Analysis and Topology

June 4, 2023

Contents

Lectures

Start of lecture 1

Chapter I

Generalizing continuity and convergence

Contents

1. Three Examples of Convergence

1.1. Convergence in $\mathbb{R}$

Recall from IA:

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

$$\forall\ \varepsilon > 0\ \exists N\ \forall\ n \ge N\ |x_n - x| < \varepsilon$$

Useful fact: for all $a, b \in \mathbb{R}$,

$$|a + b| \le |a| + |b|$$

(triangle inequality)

Recall two key theorems:

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

Recall:

Definition. A sequence (x_n) in $\mathbb{R}$ is *Cauchy* if

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

Easy exercise: prove that convergent implies Cauchy.
General principle of convergence: Any Cauchy sequence in $\mathbb{R}$ converges. (outline proof: If (x_n) Cauchy then (x_n) bounded so by Bolzano-Weierstrass it has a convergent subsequence, say $x_{n_j} \to x$. But since Cauchy, $x_n \to x$.)

1.2. Convergence in $\mathbb{R}^2$

Remark. 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 $|\bullet|$ by $\|\bullet\|$.

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. 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

$$\forall\, \varepsilon > 0 \; \exists\, N \; \forall\, n \geq N \quad \|z_n - z\| < \varepsilon.$$

Equivalently, $z_n \to z$ if and only if $\|z_n - z\| \to 0$.

Example. 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.

$$\|(z_n + w_n) - (z + w)\| \leq \|z_n - z\| + \|w_n - w\| \to 0 + 0 = 0$$

(by results from IA) □

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

Proposition 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 $z_n \to z$ if and only if $x_n \to x$ and $y_n \to y$.

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

$\Leftarrow$ 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. A sequence (z_n) in $\mathbb{R}^2$ is bounded if $\exists\, M \in \mathbb{R}$ such that for all n, $\|z_n\| \leq M$.

Theorem 2 (Bolzano Weierstrass in $\mathbb{R}^2$)**.** A bounded sequence in $\mathbb{R}^2$ must have a convergent subsequence.

Proof. Let (z_n) be a bounded sequence in $\mathbb{R}^2$. Write $z_n = (x_n, y_n)$. Now for all n, $|x_n| \le \|z_n\|$ so (x_n) is a bounded sequence in $\mathbb{R}$. So by Bolzano Weierstrass, it has a convergent subsequence, say $x_{n_j} \to x \in \mathbb{R}$. Similarly, (y_{n_j}) is a bounded sequence in $\mathbb{R}$ so has a convergent subsequence $y_{n_{j_k}} \to y$. Now also $x_{n_{j_k}} \to x$. Hence $z_{n_{j_k}} \to z = (x, y)$. $\square$

Definition. A sequence (z_n) in $\mathbb{R}^2$ is *Cauchy* if

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

Easy exercise: Convergent implies Cauchy.

Theorem 3 (General Principle of Convergence 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, $|x_m - x_n| \le \|z_m - z_n\|$ so (x_n) is Cauchy sequence in $\mathbb{R}$, so converges by General Principle of Convergence. Similarly for (y_n). So by proposition 1, (z_n) converges. $\square$

Start of lecture 2

1.3. Convergence of Functions

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

(Mostly can think of $X = \mathbb{R}$ or some interval).

Obvious idea:

Definition (Convergence of functions)**.** 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 $n \to \infty$.

Advantages:

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

Disadvantages:

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

Examples

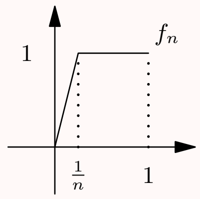
In all three examples, have $X = [0, 1]$, $f_n \to f$ pointwise.

Example (Limit of continuous functions not continuous)**.** Consider:

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

and

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



Clearly f_n continuous for all n but f not continuous.

Proving that $f_n \to f$:

- If $x = 0$, then $\forall\ n$, $f_n(0) = 0 = f(0)$.
- If $x > 0$, for sufficiently large n, $f_n(x) = 1 = f(x)$, so $f_n(x) \to f(x)$.

Note. As in IA, "integrable" means "Riemann integrable".

Consider

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

Enumerate the rationals in $[0, 1]$ as $q_1, q_2, \dots$. For $n \ge 1$, set

$$f_n(x) = \begin{cases} 1 & x = q_1, \dots, q_n \\ 0 & \text{otherwise} \end{cases}$$

Example (Functions and limit are integrable, but integral doesn't converge)**.** Let $f(x) = 0$ for all x, so $\int_0^1 f = 0$. Define f_n such that $\int_0^1 f_n = 1$ for all n:

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

Now $f_n \to f$ but clearly $\int_0^1 f_n \not\to \int_0^1 f$.

Now we try to make a "better" definition so that more of these properties might be able to hold.

Definition. 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

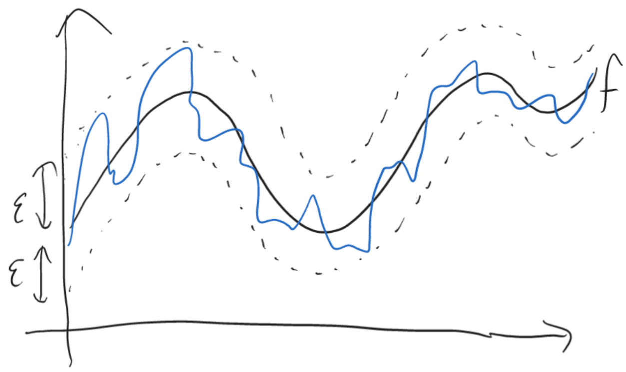
$$\forall \varepsilon > 0 \; \exists N \; \forall x \; \in X \; \forall n \geq N \quad |f_n(x) - f(x)| < \varepsilon$$

The definition for pointwise convergence can be restated as:

$$\forall \varepsilon > 0 \; \forall x \in X \; \exists N \; \forall n \geq N \quad |f_n(x) - f(x)| < \varepsilon$$

In particular $f_n \to f$ uniformly $\implies$ $f_n \to f$ pointwise.

Example (Limit of integrable not integrable)**.**



Equivalently, $f_n \to f$ uniformly if for sufficiently large n $f_n - f$ is bounded and

$$\sup_{x \in X} |f_n(x) - f(x)| \to 0$$

Theorem 4. 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 continuous.

"A uniform limit of continuous functions is continuous."

Proof. Let $x \in X$. Let $\varepsilon > 0$. As $f_n \to f$ uniformly, can find N such that $\forall n \geq N$, $\forall y \in X$, $|f_n(y) - f(y)| < \varepsilon$. In particular, $\forall y \in X$, $|f_N(y) - f(y)| < \varepsilon$. As f_N is continuous, can find $\delta > 0$ such that $\forall y \in X$,

$$|y - x| < \delta \implies |f_N(y) - f_N(x)| < \varepsilon$$

Now let $y \in X$ with $|y - x| < \delta$. Then

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

But 3ε can be made arbitrarily small by taking ε arbitrarily small. Hence f is continuous. $\square$

Remark. This is often called a "3ε proof" (or an "$\frac{\varepsilon}{3}$ proof" if written in a different style).

Theorem 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, can pick a sufficiently large n such that $f_n - f$ 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 such that $\forall n \geq N$, $\forall x \in [a, b]$ 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 $\mathcal{D} = \{x_0, x_1, \dots, x_k\}$, where $a = x_0 < x_1 < \cdots < x_k = b$. Now

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

That is,

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

Similarly

$$s(f, \mathcal{D}) \ge s(f_N, \mathcal{D}) - (b - a)\varepsilon$$

Hence

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

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

Now, for any n sufficiently large such that $f_n - f$ is bounded,

$$\begin{aligned}
\left| \int_a^b f_n - \int_a^b f \right| &= \left| \int_a^b (f_n - f) \right| \\
&\le \int_a^b |f_n - f| \\
&\le (b - a) \sup_{x \in [a,b]} |f_n(x) - f(x)| \\
&\to 0
\end{aligned}$$

as $n \to \infty$ since $f_n \to f$ uniformly. □

Start of lecture 3

What about differentiation?
Here, even uniform convergence is not enough.

Example. $f_n : [-1,1] \to \mathbb{R}$, each function differentiable, $f_n \to f$ uniformly but f not differentiable. We will let $f(x) = |x|$. Consider:

$$f_n(x) = \begin{cases} |x| & |x| \geq \frac{1}{n} \\ \frac{n}{2}x^2 + \left(\frac{1}{n} - \frac{1}{2n^2}\right) & |x| < \frac{1}{n} \end{cases}$$

By straightforward calculations, $f_n \to f$ uniformly, and all the f_n are differentiable.

In fact we need uniform convergence of the *derivatives*.

Theorem 6 (Limit of differentiable functions)**.** Let $f_n : (u,v) \to \mathbb{R}$ $(n \geq 1)$ and $f : (u,v) \to \mathbb{R}$ with $f_n \to f$ pointwise. Suppose further each function is continuously differentiable and that $f_n' \to g$ uniformly. Then f is differentiable with $f' = g$.

Proof. Fix $a \in (u,v)$. Let $x \in (u,v)$. By Fundamental theorem of calculus we have each f_n' is integrable over $[a,x]$ and $\int_a^x f_n' = f_n(x) - f_n(x)$. But $f_n' \to g$ uniformly so by Theorem 5, g is integrable over $[a,x]$ and $\int_a^x g = \lim_{n\to\infty} \int_a^x f_n'(x) = f(x) - f(a)$. So we have shown that for all $x \in (u,v)$,

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

By Theorem 4, g is continuous so by Fundamental theorem of calculus, f is differentiable with $f' = g$. $\square$

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

Definition. 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

$$\forall \varepsilon > 0 \; \exists N \; \forall m,n \geq N \; \forall x \in X \quad |f_m(x) - f_n(x)| < \varepsilon$$

Exercise: uniformly convegent $\implies$ uniformly Cauchy.

Theorem 7 (General Principle of Uniform Convergence)**.** 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

$$\exists N \; \forall m,n \geq N \; \forall y \in X \quad |f_m(y) - f_n(y)| < \varepsilon.$$

In particular, $\forall m, n \geq N$, $|f_m(x) - f_n(x)| < \varepsilon$. So $(f_n(x))_{n \geq 1}$ is a Cauchy sequence in $\mathbb{R}$ so by general principle of convergence it converges, say $f_n(x) \to f(x)$ as $n \to \infty$.
We have now constructed $f : X \to \mathbb{R}$ such that $f_n \to f$ pointwise.
Let $\varepsilon > 0$. Then we can find an N such that

$$\forall m, n \geq N \ \forall y \in X \quad |f_m(y) - f_n(y)| < \varepsilon.$$

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 $\forall m \geq N$, $|f_m(y) - f(y)| \leq \varepsilon$. But y was arbitrary so

$$\forall x \in X \ \forall m \geq N \quad |f_m(x) - f(x)| \leq \varepsilon.$$

So $f_n \to f$ uniformly. $\square$

Definition. 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 $\forall x \ \exists M \ \forall n$, $|f_n(x)| \leq M$. We say (f_n) is *uniformly bounded* if $\exists M \ \forall x \ \forall n$, $|f_n(x)| \leq M$.

What would uniform Bolzano Weierstrass say? "If (f_n) is a uniformly bounded sequence of functions then it has a uniformly convergent subsequence."
But this is *not* true.

Example. $f_n : \mathbb{R} \to \mathbb{R}$ defined by

$$f_n(x) = \begin{cases} 1 & x = n \\ 0 & x \neq n \end{cases}$$

Obviously 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 convergent.

Application to power series

Recall that if $\sum_n a_n x^n$ is a real power series with radius of convergence $R > 0$ then can differentiate / integrate it term-by-term within $(-R, R)$.

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

Can apply theorem 4 to 6 to get for example if conditions hold with f_n continuously differentiable and uniform convergence then $\sum f_n$ has derivative $\sum f_n'$.

We hope to prove that $\sum a_n x^n$ converges uniformly on $(-R, R)$ then hit it with earlier theorems, but this is not quite true.

Example. $\sum_{n=0}^{\infty} x^n$ with radius of convergence 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, $\forall x \in (-1,1)$, $|F_n(x)| \le n+1$. Fix any n. We can find a point $x \in (-1,1)$ where $f(x) \ge n+2$ and so $|f(x) - F_n(X)| \ge 1$. So clearly we can't have that $F_n \to f$ uniformly.

Back up plan: it does work if we look at 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 nice at x. 'Local uniform convergence of power series.'

Start of lecture 4

Lemma 8. Let $\sum a_n x^n$ be a real power series with radius of convergence $R > 0$. Let $0 < r < R$. Then $\sum a_n x^n$ converges uniformly on $(-r,r)$.

Proof. Define $f, f_m : (-r,r) \to \mathbb{R}$ 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 $x \in (-r,r)$. Then

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

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

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

as $m \to \infty$ by absolute convergence at r. $\square$

Theorem 9. Let $\sum a_n x^n$ be a real power series with radius of convergence $R > 0$. Define $f : (-R,R)$ by $f(x) = \sum_{n=0}^{\infty} a_n x^n$. Then

(i) f is continuous;

(ii) 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 such that $x < r < R$. By Lemma 8, $\sum a_n y^n$ 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 continuous functions on $(-r, r)$. Hence by Theorem 4, $f_{(-r,r)}$ is continuous. Hence f is continuous at x. Thus f is a continuous function on $(-R, R)$. More over, $[0, x] \subset (-r, r)$ so also have $\sum a_n y^n$ converges uniformly on $[0, x]$. Each partial sum function on $[0, x]$ is a polynomial so can be integrated with

$$\int_0^x \sum_{n=0}^m a_n y^n \mathrm{d}y = \sum_{n=0}^m \int_0^x a_n y^n \mathrm{d}y = \sum_{n=0}^m \frac{a_n}{n+1} x^{n+1}$$

Hence by Theorem 5, f is integrable over $[0, x]$ with

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

□

For differentiation, need technical lemma:

Lemma 10. Let $\sum a_n x^n$ be a real power series with radius of convergence $R > 0$. Then the power series $\sum_{n \geq 1} n a_n x^{n-1}$ has radius of convergence at least R.

Proof. Let $x \in \mathbb{R}$, $0 < x < \mathbb{R}$. Pick w with $x < w < R$. Then $\sum a_n w^n$ is absolutely convergent, so $a_n w^n \to 0$ so $\exists M$ such that $\forall n$, $|a_n w^n| \leq M$. For each n,

$$|n a_n x^{n-1}| = |a_n w^n| \left|\frac{x}{w}\right|^n \frac{1}{|x|} n$$

Fix n. Let $\alpha = \left|\frac{x}{w}\right| < 1$. Let $c = \frac{M}{|x|}$ be a constant. Then $|n a_n x^{n-1}| \leq c n \alpha^n$. By comparison test, sufficient to show $\sum n a^n$ converges. Note

$$\left|\frac{(n+1)\alpha^{n+1}}{n \alpha^n}\right| = \left(1 + \frac{1}{n}\right) \alpha \to \alpha < 1$$

as $n \to \infty$ so done by ratio test. □

Theorem 11. Let $\sum a_n x^n$ be a real power series with radius of convergence $R > 0$. Let $f : (-R, R) \to \mathbb{R}$ be defined by $f(x) = \sum_{n=0}^\infty a_n x^n$. Then f is differentiable and $\forall x \in (-R, R)$, $f'(x) = \sum_{n=1}^\infty n a_n x^{n-1}$.

Proof. Let $x \in (-R, R)$. Pick r with $|x| < r < R$. Then $\sum a_n y^n$ converges uniformly on $(-r, r)$. Moreover, the power series $\sum_{n\geq 1} n a_n y^{n-1}$ has radius of convergence at least R and so also converges uniformly on $(-r, r)$. The partial sum functions $f_m(y) = \sum_{n=0}^m a_n y^n$ are polynomials so differentiable with $f_m'(y) = \sum_{n=1}^m n a_n y^{n-1}$.
We now have f_m' converging uniformly on $(-r, r)$ to the function $g(y) = \sum_{n=1}^\infty n a_n y^{n-1}$. Hence by Theorem 6, $f_{(-r,r)}$ 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 described. □

1.4. Uniform Continuity

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

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

Definition (Uniform continuity)**.** We say that f is *uniformly continuous* if

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

Remark. Clearly if f is uniformly continuous then f is continuous.

We would suspect that f is continuous doesn't imply that f is uniformly continuous.

Example. A function $f : \mathbb{R} \to \mathbb{R}$ that is continuous but not uniformly continuous. Consider $f(x) = x^2$. We know f is continuous (as it's a polynomial). Suppose $\delta > 0$. Then

$$f(x + \delta) - f(x) = (x + \delta)^2 - x^2 = 2\delta x + \delta^2 \to \infty$$

as $x \to \infty$. So in particular, $\forall \delta > 0$, $\exists x, y \in \mathbb{R}$ such that $|x-y| < \delta$ but $f(x)-f(y)| \geq 1$. So condition for uniformly continuous fails for $\varepsilon = 1$. So f is not uniformly continuous.

Example. Make domain bounded, and we can still fail. Consider $f : (0, 1) \to \mathbb{R}$, $f(x) = \frac{1}{x}$. Clearly continuous. Check not uniform continuity as an exercise.

Start of lecture 5

Theorem 12. 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 continuous but not uniformly continuous. Then we can find an $\varepsilon > 0$ such that for all $\delta > 0$ there exists $x, y \in [a,b]$ with $|x-y| < \delta$ but $|f(x) - f(y)| \geq \varepsilon$. In particular taking $\delta = \frac{1}{n}$ for $n = 1, 2, 3, \ldots$, 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 Bolzano Weierstrass it has a convergent subsequence $x_{n_j} \to x$ say. And $[a,b]$ is a closed interval so $x \in [a,b]$. Then $x_{n_j} - y_{n_j} \to 0$ so also $y_{n_j} \to x$. But f is continuous at x, so there exists $\delta > 0$ such that for all $y \in [a,b]$, $|y - x| < \delta$ implies that $|f(y) - f(x)| < \frac{\varepsilon}{2}$. Take such a δ. As $x_{n_j} \to x$ we can find J_1 such that $j \geq J_1$ implies that $|x_{n_j} - x| < \delta$. Similarly can find J_2 such that $j \geq J_2$ implies $|y_{n_j} - x| < \delta$. Now let $j = \max\{J_1, J_2\}$. Then $|x_{n_J} - x| < \delta$ and $|y_{n_j} - x| < \delta$ so we have $|f(x_{n_J}) - f(x)| < \frac{\varepsilon}{2}$ and $|f(y_{n_j}) - f(x)| < \frac{\varepsilon}{2}$. Then

$$|f(x_{n_j}) - f(y_{n_j})| \leq |f(x_{n_j}) - f(x)| + |f(x) - f(y_{n_j})| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

contradiction. □

Corollary 13. A continuous real-valued function on a closed bounded interval is bounded.

Proof. Let $f : [a,b] \to \mathbb{R}$ be continuous, and so uniformly continuous by Theorem 12. Then can find $\delta > 0$ such that

$$\forall x, y \in [a,b] \quad |x - y| < \delta \implies |f(x) - f(y)| < 1$$

Let $M = \lceil \frac{b-a}{\delta} \rceil$. Now 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

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

□

Corollary 14. A continuous real-valued function on a closed bounded interval is integrable.

Proof. Let $f : [a, b] \to \mathbb{R}$ be continuous and so uniformly continuous by Theorem 12. Let $\varepsilon > 0$. Then can find $\delta > 0$ such that for all $x, y \in [a, b]$, $|x-y| < \delta \implies |f(x)-f(y)| < \varepsilon$. Let $\mathcal{D} = \{x_0 < x_1 < \cdots < x_n\}$ be a dissection such that 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:

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

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

2. Metric Spaces

2.1. Definitions and Examples

Can we think about convergence in a more general setting? What do we really need? - A notion of distance.

In $\mathbb{R}$: distance x to y is $|x - y|$.
In $\mathbb{R}^2$: distance x to y is $\|x - y\|$.
For functions distance x to y is

$$\sup \sup_{x \in X} |f(x) - g(x)|$$

(where this exists, i.e. if $f - g$ is bounded).

Triangle inequality was often important.

Definition (Metric space)**.** A *metric space* is a set X endowed with a *metric* d, i.e. a function $d : X^2 \to \mathbb{R}$ satisfying:

(i) $d(x, y) \geq 0$ for all $x, y \in X$ with equality if and only if $x = y$;

(ii) $d(x, y) = d(y, x)$ for all $x, y \in X$;

(iii) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

Could define a metric space as an ordered pair (X, d). If it is obvious what d is, sometimes write "The metric space $X \ldots$"

Examples

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

(2) $X = \mathbb{R}^n$ with the *Euclidean metric*

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

(3) Let $Y \subset \mathbb{R}$. Take

$$X = B(Y) = \{f : Y \to \mathbb{R} | f \text{ is bounded}\}$$

now we can use the *uniform metric*

$$d(f, g) = \sup_{x \in Y} |f(x) - g(x)|$$

(we need the bounded condition for this supremum to necessarily exist). Check triangle inequality: let $f, g, h \in B(Y)$. Let $x \in Y$. Then

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

Take sup over all $x \in Y$ we get

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

Remark. Suppose (X, d) is 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.

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

(5) Recall that a continuous function on a closed bounded interval is bounded. Define $C([a, b]) = \{f : [a, b] \to \mathbb{R} | f \text{ is continuous}\}$. This is a subspace of $B([a, b])$.

(6) The empty metric space $X =$ with the empty metric.

(7) Can define different metrics on the same set, for example the l_1 metric on $\mathbb{R}^n$:

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

(8) The l_∞ metric on $\mathbb{R}^n$:

$$d(x, y) = \max_i |x_i - y_i|$$

(proof of triangle inequality is the same as for uniform metric in example 3).

(9) On $C([a, b])$ we can define the L_1 metric

$$d(f, g) = \int_a^b |f - g|$$

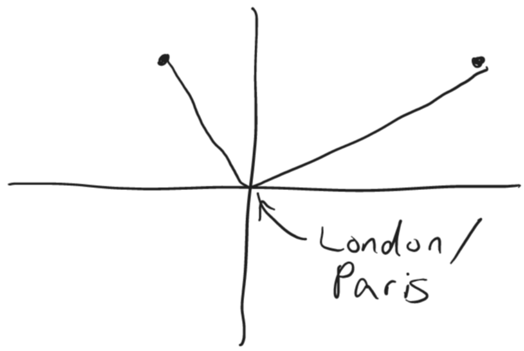
(10) $X = \mathbb{C}$ with

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

triangle inequality? Need $d(u, w) \leq d(u, v) + d(v, w)$

- if $u = w$, $LHS = 0$
- If $u = v$ or $v = w$ then $LHS = RHS$

- If u, v, w all distinct:
$$|u| + |w| < |u| + |w| + 2|w|$$
"British rail metric" or "SNCF metric":



Start of lecture 6

(11) 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 *discrete metric* on X.

(12) Let $X = \mathbb{Z}$. 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 the difference is divisible by a large power of p". Triangle inequality:

- Easy if any two of x, y, z are the same, so assume x, y, z are all distinct.
- Let $x - y = p^a m$ and $y - z = p^b n$ where $p \nmid m$, $p \nmid n$ and without loss of generality $a \le b$. So $d(x, y) = p^{-a}$ and $d(y, z) = p^{-b}$. Now:
$$\begin{aligned} x - z &= (x - y) + (y - z) \\ &= p^a m + p^b n \end{aligned}$$
$$p^a(m + p^{b-a} n)$$
so $p^a \mid x - z$ so $d(x, z) \le p^{-a}$. But $d(x, y) + d(y, z) \ge d(x, y) = p^{-a}$, so triangle inequality does hold.

Definition. 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

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

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

Proposition 15. 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,

$$\begin{aligned} d(x, y) &\leq d(x, x_n) + d(x_n, y) \\ &\leq d(x_n, x) + d(x_n, y) \\ &\to 0 + 0 = 0 \end{aligned}$$

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

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

Remarks on the definition

(1) Constant sequences obviously converge. Moreover, 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 the converse is false. For example, 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 $\left(\frac{1}{n}\right)_{n \geq 1}$ is a sequence in $\mathbb{R} \setminus \{0\}$ but it doesn't converge in $\mathbb{R} \setminus \{0\}$. This is because by uniqueness of limits, it would have to converge to 0, but $0 \notin \mathbb{R} \setminus \{0\}$.

Examples

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

What about other metrics on $\mathbb{R}^n$? For example let d_∞ be the uniform metric

$$d_\infty(x, y) = \max_i |x_i - y_i|$$

which sequences converge in $(\mathbb{R}^n, d_\infty)$? Note that

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \leq \sqrt{\sum_{i=1}^n d_\infty(x,y)^2}$$

so $d(x,y) \leq \sqrt{n} d_\infty(x,y)$. But also $d_\infty(x,y) \leq d(x,y)$. Now suppose (x_n) is a sequence in $\mathbb{R}^n$. Then

$$d(x_n, x) \to 0 \iff d_\infty(x_n, x) \to 0$$

So exactly the same sequences converge in $(\mathbb{R}^n, d)$ and $(\mathbb{R}^n, d_\infty)$. What about the l_1 metric d_1?

$$d_1(x,y) = \sum_{i=1}^n |x_i - y_i|$$

Similarly $d_\infty(x,y) \leq d_1(x,y) \leq n d_\infty(x,y)$. So again, exactly the same sequences converge in $(\mathbb{R}^n, d)$.

(2) Let $X = C([0,1]) = \{f : [0,1] \to \mathbb{R} \mid f \text{ is continuous}\}$. Let d_∞ be the uniform metric on X:

$$d_\infty(f,g) = \sup_{x \in [0,1]} |f(x) - g(x)|$$

Now note that

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

We also had the L_1 metric d_1 on X:

$$d_1(f,g) = \int_0^1 |f - g|$$

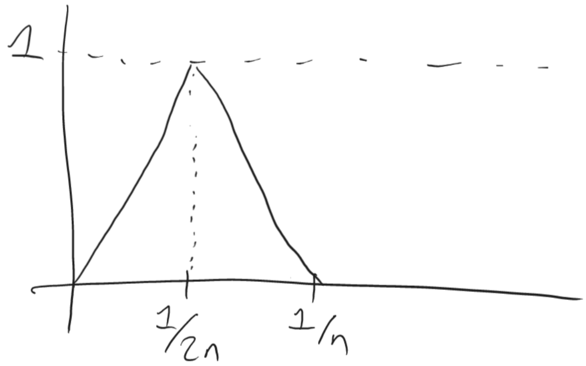
Now

$$\begin{aligned} d_1(f,g) &= \int_0^1 |f-g| \\ &\leq \int_0^1 d_\infty(f,g) \\ &= d_\infty(f,g) \end{aligned}$$

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 such that $f_n \to 0$ in d_1 metric by f_n doesn't converge in the d_∞ metric. So we want (f_n) such that $\int_0^1 |f_n| \to 0$ as $n \to \infty$, but (f_n) does not converge uniformly. We can just take functions like this:



Then clearly $f_n \to f$ in the d_1 metric but not in the d_∞ metric.

(3) Let (X, d) be a discrete metric space;

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

When do we have $x_n \to x$ in (X, d)? Suppose $x_n \to x$, i.e.

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

Setting $\varepsilon = 1$ in this, can find N such that

$$\forall n \geq N \; d(x_n, x) < 1$$

i.e. $\forall n \geq N$, $d(x_n, x) = 0$, i.e. $\forall n \geq N$, $x_n = x$. Thus (x_n) is eventually constant. But we know that in *any* metric space, eventually constant sequences converge. So in this space, (x_n) converges if and only if (x_n) eventually constant.

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

(i) Let $a \in X$ and $b \in Y$. We say $f(x) \to b$ as $x \to a$ if

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in X$$

$$0 < d(x, a) < \delta \implies e(f(x), b) < \varepsilon$$

(ii) Let $a \in X$. We say f is *continuous* at a if $f(x) \to f(a)$ as $x \to a$. That is

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in X$$

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

(iii) If $\forall x \in X$, f is continuous at a, we say f *is a continuous function* or simply f *is continuous*.

(iv) We say f *is uniformly continuous* if

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x, y \in X$$

$$d(x, y) < \delta \implies e(f(x), f(y)) < \varepsilon$$

(v) Suppose $W \subset X$. We say f is *continuous on* W (similarly for *uniformly continuous on* W) if the function $f|_W$ is continuous, as a function $W \to Y$ where now thinking of W as a subspace of X.

Start of lecture 7

Remarks

(1) Don't have a nice rephrasing of (i) in terms of similar concepts in the reals. Would want to write "$e(f(x), b) \to 0$ as $d(x, a) \to 0$", but this is meaningless.

(2) (i) says nothing about what happens at the point a itself. For example, 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 that f is continuous then

$$d(x, a) = 0 \implies x = a \implies f(x) = f(a) = e(f(x), f(a)) = 0$$

so we can drop '$0 <$' from the definition when we come to define continuity.

(3) Can rewrite definition (v): f is continuous on W if and only if $f|_W$ is a continuous function $f|_W \colon W \to Y$ thinking of W as a subspace of X. That is

$$\forall a \in W \; \forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in W \quad d(x, a) < \delta \implies e(f(x), f(a)) < \varepsilon$$

In particular, note the subtlety that this *only* mentions points of W. So, under this definition, for example $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$$

then f is continuous on $[0, 1]$, but f is not continuous at points 0 and 1.

Proposition 16. Let (X, d), (Y, e) be metric spaces. Let $f : X \to Y$ and $a \in X$. Then f is continuous at a if and only if whenever (x_n) is a sequence in X with $x_n \to a$ then $f(x_n) \to f(a)$.

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

$\Leftarrow$ Suppose f is not continuous at a. Then there is some $\varepsilon > 0$ such that $\forall \delta.0$, $\exists x \in X$ with $d(x, a) < \delta$ but $e(f(x), f(a)) \geq \varepsilon$. Now take $\delta = \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \dots$ we obtain a sequence (x_n) with, for each n

$$d(x_n, a) < \frac{1}{n} \quad \text{and} \quad e(f(x_n), f(a)) \geq \varepsilon.$$

Hence $x_n \to a$ but $f(x_n) \not\to f(a)$.

$\square$

Proposition 17. Let $(W, c), (X, d), (Y, e)$ be metric spaces, let $f : W \to X$, let $g : X \to Y$ and let $a \in W$. Suppose f is continuous at a and g is continuous at $f(a)$. Then $g \circ f$ is continuous at a.

Proof. Let (x_n) be a sequence in W with $x_n \to a$. Then by proposition 6, $f(x_n) \to f(a)$ an so also $g(f(x_n)) \to g(f(a))$. So by proposition 6 $g \circ f$ continuous at a. $\square$

Examples

(1) $\mathbb{R} \to \mathbb{R}$ with usual metric. This is the same definition as when did it directly for $\mathbb{R}$ only. So already know lots of continuous functions $\mathbb{R} \to \mathbb{R}$, for example polynomials, sin, exp, ...

(2) 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 *identity function on* X) then f is continuous.

(3) Consider $\mathbb{R}^n$ with the Euclidean 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 then this is the same as convergence in each coordinate. Let's denote a sequence in $\mathbb{R}^n$ by $(x^{(m)})_{m\geq 1}$. So for example, $x_5^{(3)}$ is the 5th coordinate of the 3rd term. We know $x^{(m)} \to x$ if and only if for each $x_i^{(m)} \to x_i$ i.e. for each i, $\pi_i(x^{(m)}) \to \pi_i(x)$. So by proposition 16 each π_i is continuous.) Similarly, suppose $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$. Let $f : \mathbb{R} \to \mathbb{R}^n$ be defined by $f(x) = (f_1(x), \ldots, f_n(x))$. Then f is continuous at a point if and only if all of $f_1, \ldots, f_n$ are.

Using these facts, and using example 1 and proposition 17 we have many continuous functions $\mathbb{R}^n \to \mathbb{R}^m$. For example consider $f : \mathbb{R}^3 \to \mathbb{R}^2$,

$$f(x, y, z) = (e^{-x} \sin y, 2x \cos z)$$

is continuous. (Why? Take $W = (x, y, z) \in \mathbb{R}^3$, 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 continuous so f continuous.)

(4) Recall that if we have the Euclidean metric, the l_1 metric or the l_∞ metric on $\mathbb{R}^n$ then convergent sequences are same in each case. So by proposition 16, the continuous functions $X \to \mathbb{R}^n$ or from $\mathbb{R}^n \to Y$ are the same with each of these three metrics.

(5) Let (X, d) be a discrete metric space and let (Y, e) be any metric space. Which functions $f : X \to Y$ are continuous? 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 continuous.

2.2. Completeness

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

Definition (Cauchy sequence)**.** Let (X, d) be a metric space and let (x_n) be a sequence in X. We say (x_n) is *Cauchy* if

$$\forall \varepsilon > 0\ \exists N\ \forall m, n \geq N \quad d(x_m, x_n) < \varepsilon$$

Exercise: (x_n) convergent implies that (x_n) Cauchy.
But the converse is not true in general.

Example. Let $X = \mathbb{R} \setminus \{0\}$ with the usual metric and let $x_n = \frac{1}{n}$. We saw 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. Take $\mathbb{Q}$ with the usual metric, and take a sequence $x_n \to \sqrt{2}$. Then x_n is Cauchy but not convergent (in $\mathbb{Q}$).

This example with $\mathbb{Q}$ is the main motivation for the following definition.

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

Examples

(1) Example above says $\mathbb{R} \setminus \{0\}$ with usual metric is not complete. Similarly $\mathbb{Q}$ with usual metric is not complete.

(2) General principle of convergence says $\mathbb{R}$ with usual metric is complete.

General principle of convergence for $\mathbb{R}^n$ says $\mathbb{R}^n$ with Euclidean metric is complete.

(3) General principle of uniform convergence (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.

Start of lecture 8

Proof. Let (f_n) be a Cauchy sequence in $B(X)$. Then (f_n) is uniformly Cauchy so by general principle of uniform convergence 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 sufficiently 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 so $d(f_n, f) \to 0$ i.e. $f_n \to f$ in $(B(X), d)$. □

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

(i) Take (x_n) Cauchy in X;

(ii) Constant / find a putative limit object x where it seems (x_n) converges to x in some sense;

(iii) Show $x \in X$,

(iv) 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 need to take care as, for example, it's tempting to talk about $d(x_n, x)$ while doing (ii) or (iii); but this makes no sense to write '$d(x_n, x)$' until we've completed (iii) as d only defined on X^2.

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

Proposition 18. A closed subset of a complete metric space is complete.

Remark. This *does* make sense: if $Y \subset X$ then Y is itself a metric space as a subspace of X so can say for example 'Y is complete' to mean the metric space Y (as a subspace of X) is complete. Could do exactly the same with any further 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.

□

(5) Define

$$l_1 = \{(x_n)_{n\geq 1} \in \mathbb{R}^{\mathbb{N}} \mid \sum_{n=1}^{\infty} |x_n| \text{ converges}\}$$

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 and for each n, $|x_n - y_n| \le |x_n| + |y_n|$ 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 $(x^{(n)})_n \ge 1$ be a Cauchy sequence in l_1, so for each n, $(x_i^{(n)})_{i\ge 1}$ is a sequence in $\mathbb{R}$ with $\sum_{i=1}^\infty |x_i^{(n)}|$ convergent.

(ii) For each i, $(x_i^{(n)})_{n\ge 1}$ is a Cauchy sequence in $\mathbb{R}$, since if $y, z \in l_1$, then $|y_i - z_i| \le d(y, z)$. But $\mathbb{R}$ is complete, so for each i we can find $x_i \in \mathbb{R}$ with $x_i^{(n)} \to x_i$ as $n \to \infty$. Let $x = (x_1, x_2, x_3, \dots) \in \mathbb{R}^{\mathbb{N}}$.

(iii) We next show $x \in l_1$, i.e. that $\sum_{i=1}^\infty |x_i|$ converges. Given $y \in l_1$, 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,

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

So

$$\sigma(x^{(m)}) - \sigma(x^{(n)}) \le d(x^{(m)}, x^{(n)}).$$

But we can find a similar inequality by swapping m and n, so

$$|\sigma(x^{(m)}) - \sigma(x^{(n)})| \le d(x^{(m)}, x^{(n)})$$

Hence $(\sigma(x^{(m)}))_{m\ge 1}$ is a Cauchy sequence in $\mathbb{R}$, and so by general principle of convergence it converges, say $\sigma(x^{(m)}) \to K$ as $m \to \infty$.

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

Proof. As $\sigma(x^{(n)}) \to K$ as $n \to \infty$ we can find N_1 such that for all $n \ge N_1$,

$$\sum_{i=1}^\infty |x_i^{(n)}| \le K + 1$$

This also implies that for all $n \ge N_1$,

$$\sum_{i=1}^I |x_i^{(n)}| \le K + 1$$

Next, for each $i \in \{1, 2, \dots, I\}$ we have $x_i^{(n)} \to x_i$ as $n \to \infty$ so can find N_2 such that

$$n \ge N_2 \implies \forall i \in \{1, \dots, I\} |x_i^{(n)} - x_i| < \frac{1}{I}$$

Now let $n = \max\{N_1, N_2\}$. Then

$$\begin{aligned}\sum_{i=1}^{I} |x_i| &\le \sum_{i=1}^{I} + \sum_{i=1}^{I} |x_i^{(n)} - x_i| \\ &\le K + 1 + I \times \frac{1}{I} \\ &= K + 2 \end{aligned}$$ $\square$

Now the partial sums of $\sum |x_i|$ are increasing and bounded above so $\sum |x_i|$ converges. That is, $x \in l_1$.

(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 $x \in l_1$) so can pick I_1 such that $\sum_{i=I_1+1}^{\infty} |x_i| < \varepsilon$. As $(x^{(n)})$ is Cauchy, we can find N_1 such that

$$m, n \ge N_1 \implies d(x^{(m)}, x^{(n)}) < \varepsilon$$

As $\sum_i |x_i^{(N_1)}|$ converges, can find I_2 such that $\sum_{i=I_2+1}^{\infty} |x_i^{(N_1)}| < \varepsilon$. Then

$$\begin{aligned} n \ge N_1 \implies & \sum_{i=I_2+1}^{\infty} \\ &\le \sum_{i=I_1+1}^{\infty} |x_i^{(N_1)}| + \sum_{i=I_2+1}^{\infty} |x_i^{(n)} - x_i^{(N_1)}| \\ &< \varepsilon + d(x^{(n)}, x^{(N_1)}) \\ &< 2\varepsilon \end{aligned}$$

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 such that

$$n \ge N_2 \implies \sum_{i=1}^{I} |x_i^{(n)} - x_i| < \varepsilon$$

Let $N = \max\{N_1, N_2\}$ and let $n \geq N$. Then

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

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.*

□

Start of lecture 9

Now we can move on to the main theorem on completeness:

Definition. Let (X, d) be a metric space and $f : X \to X$. We say f is a *contraction* if $\exists \lambda \in [0, 1)$ such that for all $x, y \in X$,

$$d(f(x), f(y)) \leq \lambda d(x, y)$$

Theorem 19 (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

$$\forall x, y \in X \quad d(f(x), f(y)) \leq \lambda d(x, y)$$

Let $x_0 \in X$. Recursively define $x_n = f(x_{n-1})$ for $n \geq 1$. Let $\Delta = d(x_), x_1)$. Then, by induction, $d(x_n, x_{n+1}) \leq \lambda^n \Delta$ for all n. Now suppose $N \leq m < n$. Then

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

as $N \to \infty$. So for all $\varepsilon > 0$, there exists N such that for all $m, n \geq N$, $d(x_m, x_n) < \varepsilon$. (we take N such that $\frac{\lambda^N \Delta}{1-\lambda} < \varepsilon$). Thus (x_n) is Cauchy, so by completeness it converges, say $x_n \to x \in X$. But also $x_n = f(x_{n-1}) \to f(x)$ because f is continuous. So by uniqueness of limits, $f(x) = x$.

Suppose also $f(y) = y$ for some $y \in X$. Then

$$d(x, y) = d(f(x), f(y)) \leq \lambda d(x, y)$$

with $\lambda < 1$. So $d(x, y) = 0$, i.e. $x = y$. □

Remark. (1) Why is f continuous? We have, for all $x, y \in X$, $d(f(x), f(y)) \leq d(x, y)$. So for all $\varepsilon > 0$, $d(x, y) < \varepsilon \implies d(f(x), f(y)) < \varepsilon$. (so we can take $\delta = \varepsilon$ in the definition of continuity). In particular this shows that f is uniformly continuous.

(2) We have proved more than claimed. Not only does f have a unique fixed point, but start 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.

Application

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

Think of $\cos : [-1, 1] \to [-1, 1]$. Suppose $x, y \in [-1, 1]$. Then using MVT, there is $z \in [x, y]$

$$\begin{aligned} |\cos x - \cos y| &= |x - y| |\cos' z| \\ &= |x - y| |-\sin z| \\ &\leq |x - y| \sin 1 \end{aligned}$$

But $0 \leq \sin 1 < 1$ so cos is a contraction of $[-1, 1]$. So by 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. Calculate cos iterated many times to 0 say, and we have rapid (exponential) convergence to the root.

Two major applications of contraction mapping theorem later.

2.3. Sequential compactness

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

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

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

Remark. Easy to check by triangle inequality that X bounded is equivalent to $X = \emptyset$ or $\exists M \in \mathbb{R}, \exists x \in X$ such that $\forall y \in X, d(x, y) \le M$.

So definition agrees with earlier definition for subsets of $\mathbb{R}^n$.

Recall: 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 actually $x \in Y$.

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

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

Theorem 20. Let $X \subset \mathbb{R}^n$ with the Euclidean metric. Then X is sequentially compact if and only if X is closed and bounded.

Proof. $\Leftarrow$ Suppose X is closed and bounded. Let (x_n) be a sequence in X. Then (x_n) be a sequence in X. Then (x_n) is a bounded sequence in $\mathbb{R}^n$ so by Bolzano Weierstrass, 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.

$\Rightarrow$ Suppose X is not closed. Then we can find a sequence (x_n) in X such that 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 can find a sequence (x_n) in X with for all n, $\|x\| \ge n$, i.e. $\|x_n\| \to \infty$ as $n \to \infty$. Suppose we have a subsequence $x_{n_j} \to x \in X$. Then $\|x_{n_j}\| \to \|x\|$ but $\|x_{n_j}\| \to \infty$, contradiction. So, again, X is not sequentially compact.

$\square$

Start of lecture 10

Can this theorem be generalised to any metric space? Obviously not: for example in $\mathbb{R}\setminus\{0\}$ with usual metric, the set $[-1,0)\cup(0,1]$ is closed and bounded but the sequence $\left(\frac{1}{n}\right)_{n\geq 1}$ has no convergent subsequence.

Problem: space not complete. Maybe complete + bounded could imply sequentially compact?
Even this doesn't work. Recall example from section 1: let

$$X = \{f \in B(\mathbb{R}) \mid \sup_{x\in\mathbb{R}} |f(x)| \leq 1\}$$

with uniform metric. Then X is complete (closed subset of complete space $B(\mathbb{R})$) and bounded (if $f, g \in X$, $d(f,g) \leq 2$). 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. So the problem is that X is 'too big'. So we need a stronger concept of boundedness.

Definition. Let (X, d) be a metric space. We say X is *totally bounded* if for all $\delta > 0$ we can find a *finite* set $A \subset X$ such that $\forall x \in X$ $\exists a \in A$ with $d(x, a) < \delta$.

Theorem 21. A metric space is sequentially compact if and only if it is complete and totally bounded.

Proof. $\Leftarrow$ 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, can find finite $A_1 \subset X$ such that $\forall x \in X$ there exists $a \in A_1$ with $d(x, a) < 1$. In particular there is an infinite set $N_1 \subset \mathbb{N}$ and a point $a_1 \in A_1$ such that $\forall n \in N_1$, $d(x_N, a_1) < 1$. Hence $\forall m, n \in N_1$, $d(x_m, x_n) < 2$. Similarly, we can find finite $A_2 \subset X$ such that $\forall x \in X$, $\exists a \in A_2$, $d(x, a) < \frac{1}{2}$. In particular, there is an infinite $N_2 \subset N_1$ such that $\forall n \in N_2$, $d(x_n, a_2) < \frac{1}{2}$ and thus $\forall m, n \in N_2$, $d(x_m, x_n) < 1$.

Keep going. We get a sequence $N_1 \supset N_2 \supset N_3 \supset \cdots$ of infinite subsets of $\mathbb{N}$ such that $\forall i$, $\forall m, n$, $m, n \in \mathbb{N}_i \implies d(x_m, x_n) < \frac{2}{i}$.

Now pick $n_1 \in N_2$. Then pick $n_2 \in N_2$ with $n_2 > n_1$. Then pick $n_3 \in N_3$ with $n_3 > n_2$ and so on. We obtain a subsequence (x_{n_j}) of (x_n) such that for all j, $x_{n_j} \in N_j$. Thus if $i \le j$ 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. So X is sequentially compact.

$\Rightarrow$ 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 $x_n \to x$ (exercise). Contradiction.

Suppose instead X is not totally bounded. Then there is some $\delta > 0$ such that whenever $A \subset X$ is finite, there exists $x \in X$ such that $\forall a \in A$, $d(x, a) \ge \delta$. So pick $x_1 \in X$. Pick $x_2 \in X$ such that $d(x_1, x_2) \ge \delta$. Pick $x_3 \in X$ such that $d(x_1, x_3) \ge \delta$ and $d(x_2, x_3) \ge \delta$. Continue. Then we get a sequence (x_n) in X such that for all i, j with $i \ne j$, $d(x_i, x_j) \ge \delta$. Then (x_n) has no convergent subsequence.

$\square$

Exercise: A continuous function on a sequentially compact metric space is uniformly continuous. If the function 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, for example in $\mathbb{R}^n$ we have the very different concepts of distance given by the Euclidean, l_1 and l_∞ metrics. But all give same concept of convergence and continuity.

Definition. Let (X, d) and (Y, e) be metric spaces. Let $f : X \to Y$. We say f is a *homeomorphism* and that X, Y are *homeomorphic* if f is a continuous bijection with continuous inverse.

Remark. Homeomorphism is an equivalence 'relation'.

Examples

(1) If $x, y \in \mathbb{R}^n$:

$$d_\infty(x, y) \le d_1(x, y) \le n d_\infty(x, y)$$

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 inverse map $(\mathbb{R}^n, d_\infty) \to (\mathbb{R}^n, d_1)$ is continuous. So it's a homeomorphism. Similarly, $\mathbb{R}^n$ with Euclidean metric is homeomorphic to both these spaces.

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

  (i) $\exists A\ \forall x, y \in X,\ e(f(x), f(y)) \le Ad(x, y)$.

  (ii) and $\exists B\ \forall x, y \in X,\ d(x, y) \le Be(f(x), f(y))$.

  Then f, f^{-1} are continuous so X, Y are homeomorphic.

(3) Define $f : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$ by $f(x) = \tan x$. Then f is a homeomorphism (usual metric in each case). But there is no constant A such that

$$\forall x, y \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) |\tan x - \tan y| \le A|x - y|$$

Proposition 22. Let (V, b), (W, c), (X, d), (Y, e) be metric spaces and $f : X \to V$, $g : Y \to W$ be homeomorphisms. Then

(i) In X, $x_n \to x$ if and only if in V, $f(x_n) \to f(x)$;

(ii) and a function $g : X \to Y$ is continuous at $a \in X$ if and only if $g \circ h \circ f^{-1}$ is continuous at $f(a) \in V$.

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

(ii) h continuous implies $g \circ h \circ f^{-1}$ because composition of continuous functions is continuous.
And $g \circ h \circ f^{-1}$ continuous implies $h = g^{-1} \circ (g \circ h \circ f^{-1}) \circ f$ is continuous because composition of continuous functions is continuous.

□

We now have examples of metric spaces that look very different but behave identically with respect to convergence / continuity.
Thought: Could we disperse with distance altogether?
Another way to think about continuity:

Definition. 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

$$B_\varepsilon(a) = \{x \in X \mid d(x, a) < \varepsilon\}$$

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

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

So we have redefined continuity in terms of open balls. But open balls have radii, so this still uses a notion of distance.

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

Start of lecture 11

Remarks

(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 triangle 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 neighbourhood of a:

$$a \in \mathcal{N} \subset \mathcal{N}.$$

However, a neighbourhood of a need not be open. For example in $\mathbb{R}$ with the usual metric then $[-1,1]$ is a neighbourhood of 0:

$$0 \in (-1,1) \subset [-1,1]$$

And $[-1,1] \cup \{396\}$ is a neighbourhood of 0 (throwing in extra stuff doesn't stop something from being a neighbourhood).

(4) $\mathcal{N}$ is a neighbourhood of a if and only if $\exists \varepsilon > 0$ with $B_\varepsilon(a) \subset \mathcal{N}$:

$\Leftarrow$ $a \in B)\varepsilon(a) \subset \mathcal{N}$.

$\Rightarrow$ $a \in G \subset \mathcal{N}$ and $\exists \varepsilon > 0$ such that $B_\varepsilon(a) \subset G$.

(5) A set G is open if and only if it's a neighbourhood of each of its points.

Proposition 23 (Generalising continuity)**.** Let (X, d), (Y, e) be metric spaces and let $f : X \to Y$.

(i) f is continuous at $a \in X$ if and only if whenever $\mathcal{N} \in Y$ is a neighbourhood of $f(a)$ we have $f^{-1}(\mathcal{N}) \subset X$ a neighbourhood of a;

(ii) f is a continuous function if and only if whenever $G \subset Y$ open we have $f^{-1}(G) \subset X$ open.

Proof. (i) $\Rightarrow$ Suppose f is continuous at $a \in X$. Let $\mathcal{N}$ be a neighbourhood of $f(a)$. Then $\exists \varepsilon > 0$ such that $B_\varepsilon(f(a)) \subset \mathcal{N}$. But f continuous at a so $\exists \delta > 0$ such that $B_\delta(a) \subset f^{-1}(B_\varepsilon(f(a))) \subset f^{-1}(\mathcal{N})$. So $f^{-1}(\mathcal{N})$ is a neighbourhood of a.

$\Leftarrow$ Suppose $f^{-1}(\mathcal{N})$ is a neighbourhood of a for every neighbourhood $\mathcal{N}$ of $f(a)$. Let $\varepsilon > 0$. In particular, $B_\varepsilon(f(a))$ is a neighbourhood of $f(a)$ so $f^{-1}(B_\varepsilon(f(a)))$ is a neighbourhood of a so $\exists \delta > 0$ such that $B_\delta(a) \subset f^{-1}(B_\varepsilon(f(a)))$. So f is continuous at a.

(ii) $\Rightarrow$ Suppose f is a continuous 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 neighbourhood of $f(a)$. Moreover, f is continuous at a so by (i) we have $f^{-1}(G)$ a neighbourhood of a. Hence $\exists \delta > 0$ such that $B_\delta(a) \subset f^{-1}(G)$. So $f^{-1}(G)$ is open.

$\Leftarrow$ Suppose $f^{-1}(G)$ open whenever G is open in Y. Let $a \in X$. Let $\mathcal{N} \subset Y$ be a neighbourhood of $f(a)$. Then $\exists G \subset Y$ open such that $f(a) \in G \subset \mathcal{N}$. 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 a neighbourhood of a. So by (i), f is continuous at a. So f is a continuous function.

□

Remarks

(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 23(ii) says homeomorphisms also preserve open sets: to be precise, if $f : X \to Y$ is a homeomorphism then $G \subset X$ is open if and only if $f(G) \subset Y$ is open. (Why? $G = f^{-1}(f(G))$ with f continuous and $f(a) = (f^{-1})^{-1}(G)$ with f^{-1} continuous.)

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.
So if X, Y homoeomorphic spaces, then

$$X \text{ sequentially compact} \iff Y \text{ sequentially compact}$$

'Sequential compactness is a *topological property*'

If X, Y are homeomorphic, and one of them has a particular topological property, then so does the other.

What about completeness? Not so good.

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

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

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

Just seen sequential compactness is a topological property. Can define sequential compactness just in terms of neighbourhoods / open sets:

$$\text{sequentially compact} \leftarrow \text{convergence of sequences} \leftarrow \text{neighbourhoods}$$

Is there a 'nicer' way to do this?

Definition. Let X be a metric space. An *open cover* of X is a collection $\mathcal{C}$ of open subsets of X such that

$$X = \bigcup_{G \in \mathcal{C}} G$$

A *subcover* of $\mathcal{C}$ is an open cover $\mathcal{B}$ of X with $\mathcal{B} \subset \mathcal{C}$. We say X is *compact* if every open cover of X has a finite subcover.

Example (The Heine-Borel theorem)**.** $[0,1]$ with the usual metric is compact.

Proof. Let $\mathcal{C}$ be an open cover of $[0,1]$. Let

$$A = \{x \in [0,1] \mid \exists \mathcal{B} \subset \mathcal{C} \text{finite with } [0,x] \subset \bigcup_{G \in \mathcal{B}} G$$

We know $\exists G \in \mathcal{C}$ with $0 \in G$. So $0 \in A$ so $A \neq \emptyset$. Clearly A bounded above by 1. So A has a supremum, $\sigma = \sup A$, say.
As G is open, $\exists \varepsilon > 0$ such that $[0,\varepsilon) = B_\varepsilon(0) \subset G$. So $\frac{\varepsilon}{2} \in A$ so $\sigma > 0$.
Suppose $\sigma < 1$. Can find $G' \in \mathcal{C}$ with $\sigma \in G$. As $\sigma = \sup A$, we can find $x \in A$ with $x \in G'$. So have $\mathcal{B} \subset \mathcal{C}$ finite with $[0,x] \subset \bigcup_{G \in \mathcal{B}} G$. But $\exists \varepsilon > 0$ such that $(\sigma - \varepsilon, \sigma + \varepsilon) = B_\varepsilon(\sigma \subset G$. So

$$\left[0, \sigma + \frac{\varepsilon}{2}\right] \subset \bigcup_{G \in \mathcal{B} \cup \{G'\}} G$$

So $\sigma + \frac{\varepsilon}{2} \in A$. Contradiction. Hence $\sigma = 1$. Can find $G'' \in \mathcal{C}$ such that $1 \in G''$. As G'' open, can find $\varepsilon > 0$ such that $(1-\varepsilon, 1] = B_\varepsilon(1) \subset G''$. As $1 = \sup A$ 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 $\mathcal{B} \cup \{G''\}$ is an open cover of $[0,1]$ and so a subcover of $\mathcal{C}$. So $[0,1]$ is compact. □

Start of lecture 12

Theorem 24. Let X be a metric space. Then the following are equivalent:

(i) X is compact;

(ii) X is sequentially compact;

(iii) X is complete and totally bounded;

and, if X is a subspace of $\mathbb{R}^n$ with the Euclidean metric

(iv) $X \subset \mathbb{R}^n$ is closed and bounded.

Proof. Done (ii) $\iff$ (iii) ($\iff$ (iv) if appropriate) in section 2.3. So only remains to show (i) $\iff$ (ii).

$\Rightarrow$ 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 neighbourhood 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_1 such that $x_{n_1} \in B_1(a)$, then $n_2 > n_1$ such that $x_{n_2} \in B_{\frac{1}{2}}(a)$, and so on, then $x_{n_j} \to a$, contradiction).

Now, let $\mathcal{C} = \{G_a \mid a \in X\}$. 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 $\bigcup_{G \in \mathcal{D}} G \neq X$. So $\mathcal{C}$ has no finite subcover. Hence X is not compact.

$\Leftarrow$ Suppose X is sequentially compact. Let $\mathcal{C}$ be an open cover of X. Then we claim that there exists $\delta > 0$ such that for all $a \in X$, there exists $G \in \mathcal{C}$ such that $B_\delta(a) \subset G$.

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 such that 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 $x_{n_j} \to a \in X$, say. Pick $G \in \mathcal{C}$ such that $a \in G$. As G open, can pick $\varepsilon > 0$ such that $B_\varepsilon(a) \subset G$. Pick j sufficiently large that $x_{n_j} \in B_{\frac{\varepsilon}{2}}(a)$ and also $\frac{1}{n_j} < \frac{\varepsilon}{2}$. Then

$$B_{\frac{1}{n_j}}(x_{n_j}) \subset B_\varepsilon(a) \subset G$$

contradiction. So such a δ does exist.

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

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

Proposition 25. Let X be a metric space and $G \subset X$. Then G is open if and only if $F = X \setminus G$ is closed.

Proof. $\Rightarrow$ Suppose F not closed. Then there is a sequence (x_n) in F with $x_n \to x \in G$. Suppose $\mathcal{N}$ is a neighbourhood of x. Then there exists N such that for all $n \geq N$, $x_n \in \mathcal{N}$. But for all n, $x_n \notin G$. So $\mathcal{N} \neq G$. So G is not a neighbourhood of x. So G is not open.

$\Leftarrow$ Suppose G is not open. Then there is some $x \in G$ such that $\forall \varepsilon > 0$, $B_\varepsilon(x) \not\subset G$. That is, $B_\varepsilon(x) \cap F \neq \emptyset$. So for $n = 1, 2, 3, \dots$ 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 26. Let X be a metric space and let $\tau = \{G \subset X \mid G \text{ open}\}$. Then:

(i) $\emptyset \in \tau$ and $X \in \tau$,

(ii) if $\sigma \subset \tau$ then

$$\bigcup_{G \in \sigma} G \in \tau$$

"any union of open sets is open"

(iii) if $G_1, G_2, \dots, G_n \in \tau$ then

$$\bigcap_{i=1}^{n} G_i \in \tau$$

"a finite intersection of open sets is open".

Remark. Do need finiteness in (iii). For example for all $n \in \mathbb{N}$, $\left(-\frac{1}{n}, \frac{1}{n}\right)$ is open in $\mathbb{R}$ with usual metric. But $\bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, \frac{1}{n}\right) = \{0\}$ is not.

Proof. (i) Obvious

(ii) 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 neighbourhood of a (as G open) so H is a neighbourhood of a (as $G \subset H$). Hence H is open, i.e. $H \in \tau$.

(iii) Suppose $G_1, \dots, G_n \in \tau$ 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 $\exists \delta_i > 0$ such that $B_{\delta_i}(a) \subset G_i$. Let $\delta = \min\{\delta_1, \dots, \delta_n\}$. 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. $J \in \tau$.

□

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

(i) $\emptyset \in \tau$ and $X \in \tau$;

(ii) if $\sigma \subset \tau$ then
$$\bigcup_{G \in \sigma} G \in \tau$$

(iii) if $G_1, \dots, Gn \in \tau$ then
$$\bigcap_{i=1}^{n} G_i \in \tau$$

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

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

Example. Let (X, d) be a metric space. Let $\tau = \{G \subset X \mid G \text{ open}\}$. Then by proposition 26, τ is a topology on X. We say τ is the topology induced by the metric d.

Want to define open / closed / continuous etc for topological spaces. As metric spaces are topological spaces we want to try to make sure it's 'backwards compatible', so to say new definitions don't contradict old metric space ones. So in making new definitions, we'll be guided by section 2.4.

Start of lecture 13

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

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

Definition. Let (X, τ) and (Y, σ) be topological spaces. 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; that is, f is continuous if $\forall G \in \sigma$, $f^{-1}(G) \in \tau$.

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

Definition. Let (X, τ), (Y, σ) be topological spaces. We say that a property is *topological* if it is preserved by homomorphisms; that is to say, if X, Y are homeomorphic then X has the property if and only if Y does.

Remarks

(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 is open if and only if $G \in \tau$, often don't need to explicitly name the topology. For example, 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 if and only if G is a neighbourhood of each of its points.

Proposition 27. Let X, Y be topological spaces and let $f : X \to Y$. Then f is continuous if and only if for all $a \in X$, f is continuous at a.

Proof. $\Rightarrow$ 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 continuous, $f^{-1}(G) \subset X$ open. Now $a \in f^{-1}(G) \subset f^{-1}(\mathcal{N})$ with $f^{-1}(G)$ open.

$\Leftarrow$ 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, $f^{-1}(G) \subset X$ is open. Hence f is continuous.

$\square$

Proposition 28. Let (X, τ), (Y, σ), (Z, ρ) be topological spaces, let $f : X \to Y$ be continuous and let $g : Y \to Z$ 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 $g \circ f$ is continuous. $\square$

Examples

(1) The discrete topology: Let X be any set and $\tau = \mathcal{P}(X)$. 'Every set is open'. However, this is not new: 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.

(2) The indiscrete topology: Let X be any set and $\tau = \{\emptyset, X\}$. 'Only open sets are $\emptyset$ and the whole space'. This is genuinely new: τ cannot be induced by a metric (if $|X| \geq 2$). Indeed, suppose $|X| \geq 2$ and that τ is induced by a metric d. Let $x, y \in X$ with $x \neq y$, so $d(x,y) = \delta > 0$, say. Then $B_\delta(x)$ is open with $x \in B_\delta(x)$ and $y \notin B_\delta(x)$, contradiction.

(3) The cofinite topology: Let X be any infinite set and let

$$\tau = \{G \subset X \mid X \setminus G \text{ is finite}\} \cup \{\emptyset\}$$

Check this is a topology:

(i) $\emptyset \in \tau$, $X \setminus X = \emptyset$ is finite so $X \in \tau$.

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

$$\left(X \setminus \bigcup_{G \in \sigma} G \right) = \bigcap_{G \in \sigma} (X \setminus G) \subset X \setminus H$$

is finite. So $\bigcup_{G\in\sigma} G \in \tau$.

(iii) Let $G, H \in \tau$. If $G = \emptyset$ or $H = \emptyset$ then $G \cap H = \emptyset \in \tau$. Otherwise $X \setminus G$, $X \setminus H$ are finite and 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.
Is it induced by a metric d? No: observe first that if G, H are open and non-empty then $G \cap H \neq \emptyset$. Now suppose $x, y \in X$ with $x \neq y$. Then $d(x, y) = \delta > 0$ so $B_{\frac{\delta}{2}}(x)$, $B_{\frac{\delta}{2}}(y)$ are non-empty disjoint open sets. So d doesn't induce τ.

(4) The cocountable topology: Let X be any uncountable set and let

$$\tau = \{G \subset X \mid X \setminus G \text{ countable}\} \cup \{\emptyset\}$$

Then, very similarly to (3), this is a topology that is not induced by any metric.

3.2. Sequences and Hausdorff spaces

Definition. 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 there exists N such that for all $n \geq N$, $x_n \in \mathcal{N}$.

Examples

(1) Let X be an uncountable set with the cocountable topology. Which sequences converge in X? Suppose $x_n \to x$. Then let

$$\mathcal{N} = (X \setminus \{x_n \mid n \in \mathbb{N}\}) \cup \{x\}$$

Then $\mathcal{N}$ is open and $x \in \mathcal{N}$ so $\mathcal{N}$ is a neighbourhood of x. So there exists N such that for all $N \geq N$, $x_n \in \mathcal{N}$. So there exists N such that for all $n \geq N$, $x_n = x$. Obviously if there exists N such that for all $n \geq N$, $x_n = x$ then $x_n \to x$. So the only convergent sequences in this space are eventually constant.

(2) Let $X = \{1, 2, 3\}$ with the indiscrete topology. Let $x_n = i \in X$ with $i \equiv n \pmod 3$. So the sequence is $1, 2, 3, 1, 2, 3, 1, 2, 3, \ldots$. Then we claim that $x_n \to 2$. Let $\mathcal{N}$ be a neighbourhood of 2. Then there exists G open such that $2 \in G \subset \mathcal{N}$. But the only open sets are $\emptyset$ or $\{1, 2, 3\}$. So $G = \{1, 2, 3\}$. So $\mathcal{N} = \{1, 2, 3\}$ so for all n, $x_n \in \mathcal{N}$. So $x_n \to 2$. Similarly $x_n \to 1$ and $x_n \to 3$. So:

Warning. LIMITS OF CONVERGENT SEQUENCES NEED NOT BE UNIQUE. So we can't write $\lim_{n\to\infty} x_n$, unless we prove that the limit exists *and* is unique. Note the above proof shows that in any indiscrete space every sequence converges to every point of the space.

Start of lecture 14

Definition (Hausdorff Space)**.** 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 $y \in H$.

Examples

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

(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 $G, h \neq \emptyset$ so $X \setminus G$, $X \setminus H$ are finite and so
$$X \setminus (G \cap H) = (X \setminus G) \cup (X \setminus H)$$
is finite. In particular $G \cap H \neq \emptyset$. Similarly, the cocountable topology is not Hausdorff.

Proposition 29. 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 $G \cap H = \emptyset$. Now G is a neighbourhood of a so there is some N_1 such that $\forall n \geq N_1$, $x_n \in G$. Similarly, there is some N_2 such that $\forall n \geq N_2$, $x_n \in H$. Take $n = \max\{N_1, N_2\}$. Then $x_n \in G \cap H = \emptyset$, contradiction. Hence $a = b$. □

Relationship to continuity?

Proposition 30. 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 is continuous at a we know $f^{-1}(\mathcal{N})$ is a neighbourhood of a. As $x_n \to a$ we can fined N such that $\forall n \geq N$, $x_n \in f^{-1}(\mathcal{N})$. Then for all $n \geq N$, $f(x_n) \in \mathcal{N}$. So $f(x_n) \to f(a)$. □

Warning. CONVERSE NOT TRUE IN GENERAL

Example. Let $X = Y = \mathbb{R}$, X with cocountable topology, Y with usual topology and $f : X \to Y$ 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 $f^{-1}((-1, 1)) = (-1, 1) \subset X$ is not a neighbourhood of 0 in X. So f is not continuous at 0.

Here, even imposing condition that the spaces are Hausdorff is not enough.

Example. Take example as above but replace topology on X by

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

This is a topology (check-exercise). And it 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 the 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. In a metric space, the topology is completely determined by convergence of sequences. Not true for a general topological space. Hence we'll tend to concentrate more on continuity than convergence of sequences.

3.3. Subspaces

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

$$\sigma = \{G \cap Y \mid G \in \tau\}$$

Easy to check that this is a topology. Need to check backward compatibility with metric space definition:

Proposition 31. Let (X, d) be a metric space with topology τ 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 $d\mid_{Y^2}$. Suppose $G \in \tau$. Let $y \in G \cap Y$. As $y \in G$ and G open in X we can find $\delta > 0$ such that $\forall x \in X$, $d(x, y) < \delta$ implies $x \in G$. Then $\forall x \in Y$, $d(x, y) < \delta \implies x \in G \cap Y$. So $G \cap Y$ is a neighbourhood of y. So $G \cap Y \in \sigma$.

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

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

$(y \in H)$. Each $B_{\delta_y}(y)$ is open , for each $y \in H$, $y \in B_{\delta_y(y)}$ and $B_{\delta_y}(y) \cap Y \subset H$. 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 32. A subspace of a Hausdorff space is Hausdorff.

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

3.4. Completeness

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

Definition. Let (X, τ) be a topological space and $\mathcal{C}$ and open cover of X. A *subcover* of $\mathcal{C}$ is a subset $\mathcal{D}$ of $\mathcal{C}$ which is itself an open cover.

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

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

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

Remark. Traditional wording: here and elsewhere, if no topology is specified $\mathbb{R}$ is generally assumed to have the usual topology. Proof similar to metric case.

We've seen for a metric space that compact and sequentially compact are equivalent.

Warning. This equivalence is not true for a general topological space.

There exists a compact space that isn't sequentially compact, and a sequentially compact space that isn't compact, but both examples are beyond the scope of this course.

Observe compactness and sequential compactness are both topological properties, since they both use only open sets and convergence of sequences.

Given 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.

Start of lecture 15

Remark. If X is a topological space and $K \subset X$ we might want to say 'K is compact'. Clearly meaningful since K is topological space with subspace topology. Think further:
Let τ be the topology on X. Then K is compact if and only if 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 if and only if 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 $K \subset \bigcup_{G \in \mathcal{D}} G$. So sometimes refer to $\mathcal{C}$ as being open cover of K (in X).

Examples

(1) $[0, 1]$ with the usual topology is compact. (section 2.4 "Heine Borel Theorem" - 'creeping along proof'). More generally, $S \subset \mathbb{R}^n$ is compact if and only if S is closed and bounded.

(2) A metric space is compact if and only if it is complete and totally bounded.

(3) Suppose X is a discrete topological space. Then $\{\{x\} \mid x \in X\}$ is an open cover. So X is compact if and only if X is finite. (Note any finite space is compact).

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

Theorem 33. 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 $\{G_n \mid n \in \mathbb{N}\}$ is an open cover of X and so, as X compact, have a finite subcover $\{G_{n_1}, \dots, G_{n_k}\}$.For all $x \in G_{n_i}$, $|f(x)| < n_i$. Hence for all $x \in X$, $|f(x)| < \max_{1 \le i \le k} n_i$. Hence f is bounded.

Let $\sigma = \sup_{x \in X} f(x)$, and suppose σ not attained by f. Then can 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 as $\sigma = \sup_{x \in X} f(x)$, so given $\varepsilon > 0$ we can find x such that $\sigma - f(x) < \varepsilon$ so $g(x) > \frac{1}{\varepsilon}$. Contradiction. Similarly, $\in_{x \in X} f(x)$ is attained. □

Remark. Think of compactness as a 'smallness' condition - next best thing to finiteness. For example: a real-valued function on a finite space is bounded (obvious). Here we have a continuous function on a compact space - how do we show boundedness? Use compactness to show space not 'too big' - we can cover it with finitely many sets on each of which f is bounded. (Then it becomes obvious).

More generally:

Theorem 34. 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 $\{f^{-1}(G) \mid G \in \mathcal{C}\}$ is an open cover of X, so by compactness, there is a finite $\mathcal{D} \subset \mathcal{C}$ such that $\{f^{-1}(G) \mid G \in \mathcal{D}\}$ is an open cover of X. Then $\mathcal{D}$ is an open cover of K in Y. So K is compact. □

Remark. This together with the fact that compact subsets of $\mathbb{R}$ are closed and bounded gives an alternative proof of Theorem 33.

Lemma. (a) A closed subsets of a compact space is compact.

(b) A compact subset of a Hausdorff space is closed.

Proof. (a) 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. Then $X \setminus F$ is open so $\mathcal{C}' = \mathcal{C} \cup \{X \setminus F\}$ then $\mathcal{C}'$ is an open cover of X. X is compact so $\mathcal{C}'$ 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.

(b) 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 $y \in X \setminus K$. Given $x \in K$, $x \neq y$ so as X Hausdorff, can find disjoint open $U_x, V_x \subset X$ with $x \in U_x$ and $y \in V_x$. Then $\{U_x \mid x \in X\}$ is an open cover of K in X so it has a finite subcover $\{U_{x_1}, U_{x_2}, \dots, U_{x_n}\}$. Let

$$U = \bigcup_{i=1}^{n} U_{x_i} \quad V = \bigcap_{i=1}^{n} V_{x_i}$$

We have U, V open, $K \subset U$, $y \in V$ and $U \cap V = \emptyset$. 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, so it's open, so K is closed.

□

Theorem 35. A continuous bijection from a compact space to a Hausdorff space is a homeomorphism.

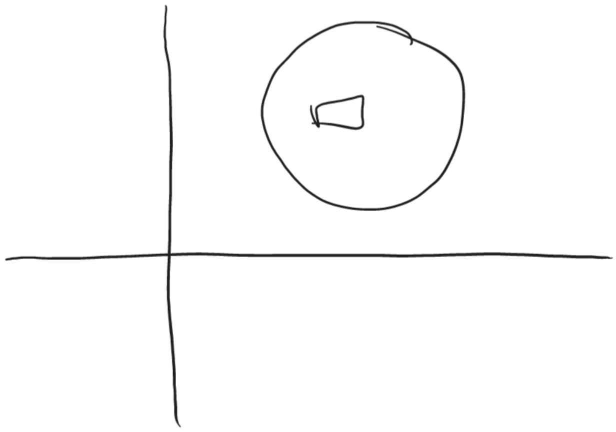
Proof. Let $f : X \to Y$ be a continuous bijection, X compact, Y Hausdorff. Aim is to show $f^{-1} : Y \to X$ is continuous.

Let $G \subset X$ open. Then $X \setminus G$ is closed, so by Lemma 35(a), $X \setminus G$ is compact. Hence by Theorem 34, $f(X \setminus G)$ is compact and so by Lemma 35(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. So f^{-1} is continuous. □

Start of lecture 16

3.5. Products

Have $\mathbb{R}$ with the usual topology. Would like $\mathbb{R} \times \mathbb{R}$ to be $\mathbb{R}^2$ with the Euclidean topology.
In general, if (X, τ) and (Y, σ) 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. For example, $\mathbb{R} \times \mathbb{R}$



Open ball not in $\tau \times \sigma$, but each point in ball *is* in some set in $\tau \times \sigma$ confined in the ball. But open ball *is* union of some sets in $\tau \times \sigma$.

In general, the product topology from τ, σ is going to be collection of all unions of sets in $\tau \times \sigma$.

Definition. 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 36. Let Π be a π-system on a set X. Then

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

is a topology on X.

Proof. Clearly $\emptyset, X \in \tau$ and it's closed under arbitrary unions. Now suppose $G, H \in \tau$. If $G = \emptyset, X$ or $H = \emptyset, X$, 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

$$G \cap H = \bigcup_{\substack{A \in \Phi \\ B \in \Theta}} (A \cap B) = \bigcup_{C \in \Sigma} C$$

where $\Sigma = \{A \cap B \mid A \in \Phi, B \in \Theta\} \subset \Pi$. $\square$

We call τ the topology *generated by* Π.

Proposition 37. Let $(X, \tau), (Y, \sigma)$ be topological spaces. Then $\tau \times \sigma$ is a π-system on $X \times Y$.

Proof. $\emptyset = \emptyset \times \emptyset \in \tau \times \sigma$. So $\tau \times \sigma \neq \emptyset$. 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. Let $(X, \tau), (Y, \sigma)$ be topological spaces. The *product topology* on $X \times Y$ is the topology generated be the π-system

$$\{U \times V : U \in \tau, V \in \sigma\}$$

Exercise: If $X = Y = \mathbb{R}$, $\tau = \sigma =$ usual topology. Then the product topology on $\mathbb{R}^2$ is the Euclidean topology. (Example sheet 3 with guidance).

Theorem 38. (a) A product of Hausdorff spaces is Hausdorff.

(b) A product of compact spaces is compact.

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

(a) Suppose X, Y Hausdorff. Let $(x, y), (z, w) \in X \times Y$ with $(x, y) \neq (z, w)$. WLOG $x \neq z$. As X is Hausdorff, can find $G, H \in \tau$ with $G \cap H = \emptyset$, $x \in G$, $z \in H$. Then $G \times Y, H \times Y \in \rho$ with $(G \times Y) \cap (H \times Y) = \emptyset$ and $(x, y) \in G \times Y$, $(z, w) \in H \times Y$. So $X \times Y$ is Hausdorff.

(b) Suppose X, 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 $U_y \in \sigma$ and $V_y \in \tau$ 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 $\{V_y \mid y \in Y\} \subset \sigma\}$ is an open cover of Y. So, as Y is compact, it has a finite subcover $\{V_{y_1}, \dots, V_{y_n}\}$, 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

$$G_{y_i}^{(x)} = G_{y_i}$$

$(1 \leq i \leq n_x)$ as above. Then $\{W_x \mid x \in X\} \subset \tau$ is an open cover of X. So, as X is compact, it has a finite subcover, $\{W_{x_1}, \dots, W_{x_m}\}$, say. Now $X = \bigcup_{j=1}^m W_{x_j}$ and, for each j,

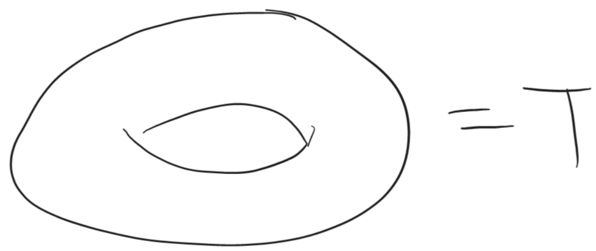
$$W_{x_J} \times Y \subset \bigcup_{i=1}^{n_{x_j}} G_{y_i}^{(x_j)}$$

Thus $\{G_{y_1}^{(x_j)} \mid 1 \leq j \leq m, 1 \leq i \leq n_{x_j}\}$ is an open cover of $X \times Y$ and hence a finite subcover of $\mathcal{C}$. Thus $X \times Y$ is compact.

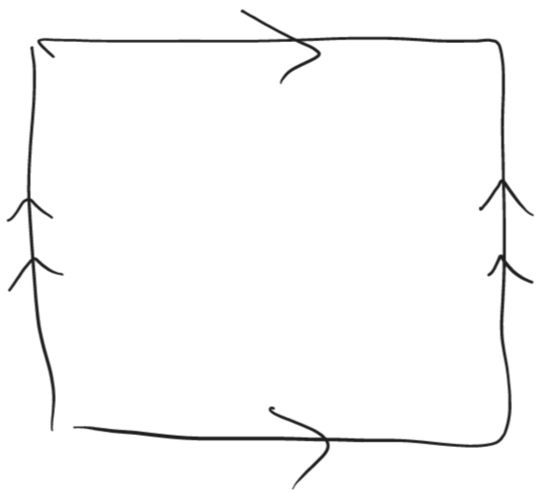
$\square$

3.6. Quotients

Consider the surface of a torus in $\mathbb{R}^3$:



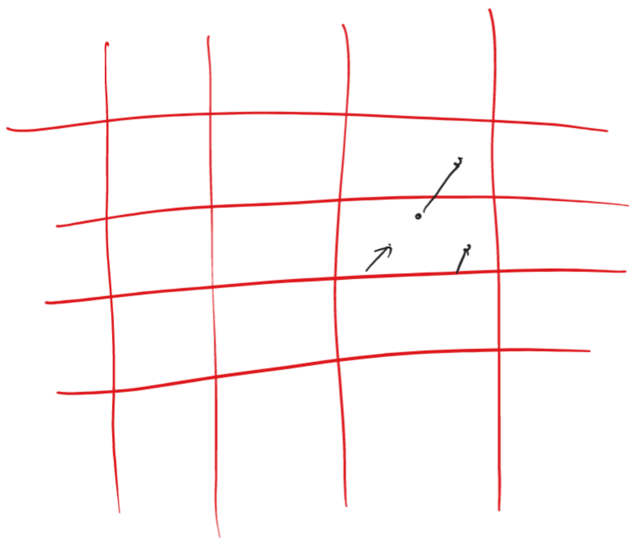
We might be interested in for example the set of continuous functions $T \to T$. Analysis is likely to be unpleasant, for example what is the equation defining T? What do we care about? Continuity and convergence. So if we replace T by a space homemorphic to T then we're happy, particularly if the new space is analytically easier to work with. For example, take closed unit square $[0,1] \times [0,1]$ with Euclidean 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 formally, defined an equivalence relation on $[0,1] \times [0,1]$, $\sim$ say, with equivalence classes:

$$\{(x,y)\} \quad (0 < x, y < 1)$$
$$\{(x,0),(x,1)\} \quad (0 < x < 1)$$
$$\{(0,y),(1,y)\} \quad (0 < y < 1)$$
$$\{(0,0),(0,1),(1,0),(1,1)\}$$

Essentially, we could define $T = [0,1]^2/\sim$, the set of equivalence classes. Maybe better way to do this?



Instead we could define an equivalence relation $\sim$ on $\mathbb{R}^2$ by $(x,y) \sim (z,w) \iff x - z \in \mathbb{Z}$ and $y - w \in \mathbb{Z}$. Again, hopefully could define $T = \mathbb{R}^2/\sim$.
But: What is the topology?

Start of lecture 17

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

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

Remarks

(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.

Examples

(1) Take $\mathbb{R}$ with the usual topology and $x \sim y$ if and only if $x - y \in \mathbb{Z}$. Then $\mathbb{R}/\sim$ 'is' S', the unit circle (a subspace of $\mathbb{R}^2$ with Euclidean topology). 'is' means 'is homeomorphic to'. (Proof later).

(2) As above but now $x \sim y$ if and only if $x - y \in \mathbb{Q}$. What is quotient topology on $\mathbb{R}/\sim$? Suppose $G \subset \mathbb{R}/\sim$ is open, $G \neq \emptyset$. 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 there exists $y \in (a, b)$ with $x - y \in \mathbb{Q}$ i.e. $x \sim y$. Then $q(x) = [x]_\sim = [y]_\sim = G$. Hence $G = \mathbb{R}/\sim$.
So the quotient topology on $\mathbb{R}/\sim$ is the indiscrete topology.

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

- Quotients of metrizable spaces need not be metrizable.
- Quotients of Hausdorff spaces need not be Hausdorff.

Basics on equivalence relations and quotients

Suppose X is a set and $\sim$ an equivalence relation on X. We have $X/\sim = \{[x]_\sim \mid x \in X\}$ 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. $x \sim y \implies f(x) = f(y)$.

$$\begin{tikzcd} X \arrow[rr, "f"] \arrow[dr, "q"'] & & Y \\ & X/\sim \arrow[ur, dashed, "\overline{f}"'] & \end{tikzcd}$$

Then there is a unique function $\overline{f} : X/\sim \to Y$ such that $f = \overline{f} \circ q$. Indeed, must have $\forall x \in X$,

$$\overline{f}([x]_\sim) = \overline{f}(q(x)) = f(x)$$

As f respects $\sim$, this is well-defined:

$$[x]_\sim = [y]_\sim \implies x \sim y \implies f(x) = f(y)$$

Example. Suppose G is a group, H another group, $\theta : G \to H$ is a homomorphism. Let $K = \ker \theta$, and define $\sim$ on G by $g \sim h \iff g^{-1}h \in K$. Then $G/K = G/\sim$ and

$$\begin{tikzcd} G \arrow[rr, "\theta"] \arrow[dr, "q"'] & & H \\ & G/K \arrow[ur, dashed, "\overline{\theta}"'] & \end{tikzcd}$$

Can check $\overline{\theta}$ is a homomorphism and injective so isomorphism onto $\theta(G)$. This is the first isomorphism theorem.

Proposition 39. Let (X, τ) be a topological space and $\sim$ 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 $\overline{f} : X/\sim \to Y$ such that $f = \overline{f} \circ q$, where $q : X \to X/\sim$ is the quotient map.

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

Remark. This is what makes quotients useful. For example recall torus $T = \mathbb{R}^2/\sim$ for appropriate relation $\sim$. Hopefully T is homeomorphic to genuine torus as a subspace in $\mathbb{R}^3$. T is nasty. $\mathbb{R}^2$ is nice. So always work 'upstairs' in $\mathbb{R}^2$ rather then 'downstairs' in T. For example if you want to think about a continuous function on T - instead think about an appropriate continuous function on $\mathbb{R}^2$ respecting $\sim$.

Example. Recall we had $\mathbb{R}$ with usual topology, $x \sim y$ if and only if $x - y \in \mathbb{Z}$ and $S' = \{x \in \mathbb{R}^2 \mid \|x\| = 1\}$ with subspace topology inherited from Euclidean topology on $\mathbb{R}^2$. We claimed $\mathbb{R}/\sim$ is homeomorphic to S'. Define $f : \mathbb{R} \to S'$ by

$$f(x) = (\sin 2\pi x, \cos 2\pi x)$$

Clearly f is a continuous surjection, and it respects $\sim$. By proposition 40 there is a unique continuous $\overline{f} : \mathbb{R}/\sim \to S'$ with $\overline{f} \circ q = f$. Clearly $\overline{f}$ is a continuous bijection (for injectivity, note each $x \in S'$ is $f(a)$ for a 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 so is compact. And S' is Hausdorff. Any continuous bijection from a compact space to a Hausdorff space is a homeomorphism. So done: $\overline{f}$ is a homeomorphism from $\mathbb{R}/\sim$ to S'.

3.7. Connectedness

Recall the *Intermediate Value Theorem*:
If $f : [a, b] \to \mathbb{R}$ is continuous, and without loss of generality $f(a) < f(b)$, then

$$[f(a), f(b)] \subset f([a, b])$$

Moreover, if $x, d \in f([a, b])$ with $c < d$ then $[c, d] \subset f([a, b])$. Doesn't work more generally, for example if we replace by $[a, b]$ by $[-1, 0) \cup (0, 1] = X$. Define: $X \to \mathbb{R}$ by

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

Then f is continuous on X, $0 \in f(X)$, $1 \in f(X)$ but for example $\frac{1}{2} \notin f(X)$ so $[0, 1] \not\subset f(X)$.
What's gone wrong? $[-1, 0) \cup (0, 1]$ is 'disconnected'.

Definition. 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.

Remarks

(1) Recall the $U \subset X$ is closed if and only if its complement is open. So

$$X \text{ disconnected} \iff \text{if there exist disjoint, non-empty closed sets } U, V \text{ with } X = U \cup V.$$

$$X \text{ connected} \iff \text{the only subsets of } X \text{ that are both open and closed are } \emptyset, X$$

Start of lecture 18

X connected if and only if whenever $U, V \subset X$ are open and disjoint with $X = U \cup V$ then $U = \emptyset$ or $V = \emptyset$.

Again, 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 $S \subset X$, X 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 if and only if there exist open sets $U, V \subset X$ such that $S \cap U \cap V = \emptyset$ and $S \subset U \cup V$, $S \cap U \neq \emptyset$, $S \cap V \neq \emptyset$. Again, we say U, V disconnect S.

Warning. We *don't* necessarily need to have $U \cap V = \emptyset$: for example in $\mathbb{N}$ with the 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

$$(\mathbb{N} \setminus \{1\}) \cap (\mathbb{N} \setminus \{2\}) \cap \{1, 2\} = \emptyset$$

but

$$(\mathbb{N} \setminus \{1\}) \cap (\mathbb{N} \setminus \{2\}) \neq \emptyset.$$

Indeed, if $U, V \subset \mathbb{N}$ are open and non-empty then $U \cap V \neq \emptyset$.

S connected if and only if whenever $U, V \subset X$ are open and $S \subset U \cap V$ and $S \cap U \cap V = \emptyset$ then either $S \cap U = \emptyset$ or $S \cap V = \emptyset$.

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

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

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

Proposition 40. Let $I \subset \mathbb{R}$ with the usual topology. Then I is connected if and only if I is an interval.

Proof. $\Rightarrow$ Suppose I 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 $\mathbb{R}$.

$\Leftarrow$ Suppose I is an interval. Work in subspace topology on I. Let $S \subset I$ be open, closed and non-empty. Let $a \in S$.

Suppose we have $b \in I \setminus S$. Without loss of generality $b > a$. 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 there exists $\delta > 0$ such that $(c - \delta, c + \delta) \subset S$, without loss of generality $\delta < b - c$. Then $c + \frac{\delta}{2} \in S \cap [a, b]$, contradiction to the supremum property. So in fact, $S = I$. So I is connected. $\square$

Another equivalent version of the definition of connectedness:

Theorem 41. Let X be a topological space. Then X connected if and only if every continuous function $f : X \to \mathbb{Z}$ (with usual topology) is constant.

Proof. $\Rightarrow$ 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 $f^{-1}(\{n\}) = \emptyset$ or $f^{-1}(\{n\}) = X$. So f is constant.

$\Leftarrow$ 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 $A \subset \mathbb{Z}$, $f^{-1}(A) = \emptyset, X, U$ or V, so $f^{-1}(A)$ is open. So f is continuous and non-constant. $\square$

Remarks

(1) Theorem 42 together with Intermediate value theorem can provide an alternative proof of Proposition 41.

(2) Theorem 42 remains true with same proof if $\mathbb{Z}$ is replaced by any discrete topological space Y with $|Y| \geq 2$.

Proposition 42. 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 $U \cap V \cap f(X) = \emptyset$. As f continuous, $f^{-1}(U), f^{-1}(V) \subset X$ are open. Also $X = f^{-1}(U0 \cup f^{-1}(V)$ and $f^{-1}(U) \cap f^{-1}(V) = \emptyset$. As X connected without loss of generality we have $f^{-1}(U) = \emptyset$. Then $U \cap f(X) = \emptyset$. So $f(X)$ is connected. □

Proposition 43. 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 $U \cap V = \emptyset$ and $U \cap V = X \times Y$. We want to show $U = \emptyset$, $V = X \times Y$ or $U = X \times Y$, $V = \emptyset$.
Fix $x \in X$. Then $\{x\} \times Y$ is homeomorphic to Y (Exercise - Example sheet 3). In particular, $\{x\} \times Y$ is connected. Then $\{x\} \times Y \subset U$ or $\{x\} \times Y \subset V$ as otherwise U, W would disconnect $\{x\} \times Y$ in $X \times Y$. So let

$$A = \{x \in X \mid \{x\} \times Y \subset U\}$$

and

$$B = \{x \in X \mid \{x\} \times Y \subset V\}$$

Clearly $A \cap B = \emptyset$ (as $U \cap V = \emptyset$). And we've just proved $X = A \cup B$.

Suppose $x \in A$. So $\{x\} \times Y \subset U$. Then (assuming $U \neq \emptyset$, which we can do without loss of generality, since if $Y = \emptyset$ then $X \times Y$ is clearly connected) 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 $\{w\} \times Y \subset U$, 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. But X is connected, so $A = \emptyset$ giving $U = \emptyset$ or $B = \emptyset$ giving $V = \emptyset$. Hence $X \times Y$ is connected. □

Example. Recall $[-1, 0) \cap (0, 1]$ is not connected. But it *is* a disjoint union of connected sets: $[-1, 0) \cup (0, 1]$. Moreover, any proper superset of $[-1, 0)$ or $(0, 1]$ in $[-1, 0) \cup (0, 1]$ is disconnected.

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

Theorem 44. The connected components of a topological space X form a partition of X.

Proof. Define $\sim$ on X by $x \sim y$ if and only if $\exists A \subset X$ connected with $x, y \in A$. Clearly $\sim$ is reflexive ($\{x\}$ is connected) and it is symmetric by the way we defined it. So we want to show $\sim$ 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. Without loss of generality $y \in U$. Pick $w \in V \cap (A \cup B)$. Without loss of generality $w \in A$. But also $y \in A$ so U, V disconnect A. So $A \cup B$ is connected so $x \sim z$.
Hence $\sim$ is an equivalence relation. Suppose S is an equivalence class of $\sim$. Suppose U, V disconnect S. Then can find $x \in U \cap S$, $y \in V \cap S$ and $U \cap V \cap S = \emptyset$. Then $x \sim y$ so there is a connected $A \subset X$ with $x, y \in A$. For all $z \in A$, $x, z \in A$ connected so $x \sim z$ so $z \in S$. So $A \subset S$ and so $U \cap V \cap A = \emptyset$ so U, V disconnect A, contradiction. So S is connected.
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 $S = T$. So S is a connected component.

Finally, let R be a connected component. Let $x, y \in R$. Then, as R is connected, $x \sim y$. So R is connected in some equivalence class Q of $\sim$. But $R \subset Q$ with Q connected so $R = Q$.

So the equivalence classes of $\sim$ are precisely the connected components. $\square$

Start of lecture 19

Remark. This is what tells us connected components exist.

Another concept of connectedness:

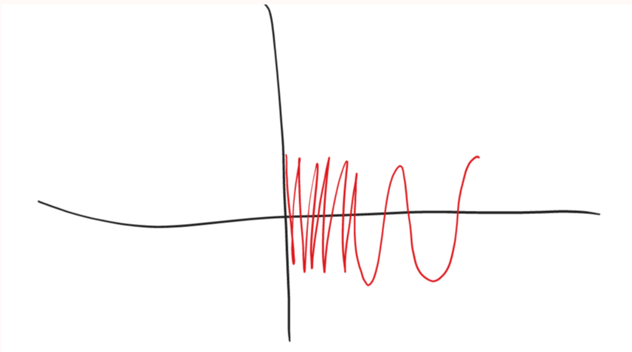
Definition. A *path* from x to y in a topological space X is a continuous function $\varphi : [0, 1] \to X$ with $\varphi(0) = x$, $\varphi(1) = y$.
X is *path-connected* if for all $x, y \in X$ there is a path from x to y.

Proposition 45. 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 $\varphi([0, 1])$. $\square$

However, the converse is not true in general.

Example. Consider



Let $A = \{(0, y) \mid -1 \le y \le 1\}$ and $B = \{(x, \sin\frac{1}{x} \mid 0 < x \le 1\}$. Let $X = a \cup B \subset \mathbb{R}^2$. X connected: Clearly A, B path-connected hence 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 open so have some $\delta > 0$ such that $B_\delta((0, 0)) \subset U$. Pick n such that $\frac{1}{2n\pi} < \delta$. Then $\left(\frac{1}{2n\pi}, 0\right) \in U \cap B$, contradiction.
X not path-connected: Suppose φ is a path from $(0, 0)$ to $(1, \sin 1)$ in X. Let $\sigma = \sup\{t \in [0, 1] \mid \varphi_1(t) = 0\}$. Let $y = \varphi_2(\sigma)$. Then, as φ 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, $\varphi_1\left(\sigma + \frac{\delta}{2}\right) = x > 0$. Choose $w \in (0, x)$ such that $\left|\sin\frac{1}{w} - y\right| \ge 1$. 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

$$\|\varphi(\sigma) - \varphi(t)\| \ge |\varphi_2(\sigma) - \varphi_2(t)| = \left|\sin\frac{1}{w} - y\right| \ge 1$$

contradiction.

BUT:

Proposition 46. An open, connected subset of Euclidean space is path-connected.

Proof. Let $X \subset \mathbb{R}^n$ be open and connected. If $X = \emptyset$ then done. So assume $X \neq \emptyset$. Fix $a \in X$. Let

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

- $U \neq \emptyset$ since $a \in U$ (constant path from a to a)
- U open in X. Suppose $b \in U$. X open so 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

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

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

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

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

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

Remark. Recall that don't always specify the topology when defining a topological space - should always assume it's standard one. In particular:

- $\mathbb{R}$ comes with the usual topology.
- $\mathbb{R}^n$ comes with the Euclidean topology.

$X \subset \mathbb{R}, \mathbb{R}^n$ comes with the subspace topology from the above. Products, quotients come with the product / quotient topology respectively.

Chapter II

Differentiation

Contents

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

$$\frac{f(a+h) - f(a)}{h} \to A$$

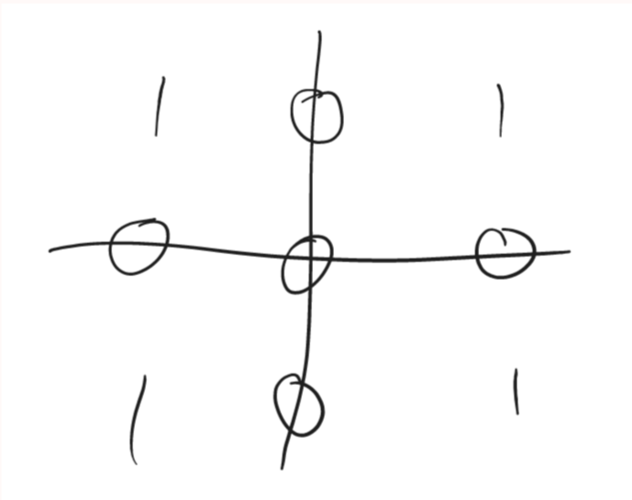
as $h \to 0$. We write $f'(a) = A$.
Want to generalise to $f : \mathbb{R}^n \to \mathbb{R}^m$.
Easy: $n = 1$. Exactly same definition works.
Problem: if $n \geq 2$, dividing by $b \in \mathbb{R}^n$ makes no sense.

Definition. If $f : \mathbb{R}^n \to \mathbb{R}^m$, 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 $e_1, \dots, e_n$ is standard basis of $\mathbb{R}^n$.

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



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

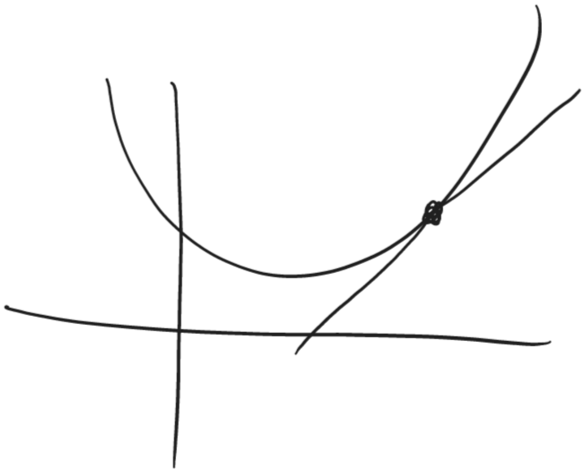
Both partial derivatives exist at $(0,0)$:

$$D_1 f(0,0) = 0 \qquad D_2 f(0,0) = 0$$

But f not continuous at $(0,0)$.

Better definition? Return to $f : \mathbb{R} \to \mathbb{R}$:

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



'small changes in a produce approximately linear changes in $f(a)$'

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

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

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

Proposition 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

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

$$f(a + h) = f(a) + \beta(h) + \eta(h)\|h\|$$

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

Definition. Once we've proved Proposition 1, we know the α in $(*)$ is unique. We say α is the *derivative* of f at a, and write $Df|_a = \alpha$. So if f differentiable at a,

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

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

Remark. If $f : \mathbb{R} \to \mathbb{R}^m$,

$$Df|_a(h) = f'(a)h$$

Start of lecture 20

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

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

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

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

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

Remarks

(1) To consider differentiability of f at a, 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 $\mathcal{N} \subset \mathbb{R}^m$ 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 rest of $\mathbb{R}^n$ and makes no difference).

(2) We can define the l_1 and l_∞ norms by

$$\|x\|_1 = d_1(0, x) = \sum_{i=1}^n |x_i|$$

and $\|x\|_\infty = d_\infty(0, x) = \max_i |x_i|$. Note $\|x\|_1 \geq 0$ with equality if and only if $x = 0$,

$$\|\lambda x\|_1 = |\lambda|\|x\|_1; \quad \|x + y\|_1 \leq \|x\|_1 + \|y\|_1$$

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

$$\|x\|_\infty \leq \|x\| \leq \sqrt{n}\|x\|_\infty$$

and
$$\|x\|_\infty \le \|x\|_1 \le n\|x\|_\infty$$
So we can replace $\| \bullet \|$ in the definition of derivative by $\| \bullet \|_1$ or $\| \bullet \|_\infty$ and the definition 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 $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \sim \mathbb{R}^{mn}$ with the obvious isomorphism (write map as a matrix with respect to standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$).
So could think about Euclidean norm of a linear map. But seems a bit unnatural.

Definition. The *operator norm* on $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is defined by
$$\|\alpha\| = \sup\{\|\alpha x\| \mid \|x\| = 1\}$$
$\alpha x \in \mathbb{R}^m$ so $\| \bullet \|$ is normal Euclidean norm.

Proposition 2. Let $\| \bullet \|$ be the operator norm on $V = \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Let $\alpha, \beta \in V$. Then

(i) $\|\alpha\| \ge 0$ with equality if and only if $\alpha = 0$;

(ii) $\forall \lambda \in \mathbb{R}$, $\|\lambda\alpha\| = |\lambda|\|\alpha\|$;

(iii) $\|\alpha + \beta\| \le \|\alpha\| + \|\beta\|$;

(iv) $\forall x \in \mathbb{R}^n$, $\|\alpha x\| \le \|\alpha\|\|x\|$;

(v) $\|\alpha\beta\| \le \|\alpha\|\|\beta\|$;

(vi) If $\| \bullet \|'$ is the Euclidean norm on $V \sim \mathbb{R}^{mn}$ with the standard isomorphism then there are constants $c, d > 0$ (depending on m, n but independent of α) such that
$$c\|\alpha\| \le \|\alpha\|' \le d\|\alpha\|$$

Remarks

(1) A linear map from $\mathbb{R}^n \to \mathbb{R}^m$ is continuous and $\{x \in \mathbb{R}^n \mid \|x\| = 1\}$ is compact so the operator norm is well-defined.

(2) Notation: standard is $\| \bullet \|$ refers to operator norm if applied to a linear map and Euclidean norm if applied to a point of $\mathbb{R}^n$, unless otherwise stated.

Proof. (i) to (iii) Exercise.

(iv) Let $x \in \mathbb{R}^n$. If $x = 0$ then done. Otherwise, $\alpha(x) = \|x\|\alpha\left(\frac{x}{\|x\|}\right)$ with $\left\|\frac{x}{\|x\|}\right\| = 1$. So

$$\|\alpha(x)\| \le \|x\|\|\alpha\|$$

(v) Let $x \in \mathbb{R}^n$ with $\|x\| = 1$. Then

$$\|\alpha\beta x\| \le \|\alpha\|\|\beta x\| \le \|\alpha\|\|\beta\|\|x\| = \|\alpha\|\|\beta\|$$

(by (iv) twice). So $\|\alpha\beta\| \le \|\alpha\|\|\beta\|$.

(vi) Let $x \in \mathbb{R}^n$ with $\|x\| = 1$.

$$\|\alpha x\| \le \sqrt{m} \max_{1 \le i \le m} |(\alpha x)_i|$$

Let A be the matrix of α with respect to the standard bases $e_1, \dots, e_n$ of $\mathbb{R}^n$ and $f_1, \dots, f_m$ of $\mathbb{R}^n$. Then

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

Hence

$$\|\alpha\| \le n\sqrt{m}\|\alpha\|'$$

On the other hand, pick i, j that maximise $|A_{ij}|$. Then

$$\|\alpha e_j\| \ge \|(\alpha e_j)_i\| = |A_{ij}|$$

But

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

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

Proposition 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 $\varepsilon(h) \to 0$ as $h \to 0$. Also, $\|h\| \to 0$ as $h \to 0$, and $Df|_a$ is linear so continuous so $Df|_a(h) \to Df|_a(0) = 0$ as $h \to 0$. Thus $f(a+h) \to f(a)$ as $h \to 0$. □

Proposition 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 $f+g$, λf 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
$$f(a+h) = f(a) + Df|_a(h) + \varepsilon(h)\|h\|$$
$$g(a+h) = g(a) + Dg|_a(h) + \eta(h)\|h\|$$
$$\lambda(a+h) = \lambda(a) + D\lambda|_a(h) + \zeta(h)\|h\|$$
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,
$$(\lambda f)(a+h) = (\lambda f)(a) + \lambda(a)Df|_a(h) + D\lambda|_a(h)f(a) + \zeta(h)\|h\|$$
where $h \mapsto \lambda(a)Df|_a(h) + D\lambda|_a(h)f(a)$ is a linear map, and
$$\xi(h) = \zeta(h)f(a) + Df|_a(h)D\lambda|_a(h)\frac{1}{\|h\|} + Df|_a(h)\zeta(h) + \lambda(a)\varepsilon(h) + D\lambda|_a(h)\varepsilon(h) + \varepsilon(h)\zeta(h)\|h\|$$
which we claim tends to 0 as $h \to 0$ since:

- $\varepsilon(h), \zeta(h) \to 0$ as $h \to 0$ so $\zeta(h)f(a)$, $Df|_a(h)\zeta(h)$, $\lambda(a)\varepsilon(h)$, $D\lambda|_a(h)\varepsilon(h)$ and $\varepsilon(h)\zeta(h)\|h\|$ all tend to 0 as $h \to 0$.

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

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

  as $h \to 0$.

so $\xi(h) \to 0$ as $h \to 0$. □

Start of lecture 21

Partial derivatives *can* still be useful for computation:

Proposition 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

(a) f is differentiable at a if and only if each f_i is differentiable at a, in which case

$$Df|_a = \begin{pmatrix} Df_1|_a \\ \vdots \\ Df_m|_a \end{pmatrix}$$

and

(b) 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. (a) $\Rightarrow$ 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 and $|\varepsilon_i(h)| \le \|\varepsilon(h)\| \to 0$ as $h \to 0$.

$\Leftarrow$ For each i, write

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

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

(b) 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 of $\mathbb{R}^n$. Then

$$\frac{f(a+ke_j) - f(a)}{k} = \frac{Df|_a(ke_j) + \varepsilon(ke_j)\|ke_j\|}{k} = Df|_a(e_j) + \varepsilon(ke_j) \to Df|_a(e_j)$$

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

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

Theorem 6 (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. In principle this should be obvious: if f 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 obvious thing.
Proof looks a bit messy - calculation to make sure error terms behave.

Proof. Write

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

and

$$g(f(a)+k) = g(f(a)) + \beta(k) + \eta(k)\|k\|$$

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

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

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 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 g continuous at 0. Then $\eta(\alpha(h) + \varepsilon(h)\|h\| \to 0$ as $h \to 0$. Finally,

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

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

Examples

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

(2) Suppose f is linear. Then

$$f(a+h) = f(a) + f(h) + 0\|h\|$$

so f everywhere differentiable with $Df|_a = f$ for all a.

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

(4) Using the above together with Chain Rule, get lots of differentiable functions, for example

$$f : \mathbb{R}^2 \to \mathbb{R}^2, \qquad f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} e^{x+y} \\ \cos(xy) \end{pmatrix}$$

is differentiable. Why? The projection maps $\pi_1, \pi_2 : \mathbb{R}^2 \to \mathbb{R}$, $\pi_s(x,y) = x, \pi_2(x,y) = y$ are linear so differentiable. So by Chain Rule:

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

are differentiable. So by Proposition 5(a), f is differentiable.
What is derivative of f at $z = (x, y)$? It's some linear map $\mathbb{R}^2 \to \mathbb{R}^2$. By Proposition 5(b), the matrix of the derivative is given by the partial derivatives:

$$Df|_{(x,y)} = \begin{pmatrix} e^{x+y} & e^{x+y} \\ -y \sin xy & -x \sin xy \end{pmatrix}$$

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

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

where

$$\left\| \frac{H^2}{\|H\|} \right\| \le \frac{\|H\|^2}{\|H\|} = \|H\| \to 0$$

as $H \to 0$. So f everywhere differentiable and

$$Df|_A(H) = AH + HA$$

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

$$\begin{aligned} \det(I + H) &= \begin{vmatrix} 1 + H_{11} & H_{12} & \cdots & H_{1n} & & & \\ H_{21} & 1 + H_{22} & \cdots & H_{2n} & & & \\ \vdots & \vdots & \ddots & \vdots H_{n1} & H_{n2} & \cdots & 1 + H_{nn} \end{vmatrix} \\ &= \underbrace{1}_{\det I} + \underbrace{\mathrm{Tr}(H)}_{\text{linear in } H} + \underbrace{\text{other terms involving two or more } H_{ij} \text{ multiplied together.}}_{\text{higher order}} \end{aligned}$$

Note

$$\left| \frac{H_{ij} H_{kl}}{\|H\|_2} \right| \le |H_{kl}| \to 0$$

as $H \to 0$. ($\|H\|_2$ is the Euclidean norm). So det is differentiable at I with $D \det |_i(H) = \mathrm{Tr}(H)$. Suppose $A \in \mathcal{M}_n$ invertible. Then

$$\begin{aligned} \det(A + H) &= \det(A) \det(I + A^{-1}H) \\ &= \det A(1 + \mathrm{Tr}(A^{-1}H) + \varepsilon(A^{-1}H)\|A^{-1}H\|) \\ &= \det A + (\det A)(\mathrm{Tr}\, A^{-1}H) + (\det A)\varepsilon(A^{-1}H)\|A^{-1}H\| \end{aligned}$$

where $\varepsilon(K) \to 0$ as $K \to 0$. And

$$\begin{aligned} \left| \frac{(\det A)\varepsilon(A^{-1}H)\|A^{-1}H\|}{\|H\|} \right| &\le |(\det A)\varepsilon(A^{-1}H)\|A^{-1}\|| \\ &\to 0 \end{aligned}$$

as $H \to 0$. So det differentiable at A with $D \det |_A(H) = (\mathrm{Tr}\, A^{-1}H)(\det A)$.

Start of lecture 22

Recall: If $f : \mathbb{R} \to \mathbb{R}$ differentiable with zero derivative everywhere then f is constant. This followed from the mean value theorem.

Theorem 7 (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 that

$$[a, b] = \{a + t(b - a) \mid 0 \le t \le 1\} \subset X.$$

Then

$$\|f(b) - f(a)\| \le \|b - a\| \sup_{z \in (a,b)} \|Df|_z\|$$

where $(a, b) = [a, b] \setminus \{a, b\}$.

Proof. Define $\phi : [0, 1] \to \mathbb{R}$ by $\phi(t) = f(a + t(b - a)) \cdot (f(b) - f(a))$. Then $\phi = \alpha \circ f \circ \beta$ where $\beta : [0, 1] \to \mathbb{R}^n$, $\alpha(x) = x \cdot (f(b) - f(a))$. 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

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

That is,

$$\phi'(t) = (f(b) - f(a)) \cdot Df|_{a+t(b-a)}(b - a)$$

So, by the Mean Value Theorem,

$$\begin{aligned} \|f(b) - f(a)\|^2 &= (f(b) - f(a)) \cdot f(b) - (f(b) - f(a)) \cdot f(a) \\ &= \phi(1) - \phi(0) \\ &= \phi'(t) && \text{for some } t \in (0, 1) \\ &= (f(b) - f(a)) \cdot Df|_{a+t(b-a)}(b - a) \\ &\le \|f(b) - f(a)\| \|Df|_{a+t(b-a)}(b - a)\| && \text{Cauchy Schwartz} \\ &\le \|f(b) - f(a)\| \|Df|_{a+t(b-a)}\| \|b - a\| \end{aligned}$$

Hence

$$\|f(b) - f(a)\| \le \|b - a\| \|Df|_{a+t(b-a)}\|$$

□

Corollary 8. 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 Mean Value Inequality, 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 if and only if U is open in $\mathbb{R}^n$. If $x = \emptyset$ then done. Suppose not. Fix $a \in X$. Let

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

- $U \neq \emptyset$ because $a \in U$.
- 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 $b \in X \setminus U$ then there is some $\delta > 0$ such that $B_\delta(b) \subset X$ and f constant on $B_\delta(b)$ so $B_\delta(b) \subset X \setminus U$. So $X \setminus U$ open in $\mathbb{R}^n$, so open in X. So U is closed in X.

But X is connected, so $U = X$. □

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

Theorem 9 (Continuous Partial Derivatives Implies Differentiable)**.** 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 $D_i f$ $(1 \le i \le n)$ all exist and are continuous at a. Then f is differentiable at a.

How can we prove this? For simplicity, we'll just prove this when $n = 2$, $m = 1$. So $f : \mathbb{R}^2 \to \mathbb{R}$. Write $a = (x, y)$. Want to think about $f(x + h, y + k)$ for small h, k. Now, by definition of partial derivatives,

$$f(x+h, y+k) = f(x+h, y) + kD_2 f(x+h, y) + o(k) \tag{$*$}$$

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 is nonsense. In particular, the $o(k)$ in $(*)$ 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 only know for each h, $\frac{\eta(h,k)}{k} \to 0$ as $k \to 0$, and this is weaker.
In fact, to write a proof that actually works, we need 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 $\theta_{h,k} \in (0,1)$. Again by MVT,

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

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

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

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

$$D_2 f(x+y, y+\theta_{h,k} k) \to D_2 f(x,y)$$

and

$$D_1 f(x + \phi_h h, y) \to D_1 f(x,y)$$

Write $D_2 f(x+h, y+\theta_{h,k} k) = D_2 f(x,y) + \eta(h,k)$ and $D_1 f(x+\phi_h h, y) = D_1 f(x,y) + \zeta(h,k)$ 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 $(h,k) \mapsto hD_1 f(x,y) + kD_2 f(x,y)$ 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 $a = (x,y)$. □

Remarks

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

(2) If you try to prove something like this and don't use MVT it's probably wrong.

Start of lecture 23

2. The Second Derivative

We'll start with a result in partial derivatives

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

Theorem 10 (Symmetry of second partial derivatives)**.** Let $f : \mathbb{R}^n \to \mathbb{R}^m$, $a \in \mathbb{R}^n$ and $\varepsilon > 0$. Suppose $D_i D_j f$ and $D_j D_i f$ exist on $B_\varepsilon(a)$ and are continuous at a. Then $D_i D_j f(a) = D_j D_i f(a)$.

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

$$\begin{aligned} \Delta_h &= f(x+h, y+h) - f(x, y+h) - f(x+h, y) + f(x, y) \\ &= g(y+h) - g(y) \end{aligned}$$

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

$$\begin{aligned} \Delta_h &= hg'(y + \theta_h h) && (\theta_h \in (0,1)) \\ &= h(D_2 f(x+h, y+\theta_h h) - D_2 f(x, y + \theta_h h)) \\ &= h^2 D_1 D_2 f(x + \phi_h h, y + \theta_h h) && (\phi_h \in (0,1)) \end{aligned}$$

Similarly, $\Delta_h = h^2 D_2 D_1 f(x + \zeta_h h, y + \xi_h h)$ for some $\zeta_h, \xi_h \in (0, 1)$. Hence

$$D_1 D_2 f(x + \phi_h h, y + \theta_h h) = D_2 D_1 f(x + \zeta_h h, y + \xi_h h)$$

So let $h \to 0$ and use continuity of $D_1 D_2 f$ and $D_2 D_1 f$ at (x, y)

$$D_1 D_2 f(x, y) = D_2 D_1 f(x, y)$$ □

What is the second derivative really?
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 $F : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \sim \mathbb{R}^{nm}$ by $F(x) = Df|_x$. If F is differentiable at $a \in \mathbb{R}^n$ then we say f is *twice differentiable* at a and the *second derivative* of f at a is $D^k f|_a = DF|_a$. What is $D^2 f|_a$?

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

So $D^2 f|_a$ is a bilinear map from $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$. If f twice differentiable at a, this says

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

(Everything in this expression is a linear map), i.e.

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

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

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

So f everywhere differentiable with

$$Df|_A(K) = A^2K + AKA + KA^2$$

Now

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

So f is twice differentiable at A and

$$D^2f|_A(H,k) = AHK + HAK + AKH + HKA + KAH + KHA$$

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

How does $D^2f|_a$ relate to the $D_iD_jf(a)$? Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at $a \in \mathbb{R}^n$. Then, with $e_1, \dots, e_n$ the 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 representing the bilinear form $D^2f|_a$, we have

$$H_{ij} = D_iD_jf(a)$$

We call H the *Hessian* matrix of f. If $\mathbb{R}^n \to \mathbb{R}^m$, could do this for each $f_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \dots, m)$, or think about matrices whose entries are elements of $\mathbb{R}^m$.

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

If f is twice differentiable at a then Theorem 10 tells us that H is a symmetric matrix. Hence under this condition, $D^2 f|_a$ is a symmetric bilinear form.
An application:

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

Proposition 11. Let $f : \mathbb{R}^n \to \mathbb{R}$ and let a be a local maximum / minimum for f. Suppose f differentiable at a. Then $Df|_a$ is the zero map.

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

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

as $\lambda \to 0$. Assume WLOG a is a maximum (otherwise consider $-f$). Then

$$\frac{f(a+\lambda u) - f(u)}{\lambda} \begin{cases} \ge 0 & \text{if } \lambda < 0 \\ \le 0 & \text{if } \lambda > 0 \end{cases}$$

Hence $Df|_a(u) = 0$. $\square$

Converse of course does not hold: for example $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$, $a = 0$.

Lemma 12 (Second-order Taylor Theorem)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable at $a \in \mathbb{R}^n$. Then

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

Proof. Define $g : [0,1] \to \mathbb{R}$ by

$$g(t) = f(a+th) - f(a) - tDf|_a(h) - \frac{t^2}{2} D^2 f|_a(h,h)$$

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 Mean Value Theorem, $\exists t \in (0,1)$ such that $g(1) - g(0) = g'(t)$. Hence

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

as $h \to 0$. □

Theorem 13. Let $f : \mathbb{R}^n \to \mathbb{R}$ and $a \in \mathbb{R}^n$. Suppose f is twice differentiable at a (so, in particular, $D^2 f|_a$ is a symmetric bilinear form) and $Df|_a = 0$. Then

$$D^2 f|_a \text{ positive definite} \implies a \text{ local minimum}$$

and

$$D^2 f|_a \text{ negative definite} \implies a \text{ local maximum}$$

Proof. Suppose WLOG $D^2 f|_a$ positive definite (otherwise consider $-f$). Then with respect to some orthonormal basis $D^2 f|_a$ has diagonal matrix with strictly positive elements on the leading diagonal. Have $\forall x \in \mathbb{R}^n$, $D^2 f|_a(x,x) \geq \mu\|x\|^2$ where $\mu > 0$ is the least eigenvalue of $D^2 f|_a$. By Lemma 12,

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

as $h \to 0$. But $\frac{1}{2}\mu > 0$ so for h sufficiently small,

$$\frac{f(a+h) - f(a)}{\|h\|^2} > 0$$

so $f(a+h) - f(a) > 0$ so a is a local minimum for f. □

3. Ordinary Differential Equations

Lemma 14. Let $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$ with A compact and B closed. Let $X = \mathcal{C}(A, B) = \{f : A \to B \mid A \text{ continuous}\}$ with uniform metric

$$d(f, g) = \sup_{x \subset A} \|f(x) - g(x)\|$$

Then X is a complete metric space.

Proof. As A compact, d is well-defined. Let (f_n) be a Cauchy sequence in X. Then (f_n) is uniformly Cauchy so uniformly convergent by General Principle of Uniform Convergence 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 closed, $f(x) \in B$. So $f \in X$ and $d(f_n, f) \to 0$. □

Often we want to solve an ODE but can't find a closed-form solution.

- Numerical Methods
- Phase plane portraits

But this is silly if the ODE has no solution. So want a general result telling us under appropriate conditions ODEs have unique solutions. Typical ODE: $\frac{\mathrm{d}y}{\mathrm{d}x} = \phi(x, y)$, subject to $y = y_0$ when $x = x_0$. Useful to think about things $\mathbb{R}^n \to \mathbb{R}^n$. Want to solve the initial value problem:

$$f : \mathbb{R} \to \mathbb{R}^n$$
$$f'(t) = \phi(t, f(t))$$

for all $t \in \cdots$, and $f(t_0) = y_0$.

Start of lecture 24

Notation. If $a \in \mathbb{R}^n$ and $\delta > 0$, the *closed ball of radius δ about a* is

$$\overline{B_\delta(a)} = \{\mathbb{R}^n \mid \|x - a\| \le \delta\}$$

Theorem 15 (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 $\phi : [a, b] \times \overline{B_\delta(y_0)} \to \mathbb{R}^n$ be continuous and suppose there is some $K > 0$ such that

$$\forall t \in [a, b]\ \forall y, z \in \overline{B_\delta(y_0)} \quad \|\phi(t, y) - \phi(t, z)\| \le K\|y - z\|$$

Then there is some $\varepsilon > 0$ such that $[t_0 - \varepsilon, t_0 + \varepsilon] \subset [a, b]$ and the initial value problem

$$f'(t) = \phi(t, f(t)) \quad \text{with} \quad f(t_0) = y_0 \tag{$*$}$$

has a unique solution on $[t_0 - \varepsilon, t_0 + \varepsilon]$.

Proof. As ϕ is a continuous function on a compact set so can find M such that

$$\forall t \in [a,b]\ \forall y \in \overline{B_\delta(y_0)}\quad \|\phi(t,y)\| \leq M$$

Take $\varepsilon > 0$ such that $[t_0 - \varepsilon, t_0 + \varepsilon] \subset [a,b]$. Let $X = \mathcal{C}([t_0 - \varepsilon, t_0 + \varepsilon], \overline{B_\delta(y_0)})$. Then by Lemma 14, X is complete with the uniform metric d. And obviously $X \neq \emptyset$. For $g \in X$, define $Tg : [t_0 - \varepsilon, t_0 + \varepsilon] \to \mathbb{R}^n$ by

$$Tg(t) = y_0 + \int_{t_0}^{t} \phi(x, g(x))\mathrm{d}x$$

Note that by the Fundamental Theorem of Calculus, $Tf = f$ if and only if f is a solution of $(*)$.
Now, if $g \in X$ and $t \in [t_0 - \varepsilon, t_0 + \varepsilon]$ then

$$\begin{aligned}
\|Tg(t) - y_0\| &= \left\| \int_{t_0}^{t} \phi(x, g(x))\mathrm{d}x \right\| \\
&\leq \int_{t_0}^{t} \|\phi(x, g(x))\|\mathrm{d}x \\
&\leq M\varepsilon
\end{aligned}$$

Also, if $g, h \in X$ and $t \in [t_0 - \varepsilon, t_0 + \varepsilon]$ then

$$\begin{aligned}
\|Tg(T) - Th(t)\| &= \left\| \int_{t_0}^{t} (\phi(x, g(x)) - \phi(x, h(x)))\mathrm{d}x \right\| \\
&\leq \int_{t_0}^{t} \|\phi(x, g(x)) - \phi(x, h(x))\|\mathrm{d}x \\
&\leq \int_{t_0}^{t} K\|g(x) - h(x)\|\mathrm{d}x \\
&\leq K\varepsilon d(g,h)
\end{aligned}$$

i.e. $d(Tg, Th) \leq K\varepsilon d(g,h)$. So taking $\varepsilon = \min\left\{\frac{\delta}{M}, \frac{1}{2K}\right\}$ we have that T is a contraction of X and so has a unique fixed point by Contraction Mapping Theorem as desired. $\square$

Remark. Not that much use as stated - doesn't provide a global solution. Might or might not be one. In practice, given appropriate conditions on ϕ can often 'patch together' local solutions. Beyond scope of this course.

4. The Inverse Function Theorem

Theorem 16 (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 $f|_U$, then g is differentiable at $f(a)$ with $Dg|_{f(a)} = \alpha^{-1}$.

Proof. (We won't prove the fact about differentiability of the inverse in this course). 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 $\overline{B_\delta(a)}$. Let $W = \overline{B_\delta(a)}$, $V = B_\eta(f(a))$. Define $\phi : \mathbb{R}^n \to \mathbb{R}^n$ by $\phi(x) = f(X) - \alpha(x)$. Then for $x \in W$, ϕ is differentiable at x with

$$D\phi|_x = Df|_x - \alpha \to 0$$

as $x \to a$. Note W is a complete, non-empty metric space.
Fix $y \in V$. Define $T_y : W \to \mathbb{R}^n$ by $T_y(x) = x - \alpha^{-1}(f(x) - y)$. Note $f(x) = y \iff T_y(x) = x$. Now, given $x \in W$,

$$\begin{aligned}
\|T_y a\| &= \|\alpha^{-1}(\alpha x - f(x) + y - \alpha(a)\| \\
&= \|\alpha^{-1}(y - f(x) + \alpha(x-a))\| \\
&= \|\alpha^{-1}(y - f(a) - \varepsilon(x-a)\|x-a\|)\| \\
&\le \|\alpha^{-1}\|(\|y - f(a)\| + \|\varepsilon(x-a)\|\|x-a\|) \\
&\le \|\alpha^{-1}\|(\eta + \delta\|\varepsilon(x-a)\|)
\end{aligned}$$

Also, given $w, x \in W$,

$$\begin{aligned}
\|T_y x - T_y w\| &= \|\alpha^{-1}(\alpha x - f(x) + f(w) - \alpha(w))\| \\
&= \|\alpha^{-1}(\phi(w) - \phi(x))\| \\
&\le \|\alpha^{-1}\|\|\phi(w) - \phi(x)\| \\
&\le \|\alpha^{-1}\|\|w - x\| \sup_{z\in W} \|D\phi|_z\|
\end{aligned}$$

by Mean Value Inequality. Pick $\delta > 0$ sufficiently small such that $\forall x \in \overline{B_\delta(a)}$ we have $\|\varepsilon(x-a)\| < \frac{1}{2\|\alpha^{-1}\|}$ and also $\sup_{z\in W} \|D\phi|_z\| < \frac{1}{\alpha^{-1}}$. (Can do this since $\varepsilon(x-a) \to 0$ as $x \to a$ and $D\phi|_x \to 0$ as $x \to a$.). Take $y = \frac{\delta}{2}$. Then for each $y \in V$ we have $\forall x \in W \|T_y x - a\| < \delta$ and $\forall x, w \in W \quad \|T_y x - T_y w\| \le K\|w - x\|$ where $K < 1$ is a constant. So T_y is a contraction of W and thus by Contraction Mapping Theorem has

a unique fixed point, $x_y \in T_y(W) \subset B_\delta(a)$. 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 have open neighbourhoods U of a and V of $f(a)$ such that f maps U bijectively onto V.
Remains to show inverse function is continuous. Let $X = \mathcal{C}(V, W)$. As W is bounded, similarly to Lemma 14 we have X is a complete, non-empty metric space with the uniform metric. Define $S : X \to X$ by

$$\begin{aligned}(Sg)(y) &= g(y) - \alpha^{-1}(f(g(y)) - y)\\ &= T_y(g(y))\end{aligned}$$

Given $g, h \in X$ and $y \in V$,

$$\begin{aligned}\|(Sg)(y) - (Sh)(y)\| &= \|T_y(g(y)) - T_y(h(y))\|\\ &\leq K\|g(y) - h(y)\|\\ &\leq Kd(g, h)\end{aligned}$$

So $d(Sg, Sh) \leq Kd(g, h)$. Now S is a contraction of X so by Contraction Mapping Theorem has a unique fixed point g. And by definition of S, for each $y \in V$, we have $g(y)$ is the unique $x \in W$ with $f(x) = y$. Hence $g = (f|_U)^{-1}$. But $g \in X$ so $(f|_U)^{-1}$ is continuous and thus $f|_U$ is a homeomorphism from U onto V. $\square$