h)$.

So taking $\varepsilon = \min\{\frac{\delta}{M}, \frac{1}{2K}\}$ we have that T is a contraction of X and so has a unique fixed point by the Contraction Mapping Theorem as desired. □

[a] X contains the constant function y_0

Remark 42. Not that much use as stated, as it doesn't provide a global solution, and in fact there might not be one. In practice, given appropriate conditions on φ we can often 'patch together' local solutions, but this is beyond the scope of this course.

§4.4 The Inverse Function Theorem

Theorem 4.7 (The Inverse Function Theorem)
Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable at $a \in \mathbb{R}^n$ with $\alpha = Df|_a$ being non-singular. Then there exist open neighbourhoods U of a and V of $f(a)$ such that $f|_U$ is a homeomorphism of U onto V.

Moreover, if $g : V \to U$ is the inverse of $f|_U$, it is differentiable at $f(a)$ with $Dg|_{f(a)} = \alpha^{-1}$, but this part will not be proven.

Proof. Write

$$f(a + h) = f(a) + \alpha(h) + \varepsilon(h)\|h\|$$

where $\varepsilon(h) \to 0$ as $h \to 0$. Let $\delta, \eta > 0$ such that f is differentiable on $\overline{B_\delta(a)}$. Define $W = \overline{B_\delta(a)}$, $V = B_\eta(f(a))$. Also define $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ by $\varphi(x) = f(x) - \alpha(x)$. Then, for $x \in W$, φ is differentiable at x with

$$D\varphi|_x = Df|_x - \alpha \to 0 \text{ as } x \to a.$$

Note W is a complete, non-empty metric space. We hope to define something $W \to W$ such that it is a contraction.

Fix $y \in V$. Define $T_y : W \to \mathbb{R}^n$ by

$$T_y(x) = x - \alpha^{-1}(f(x) - y).$$

Note $f(x) = y \Leftrightarrow T_y(x) = x$. Now, given $x \in W$,

$$\begin{aligned}
\|T_y x - a\| &= \left\|\alpha^{-1}(\alpha x - f(x) + y - \alpha(a))\right\| \\
&= \left\|\alpha^{-1}(y - (f(x) - \alpha(x - a)))\right\| \\
&= \left\|\alpha^{-1}(y - f(a) - \varepsilon(x - a)\|x - a\|)\right\| \\
&\leq \left\|\alpha^{-1}\right\|(\|y - f(a)\| + \|\varepsilon(x - a)\|\|x - a\|) \\
&\leq \left\|\alpha^{-1}\right\|(\eta + \delta\|\varepsilon(x - a)\|).
\end{aligned}$$

Also, given $w, x \in W$,

$$\begin{aligned}
\|T_y x - T_y w\| &= \left\|\alpha^{-1}(\alpha(x) - f(x) + f(w) - \alpha(w))\right\| \\
&= \left\|\alpha^{-1}(\varphi(w) - \varphi(x))\right\| \\
&\leq \left\|\alpha^{-1}\right\|\|\varphi(w) - \varphi(x)\| \\
&\leq \left\|\alpha^{-1}\right\|\|w - x\| \sup_{z \in w} \|D\varphi|_z\|
\end{aligned}$$

by Mean Value Inequality. Pick $\delta > 0$ sufficiently small such that $\forall x \in W$, we have $\|\varepsilon(x - a)\| \leq \frac{1}{2\|\alpha^{-1}\|}$ and also $\sup_{z \in W} \|D\varphi|_z\| < \frac{1}{\|\alpha^{-1}\|}$. (We can do this since $\varepsilon(x - a) \to 0$ as $x \to a$ and $D\varphi|_x \to 0$ as $x \to a$.)

Take $\eta = \frac{\delta}{2}$. Then for each $y \in V$, we have $\forall x \in W$, $\|T_y x - a\| < \delta$ and $\forall x, w \in W$, $\|T_y x - T_y w\| \leq K\|w - x\|$ for some constant $K < 1$. Thus T_y is a contraction of W and by CMT has a unique fixed point, and $x_y \in T_y(W) \subset B_\delta(a)$. That is, for each $y \in V$ there is a unique $x \in W$ with $f(x) = y$, and in fact $x \in B_\delta(a)$.

Let U be the set of all such x. Let $h = f|_{B_\delta(a)}$. Then $U = h^{-1}(V)$ so U is open in $B_\delta(a)$. But $B_\delta(a)$ is open in $\mathbb{R}^n$ so U is open in $\mathbb{R}^n$. So now we have open neighbourhoods U of a and V of $f(a)$ such that f maps U bijectively onto V. Remains to check that f is continuous so that we have a homeomorphism.

Let $X = C(V, W)$. As W is bounded, similarly to Lemma 4.14 we have that X is a complete non-empty metric space. We want to define a function $X \to X$ that is a contraction. Define $S : X \to X$ by

$$S_g(y) = g(y) - \alpha^{-1}(f(g(y)) - y) = T_y(g(y)).$$

Given $g, h \in X$ and $y \in V$,

$$\begin{aligned} \|S_g(y) - S_h(y)\| &= \|T_y(g(y)) - T_y(h(y))\| \\ &\leq K\|g(y) - h(y)\| \\ &\leq Kd(g, h) \end{aligned}$$

and so by taking sup over all y, $d(S_g, S_h) \leq Kd(g, h)$. Hence S is a contraction of X and has a unique fixed point g. By definition of S, for each $y \in V$ we have $g(y)$ is the unique $x \in W$ with $f(x) = y$. Hence $g = (f|_U)^{-1}$ is continuous. Since $g \in X$, $(f|_U)^{-1}$ is continuous and thus $f|_U$ is a homeomorphism from U onto V. $\square$