Analysis & Topology Lecture Notes

Conor Rajan

Michaelmas 2020

Course Outline

1. Uniform convergence and uniform continuity
2. Metric spaces
3. Completeness and the Contraction Mapping Theorem
4. Topological spaces
5. Connectedness
6. Compactness
7. Differentiation and the Inverse Function Theorem

Pre-requisites: Analysis I
Books: Burkill & Burkill, Sutherland
Examples Sheets: Hard questions are **not exam questions**!!!

1 Uniform convergence and uniform continuity

Recall: in $\mathbb{R}$ or $\mathbb{C}$ we write $x_n \to x$ as $n \to \infty$ if

$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n \geq N \ |x_n - x| < \varepsilon$$

Aim: to define $f_n \to f$ for functions f_n and f

Definition. We have a set S and functions $f_n : S \to \mathbb{R}$, $n \in \mathbb{N}$ and $f : S \to \mathbb{R}$. Then $f_n \to f$ *pointwise* on S as $n \to \infty$ if for every $x \in S$, the real sequence $(f_n(x))_n$ converges to $f(x)$. In symbols:

$$\forall x \in S \ \forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n \geq N \ |f_n(x) - f(x)| < \varepsilon$$

Remarks:

1. N can depend on ε and x

2. Can replace $\mathbb{R}$ with $\mathbb{C}$
3. It is possible that $f_n \to f$ pointwise on some subset $T \subseteq S$.

Examples:

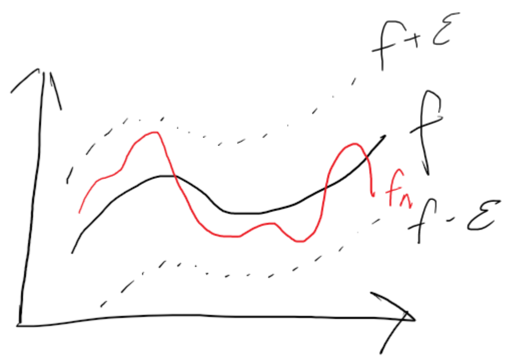
1. $f_n(x) = x^n$ for $x \in [0,1], n \in \mathbb{N}$ For $0 \leq x < 1$, we have $f_n(x) = x^n \to 0$ as $n \to \infty$. Also, $f_n(1) = 1 \to 1$ as $n \to \infty$. So $f_n \to f$ pointwise on $[0,1]$, where
$$f(x) = \begin{cases} 0 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x = 1 \end{cases}$$
2. $f_n(x) = x^2 e^{-nx}$ for $x \in [0, \infty)$, and $n \in \mathbb{N}$. For $x > 0$,
$$0 \leq f_n(x) = \frac{x^2}{e^{nx}} = \frac{x^2}{1 + nx + \frac{n^2x^2}{2} + \ldots} \leq \frac{x^2}{nx} = \frac{x}{n}$$
Hence $f_n(x) \to 0$ as $n \to \infty$. Same is true $x = 0$. Thus, $f_n \to 0$ pointwise on $[0, \infty)$

Remark: The pointwise limit of a sequence of functions (f_n) is unique if it exists.

Definition. We are given a set S and functions $f_n : S \to \mathbb{R}$, $n \in \mathbb{N}$ and $f : S \to \mathbb{R}$. Then $f_n \to f$ *uniformly* on S as $n \to \infty$ if
$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall n \geq N \; \forall x \in S \; |f_n(x) - f(x)| < \varepsilon$$

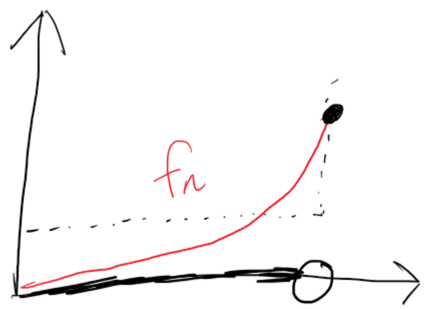
Remarks:

- N depends *only* on ε
- Uniform convergence implies pointwise convergence
- Can replace $\mathbb{R}$ with $\mathbb{C}$ and can restrict to a subset of the domain
- An equivalent definition of $f_n \to f$ uniformly on S:
$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall n \geq N \; \sup_{x \in S} |f_n(x) - f(x)| < \varepsilon$$
Even shorter: $\sup_{x \in S} |f_n(x) - f(x)| \to 0$ as $n \to \infty$



Examples:

1. $f_n(x) = x^n$ for $x \in [0,1]$, $n \in \mathbb{N}$ We know that $f_n \to f$ pointwise on $[0,1]$, where $f(x) = \begin{cases} 0 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x = 1 \end{cases}$



Let $\varepsilon = 1/2$. For any $n \in \mathbb{N}$, setting $x = \varepsilon^{1/n}$, we have $f_n(x) = \varepsilon$, and so $|f_n(x) - f(x)| \geq \varepsilon$. So $f_n \not\to f$ uniformly on $[0,1]$.

Better: Since $f_n(1) = 1$ and f_n is continuous, there exists $\delta > 0$ such taht $|f_n(x) - 1| < 1/2$ for $x \in (1-\delta, 1+\delta)$. So for any $x \in [0,1]$ with $1 - \delta < x < 1$, we have $|f_n(x) - f(x)| \geq \varepsilon$

2. $f_n(x) = x^2 e^{-nx}$ for $x \in [0, \infty)$, and $n \in \mathbb{N}$. We saw that $f_n \to f = 0$ pointwise on $[0, \infty)$.

$$0 \leq f_n(x) = \frac{x^2}{e^{nx}} = \frac{x^2}{1 + nx + \frac{n^2x^2}{2} + \dots} \leq \frac{x^2}{n^2x^2/2} = \frac{2}{n^2}$$

Thus,

$$\sup_{x \in [0,\infty)} |f_n(x) - f(x)| = \sup_{x \in [0,\infty)} f_n(x) \leq \frac{2}{n^2} \to 0 \text{ as } n \to \infty$$

So $f_n \to 0$ uniformly on $[0, \infty)$

Remark: Could have used differentiation above to find the supremum, but the above method of finding an upper bound is better.

Question: Does (f_n) converge uniformly on S?

Strategy:

- First check if (f_n) converges pointwise on S.
- If it doesn't, then (f_n) doesn not converge uniformly on S

- If it does, then compute the pointwise limit f, and then it remains to check whether
$$\sup_{x \in S} |f_n(x) - f(x)| \to 0 \text{ as } n \to \infty$$
- If the above holds, then $f_n \to f$ uniformly on S, otherwise (f_n) does not converge uniformly on S.

Remark: What does it mean that $f_n \not\to f$ uniformly on S? We have to negate the sentence

$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n \geq N \ \forall x \in S \ |f_n(x) - f(x)| < \varepsilon$$

The negation is

$$\exists \varepsilon > 0 \ \forall N \in \mathbb{N} \ \exists n \geq N \ \exists x \in S \ |f_n(x) - f(x)| \geq \varepsilon$$

Theorem 1.1. *Let S be a subset of $\mathbb{R}$ or $\mathbb{C}$. Assume $f_n \to f$ uniformly on S. if f_n is continuous for every $n \in \mathbb{N}$, then f is continuous.*

Idea: Given $a \in S$, we want that $f(x) \approx f(a)$ provided $x \approx a$. We choose large n so that $f_n(x) \approx f(x)$ for every $x \in S$. Since f_n is continuous, we have $f(x) \approx f_n(a)$ provided $x \approx a$. Thus, if $x \approx a$, then $f(x) \approx f_n(x) \approx f_n(a) \approx f(a)$.

Proof. Fix $a \in S$ and $\varepsilon > 0$. We seek $\delta > 0$ such that

$$\forall x \in S \ |x - a| < \delta \implies |f(x) - f(a)| < \varepsilon$$

Since $f_n \to f$ uniformly on S, we may fix $n \in \mathbb{N}$ so that

$$\forall x \in S \ |f_n(x) - f(x)| < \varepsilon$$

Since f_n is continuous, there exists $\delta > 0$ such that

$$\forall x \in S \ |x - a| < \delta \implies |f_n(x) - f_n(a)| < \varepsilon$$

Thus, for any $x \in S$, if $|x - a| < \delta$, then

$$|f(x) - f(a) \leq |f(x) - f_n(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)| < 3\varepsilon$$

$\square$

Remarks:

1. The result does not extend to differentiability
2. It follows from Theorem 1 that x^n does not converge uniformly on $[0, 1]$
3. The above proof is sometimes called a 3ε-proof
4. $\lim\limits_{x \to a} \lim\limits_{n \to \infty} f_n(x) = \lim\limits_{x \to a} f(x) = f(a) = \lim\limits_{n \to \infty} f_n(a) = \lim\limits_{n \to \infty} \lim\limits_{x \to a} f_n(x)$

Lemma 1.2. *Let S be any set and f_n be a bounded function on S for every $n \in \mathbb{N}$. If $f_n \to f$ uniformly on S, then f is also bounded on S.*

Proof. Fix $n \in \mathbb{N}$ so that $|f_n(x) - f(x)| \leq 1$ for every $x \in S$. We can do this, since $f_n \to f$ uniformly on S. Since f_n is bounded, there is an $M \in \mathbb{R}$ such that $|f_n(x)| \leq M$ for every $x \in S$. It follows that for every $x \in S$, we have

$$|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| \leq 1 + M$$

□

Before the next theorem, we recall some definitions and results from Analysis I. Assume $f : [a, b] \to \mathbb{R}$ is a bounded function. Given a dissection $\mathcal{D} : a = x_0 < x_1 < \ldots < x_n = b$ of $[a, b]$, the upper and lower sums of f with respect to $\mathcal{D}$ are defined as

$$U_{\mathcal{D}}(f) = \sum_{k=1}^{n} (x_k - x_{k-1}) \sup_{[x_{k-1}, x_k]} f$$

and

$$L_{\mathcal{D}}(f) = \sum_{k=1}^{n} (x_k - x_{k-1}) \inf_{[x_{k-1}, x_k]} f$$

Riemann's criterion states that f is integrable iff for every $\varepsilon > 0$ there is a dissection $\mathcal{D}$ of $[a, b]$ such that

$$U_{\mathcal{D}}(f) - L_{\mathcal{D}}(f) = \sum_{k=1}^{n} (x_k - x_{k-1}) (\sup_{[x_{k-1}, x_k]} - \inf_{[x_{k-1}, x_k]}) < \varepsilon$$

An easy exercise shows that for an interval $I \subset [a, b]$ we have

$$\sup_I f - \inf_i f = \sup_{x,y \in I} (f(x) - f(y)) = \sup_{x,y \in I} |f(x) - f(y)|$$

(This quantity is sometimes called the *ocscillation* of f on I)

Theorem 1.3. *Assume $f_n : [a, b] \to \mathbb{R}$ is Riemann-integrable for every $n \in \mathbb{N}$. If $f_n \to f$ uniformly on $[a, b]$, then f is also Riemann-integrable on $[a, b]$ and moreover*

$$\int_a^b f_n \to \int_a^b f \text{ as } n \to \infty$$

Proof. We are going to prove that f is bounded and that it satisfies Riemann's criterion. By definition of integrability, each f_n is bounded, and hence so is f by Lemma 2.

Next, fix $\varepsilon > 0$. Since $f_n \to f$ uniformly on $[a, b]$, we can fix $n \in \mathbb{N}$ so that $|f_n(x) - f(x)| < \varepsilon$ for all $x \in [a, b]$. Since f_n is integrable, it satisfies Riemann's criterion, so there is a dissection $\mathcal{D}$ of $[a,b]$ such that $U_{\mathcal{D}}(f_n) - L_{\mathcal{D}}(f_n) < \varepsilon$. If I is one of the sub-intervals of $\mathcal{D}$, then for any $x, y \in I$ we have

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &< |f_n(x) - f_n(y)| + 2\varepsilon \end{aligned}$$

It follows that

$$\sup_{x,y\in I} |f(x) - f(y)| \leq \sup_{x,y\in I} |f_n(x) - f_n(y)| + 2\varepsilon$$

Multiplying both sides with the length of I and summing over all subintervals I of $\mathcal{D}$, we obtain

$$U_{\mathcal{D}}(f) - L_{\mathcal{D}}(f) \leq U_{\mathcal{D}}(f_n) - L_{\mathcal{D}}(f_n) + 2\varepsilon(b-a) < (2(b-a)+1)\varepsilon$$

So f satisfies Riemann's criterion, and thus, f is integrable.

Finally, we estimate

$$\left| \int_a^b f_n - \int_a^b f \right| \leq \int_a^b |f_n - f| \leq (b-a) \sup_{[a,b]} |f_n - f| \to 0 \text{ as } n \to \infty$$

□

Remark: The conclusion of Theorem 3 says $\displaystyle\int_a^b \lim_{n\to\infty} f_n(x)\mathrm{d}x = \lim_{n\to\infty} \int_a^b f_n(x)\mathrm{d}x$

Corollary 1.4. *Let $f_n : [a,b] \to \mathbb{R}$ be an integrable function for each $n \in \mathbb{N}$. If $\sum_{n=1}^\infty f_n(x)$ converges uniformly on $[a,b]$, then $\sum_{n=1}^\infty f_n(x)$ defines an integrable function on $[a,b]$ and moreover*

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

Proof. Define $F_n(x) = \sum_{k=1}^n f_k(x)$ for $x \in [a,b]$ and $n \in \mathbb{N}$. To say that $\sum_{n=1}^\infty f_n(x)$ converges uniformly on $[a,b]$ means that (F_n) converges uniformly on $[a,b]$. So for each $x \in [a,b]$, the series $\sum_{n=1}^\infty f_n(x)$ is convergent, and the function $x \mapsto \sum_{n=1}^\infty f_n(x)$ is the uniform limit of (F_n) on $[a,b]$.

We know that each F_n is integrable and $\int_a^b F_n = \sum_{k=1}^n \int_a^b f_k$. So by Theorem 3, the function $\sum_{n=1}^\infty f_n(x)$ is integrable and

$$\int_a^b \sum_{k=1}^\infty f_n(x)\mathrm{d}x = \lim_{n\to\infty} \int_a^b F_n(x)\mathrm{d}x = \lim_{n\to\infty} \sum_{k=1}^n \int_a^b f_k(x)\mathrm{d}x = \sum_{k=1}^\infty \int_a^b f_k(x)\mathrm{d}x$$

□

Theorem 1.5. *Let (f_n) be a sequence of continuously differentiable functions on $[a, b]$. Assume further that*

(i) $\sum_{n=1}^{\infty} f_n'(x)$ converges uniformly on $[a, b]$

(ii) There exists $x \in [a, b]$ such that $\sum_{n=1}^{\infty} f_n(c)$ converges

Then $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly to a continuously differentiable function f on $[a, b]$ and moreover we have

$$f'(x) = \sum_{n=1}^{\infty} f_n'(x) \text{ for all } x \in [a, b]$$

Remark: Informally

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

Proof. Let $g(x) = \sum_{n=1}^{\infty} f_n'(x)$ for $x \in [a, b]$. Idea: solve the equation $f' = g$ with initial condition $f(c) = \sum_{n=1}^{\infty} f_n(c)$.

Since $\sum_{n=1}^{\infty} f_n'(x)$ converges uniformly to $g(x)$, and since f_n' is continuous for every $n \in \mathbb{N}$, by Theorem 1, g is continuous, and hence integrable. Let $\lambda = \sum_{n=1}^{\infty} f_n(c)$ and define

$$f(x) = \lambda + \int_c^x g(t)\mathrm{d}t \text{ for } x \in [a, b]$$

Since g is continuous, by the FTC f is differentiable with $f' = g$ and moreover $f(c) = \lambda$. By the FTC we also have

$$f_k(x) = f_k(c) + \int_c^x f_k'(t)\mathrm{d}t \text{ for } x \in [a, b], k \in \mathbb{N}$$

Given $\varepsilon > 0$, by our assumptions there exists $N \in \mathbb{N}$ such that

$$\left|\lambda - \sum_{k=1}^{n} f_k(c)\right| < \varepsilon \ \forall n \geq N$$

$$\left|g(t) - \sum_{k=1}^{n} f_k'(t)\right| < \varepsilon \ \forall n \geq N \ \forall t \in [a, b]$$

It follows that for all $n \geq N$ and for all $x \in [a,b]$ we have

$$\begin{aligned}\left|f(x) - \sum_{k=1}^{n} f_k(x)\right| &= \left|\lambda + \int_c^x g(t)\mathrm{d}t - \sum_{k=1}^{n}\left(f_k(c) + \int_c^x f_k'(t)\mathrm{d}t\right)\right| \\ &= \left|\lambda - \sum_{k=1}^{n} f_k(c) + \int_c^x \left(g(t) - \sum_{k=1}^{n} f_k'(t)\right)\mathrm{d}t\right| \\ &\leq \left|\lambda - \sum_{k=1}^{n} f_k(c)\right| + \left|\int_c^x \left(g(t) - \sum_{k=1}^{n} f_k'(t)\right)\mathrm{d}t\right| \\ &< \varepsilon + (b-a)\varepsilon\end{aligned}$$

This shows that $\sum_{k=1}^{n} f_k(x) \to f(x)$ uniformly on $[a,b]$. We have already seen that f is differentiable and $f' = g$ is continuous.

□

We recall from Analysis I: a scalar sequence (x_n) is Cauchy if

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall m, n \geq N\ |x_m - x_n| < \varepsilon$$

We proved the General Principle of Convergence (GPC): every Cauchy sequence is convergent.

Definition. Let (f_n) be a sequence of scalar functions on a set S. We say (f_n) is *uniformly Cauchy* on S if

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall m, n \geq N\ \forall x \in S\ |f_m(x) - f_n(x)| < \varepsilon$$

Theorem 1.6 (General Principle of Uniform Convergence (GPUC))**.** *If (f_n) is a uniformly Cauchy sequence of scalar functions on a set S, then (f_n) converges uniformly to some function f on S.*

Proof. We first show that (f_n) converges pointwise on S. Fix $x \in S$. Given $\varepsilon > 0$, since (f_n) is uniformly Cauchy, there exists $N \in \mathbb{N}$ such that

$$\forall m, n \geq N\ \forall t \in S\ |f_m(t) - f_n(t)| < \varepsilon$$

In particular for all $m, n \geq N$, we have $|f_m(x) - f_n(x)| < \varepsilon$. Thus, $(f_n(x))_{n=1}^{\infty}$ is a Cauchy sequence and hence convergent by the GPC. Set $f(x) = \lim_{n\to\infty} f_n(x)$.

Doing this for every $x \in S$, we obtain a function f on S.

We claim that $f_n \to f$ uniformly on S. Then we will be done.

Given $\varepsilon > 0$, since (f_n) is uniformly Cauchy, there exists $N \in \mathbb{N}$ such that

$$\forall m, n \geq N\ \forall x \in S\ |f_m(x) - f_n(x)| < \varepsilon$$

Fix $n \geq N$ and $x \in S$. Since $|f_m(x) - f_n(x)| < \varepsilon$ for every $m \geq N$, letting $m \to \infty$ we obtain $|f(x) - f_n(x)| \leq \varepsilon$. Since this holds for every $n \geq N$ and for every $x \in S$, we are done. □

Corollary 1.7 (The Weierstrass M-test)**.** *Let (f_n) be a sequence of scalar functions on a set S. Let $\sum_n M_n$ be a convergent series of non-negative real numbers. Assume that $|f_n(x)| \leq M_n$ for every $x \in S$ and $n \in \mathbb{N}$. Then $\sum_n f_n(x)$ converges uniformly on S.*

Proof. Set $F_n(x) = \sum_{k=1}^{n} f_k(x)$ for $x \in S$ and $n \in \mathbb{N}$. For $n \geq m$ and $x \in S$ we have

$$|F_m(x) - F_n(x)| \leq \left| \sum_{k=m+1}^{n} f_k(x) \right| \leq \sum_{k=m+1}^{n} |f_k(x)| \leq \sum_{k=m+1}^{n} M_k$$

Given $\varepsilon > 0$, choose $N \in \mathbb{N}$ such that $\sum_{k=N}^{\infty} M_k < \varepsilon$. Then by the above, for every $x \in S$ and every $n \geq m \geq N$ in $\mathbb{N}$, we have

$$|F_m(x) - F_n(x)| \leq \sum_{k=m+1}^{n} M_k < \varepsilon$$

So (F_n) is uniformly Cauchy, and hence uniformly convergent by Theorem 6. $\square$

Consider a power series $\sum_{n=0}^{\infty} c_n(z-a)^n$. Here $(c_n)_{n=0}^{\infty}$ is a complex sequence, $a, z \in \mathbb{C}$. We think of (c_n) and a as fixed, and z as a variable.

Let R be the radius of convergence of this power series. This means:

$$|z-a| < R \implies \sum_{n=0}^{\infty} c_n(z-a)^n \text{ converges absolutely}$$

$$|z-a| > R \implies \sum_{n=0}^{\infty} c_n(z-a)^n \text{ diverges}$$

Denote $D(a, R) = \{z \in \mathbb{C} : |z-a| < R\}$, the open disk of centre a and radius R. Consider the function $f : D(a, R) \to \mathbb{C}$ defined by

$$f(z) = \sum_{n=0}^{\infty} c_n(z-a)^n$$

The power series converges pointwise to f.

Question: Does the power series converge uniformly inside the radius of convergence?

Answer: In general, NO.

Examples:

1. $\sum_{n=1}^{\infty} \frac{z^n}{n^2}$ for $|z| < 1$. Consider $f_n : D(0,1) \to \mathbb{C}$ defined by $f_n(z) = z^n/n^2$. Note that $|f_n(z)| \leq 1/n^2$ for all $|z| < 1$. Moreover, $\sum_{n=1}^{\infty} \frac{1}{n^2}$. Hence by the Weierstrass M-test, the power series converges uniformly on $D(0,1)$.

2. $\sum_{n=0}^{\infty} z^n = \frac{1}{1-z}$ for $|z| < 1$.

   $\left|\sum_{n=0}^{N} z^n\right| \leq N+1$ for $|z| < 1$ and $N \in \mathbb{N}$. So the partial sum functions are bounded on $D(0,1)$. However $\frac{1}{1-z}$ is not bounded on $D(0,1)$. By Lemma 2, the convergence cannot be uniform on $D(0,1)$.

   Alternatively,

$$\sup_{|z|<1} \left|\sum_{n=0}^{N} z^n - \frac{1}{1-z}\right| = \sup_{|z|<1} \left|\frac{z^{N+1}}{1-z}\right| = \infty$$

Theorem 1.8. *Let $\sum_{n=0}^{\infty} c_n(z-a)^n$ be a power series with radius of convergence R. Then for any r with $0 < r < R$, the power series $\sum_{n=0}^{\infty} c_n(z-a)^n$ converges uniformly on $D(a, r) = \{z \in \mathbb{C} : |z-a| < r\}$.*

Proof. Fix $w \in D(a,R)$ with $r < |w-a| < R$ (e.g $w = a + \frac{r+R}{2}$). Set $\rho = \frac{r}{|w-a|}$.

Since $\sum_{n=0}^{\infty} c_n(w-a)^n$ converges, we must have $c_n(w-a)^n \to 0$ as $n \to \infty$. It follows that the sequence $(c_n(w-a)^n)_{n=0}^{\infty}$. Thus, there exists $M \geq 0$ such that $|c_n(w-a)^n| \leq M$ for all $n \geq 0$.

Now, for any $z \in D(a,r)$ and $n \in \mathbb{N}$, we have

$$|c_n(z-a)^n| = |c_n(w-a)^n| \cdot \left|\frac{(z-a)^n}{(w-a)^n}\right| \leq M \frac{r^n}{|(w-a)|^n} = M\rho^n$$

Since $\sum_{n=0}^{\infty} M\rho^n$ converges, by Corollary 6 (the Weierstrass M-test) the power series $\sum_{n=0}^{\infty} c_n(z-a)^n$ converges uniformly on $D(a,r) = \{z \in \mathbb{C} : |z-a| < r\}$. $\square$

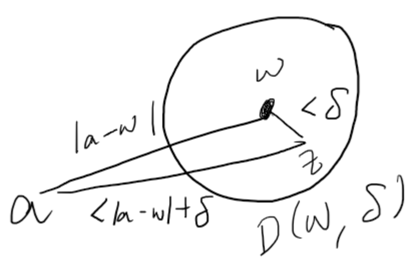
Remarks:

1. The function $f : D(a,R) \to \mathbb{C}, f(z) = \sum_{n=0}^{\infty} c_n(z-a)^n$ is continuous on $D(a,r)$ for any $0 < r < R$ being the uniform limit of continuous functions (polynomials). This is by Theorem 1. Since $D(a,R) = \bigcup_{0<r<R} D(a,r)$, f is continuous on $D(a,R)$

2. The power series $\sum_{n=1}^{\infty} n \cdot c_n(z-a)^{n-1}$ also has radius of convergence R (from Analysis I). Hence it converges uniformly on $D(a,r)$ for any $0 < r < R$. By an analogue of Theorem 5 we can deduce that f is complex differentiable on $D(a,R)$ and

$$f'(z) = \frac{\mathrm{d}}{\mathrm{d}z}\left(\sum_{n=0}^{\infty} c_n(z-a)^n\right) = \sum_{n=1}^{\infty} n \cdot c_n(z-a)^{n-1}$$

3. Given $w \in D(a,R)$, fix r with $|w-a| < r < R$ and $\delta > 0$ with $|w-a| + \delta < r$. Then $D(w,\delta) \subset D(a,r)$. Indeed, given $z \in D(w,\delta)$



$$|z-a| \leq |z-w| + |w-a| < \delta + |w-a| < r$$

So $z \in D(a,r)$. It follows that the power series $\sum_{n=0}^{\infty} c_n(z-a)^n$ converges uniformly on $D(w,\delta)$.

Definition. A subset U of $\mathbb{C}$ is *open* if for every $w \in U$ there is a $\delta > 0$ such that $D(w,\delta) \subset U$.

Definition. Let U be an open subset of $\mathbb{C}$. A sequence (f_n) of scalar functions on U converges *locally uniformly* on U if for every $w \in U$ there is a $\delta > 0$ such that $D(w,\delta) \subset U$ and $f_n \to f$ uniformly on $D(w,\delta)$.

Remarks:

1. We will return to this concept after covering compactness
2. Above we proved that a power series converges locally uniformly inside the radius of convergence.

We next turn to the topic of Uniform Continuity. We begin by recalling the notion of continuity from Analysis I.

Let U be a subset of $\mathbb{R}$ or $\mathbb{C}$ and let f be a scalar function on U. For $x \in U$, we say f is continuous at x if

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

We say f is continuous on U if f is continuous at x for every $x \in U$, i.e.

$$\forall x \in U \; \forall \varepsilon > 0 \; \exists \delta > 0 \; \forall y \in Y \; |y - x| < \delta \implies |f(y) - f(x)| < \varepsilon$$

Note that δ depends on ε and x.

Definition. Let U be a subset of $\mathbb{R}$ or $\mathbb{C}$ and let f be a scalar function on U. We say f is *uniformly continuous* on U if

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

Note δ depends only on ε. So uniform continuity implies continuity.

Examples:

1. Consider $f : \mathbb{R} \to \mathbb{R}, f(x) = -3x + 7$. Given $\varepsilon > 0$, set $\delta = \varepsilon/3$. Then for every $x, y \in \mathbb{R}$, if $|x - y| < \delta$, then

   $$|f(x) - f(y)| = 3|x - y| < 3\delta = \varepsilon$$

   So f is uniformly continuous on $\mathbb{R}$

2. Consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. Given $\varepsilon = 1$, we try to find $\delta > 0$ that works. Consider any $x > 0$ and $y = x + \delta/2$. Then $|x - y| = \delta/2 < \delta$ and

   $$|f(x) - f(y)| = (x + \delta/2)^2 - x^2 = x\delta + \delta^2/4$$

   So for any $\delta > 0$, if $x = 1/\delta$ and $y = x + \delta/2$, then $|x - y| < \delta$ but $|f(x) - f(y)| > \varepsilon$. So f is not uniformly continuous, but f is continuous on $\mathbb{R}$.

Note: A function f on a subset U of $\mathbb{R}$ or $\mathbb{C}$ is **not** uniformly continuous if

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

is **false**, that is

$$\exists \varepsilon > 0 \ \forall \delta > 0 \ \exists x, y \in U \ |x - y| < \delta \text{ and } |f(x) - f(y)| \geq \varepsilon$$

Theorem 1.9. *Let f be a scalar function on a closed, bounded interval $[a, b]$. If f is continuous on $[a, b]$, then f is uniformly continuous on $[a, b]$.*

One idea: Let $\varepsilon > 0$. For each $x \in [a, b]$ there is a $\delta_x > 0$ such that if $|y - x| < \delta_x$ then $|f(y) - f(x)| < \varepsilon$. We want to take $\delta = \inf\limits_x \delta_x$. However, we might have $\delta = 0$. Instead, we work indirectly: argue by contradiction.

Proof. Assume there is an $\varepsilon > 0$ such that

$$\forall \delta > 0 \ \exists x, y \in [a, b] \ |x - y| < \delta \text{ and } |f(x) - f(y)| \geq \varepsilon$$

In particular, for every $n \in \mathbb{N}$, there exists $x_n, y_n \in [a, b]$ such that $|x_n - y_n| < 1/n$ and $|f(x_n) - f(y_n)| \geq \varepsilon$. By Bolzano-Weierstrass there is a subsequence $(x_{k_n})_n$ of (x_n) that converges to some $x \in [a, b]$. Then

$$|y_{k_n} - x| \leq |y_{k_n} - x_{k_n}| + |x_{k_n} - x| \leq 1/n + |x_{k_n} - x| \to 0$$

Since f is continuous at x then $f(x_{k_n}) \to f(x)$ and $f(y_{k_n}) \to f(x)$. Hence $\varepsilon \leq |f(x_{k_n}) - f(y_{k_n})| \to |f(x) - f(x)| = 0$, which is a contradiction. □

Corollary 1.10. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then f is integrable.*

Proof. f is bounded since it is continuous on the closed bounded interval $[a, b]$. It remains to check Riemann's criterion.

Given $\varepsilon > 0$, since f is uniformly continuous by Theorem 9, there is a $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$. Next, choose a dissection $\mathcal{D}$ of $[a, b]$ such that every subinterval of $\mathcal{D}$ has length strictly less than δ. If I is a subinterval of $\mathcal{D}$, then for all $x, y \in I$, we have $|x - y| < \delta$, and hence $|f(x) - f(y)| < \varepsilon$. It follows that

$$\sup_I f - \inf_I f = \sup_{x, y \in I} |f(x) - f(y)| \leq \varepsilon$$

Multiplying both sides of the lenth of I, and summing over all I, we obtain

$$U_{\mathcal{D}}(f) - L_{\mathcal{D}}(f) \leq (b - a)\varepsilon$$

□

2 Metric Spaces

In $\mathbb{R}$ or $\mathbb{C}$, we measure the "closeness" of points x and y by the expression $|x-y|$. A frequently used property of this "distance" is the triangle-inequality. We will now generalize this concept.

Definition. Let M be a set. A *metric* on M is a function $d : M \times M \to \mathbb{R}$ satisfying

(i) $\forall x, y \in M \ d(x, y) \geq 0$ and $d(x, y) \iff x = y$ (positivity)

(ii) $\forall x, y \in M \ d(x, y) = d(y, x)$ (symmetry)

(iii) $\forall x, y, z \in M \ d(x, z) \leq d(x, y) + d(y, z)$ (triangle-inequality)

A *metric space* is a pair (M, d) where M is a set and d is a metric on M.

Examples:

1. $M = \mathbb{R}$ or $\mathbb{C}$, $d(x, y) = |x - y|$. We refer to this as the usual metric on M and will always be used unless otherwise stated.

2. $M = \mathbb{R}^n$ or $\mathbb{C}^n$. We define the *euclidean norm* or the *euclidean length* of a vector $x \in M$ by

$$||x|| = ||x||_2 = \left(\sum_{k=1}^{n} |x_k|^2\right)^{1/2}$$

   This satisfies the inequality $||x + y|| \leq ||x|| + ||y||$. It follows that

$$d(x, y) = d_2(x, y) = ||x - y|| = \left(\sum_{k=1}^{n} |x_k - y_k|^2\right)^{1/2}$$

   defines a metric on M called the *euclidean metric*. E.g, we check the triangle-inequality:

$$d(a, c) = ||a - c|| = ||a - b + b - c|| \leq ||a - b|| + ||b - c|| = d(a, b) + d(b, c)$$

   The resulting metric space (M, d) is called *n-dimensional real or complex euclidean space*. M will always be equipped with the euclidean metric unless otherwise stated. The metric space (M, d) is sometimes denoted by l_2^n and d_2 is also called the l_2-metric, and $|| \cdot ||_2$ is also called l_2-norm.

3. $M = \mathbb{R}^n$ or $\mathbb{C}^n$. We define the l_1-norm of a vector $x \in M$ by

$$||x||_1 = \sum_{k=1}^{n} |x_k|$$

and the corresponding metric, called the l_1-metric, is given by

$$d_1(x,y) = ||x-y||_1 = \sum_{k=1}^{n} |x_k - y_k|$$

The metric space (M, d_1) is denoted by l_1^n. It is not hard to see how to generalize this further. For any $1 \leq p < \infty$ one can define the l_p-norm by

$$||x||_p = \left(\sum_{k=1}^{n} |x_k|^p \right)^{1/p}$$

and the corresponding l_p-metric by $d_p(x,y) = ||x-y||_p$. In this course, we will only deal with $p = 1, 2$. (For the general case, see Part II Linear Analysis). How about $p = \infty$?

4. Letting $p \to \infty$ in the previous example leads to the l_∞-norm and to the l_∞-metric on $M = \mathbb{R}^n$ or $\mathbb{C}^n$:

$$||x||_\infty = \max_{1 \leq k \leq n} |x_k| \text{ and } d_\infty(x,y) = ||x-y||_\infty = \max_{1 \leq k \leq n} |x_k - y_k|$$

   The metric space (M, d_∞) is denoted by l_∞^n.

5. Let S be a set. We denote by $l_\infty(S)$ the set of all bounded scalar functions on S. We define the l_∞-norm (or *uniform norm* or *sup norm*) of a function $f \in l_\infty(S)$ by

$$||f|| = ||f||_\infty = \sup\{|f(x)| : x \in S\}$$

   (The sup exists since f is bounded). Note that for $f, g \in l_\infty(S)$ and for any $x \in S$, we have

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq ||f|| + ||g||$$

   and hence $||f+g|| \leq ||f|| + ||g||$. It follows that $d(f,g) = ||f-g||_\infty$ defines a metric, called the *uniform metric* (or l_∞-metric) on $l_\infty(S)$. Note that $l_\infty^n = l_\infty(\{1, 2, \ldots, n\})$. Also, $l_\infty(\mathbb{N})$ is often denoted simply l_∞. This is the space of bounded scalar sequences.

6. $C[a,b]$ denotes the space of continuous scalar functions on the closed bounded interval $[a,b]$. Let $p = 1$ or 2. We define the *L_p-norm* on $C[a,b]$ by

$$||f||_p = \left(\int_a^b |f(t)|^p \mathrm{d}t \right)^{1/p} \quad (f \in C[a,b])$$

   The corresponding metric $d_p(f,g) = ||f-g||_p$ is the *L_p-metric* on $C[a,b]$. E.g for $f, g \in C[a,b]$ we have

$$||f+g||_1 = \int_a^b |f+g| \leq \int_a^b (|f| + |g|) = ||f||_1 + ||g||_1$$

   which implies the triangle-inequality for d_1.

7. Let M be any set. For $x, y \in M$ we define

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

This is called the *discrete metric* on M and (M, d) is a *discrete metric space*.

8. Let G be a group generated by $S \subseteq G$. Then

$$d(x,y) = \min\{n : \exists s_1, \ldots, s_n \in S \ y = xs_1s_2\ldots s_n\}$$

is a metric on G (Geometric Group Theory).

9. Fix a prime p in $\mathbb{Z}$. For $x, y \in \mathbb{Z}$ write $x - y = p^n m$ where $m, n \in \mathbb{Z}$, $n \geq 0$ and $p \not| m$; then define

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ p^{-n} & \text{if } x \neq y \end{cases}$$

This is called the *p-adic metric* (Number Theory). This is in fact an *ultrametric* which means that for any x, y, z the following holds:

$$d(x,y) \leq \max\{d(x,y), d(y,z)\}$$

which implies the triangle-inequality. A set equipped with an ultrametric is called an *ultrametric space*.

2.1 Subspaces

Let (M, d) be a metric space and let $N \subseteq M$. Then $d_{N\times N}$ (the restriction of d to $N \times N$) is a metric on N.

N with this metric is called a *subspace* of (M, d). With a slight abuse of notation, we write d for both the metric on M and the metric $d_{N\times N}$ on N.

Examples

1. $\mathbb{Q}$ with metric $d(x,y) = |x - y|$ is a subspace of $\mathbb{R}/$

2. Since every continuous function on a closed bounded interval is bounded, $C[a,b]$ is a subset of $l_\infty([a,b])$. So $C[a,b]$ with the uniform metric is a subspace of $l_\infty([a,b])$.

2.2 Products

Let (M, d) and (M', d') be metric spaces. Then any of the following define a metric on $M \times M'$:

- $d_1((x,x'),(y,y')) = d(x,y) + d'(x',y')$
- $d_2((x,x'),(y,y')) = (d(x,y)^2 + d'(x',y')^2)^{1/2}$
- $d_\infty((x,x'),(y,y')) = \max\{d(x,y), d'(x',y')\}$

The metric space $(M \times M', d_p)$ is denoted $M \oplus_p M'$ $(p = 1, 2, \infty)$. This can be generalized to the product of any finite number of metric spaces.

Note that $d_\infty \leq d_2 \leq d_1 \leq 2d_\infty$.

Examples:

1. $\mathbb{R} \oplus_1 \mathbb{R} = l_1^2$
2. $\mathbb{R} \oplus_2 \mathbb{R} \oplus_2 \mathbb{R} = l_2^3$
3. $\underbrace{\mathbb{R} \oplus_\infty \ldots \oplus_\infty \mathbb{R}}_{n \text{ times}} = l_\infty^n$

However $\mathbb{R} \oplus_1 \mathbb{R} \oplus_2 \mathbb{R}$ makes no sense as $(\mathbb{R} \oplus_1 \mathbb{R}) \oplus_2 \mathbb{R}$ and $\mathbb{R} \oplus_1 (\mathbb{R} \oplus_2 \mathbb{R})$ are different.

2.3 Convergence

Let M be a metric space and let (x_n) be a sequence in M. Given $x \in M$, we say (x_n) *converges* to x and write $x_n \to x$ as $n \to \infty$ if

$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall n \geq N \; d(x_n, x) < \varepsilon$$

We say (x_n) is *convergent* in M if there is an $x \in M$ such that (x_n) converges to x. We say (x_n) is *divergent* if it is not convergent.

Note $x_n \to x \iff d(x_n, x) \to 0$

Lemma 2.1 (Uniqueness of limit). *Assume that $x_n \to x$ and $x_n \to y$ in a metric space M. Then $x = y$*

Proof. Suppose not. Set $\varepsilon = d(x,y)/3$. Then $\varepsilon > 0$ so we can choose $N_1, N_2 \in \mathbb{N}$ such that

$$\forall n \geq N_1 \; d(x_n, x) < \varepsilon \text{ and } \forall n \geq N_1 \; d(x_n, y) < \varepsilon$$

Fix any $n \geq \max\{N_1, N_2\}$. Then we have

$$d(x,y) \leq d(x, x_n) + d(x_n, y) < 2\varepsilon < d(x,y)$$

which is a contradiction. □

Definition. Given a convergent sequence (x_n) in a metric space M, the *limit* of (x_n) (denoted by $\lim\limits_{n\to\infty} x_n$) is the unique $x \in M$ such that $x_n \to x$ as $n \to \infty$.

Examples:

1. In $\mathbb{R}$ and $\mathbb{C}$ convergence has the usual meaning.

2. Constant sequences converge. More generally, assume (x_n) is an eventually constant sequence in a metric space M, i.e there exists $x \in M$ such that $x_n = x$ for all n sufficiently large. Then $x_n \to x$ as $n \to \infty$. The converse is clearly false: e.g $1/n \to 0$ in $\mathbb{R}$.

   However, assume $x_n \to x$ is a discrete metric space. Then there is an $N \in \mathbb{N}$ such that $d(x_n, x) < 1$ for $n \geq N$, and hence $x_n = x$ for $n \geq N$.

3. In the 3-adic metric on $\mathbb{Z}$, $3^n \to 0$ as $n \to \infty$ since $d(3^n, 0) = 3^{-n} \to 0$ as $n \to \infty$.

4. Let S be a set. In $l_\infty(S)$, $f_n \to f$ in the uniform metric if and only if $d(f_n, f) = ||f_n - f||_\infty = \sup_S |f_n - f| \to 0$ as $n \to \infty$. This is equivalent to saying that $f_n \to f$ uniformly on S.

   Note that if $f_n(x) = x + \frac{1}{n}$ and $f(x) = x$ for $x \in \mathbb{R}$ and $n \in \mathbb{N}$, then $f_n \to f$ uniformly on $\mathbb{R}$, but $f_n, f \notin l_\infty(\mathbb{R})$.

5. Consider $\mathbb{R}^{\mathbb{N}}$, the set of all real sequences. Can check that for sequences $x = (x_k)$ and $y = (y_k)$,

$$d(x, y) = \sum_{k=1}^{\infty} 2^{-k} \min\{1, |x_k - y_k|\}$$

defines a metric on $\mathbb{R}^{\mathbb{N}}$. Then a sequence $(x^{(n)})$ in $\mathbb{R}^{\mathbb{N}}$ converges to $x \in \mathbb{R}^{\mathbb{N}}$ if and only if $\mathbf{x}_k^{(n)} \to x_k$ as $n \to \infty$ for each $k \in \mathbb{N}$. $(x^{(n)} = (x_k^n)_{k=1}^{\infty})$

Question: Given any set S, is there a metric on $\mathbb{R}^{\mathbb{S}}$ such that convergence in the metric is equivalent to pointwise convergence on S?

6. Consider $f_n(x) = x^n$ for $x \in [0, 1]$ and $n \in \mathbb{N}$. Then (f_n) is a sequence in $C[0, 1]$. Recall that (f_n) converges pointwise but not uniformly on $[0, 1]$. Thus, (f_n) does not converge in the uniform metric. However,

$$d_1(f_n, 0) = ||f_n||_1 = \int_0^1 |f_n| = \frac{1}{n+1} \to 0$$

and so $f_n \to 0$ in $C[0, 1]$ in the L_1-metric.

7. Let M be a metric space, N a subspace of M and (x_n) a sequence in N. If (x_n) is convergent in N, then (x_n) is convergent in M. The converse is false. E.g $1/n \to 0$ in $\mathbb{R}$, but $(1/n)$ does not converge in $(0, \infty)$.

8. Let M and M' be metric spaces and consider $N = M \oplus_p M'$ where $p = 1, 2$ or ∞. Let (a_n) be a sequence in N and write $a_n = (x_n, x'_n)$ where $x_n \in M$ and $x'_n \in M'$. Let $a = (x, x') \in N$. Then

$$a_n \to a \text{ in } N \iff x_n \to x \text{ in } M \text{ and } x'_n \to x' \text{ in } M'$$

This follows easily from the following:

$$\begin{aligned} \max\{d(x_n, x), d'(x'_n, x')\} &= d_\infty(a_n, a) \\ &\leq d_p(a_n, a) \\ &\leq d_1(a_n, a) \\ &= d(x_n, x) + d'(x'_n, x') \end{aligned}$$

E.g in $\mathbb{R}^n$, we have $v_i = (x_{i,1}, \ldots, x_{i,n}) \to v = (x_1, \ldots x_n)$ iff $x_{i,k} \to x_k$ as $i \to \infty$ for each $k = 1, \ldots, n$.

2.4 Continuity

Consider a function $f : M \to M'$ between metric spaces M and M'. For $a \in M$ we say f is *continuous* at a if

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in M \; d(x, a) < \delta \implies d'(f(x), f(a)) < \varepsilon$$

We say f is *continuous* if f is continuous at a for every $a \in M$.

Note: δ depends on ε and a.

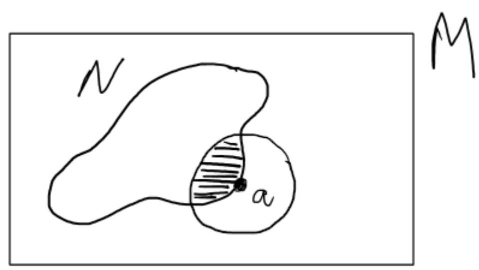
More generally, for a subset $N \subseteq M$ we say f is *continuous* on N if f is continuous at every $a \in N$.

Note: f being continuous on N is related to but is different from $f_N : N \to M'$ being continuous. The former implies the latter, but the converse fails. f continuous on N means:

$$\forall a \in N \; \forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in M \; d(x,a) < \delta \implies d'(f(x), f(a)) < \varepsilon$$

$f_N : N \to M'$ continuous means:

$$\forall a \in N \; \forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in N \; d(x,a) < \delta \implies d'(f(x), f(a)) < \varepsilon$$



Example:
$M = \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$

$N = [0, \infty)$ and then $f|_N : [0, \infty) \to \mathbb{R}$, $x \mapsto 1$, is a constant function, so clearly continuous. However, f is not continuous on N since f is not continuous at $0 \in N$.

Proposition 2.2. *Let $f : M \to M'$ be a function between metric spaces and let $a \in M$. The following are equivalent:*

(i) f is continuous at a

(ii) whenever $x_n \to a$ in M, we have $f(x_n) \to f(a)$ in M'

So if f is continuous, then for every convergent sequence (x_n) in M, $(f(x_n))$ is convergent in M', and moreover $\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n)$

Proof. To show (i) $\implies$ (ii): given $\varepsilon > 0$, since f is continuous at a, there is a $\delta > 0$ such that $d'(f(x), f(a)) < \varepsilon$ for any $x \in M$ with $d(x,a) < \delta$. If $x_n \to a$, there is an $N \in \mathbb{N}$ such that $d(x_n, a) < \delta$ for $n \geq N$. It follows that

$d'(f(x_n), f(a)) < \varepsilon$ for $n \geq N$. Thus $f(x_n) \to f(a)$ as $n \to \infty$.

Now show (ii) $\implies$ (i): assume f is not continuous at a. This means

$$\exists \varepsilon > 0\ \forall \delta > 0\ \exists x \in M\ d(x, a) < \delta \text{ and } d'(f(x), f(a)) \geq \varepsilon$$

Fix such a "bad" ε. Then for every $n \in \mathbb{N}$ there is an $x_n \in M$ such that $d(x_n, a) < 1/n$ and $d'(f(x_n), f(a)) \geq \varepsilon$ (applying the above statement with $\delta = 1/n$). So we constructed a sequence (x_n) such that $x_n \to a$ but $f(x_n) \not\to f(a)$. $\square$

Corollary 2.3. *Let f and g be scalar functions on a metric space M. Let $a \in M$ and assume f, g are continuous at a. Then $f + g$ and fg are also continuous at a.*
Moreover, setting $N = \{x \in M : g(x) \neq 0\}$ and assuming $a \in N$, the function $f/g : N \to M'$ is also continuous at a. It follows that if f, g are continuous, then so are $f + g$, fg and f/g.

Proof. Assume $x_n \to a$ in M. Since f, g are continuous at a, it follows by Proposition 2.2 that $f(x_n) \to f(a)$ and $g(x_n) \to g(a)$. Hence,

$$(f + g)(x_n) = f(x_n) + g(x_n) \to f(a) + g(a) = (f + g)(a)$$

Using Proposition 2.2 the second time, we deduce $f + g$ is continuous at a. The argument for fg and f/g is similar. $\square$

Proposition 2.4. *Let $f : M \to M'$ and $g : M' \to M''$ be functions between metric spaces and let $a \in M$. If f is continuous at a and g is continuous at $f(a)$, then the composition $g \circ f : M \to M''$ is continuous at a. It follows that if f, g are continuous, so is $g \circ f$.*

Proof. Given $\varepsilon > 0$, since g is continuous at $f(a)$, there is a $\delta > 0$ such that

$$\forall y \in M'\ d'(y, f(a)) < \delta \implies d''(g(y), g(f(a))) < \varepsilon$$

Since f is continuous at a, there is an $\eta > 0$ such that

$$\forall x \in M\ d(x, a) < \eta \implies d'(f(x), f(a)) < \delta$$

Hence, if $x \in M$ and $d(x, a) < \eta$, then $d''(g(f(x)), g(f(a))) < \varepsilon$. $\square$

Examples:

1. Constant functions are continuous: $f : M \to M'$, $f(x) = b$ for all $x \in M$. Indeed, $d'(f(x), f(a)) = d'(b, b) = 0$ for any $x, a \in M$. So given $a \in M$ and $\varepsilon > 0$, any $\delta > 0$ will do.

2. Identity functions are continuous: $f : M \to M$, $f(x) = x$ for every $x \in M$. Indeed, $d(f(x), f(a)) = d(x, a)$ for every $x, a \in M$. Given $a \in M$ and $\varepsilon > 0$, $\delta = \varepsilon$ will do.

3. Combining the previous examples with Corollary 2.3, we deduce that real and complex polynomials are continuous, as are rational functions. Using results on uniform convergence, uniform limits of such functions are also continuous, e.g. the exponential function.

4. Let (M, d) be a metric space. The metric is itself a function between metric spaces: $d : M \oplus_p M \to \mathbb{R}$ where $p = 1, 2$ or ∞. For $\mathbf{x} = (x, x')$ and $\mathbf{y} = (y, y')$ in $M \oplus_p M$ we have

$$\begin{aligned} |d(\mathbf{x}) - d(\mathbf{y})| &= |d(x, x') - d(y, y')| \leq d(x, y) + d(x', y') \\ &= d_1(\mathbf{x}, \mathbf{y}) \leq 2d_p(\mathbf{x}, \mathbf{y}) \end{aligned}$$

   It follows that d is continuous ($\delta = \varepsilon/2$ will do).

Definition. A map $f : M \to M'$ between metric spaces is

(i) *Isometric* if $d'(f(x), f(y)) = d(x, y)$ for all $x, y \in M$

(ii) *Lipschitz* if there is a $C \geq 0$ such that $f'(f(x), f(y)) \leq Cd(x, y)$ for all $x, y \in M$. We also say that f is *C-Lipschitz*.

(iii) *Uniformly continuous* if

$$\forall \varepsilon > 0 \; \exists \delta \; \forall x, y \in M d(x, y) < \delta \implies d'(f(x), f(y)) < \varepsilon$$

Note:

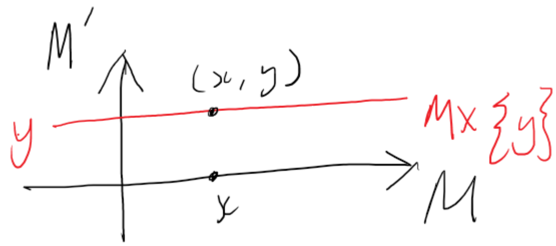
1. isometric $\implies$ Lipschitz $\implies$ uniformly continuous $\implies$ continuous

2. An isometric map is injective but not necessarily surjective. An isometric map that is also surjective is called an *isometry*. The inverse of an isometry is also an isometry. If there is an isometry $M \to M'$, we say that M and M' are *isometric* metric spaces or that M' is an *isometric copy* of M.

Examples (continued):

5. Let M, M' be metric spaces. Fix $y \in M'$ and consider the function

$$f : M \to M \oplus_p M', f(x) = (x, y)$$



For $x, z \in M$, we have $d_p(f(x), f(z)) = d_p((x, y), (z, y)) = d(x, z)$, and thus f is isometric, and $M \times \{y\}$ is an isometric copy of M.

This generalizes to any finite number of metric spaces. For example, given $a \in \mathbb{R}^n$, the map $x \mapsto (a_1, \ldots, a_{k-1}, x, a_{k+1}, \ldots, a_n) : \mathbb{R} \to \mathbb{R}^n$ is isometric to $\mathbb{R}$.

6. Let M, M' be metric spaces. Consider the functions

$$q : M \oplus_p M' \to M, (x, x') \mapsto x \text{ and } q' : M \oplus_p M' \to M', (x, x') \mapsto x'$$

For $\mathbf{x} = (x, x')$ and $\mathbf{y} = (y, y')$ in $M \oplus_p M'$, we have

$$d(q(\mathbf{x}), q(\mathbf{y})) = d(x, y) \leq \max\{d(x, y), d'(x', y')\} = d_\infty(\mathbf{x}, \mathbf{y}) \leq d_p(\mathbf{x}, \mathbf{y})$$

and so q is 1-Lipschitz. The same holds for q'.

E.g. $\mathbb{C}^n \to \mathbb{C}$, $(z_1, \ldots, z_n) \mapsto z_k$ is continuous. Using Corollary 2.3 again, we deduce that polynomials in any number of variables are also continuous.

2.5 The topology of metric spaces

In a metric space, continuity at a point x or convergence of a sequence to a point x depends on the set of points close to x. This motivates the following:

Definition. Let (M, d) be a metric space and $x \in M$. We define the *open ball* in M with *centre* x and *radius* r (the *open r-ball around* x) as the set

$$D_r(x) = \{y \in M : d(y, x) < r\}$$

We sometimes write $D_r^M(x)$ to indicate the underlying metric space M.

Note:

$$x_n \to x \text{ in } M \iff \forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n \geq N\ x_n \in D_\varepsilon(x)$$

$$f : M \to M' \text{ is continuous at } x \iff \forall \varepsilon > 0\ \exists \delta > 0\ f(D_\delta(x)) \subseteq D_\varepsilon(f(x))$$

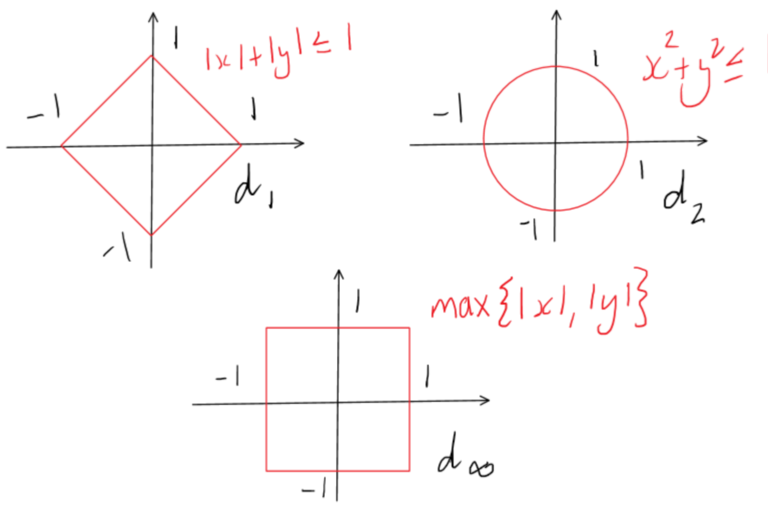
Definition. We define the *closed ball* in M with *centre* x and *radius* r (the *closed* r*-ball around* x) as the set

$$B_r(x) = \{y \in M : d(y,x) \leq r\}$$

We sometimes write $B_r^M(x)$ to indicate the underlying metric space M.

Examples:

1. In $\mathbb{R}$ we have $D_r(x) = (x-r, x+r)$ and $B_r(x) = [x-r, x+r]$.
2. In $\mathbb{C}$, $D_r(x)$ and $B_r(x)$ are the open and closed discs with centre x and radius r.
3. In $\mathbb{R}^2$ the following are pictures of $B_1(0)$ (the *closed unit ball*).



4. If M is a discrete metric space, then $D_1(x) = \{x\}$ and $B_1(x) = M$.

Note:

1. $D_r(x) \subset B_r(x) \subset D_s(x)$ whenever $r < s$
2. $d(x,y) = \min\{r \geq 0 : y \in B_r(x)\} = \inf\{r > 0 : y \in D_r(x)\}$

Definition. Let M be a metric space and $U \subseteq M$. Given $x \in M$, we say that U is a *neighbourhood* of x in M if

$$\exists r > 0 \; D_r(x) \subseteq U \text{ or equivalently } \exists r > 0 \; B_r(x) \subseteq U$$

We say U is *open* in M (or is an *open subset* of M) if

$$\forall x \in U \; \exists r > 0 \; D_r(x) \subseteq U \text{ or equivalently } \forall x \in U \; \exists r > 0 \; B_r(x) \subseteq U$$

Thus, U is open iff U is a neighbourhood of all its points.

Note: For $x \in M$ and $r > 0$, the balls $D_r(x)$ and $B_r(x)$ are both neighbourhoods of x.
Example: Consider $H = \{z \in \mathbb{C} : \mathrm{Im}(z) \geq 0\} \subset \mathbb{C}$. If $\mathrm{Im}(w) > 0$, then $D_\delta(w) \subset H$ for $\delta = \mathrm{Im}(w)$, so H is a neighbourhood of w. If $\mathrm{Im}(w) = 0$, then $w - \frac{\delta}{2}i \in D_\delta(w) \setminus H$ for any $\delta > 0$, and so H is not a neighbourhood of w.

Lemma 2.5. *Open balls are open.*

Proof. Let M be a metric space, $x \in M$ and $r > 0$. We show that $D_r(x)$ is an open set in M. To do this, we fix $y \in D_r(x)$, and then seek $\delta > 0$ such that $D_\delta(y) \subset D_r(x)$. We show that $\delta = r - d(y, x)$ works. We first observe that $\delta > 0$. Next, for any $x \in D_\delta(y)$, we have

$$d(z, x) \leq d(z, y) + d(y, x) < \delta + d(y, x) = r$$

and thus $z \in D_r(x)$. It follows that $D_\delta(y) \subset D_r(x)$, and hence $D_r(x)$ is open. $\square$

Corollary 2.6. *Let U be a subset of a metric space M and let $x \in M$. Then U is a neighbourhood of x iff there is an open subset V of M such that $x \in V \subseteq U$.*

Proof. If U is a neighbourhood of x, then by definition $D_r(x) \subseteq U$ for some $r > 0$. So taking $V = R_r(x)$, we have $x \in V \subseteq U$, and moreover V is open by Lemma 2.5.
Conversely, if $x \in V \subseteq U$ for some open subset V of M, then by definition of open set, there exists $r > 0$ such that $D_r(x) \subseteq V$. It follows that $D_r(x) \subseteq U$, and thus U is a neighbourhood of x. $\square$

Proposition 2.7. *In a metric space M the following are equivalent*

(i) $x_n \to x$

(ii) $\forall$ *neighbourhoods* U *of* x *in* M $\exists N \in \mathbb{N}\ \forall n \geq N\ x_n \in U$

(iii) $\forall$ *open sets* U *in* M *with* $x \in U$ $\exists N \in \mathbb{N}\ \forall n \geq N\ x_n \in U$

Proof. Let us first recall that (i) can be written as follows:

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n \geq N\ x_n \in D_\varepsilon(x)$$

To show (i) $\implies$ (ii): Given a neighbourhood U of x, there is an $\varepsilon > 0$ such that $D_\varepsilon(x) \subseteq U$. Since $x_n \to x$, there is an $N \in \mathbb{N}$ such that for all $n \geq N$, we have $x_n \in D_\varepsilon(x) \subseteq U$.

To show (ii) $\implies$ (iii): This is clear, since any open set U with $x \in U$ is, by definition, a neighbourhood of x.

To show (iii) $\implies$ (i): This is also clear, since for any $\varepsilon > 0$, the ball $D_\varepsilon(x)$ is an open set containing x. $\square$

Proposition 2.8. *Let $f : M \to M'$ be a function between metric spaces.*

(a) For $x \in M$ the following are equivalent:

- *(i) f is continuous at x*
- *(ii) $\forall$ neighbourhoods V of $f(x)$ in M' $\exists$ neighbourhood U of x in M with $f(U) \subseteq V$*
- *(iii) $\forall$ neighbourhoods V of $f(x)$ in M' $f^{-1}(V)$ is a neighbourhood of x in M*

(b) The following are equivalent:

- *(i) f is continuous*
- *(ii) $\forall$ open sets V in M' the set $f^{-1}(V)$ is open in M (The inverse image of open set is open).*

Proof. We begin with part (a). We first recall that (i) is the following statement

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; f(D_\delta(x)) \subseteq D_\varepsilon(f(x))$$

To show (i) $\implies$ (ii): Given a neighbourhood V of $f(x)$ in M', there is an $\varepsilon > 0$ such that $D_\varepsilon(f(x)) \subseteq V$. Since f is continuous at x, there is a $\delta > 0$ such that $f(D_\delta(x)) \subseteq D_\varepsilon(f(x))$. Set $U = D_\delta(x)$. Then U is a neighbourhood of x in M and $f(U) \subseteq D_\varepsilon(f(x)) \subseteq V$.

To show (ii) $\implies$ (iii): Given a neighbourhood V of $f(x)$ in M', by the assumption of (ii) there is a neighbourhood U of x in M such that $f(U) \subseteq V$. By definition of neighbourhood, there is a $r > 0$ such that $D_r(x) \subseteq U$. It follows that $D_r(x) \subseteq U \subseteq f^{-1}(V)$, and hence $f^{-1}(V)$ is a neighbourhood of x in M.

To show (iii) $\implies$ (i): Given $\varepsilon > 0$, the set $V = D_\varepsilon(f(x))$ is a neighbourhood of $f(x)$ in M', and hence $f^{-1}(V)$ is a neighbourhood of x in M by assumption. By definition, there is a $\delta > 0$ such that $D_\delta(x) \subseteq f^{-1}(V)$. It follows that $f(D_\delta(x)) \subseteq V = D_\varepsilon(f(x))$.

We now deduce part (b).

To show (i) $\implies$ (ii): Let V be an open set in M'. Fix $x \in f^{-1}(V)$. Then $f(x) \in V$, and so V is a neighbourhood of $f(x)$ in M'. Since f is continuous at x, it follows from (a) that $f^{-1}(V)$ is a neighbourhood of x in M. This holds for any $x \in f^{-1}(V)$, and thus $f^{-1}(V)$ is open in M.

To show (ii) $\implies$ (i): Let $x \in M$ and let $\varepsilon > 0$. Since $V = D_\varepsilon(f(x))$ is open in M', it follows that $f^{-1}(V)$ is open in M by assumption. Since $f(x) \in V$, we have $x \in f^{-1}$. Thus there is a $\delta > 0$ such that $D_\delta(x) \subseteq f^{-1}(V)$. It follows that $f(D_\delta(x)) \subseteq V = D_\varepsilon(f(x))$. This shows that f is continuous at every $x \in M$, and thus f is continuous. $\square$

Definition. The *topology* of a metric space M is the family of all open subsets of M.

Proposition 2.9. *The topology of a metric space M satsifies the following:*

(i) $\emptyset$ *and* M *are open*

(ii) If U_i *is open for all* $i \in I$ *(I some index set) then* $\bigcup_{i \in I} U_i$ *is open*

(iii) If U *and* V *are open, then* $U \cap V$ *is open.*

Proof. Part (i) is trivial. To see (ii), fix $x \in \bigcup_{i \in I} U_i$. Then for some $j \in I$, we have $x \in U_j$. Since U_j is open, there exists $r > 0$ such that $D_r(x) \subseteq U_j \subseteq \bigcup_{i \in I} U_i$.

Finally, to prove (iii), fix $x \in U \cap V$. Since U and V are open, there exists $r > 0$ and $s > 0$ such that $D_r(x) \subseteq U$ and $D_s(x) \subseteq V$. Then $t = \min\{r, s\}$ is strictly positive and

$$D_t(x) = D_r(x) \cap D_s(x) \subseteq U \cap V$$

It follows that $U \cap V$ is open. □

Definition. Let M be a metric space and $A \subseteq M$. We say A is *closed* in M (or that A is a *closed subset* of M) if for every sequence (x_n) in A that converges in M, we have $\lim_{n\to\infty} x_n$ is in A.

Lemma 2.10. *Closed balls are closed.*

Proof. Let M be a metric space, $z \in M$, $r > 0$ and $x_n \in B_r(z)$ for every $n \in \mathbb{N}$. Assume $x_n \to x$ in M. Then $d(x, z) \leq d(x, x_n) + d(x_n, z) \leq d(x, x_n) + r \to r$ as $n \to \infty$. It follows that $d(x, z) \leq r$ and hence $x \in B_r(z)$. $\square$

Examples: We work in $\mathbb{R}$.

1. $[0, 1] = B_{1/2}(1/2)$ is closed
2. $(0, 1) = D_{1/2}(1/2)$ is open
3. $(0, 1]$ is neither open nor closed. Indeed $(1/n)$ is a sequence in $(0, 1]$ converging to $0 \notin (0, 1]$, and thus $(0, 1]$ is not closed. Also, for any $\delta > 0$ we have $D_\delta(1) \not\subseteq (0, 1]$, so $(0, 1]$ is not open.

Lemma 2.11. *Let M be a metric space and $A \subseteq M$. Then A is closed in M iff $M \setminus A$ is open in M.*

Proof. Assume that A is closed and that $x \in M \setminus A$. We seek $r > 0$ such that $D_r(x) \subseteq M \setminus A$. If no such r exists, then $D_{1/n} \cap A \neq \emptyset$ for every $n \in \mathbb{N}$. So there exists $x_n \in A$ with $d(x_n, x) < 1/n$ for every n. So we found a sequence (x_n) in A such that $x_n \to x$. Since $x \notin A$, this contradicts the assumption that A is closed. So there must be an $r > 0$ such that $D_r(x) \subseteq M \setminus A$.

Conversely, assume $M \setminus A$ is open and that A is not closed. Thus, there is a sequence (x_n) in A such that $x_n \to x$ in M and $x \notin A$. Since $M \setminus A$ is open, there exists $r > 0$ such that $D_r(x) \subseteq M \setminus A$. Since $x_n \to n$, there exists $N \in \mathbb{N}$ such that $x_n \in D_r(x) \subseteq M \setminus A$ for all $n \geq N$. This contradicts the assumption that (x_n) is a sequence in A. So in fact A must be closed. $\square$

Example: Let M be a discrete metric space and let $A \subseteq M$. For any $x \in A$, we have $D_1(x) = \{x\}$, and hence A is open. Thus, every subset of M is open in M. By Lemma 11, it follows that every subset of M is closed in M.

Recall: a map $f : M \to M'$ between metric spaces is an isometry if it is bijective and isometric ($d'(f(x), f(y)) = d(x, y)$ for all $x, y \in M$). This is the correct notion of a structure-preserving map for metric spaces *provided* we are interested in metric properties. What if we are only interested in the topology that comes from the metric?

Definition. A function $f : M \to M'$ between metric spaces is a *homeomorphism* if f is a bijection and both f and f^{-1} are continuous. If there is a homeomorphism $M \to M'$, then M and M' are said to be *homeomorphic*.

Example: $(0, \infty)$ and $(0, 1)$ are homeomorphic: consider $x \mapsto \frac{1}{x+1}$ and $x \mapsto \frac{1}{x} - 1$.

Remarks:

1. Every isometry is a homeomorphism. Converse is false
2. Isometric spaces are homeomorphic. Converse is false
3. The identity map on a metric space is a homeomorphism, the invese of a homeomorphism is a homeomorphism and the composite of homeomorphisms is a homeomorphism. Hence being homeomorphic is an equivalence relation.
4. Let $f : M \to M'$ be a homeomorphism. If U is an open subset of M, then $f(U) = (f^{-1})^{-1}(U)$ is open in M', and converseluy, if V is an open subset of M', then $f^{-1}(V)$ is open in M. So we have a bijection between the topologies of M and M'.
5. Let $f : M \to M'$ be a homeomorphism. Then (x_n) is convergent in M iff $(f(x_n))$ is convergent in M'. Also $g : N \to M$ is continuous iff $f \circ g$ is continuous, and $h : N \to M'$ is continuous iff $f^{-1} \circ h$ is continuous.
6. A continuous bijection need not be a homeomorphism. E.g $\text{id} : (\mathbb{R}, \text{discrete}) \to (\mathbb{R}, \text{euclidean})$

Definition. Let d and d' be metrics on a set M. We say d and d' are *equivalent*, and write $d \sim d'$, if (M, d) and (M, d') have the same open sets (i.e the same topology).

Remark: Each of the following are equivalent to $d \sim d'$

1. The identity map $\text{id} : (M, d) \to (M, d')$ is a homeomorphism. Note that this is stronger than (M, d) and (M, d') are homeomorphic.
2. (M, d) and (M, d') have the same convergent sequences.
3. For every metric space N and every $f : N \to M$, f is continuous with respect to d iff it is continuous with respect to d'.
4. For every metric space N and every $f : M \to N$, f is continuous with respect to d iff it is continuous with respect to d'.

Definition. Let d and d' be metrics on a set M.

We say d and d' are *uniformly equivalent* (and write $d \sim_{\text{u}} d'$) if the identity maps $\text{id} : (M, d) \to (M, d')$ and $\text{id} : (M, d') \to (M, d)$ are uniformly continuous.

Definition. Let d and d' be metrics on a set M.

We say d and d' are *Lipschitz equivalent* (and write $d \sim_{\text{Lip}} d'$) if the identity maps $\text{id} : (M, d) \to (M, d')$ and $\text{id} : (M, d') \to (M, d)$ are Lipschitz maps, i.e there exist $a > 0$ and $b > 0$ such that

$$a \cdot d(x, y) \leq d'(x, y) \leq b \cdot d(x, y) \text{ for all } x, y \in M$$

Examples:

1. Given a metric space (M, d), $d'(x, y) = \min\{1, d(x, y)\}$ defines a uniformly equivalent metric on M.
2. Given metric spaces M and M', the metrics d_1, d_2 and d_∞ on $M \times M'$ are pairwise Lipschitz equivalent.
3. The uniform metric and the L_1-metric on $C[0, 1]$ are not equivalent.
4. The usual metric and the discrete metric on $\mathbb{R}$ are not equivalent.

3 Completeness and the Contraction Mapping Theorem

Recall that in $\mathbb{R}$ we have the GPC: a Cauchy sequence (x_n) is convergent. We also have the notion of a bounded sequence: (x_n) is bounded if there exists $C \geq 0$ such that $|x_n| \leq C$ for all $n \in N$.

Definition. Let (x_n) be a sequence in a metric space M. We say (x_n) is a *Cauchy sequence* if

$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall m, n \geq N \ d(x_m, x_n) < \varepsilon$$

We say (x_n) is *bounded* if

$$\exists z \in M \ \exists r > 0 \ \forall n \in \mathbb{N} \ x_n \in B_r(z)$$

Note: If $w \in M$ and $R = r + d(z, w)$, then $B_r(z) \subseteq B_R(w)$.

It follows that if $M = \mathbb{R}^n, \mathbb{C}^n$ of $C[a, b]$ with the l_p-metric or the L_p-metric, as appropriate, and $||\cdot||$ denotes the corresponding norm, then (x_n) is bounded iff there exists $r > 0$ such that $||x_n|| \leq r$ for all $n \in N$.

Lemma 3.1. *convergent* $\implies$ *Cauchy* $\implies$ *bounded*

Proof. Let (x_n) be a sequence in a metric space M. For the first implication, assume (x_n) is convergent in M and let x denote its limit. Given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon/2$ for all $n \geq N$. Then for $m, n \geq N$ we have

$$d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \varepsilon$$

Thus, (x_n) is Cauchy.

For the second implication, assume (x_n) is Cauchy. Then there exists $N \in \mathbb{N}$ such that $d(x_m, x_n) \leq 1$ for all $m, n \geq N$. It follows that $x_n \in B_1(x_N)$ for all $n \geq N$. Now let

$$r = \max\{d(x_1, x_N), d(x_2, x_N), \ldots, d(x_{N-1}, x_N, 1\}$$

Then $x_n \in B_r(x_N)$ for all $n \in \mathbb{N}$. $\square$

Remarks:

1. Bounded $\not\Rightarrow$ Cauchy, e.g $0, 1, 0, 1, 0, 1, \ldots$ in $\mathbb{R}$
2. Cauchy $\not\Rightarrow$ convergent, e.g $(1/n)$ in $(0, \infty)$.

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

Examples: $\mathbb{R}$ and $\mathbb{C}$ are complete.

Proposition 3.2. *If M and M' are complete metric spaces, then $M \oplus_p M'$ is also complete for $p = 1, 2, \infty$.*

Proof. Let (a_n) be a Cauchy sequence in $M \oplus_p M'$ where $a_n = (x_n, x'_n)$ for $n \in \mathbb{N}$. Given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $d_p(a_m, a_n) < \varepsilon$ for all $m, n \geq N$. It follows that

$$d(x_m, x_n) \leq d_p(a_m, a_n) < \varepsilon \quad \forall m, n \geq N$$

and hence (x_n), and by a similar argument (x'_n) is Cauchy. Since M and M' are complete, $x_n \to x$ and $x'_n \to x'$ for some $x \in M$ and $x' \in M'$. Finally, let $a = (x, x')$. Then

$$d_p(a_n, a) \leq d(x_n, x) + d'(x'_n, x') \to 0 \text{ as } n \to \infty$$

Thus, (a_n) is convergent in $M \oplus_p M'$. $\square$

Note: (a_n) is Cauchy iff (x_n) and (x'_n) are Cauchy

Corollary 3.3. *$\mathbb{R}^n$ and $\mathbb{C}^n$ are complete in the l_p-metric for $p = 1, 2, \infty$. In particular, n-dimensional real or complex euclidean space is complete.*

Proof. Trivial by the previous Proposition. $\square$

Theorem 3.4. *Let S be an arbitrary set. Then the space $l_\infty(S)$ is complete in the uniform metric D.*

Proof. Let (f_n) be a Cauchy sequence in $l_\infty(S)$. Then given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $m, n \geq N$, we have

$$F(f_m, f_n) = \sup_{x \in S} |f_m(x) - f_n(x)| < \varepsilon \implies \forall x \in S \ |f_m(x) - f_n(x)| < \varepsilon$$

Thus, (f_n) is uniformly Cauchy. By Theorem 1.6 it follows that (f_n) converges uniformly to a scalar function f on S. By Lemma 1.2, it follows that f is bounded, i.e $f \in l_\infty(S)$. Finally, since $f_n \to f$ uniformly on S, given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\forall x \in S \ |f_n(x) - f(x)| < \varepsilon \implies D(f_n, f) = \sup_{x \in S} |f_n(x) - f(x)| \leq \varepsilon$$

This shows that $f_n \to f$ in the uniform metric. □

Proposition 3.5. *Let N be a subspace of a metric space M.*

(i) If N is complete, then N is closed in M

(ii) If M is complete and N is closed in M, then N is complete

In particular, in a complete metric space, a subspace is closed iff it is complete.

Proof.

(i) Given (x_n) in N, assume $x_n \to x$ in M. Then (x_n) is Cauchy in M by Lemma 3.1, and hence Cauchy in N. Since N is complete, then $x_n \to y$ for some y in N, and hence $x_n \to y$ in M. By uniqueness of limit (Lemma 2.1), $x = y$ and hence $x \in N$ as required.

(ii) Let (x_n) be a Cauchy sequence in N. Then (x_n) is Cauchy in M, and hence $x_n \to x$ for some $x \in M$. Since N is closed, $x \in N$. It follows that $x_n \to x$ in N.

□

Definition. Let (M, d) be a metric space. We define

$$C_b(M) = \{f \in l_\infty(M) : f \text{ is continuous}\}$$

So in the real case, we have

$$C_b(M) = \{f : M \to \mathbb{R} : f \text{ is bounded and continuous}\}$$

Note: $C_b(M)$ is a subspace of $l_\infty(M)$ in the uniform metric D.

Theorem 3.6. *$C_b(M)$ is complete in the uniform metric for any metric space M.*

Proof. By Theorem 3.4 and Proposition 3.5 (ii), it is enough to show that $C_b(M)$ is closed in $l_\infty(M)$. So let (f_n) be a sequence in $C_b(M)$, let $f \in l_\infty(M)$, and assume that $f_n \to f$. We need to show that $f \in C_b(M)$, i.e that f is continuous. So fix $a \in M$ and $\varepsilon > 0$. Since $f_n \to f$, we can choose a large $n \in \mathbb{N}$ such that

$$D(f_n, f) = \sup_{x \in M} |f_n(x) - f(x)| < \varepsilon$$

Since f_n is continuous at a, there exists $\delta > 0$ such that

$$\forall x \in M \; d(x,a) < \delta \implies |f_n(x) - f_n(a)| < \varepsilon$$

It follows that for any $x \in M$, if $d(x,a) < \delta$, then

$$|f(x) - f(a)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)| \leq 3\varepsilon$$

This shows that f is continuous at a, and hence continuous. □

Corollary 3.7. *For any closed, bounded interval $[a,b]$ in $\mathbb{R}$, the space $C[a,b]$ of continuous scalar functions on $[a,b]$ is complete in the uniform metric.*

Proof. Trivial by the previous theorem. □

Definitions. Let S be a set and (N,e) be a metric space. Define

$$l_\infty(S,N) = \{f : S \to N : f \text{ is bounded}\}$$

We say f is *bounded* if:

$$\exists y \in N \; \exists r > 0 \; \forall x \in S \; f(x) \in B_r(y)$$

Now assume that $g : S \to N$ is another bounded function: there exists $z \in N$ and $s > 0$ such that $g(x) \in B_s(z)$ for all $x \in S$. Then

$$\forall x \in S \; e(f(x), g(x)) \leq e(f(x), y) + e(y,z) + e(z, g(x)) \leq r + e(y,z) + s$$

It follows that $\sup_{x \in S} e(f(x), g(x))$ exists and will be denoted by $D(f,g)$. It is routine to verify that D is a metric on $l_\infty(S,N)$ called the *uniform metric.*

Now assume further that (M,d) is a metric space. Define

$$C_b(M,N) = \{f : M \to N : f \text{ is bounded and continuous}\}$$

which is a subspace of $l_\infty(M,N)$.

Theorem 3.8. *Let S be a set, (m,d) and (N,e) be metric spaces, and assume N is complete. Then*

(i) $l_\infty(S,N)$ is complete in the uniform metric

(ii) $C_b(M,N)$ is complete in the uniform metric

Proof.

(i) Let (f_k) be a Cauchy sequence in $l_\infty(S, N)$. We first show that (f_k) is pointwise convergent. So fix $x \in S$ and $\varepsilon > 0$, since (f_k) is Cauchy, there exists $K \in \mathbb{N}$ such that $D(f_i, f_j) < \varepsilon$ for all $i, j \geq K$. It follows that

$$e(f_i(x), f_j(x)) \leq D(f_i, f_j) < \varepsilon \text{ for all } i, j \geq K$$

Hence $(f_k(x))$ is Cauchy in N, and thus convergent in N, since N is complete. We can now define

$$f : S \to N \text{ by } f(x) = \lim_{k\to\infty} f_k(x)$$

We next show that f is bounded. By Lemma 3.1, the sequence (f_k) is bounded in the metric space $l_\infty(S, N)$. So there exists $g \in l_\infty(S, N)$ and $r > 0$ such that $f_k \in B_r(g)$ for all $k \in \mathbb{N}$. Since g is bounded, there exists $y \in N$ and $s > 0$ such that $g(x) \in B_s(y)$ for all $x \in S$. It follows that for any $x \in S$ and $k \in \mathbb{N}$, we have

$$e(f_k(x), y) \leq e(f_k(x), g(x)) + e(g(x), y) \leq D(f_k, g) + s \leq r + s$$

and thus $f_k(x) \in B_{r+s}(y)$. By Lemma 2.10, the set $B_{r+s}(y)$ is closed in N. It follows that $f(x) = \lim_{k\to\infty} f_k(x) \in B_{r+s}(y)$ for all $x \in S$, and so f is bounded, i.e $f \in l_\infty(S, N)$.

It remains to show that $f_k \to f$ is the uniform metric in $l_\infty(S, N)$. Let $\varepsilon > 0$. Since (f_k) is Cauchy, there exists $K \in \mathbb{N}$ such that $D(f_i, f_j) < \varepsilon$ for all $i, j \geq K$. Fix $x \in S$ and $i \geq K$. Then

$$e(f_i(x), f_j(x)) \leq D(f_i, f_j) < \varepsilon \text{ for all } j \geq K$$

Since $f_j(x) \to f(x)$, and since the metric e is continuous, it follows that $e(f_i(x), f(x)) \leq \varepsilon$. Since this holds for any $x \in S$, we have $D(f_i, f) \leq \varepsilon$, and this is true for any $i \geq K$. This proves that $f_i \to f$ in the uniform metric D.

(ii) Since $l_\infty(M, N)$ is complete, by Proposition 3.5 it is sufficient to prove that $C_b(M, N)$ is closed in $l_\infty(M, N)$. So let (f_k) be a sequence in $C_b(M, N)$, let $f \in l_\infty(M, N)$, and assume that $f_k \to f$ in the uniform metric. We need to show that $f \in C_b(M, N)$, i.e that f is continuous.

So fix $a \in M$ and $\varepsilon > 0$. Since $f_k \to f$, we can choose a large $k \in \mathbb{N}$ such that

$$D(f_k, f) = \sup_{x\in M} e(f_k(x), f(x)) < \varepsilon$$

Since f_k is continuous at a, there exists $\delta > 0$ such that

$$\forall x \in M \ d(x, a) < \delta \implies e(f_k(x), f_k(a)) < \varepsilon$$

It follows that for any $x \in M$, if $d(x, a) < \delta$, then

$$\begin{aligned} e(f(x), f(a)) &\leq e(f(x), f_k(x)) + e(f_k(x), f_k(a)) + e(f_k(a), f(a)) \\ &\leq \varepsilon + \varepsilon + \varepsilon \leq 3\varepsilon \end{aligned}$$

This shows that f is continuous at a, and hence continuous.

$\square$

Definition. A map $f : M \to M'$ between metric spaces is called a *contraction mapping* if

$$\exists \lambda < 1 \ \forall x, y \in M \ d'(f(x), f(y)) \leq \lambda d(x, y)$$

Note: Equivalently, f is a Lipschitz map with constant $\lambda < 1$. Thus, a contraction mapping is continuous.

Theorem 3.9 (Contraction Mapping Theorem)**.** *Let M be a non-empty, complete metric space and let $f : M \to M$ be a contraction mapping. Then f has a unique fixed point, i.e there is a unique $z \in M$ such that $f(z) = z$.*

Proof. Fix $\lambda < 1$ such that $d(f(x), f(y)) \leq \lambda d(x, y)$ for all $x, y \in M$.

Uniqueness: assume $f(z) = z$ and $f(w) = w$. Then

$$d(z, w) = d(f(z), f(w)) \leq \lambda d(z, w) < d(z, w)$$

Which is a contradiction.

Existence: fix an arbitrary $x_0 \in M$ and define $(x_n)_{n=1}^{\infty}$ recursively by $x_n = f(x_{n-1})$. Thus $x_n = f(f(\ldots f(x_0)\ldots))$ (with f applied n times). Idea: if $z = f(f(f(\ldots(x_0)\ldots)))$ (with infinitely many f's), then $f(z) = z$.

For any $n \in \mathbb{N}$, we have

$$d(x_n, x_{n+1}) = d(f(x_{n-1}), f(x_n)) \leq \lambda d(x_{n-1}, x_n) \leq \ldots \leq \lambda^n d(x_0, x_1)$$

It follows that for any $m, n \in \mathbb{N}$ with $n \geq m$, we have

$$\begin{aligned} d(x_m, x_n) &\leq \sum_{k=m}^{n-1} d(x_k, x_{k+1}) \\ &\leq \sum_{k=m}^{n-1} \lambda^k d(x_0, x_1) = \frac{\lambda^m - \lambda^n}{1 - \lambda} d(x_0, x_1) \leq \frac{\lambda^m}{1 - \lambda} d(x_0, x_1) \end{aligned}$$

Since $\frac{\lambda^m}{1-\lambda} \to 0$ as $m \to \infty$, given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $\frac{\lambda^m}{1-\lambda} d(x_0, x_1) < \varepsilon$ for all $m \geq N$, and then by the above $d(x_m, x_n) < \varepsilon$ whenever $n \geq m \geq N$. This shows that (x_n) is Cauchy. Since M is complete, $x_n \to z$ for some $z \in M$.

Since f is continuous, $f(x_n) \to f(z)$ as $n \to \infty$. On the other hand, $f(x_n) = x_{n+1} \to z$ also, and hence $f(z) = z$. $\square$

Remarks:

1. The proof not only shows the existence of a fixed point, but also a way of approximating it. Letting $n \to \infty$ in the inequality for $d(x_m, x_n)$, we obtain
$$d(x_m, z) \leq \frac{\lambda^m}{1-\lambda} d(x_0, x_1) \text{ for all } m \in \mathbb{N}$$
Thus the convergence is exponentially fast!

2. $f : \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}$, $x \mapsto x/2$ is a contraction ($\lambda = 1/2$) but has no fixed point.

3. $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x + 1$ is isometry ($\lambda = 1$) but has no fixed point.

4. $f : [1, \infty), [1, \infty),\ x \mapsto x + \frac{1}{x}$ satisfies $|f(x) - f(y)| < |x - y|$ for all $x, y \in [1, \infty)$ but has no fixed point.

An application: Given $y_0 \in \mathbb{R}$ the initial value problem (IVP)

$$f'(t) = f(t^2),\ f(0) = y_0$$

has a unique solution on $[0, \frac{1}{2}]$.

Step 1: f is a solution of the IVP iff

$$f \in C\left[0, \frac{1}{2}\right] \text{ and } f(t) = y_0 + \int_0^t f(s^2)\mathrm{d}s \text{ for all } t \in \left[0, \frac{1}{2}\right]$$

Step 2: $M = C[0, \frac{1}{2}]$ is a non-empty, complete metric space in the uniform metric D. Define $T : M \to M$ by

$$T(g)(t) = y_0 + \int_0^t g(s^2)\mathrm{d}s \text{ for } t \in \left[0, \frac{1}{2}\right] \text{ and } g \in M$$

$T(g)(t)$ is well-defined since $s \mapsto g(s^2)$ is continuous. $T(g)$ is continuous, and indeed also differentiable by the FTC with $(Tg)'(t) = g(t^2)$ for all $t \in [0, \frac{1}{2}]$. Thus $T(g) \in M$.

By Step 1, f is a solution of the IVP iff $f \in M$ and $T(f) = f$.

Step 3: T is a contraction mapping. For $g, h \in M$ we estimate

$$|(Tg)(t) - (Th)(t)| \leq \left|\int_0^t g(s^2) - h(s^2)\mathrm{d}s\right| \leq t \sup_{s\in[0,\frac{1}{2}]} |g(s^2) - h(s^2)| \leq \frac{1}{2}D(g, h)$$

Taking sup over all $t \in [0, \frac{1}{2}]$, we get $D(Tg, Th) \leq \frac{1}{2}D(g, h)$. This T is a contraction mapping on M.

Step 4: By the Contraction Mapping Theorem (Theorem 3.9), T has a unique fixed point, and thus by Step 2, the IVP has a unique solution.

Remark: The same proof shows that for any $\delta \in (0, 1)$, there is a unique solution f_δ to the IVP on $[0, \delta]$. For $\delta < \mu < 1$, we have $f_\mu|_{[0,\delta]} = f_\delta$ by uniqueness. Hence there is a unique solution to the IVP on $[0, 1)$.

Theorem 3.10 (Lindelöf-Picard)**.** *We are given $n \in \mathbb{N}$, $a, b, R \in \mathbb{R}$ with $a < b$ and $R > 0$, $y_0 \in \mathbb{R}^n$ and a continuous function*

$$\phi : [a, b] \times B_R(y_0) \to \mathbb{R}^n$$

Assume that there exists $K > 0$ such that

$$||\phi(t, x) - \phi(t, y)|| \leq K||x - y|| \textit{ for all } t \in [a, b], x, y \in B_R(y_0)$$

Then there exists $\varepsilon > 0$ such that for any $t_0 \in [a, b]$ the IVP

$$f'(t) = \phi(t, f(t)),\ f(t_0) = y_0$$

has a unique solution on $[c, d] = [t_0 - \varepsilon, t_0 + \varepsilon] \cap [a, b]$*. In other words, there is a unique differentiable function* $f : [c, d] \to \mathbb{R}^n$ *that solves the above IVP.*

Notes:

1. The statement that $f : [c, d] \to \mathbb{R}^n$ solves the IVP includes implicitly the condition that f takes values in $B_R(y_0)$ (otherwise the differential equation would not make sense).

2. The assumption on ϕ is called a Lipschitz condition in the second variable.

3. Given a function $f : [c, d] \to \mathbb{R}^n$, we let $f_k = q_k \circ f : [c, d] \to \mathbb{R}$, where $q_k : \mathbb{R}^n \to \mathbb{R}$ is the kth coordinate mapping $(y_1, \dots, y_n) \mapsto y_k$ for $k = 1, 2, \dots, n$. Thus, $f(t) = (f_1(t), f_2(t), \dots, f_n(t))$ for every $t \in [c, d]$. Now, f is differentiable if each f_k is differentiable, and

   $$f'(t) = (f_1'(t), \dots, f_n'(t)) \text{ for every } t \in [c, d]$$

   Also, if f is continuous, then each f_k is continuous, and hence integrable on $[c, d]$. We can then define the integral of f coordinate-wise: $\int_c^d f(t)\mathrm{d}t$ is the element $v \in \mathbb{R}^n$ with kth coordinate $v_k = \int_c^d f_k(t)\mathrm{d}t$. Observe that

   $$\begin{aligned} ||v||^2 = \sum_{k=1}^n v_k^2 = \sum_{k=1}^n v_k \int_c^d f_k(t)\mathrm{d}t &= \int_c^d \sum_{k=1}^n v_k f_k(t)\mathrm{d}t \\ = \int_c^d v \cdot f(t)\mathrm{d}t \le \int_c^d ||v||||f(t)||\mathrm{d}t &= ||v|| \int_c^d ||f(t)||\mathrm{d}t \end{aligned}$$

   So we have proved that $|| \int_c^d f(t)\mathrm{d}t|| \le \int_c^d ||f(t)||\mathrm{d}t$.

Proof. By Lemma 2.10, $B_R(y_0)$ is a closed subset of $\mathbb{R}^n$. So ϕ is a continuous function on a closed and bounded subset $[a, b] \times B_R(y_0)$ of $\mathbb{R}^{n+1}$, and hence it is bounded. Set $C = \sup\{||\phi(t, x)|| : t \in [a, b], x \in B_R(y_0)\}$ and then set $\varepsilon = \min(\frac{R}{C}, \frac{1}{2K})$. We are going to show that this ε works.

Given $t_0 \in [a, b]$, set $[c, d] = [t_0 - \varepsilon, t_0 + \varepsilon] \cap [a, b]$. We need to prove that there is a unique differentiable function $f : [c, d] \to \mathbb{R}^n$ such that $f(t_0) = y_0$ and $f'(t) = \phi(t, f(t))$ for all $t \in [c, d]$.

Since $B_R(y_0)$ is closed in $\mathbb{R}^n$, and since $\mathbb{R}^n$ is complete (Corollary 3.3), it follows that $B_R(y_0)$ is complete (Proposition 3.5). By Theorem 3.8, the space $M = C([c, d], B_R(y_0))$ is complete in the uniform metric D. Also $M \neq \emptyset$.

Next, note that f is a solution of the IVP on $[c, d]$ iff

$$f \in M \text{ and } f(t) = y_0 + \int_{t_0}^t \phi(s, f(s))\mathrm{d}s \text{ for all } t \in [c, d]$$

This follows from the FTC applied coordinate-wise.

We next define

$$T : M \to M \text{ by } T(g)(t) = y_0 + \int_{t_0}^{t} \phi(s, g(s))\mathrm{d}s \text{ for } t \in [c, d]$$

Is T well defined? First, the definition of $T(g)(t)$ makes sense, since $s \mapsto \phi(s, g(s))$ is continuous. Next, $T(g)$ is a continuous function $[c, d] \to \mathbb{R}^n$. Indeed, $T(g)$ is differentiable with $(Tg)'(t) = \phi(t, g(t))$ by the FTC. Finally, $T(g)$ takes values in $B_R(y_0)$ since

$$||T(g)(t) - y_0|| = \left|\left|\int_{t_0}^{t} \phi(s, g(s))\mathrm{d}s\right|\right| \leq |t - t_0| \sup_{s\in[c,d]} ||\phi(s, g(s))|| = \varepsilon C \leq R$$

for all $t \in [c, d]$. This shows that $T(g) \in M$.

Now by the earlier observation, f is a solution of the IVP iff $f \in M$ and $T(f) = f$. So the proof is complete if we can show that T is a contraction mapping.

Let $g, h \in M$. Note that

$$||\phi(s, g(s)) - \phi(s, h(s))|| \leq K||g(s) - h(s)|| \leq KD(g, h) \text{ for all } s \in [c, d]$$

It follows that for every $t \in [c, d]$ we have

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

Taking the sup over $t \in [c, d]$, we get $D(Tg, Th) \leq \varepsilon KD(g, h) \leq \frac{1}{2}D(g, h)$. Thus, $T : M \to M$ is a contraction mapping, as required.

□

Remarks:

1. In general, it is not possible to get a general solution, i.e one on $[a, b]$.
2. This Theorem handles nth order ODEs as well. We do this next.

A special case of Lindelöf-Picard

We are given $n \in \mathbb{N}$, $a, b, R \in \mathbb{R}$ with $a < b$ and $R > 0$, $z = (z_0, z_1, \ldots z_{n-1}) \in \mathbb{R}^n$ and a continuous function

$$\psi : [a, b] \times B_R(z) \to \mathbb{R}$$

Assume that for some $K > 0$ we have

$$|\psi(t,x) - \psi(t,y)| \leq K||x-y|| \text{ for all } t \in [a,b] \text{ and all } x,y \in B_R(z)$$

Then there exists $\varepsilon > 0$ such that for any $t_0 \in [a,b]$ the nth order IVP

$$\begin{aligned} &g^{(n)}(t) = \psi(t, g(t), g^{(1)}(t), g^{(2)}(t), \ldots, g^{(n-1)}(t)) \qquad (*) \\ &g^{(j)}(t_0) = z_j \text{ for } 0 \leq j \leq n-1 \end{aligned}$$

has a unique solution on $[c,d] = [t_0 - \varepsilon, t_0 + \varepsilon] \cap [a,b]$.

This means that there is a unique n-times differentiable function

$$g : [c,d] \to \mathbb{R}$$

that satisfies $(*)$ for all $t \in [c,d]$. This implicitly includes the assumption that

$$(g(t), g^{(1)}(t), g^{(2)}(t), \ldots, g^{(n-1)}(t)) \in B_R(z)$$

for all $t \in [c,d]$.

Proof. Let us define $\phi : [a,b] \times B_R(z) \to \mathbb{R}^n$ by setting

$$\phi(t, x_0, x_1, \ldots, x_{n-1}) = (x_1, \ldots, x_{n-1}, \psi(t, x_0, x_1, \ldots, x_{n-1}))$$

for $t \in [a,b]$ and $x = (x_0, x_1, \ldots x_{n-1}) \in B_R(z)$. Then ϕ is continuous and satisfies

$$||\phi(t,x) - \phi(t,y)|| \leq (K+1)||x-y|| \text{ for all } t \in [a,b] \text{ and all } x,y \in B_R(z)$$

By Linedlöf-Picard (Theorem 3.10), there exists $\varepsilon > 0$ such that the IVP

$$f'(t) = \phi(t, f(t)),\ f(t_0) = z \qquad (**)$$

has a unique solution on $[c,d] = [t_0 - \varepsilon, t_0 + \varepsilon] \cap [a,b]$. Let f be this unique solution. Thus $f : [c.d] \to B_R(z)$ is a differentiable function with $f(t_0) = z$ and $f'(t) = \phi(t, f(t))$ for all $t \in [c,d]$. Let $f_0, f_1, \ldots, f_{n-1}$ be the components of f. Since f is a solution of $(**)$, each f_j is differentiable and

$$\begin{aligned} &(f_0'(t), f_1'(t), \ldots, f_{n-1}'(t)) = f'(t) = \phi(t, f(t)) \\ &= (f_1(t), f_2(t), \ldots, f_{n-1}(t), \psi(t, f_0(t), f_1(t), \ldots, f_{n-1}(t))) \qquad (\dagger) \end{aligned}$$

for all $t \in [c,d]$. Set $g = f_0$. Comparing coordinates in $(\dagger)$ shows that g is an n-times differentiable function $[c,d] \to \mathbb{R}$ with $g^{(j)} = f_j$ for $0 \leq j < n$, and

$$g^{(n)} = f_{n-1}'(t) = \psi(t, f_0(t), f_1(t), \ldots, f_{n-1}(t)) = \psi(t, g(t), g^{(1)}(t), \ldots, g^{(n-1)}(t))$$

for all $t \in [c,d]$.

Finally, since $f(t_0) = z$, we have $g^{(j)}(t_0) = f_j(t_0) = z_j$ for $0 \leq j \leq n-1$. Thus, g satisfies the IVP $(*)$, which completes the proof of existence.

To prove uniqueness, assume that $\tilde{g}$ is another solution to $(*)$ on $[c,d]$. Define $\tilde{f} : [c,d] \to B_R(z)$ by setting $\tilde{f} = (\tilde{g}(t), \tilde{g}^{(1)}(t), \ldots, \tilde{g}^{(n-1)}(t))$. It is straightforward to verify $\tilde{f}$ is a solution to $(**)$. It follows that $\tilde{f} = f$ and $\tilde{g} = g$. $\square$

4 Topological Spaces

Definition. Let X be a set. A *topology* on X is a family τ of subsets of X (i.e $\tau \subset \mathcal{P}(X)$) such that

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

(ii) If $U_i \in \tau$ for all $i \in I$ then $\bigcup_{i \in I} U_i \in \tau$

(iii) If $U, V \in \tau$, then $U \cap V \in \tau$

Members of τ are called *open sets*. Thus, for $U \subseteq X$, we say U is *open* in X if $U \in \tau$. We might also say that U is τ*-open* to emphasize the topology τ.

A *topological space* is a pair (X, τ), where X is a set, and τ is a topology on X.

Note: If $n \in \mathbb{N}$ and $U_1, \ldots, U_n$ are open sets in a topological space, then $\bigcap_{i=1}^n U_i$ is also open.

Examples:

1. Metric topologies.
   Let (M, d) be a metric space. Recall from chapter 2 that $U \subseteq M$ is defined to be open in the metric space M if

   $$\forall x \in U \ \exists r > 0 \ B_r(x) \subseteq U$$

   We sometimes call such a set U d*-open* to emphasize that it is the metric notion of being open. By Proposition 2.9, the family of d-open sets is a topology on M, which we called the metric topology of M.

   Unless mentioned otherwise, a metric space will always be equipped with its metric topology.

Definition. We say that a topological space (X, τ) is *metrizable* if there is a metric d on X such that τ is the metric topology induced by d.

Thus, for $U \subseteq X$, we have U is τ-open iff U is d-open. Note that in this case any metric d' equivalent to d induces the same topology on τ.

Examples (continued):

2. The *indiscrete topology* on a set X is the topology $\{\emptyset, X\}$. If $|X| \geq 2$, then this is not metrizable.

   To see this, consider any metric d on X, fix $x \neq y$ in X, set $r = d(x, y)$, and note that $U = D_r(x)$ is d-open (Lemma 2.5), non-empty and $U \neq X$ (as $y \notin U$).

Definition. Let τ_1 and τ_2 be topologies on a set X. We say that τ_1 is *coarser* than τ_2 (or that τ_2 is *finer* than τ_1) if $\tau_1 \subseteq \tau_2$.

E.g the indiscrete topology on a set is the coarsest topology on that set.

Examples (continued):

3. The *discrete topology* on a set X is the power set $\mathcal{P}(X)$ of X. This is the finest topology on X. The discrete topology is metrizable: by the discrete metric.

4. The *cofinite topology* on a set X is

$$\tau = \{\emptyset\} \cup \{U \subseteq X : U \text{ is confinite in } X\}$$

We say $U \subseteq X$ is *cofinite* in X if $X \setminus U$ is finite.

Note that if X is a finite set, then this is the discrete topology. However, we will see that if X is infinite, then the confinite topology is not metrizable. Observe that if $x \neq y$ in X and U, V are open sets with $x \in U$ and $y \in V$, then $U \cap V \neq \emptyset$.

Definition. A topological space X is called *Hausdorff* if for all $x \neq y$ in X, there exist disjoint open sets U and V in X such that $x \in U$ and $y \in V$.

Note: An infinite set with the cofinite topology is not Hausdorff. On the other hand we have the following result.

Proposition 4.1. *Every metric space is Hausdorff.*

Proof. Given distinct elements x and y in a metric space M, fix $r > 0$ with $2r < d(x, y)$. Then $U = D_r(x)$ and $V = D_r(y)$ are open sets in M with $x \in U$ and $y \in V$. Moreover, $U \cap V = \emptyset$, otherwise if $z \in U \cap V$, then

$$d(x, y) \leq d(x, z) + d(z, y) < 2r < d(x, y)$$

which is a contradiction. □

Remark: It follows that the cofinite topology on an infinite set is not metrizable.

Definition. A subset A of a topological space (X, τ) is *closed* in X if $X \setminus A$ is open in X.

Note: In a metric space, this agrees with the earlier definition of a closed set by Lemma 2.11.

Proposition 4.2. *The collection of closed sets in a topological space X satisfies:*

(i) $\emptyset, X$ *are closed.*

(ii) If A_i, $i \in I$ are closed, then $\bigcap_{i\in I} A_i$ is closed.

(iii) If A, B are closed, then $A \cup B$ is closed.

Proof. Trivial. $\square$

Examples:

1. In a discrete topological space, every subset is closed.
2. In the cofinite topology on a set X, a subset A is closed iff $A = X$ or A is finite.

Definition. Let X be a topological space, $U \subseteq X$ and $x \in X$. We say U is a *neighbourhood* of x in X if there exists an open subset V of X such that $x \in V \subseteq U$.

Note: In a metric space, this agrees with the earlier definition by Corollary 2.6. The following result generalizes an earlier observation in a metric space.

Proposition 4.3. *Let U be a subset of a topological space X. Then U is open in X iff U is a neighbourhood of x for all $x \in U$.*

Proof. Assume U is open and $x \in U$. Then $V = U$ is open, and $x \in V \subseteq U$. Thus U is a neighbourhood of x.

Conversely, assume that for every $x \in U$, there exists some open set V_x in X such that $x \in V_x \subseteq U$. Then $U = \bigcup_{x\in U} V_x$ and hence U is open in X. $\square$

Definition. Let (x_n) be a sequence in a topological space and let $x \in X$. We say (x_n) *converges* to x (and write $x_n \to x$) if

$$\forall \text{ neighbourhoods } U \text{ of } x\ \exists N \in \mathbb{N}\ \forall n \geq N\ x_n \in U$$

or equivalently,

$$\forall \text{ open sets } U \text{ with } x \in U\ \exists N \in \mathbb{N}\ \forall n \geq N\ x_n \in U$$

Note: In a metric space, this agrees with the previous definition of convergence by Proposition 2.7.

Examples:

1. Eventually constant sequences: if (x_n) is a sequence in some topological space, and for some $N \in \mathbb{N}$, we have $x_n = z$ for $n \geq N$, then $x_n \to z$.
2. In an indiscrete space, every sequence converges to every point.
3. Let X be a set equipped with the cofinite topology. Assume that $x_n \to X$ in X. For any $y \neq x$, the set $U = X \setminus \{y\}$ is a neighbourhood of x. It follows that $N_y = \{n \in \mathbb{N} : x_n = y\}$ is finite.

   Conversely, if for some $x \in X$, the set N_y is finite for all $y \neq x$, then $x_n \to x$. So if N_y is finite for all $y \in X$, then $x_n \to x$ for all $x \in X$.

Proposition 4.4. *Suppose $x_n \to x$ and $x_n \to y$ in a Hausdorff space. Then $x = y$.*

Proof. Assume $x \neq y$. Then there exist disjoint open sets U and V such that $x \in U$ and $y \in V$. Since $x_n \to x$, there exists $N_1 \in \mathbb{N}$ such that $x_n \in U$ for every $n \geq N_1$. Similarly, since $x_n \to y$, there exists $N_2 \in \mathbb{N}$ such that $x_n \in V$ for every $n \geq N_2$. Then for $n = \max\{N_1, N_2\}$, we have $x_n \in U \cap V$. Contradiction. $\square$

Remark: So in a Hausdorff space limits are unique. If $x_n \to x$, we will sometimes denote this unique limit by $\lim_{n\to\infty} x_n$.

Remark: Recall that a subset A of a metric X is closed in $X \iff$ whenever (x_n) is a sequence in A converging to a point x in X, the limit belongs to A. The implication " $\implies$ " is true also in an arbitrary topological space. Hoever the reverse implication is not in general.

Definition. Let A be a subset of a topological space X.

The *interior* of A in X, denoted $\text{int}(A)$ or A°, is the set

$$\text{int}(A) = A^\circ = \bigcup\{U \subseteq X : U \text{ is open in } X \text{ and } U \subseteq A\}$$

The *closure* of A in X, denoted $\text{Cl}(A)$ or $\hat{A}$, is the set

$$\text{Cl}(A) = \hat{A} = \bigcap\{F \subseteq X : F \text{ is closed in } X \text{ and } A \subseteq F\}$$

(Note that X is closed in X and $A \subseteq X$)

Remarks:

1. A° is open, $A^\circ \subseteq A$ and if U is open with $U \subseteq A$, then $U \subseteq A^\circ$. Thus, A° is the largest open set contained in A. It follows that A is open $\iff A = A^\circ$.
2. $\bar{A}$ is closed, $A \subseteq \bar{A}$ and if F is closed with $A \subseteq F$, then $\bar{A} \subseteq F$. Thus, $\bar{A}$ is the smallest closed set containing A. It follows that A is closed $\iff A = \bar{A}$.

Proposition 4.5. *Let A be a subset of a topological space X. Then*

(a) $A^\circ = \{x \in X : A \textit{ is a ngbd of } x\}$

(b) $\bar{A} = \{x \in X : \forall \textit{ ngbds } U \textit{ of } x, U \cap A \neq \emptyset\}$

Proof.

(a) For $x \in X$ we have

$$\begin{aligned} x \in A^\circ &\iff \exists \text{ open } U \text{ with } U \subseteq A \text{ and } x \in U \\ &\iff A \text{ is a ngbd of } x \end{aligned}$$

(b) If $x \notin \bar{A}$, then $U = X \setminus \bar{A}$ is an open set with $x \in U$, so U is a ngbd of x and $U \cap A = \emptyset$.

Conversely, if there is a ngbd U of x with $U \cap A = \emptyset$, then there is an open set V with $x \in V \subseteq U$. Then $V \cap A = \emptyset$, and hence $A \subseteq X \setminus V$. Since $X \setminus V$ is closed, it follows that $\bar{A} \subseteq X \setminus V$, and thus $x \notin \bar{A}$.

□

Examples:

1. In $\mathbb{R}$, if $A = [0, 1) \cup \{2\}$, then $A^\circ = (0, 1)$ and $\bar{A} = [0, 1] \cup \{2\}$.
2. $\mathbb{Q}^\circ = \emptyset, \bar{\mathbb{Q}} = \mathbb{R}, \mathbb{Z}^\circ = \emptyset$ and $\bar{\mathbb{Z}} = \mathbb{Z}$

Note: In a metric space, $x \in \bar{A} \iff$ there is a sequence (x_n) in A such that $x_n \to x$. In a topological space, $\Leftarrow$ is still true but $\Rightarrow$ is not in general. In a metric space, convergent sequences determine the topology. This fails in topological spaces in general.

Definitions. Let X be a topological space. A subset A of X is *dense* in X if $\bar{A} = X$. We say X is *seperable* if there is a countable subset A of X that is dense in X.

Examples:

1. $\mathbb{R}$ is seperable, since $\mathbb{Q}$ is dense in $\mathbb{R}$. Similarly $\mathbb{R}^n$ is seperable for any $n \in \mathbb{N}$. On the other hand, an uncountable discrete topological space is not seperable.

4.1 Subspaces

Let (X, τ) be a topological space, and let $Y \subseteq X$. The *subspace topology* on Y (or the *relative topology* on Y) is the topology $\{U \cap Y : U \in \tau\}$. This is also called the topology on Y *induced* by the topology on X.

So for $U \subseteq Y$, we have U is open in $Y \iff \exists$ open V in X with $U = V \cap Y$.

Example:
Consider $X = \mathbb{R}$, $Y = [0, 2]$ and $U = (1, 2]$. Then $U \subseteq Y \subseteq X$, and U is open in Y since $U = (1, 3) \cap Y$. However U is not open in X.

Remarks:

1. A subset of a topological space will always be assumed to be equipped with the relative topology

2. A subspace of a subspace is a subspace: Let X be a topological space and $Z \subseteq Y \subseteq X$. Then the topology on Z induced by the topology of X is the same as the topology on Z induced by the relative topology on Y.

3. Let (M, d) be a metric space and $N \subseteq M$. The metric d induces the metric topology on M. In turn, this induces the relative topology on N.

   On the other hand, the restriction of d to N is a metric on N, which in turn induces the metric topology on N. These two topologies are fortunately the same. This follows easily from the observation that $D_r^N(x) = D_r^M(x) \cap N$ for $x \in N$ and $r > 0$.

Proposition 4.6. *Let X be a topological space, and let $A \subseteq Y \subseteq X$.*

(i) A is closed in $Y \iff \exists$ a closed set B in X with $A = B \cap Y$.

(ii) The closure of A in Y is the intersection with Y of the closure of A in X.

Remark: The analogue of part (ii) for interiors is false. E.g $X = \mathbb{R}$ and $A = Y = \{0\}$. Then the interior of A in Y is A, and the interior of A in X is $\emptyset$.

Proof.

(i) If A is closed in Y, then $Y \setminus A$ is open in Y. It follows that $Y \setminus A = U \cap Y$ for some set U open in X. Then $A = (X \setminus U) \cap Y$ and $X \setminus U$ is closed in X, which proves $\Rightarrow$.

Conversely, assume that $A = B \cap Y$ where B is closed in X. Then $Y \setminus A = (X \setminus B) \cap Y$ and $X \setminus B$ is open in X. It follows that $Y \setminus A$ is open in Y, and thus A is closed in Y.

(ii) Let us write $\bar{A}^Y$ and $\bar{A}^X$ for the closure of A in Y and X respectively. Then $\bar{A}^X \cap Y \supseteq A$, and $\bar{A}^X \cap Y$ is closed in Y by part (i). It follows that $\bar{A}^X \cap Y \supseteq \bar{A}^Y$.

Conversely, since $\bar{A}^Y$ is closed in Y, by part (i) there is a closed set V in X such that $\bar{A}^Y = V \cap Y$. Then $V \supseteq A$, and thus $V \supset \bar{A}^X$ since V is closed in X. It follows that $\bar{A}^Y = V \cap Y \supseteq \bar{A}^X \cap Y$.

□

Definitions. Let (X, τ) be a topological space. A *base* for τ is a family $\mathcal{B} \subseteq \tau$ such that for every $U \in \tau$, there exists $\mathcal{C} \subseteq \mathcal{B}$ such that $U = \bigcup_{B \in \mathcal{C}} B$.

Note:

1. If $\mathcal{B}$ is a base for τ, then $\tau = \{\bigcup_{B \in \mathcal{C}} B : \mathcal{C} \subseteq \mathcal{B}\}$. So $\mathcal{B}$ determines τ.
2. If $\mathcal{B}$ is a base for τ, then for every $U \subseteq C$, we have:

$$U \in \tau \iff \text{for all } x \in U \text{ there exists } B \in \mathcal{B} \text{ such that } x \in B \subseteq U$$

Examples:

1. The family of open balls in a metric space is a base for the metric topology.
2. In $\mathbb{R}$ the family $\{(a, b) : a < b\}$ of all open intervals is a base for the usual topology.
3. In a discrete space X, any base must contain $\{\{x\} : x \in X\}$, and moreover this family is a base for X.

Lemma 4.7. *Let X be a set and $\mathcal{B} \subseteq \mathcal{P}(X)$. If $X \in \mathcal{B}$ and for all $B_1, B_2 \in \mathcal{B}$ we have $B_1 \cap B_2 \in \mathcal{B}$, then there is a unique topology τ on X such that $\mathcal{B}$ is a base for τ.*

Proof. If such a topology τ exists, then by the definition of a base, we must have

$$\tau = \{U \subseteq X : \forall x \in U \ \exists B \in \mathcal{B} \ x \in B \subseteq U\}$$

It is clear that $\mathcal{B} \subseteq \tau$: if $U \in \mathcal{B}$, then for each $x \in U$, one can take $B = U$ and get $x \in B \subseteq U$. So it is enough to check that the family τ defined above is a topology on X.

Clearly $\emptyset \in \tau$, and $X \in \tau$ since $X \in \mathcal{B}$.

Let $U_i \in \tau$ for every i in some set I, and let $x \in \bigcup_{i \in I} U_i$. Then $x \in U_j$ for some $j \in I$, and hence there exists $B \in \mathcal{B}$ with $x \in B \subseteq U_j \subseteq \bigcup_{i \in I} U_i$. It follows that $\bigcup_{i \in I} U_i \in \tau$.

Finally, let $U_1, U_2 \in \tau$, and let $x \in U_1 \cap U_2$. Then for each $i = 1, 2$, since $x \in U_i \in \tau$, there exists $B_i \in \mathcal{B}$ with $x \in B_i \subseteq U_i$. It follows that $x \in B_1 \cap B_2 \subseteq U_1 \cap U_2$. By assumption, $B_1 \cap B_2 \in \mathcal{B}$, and hence we have showed that $U_1 \cap U_2 \in \tau$. $\square$

Remark: The conditions in Lemma 4.7 are sufficient but not necessary. It follows easily from the proof that the two conditions below are also sufficient

(i) $\bigcup_{B \in \mathcal{B}} B = X$

(ii) $\forall U, V \in \mathcal{B} \ \forall x \in U \cap V \ \exists W \in \mathcal{B} \ x \in W \subseteq U \cap V$

These are also clearly necessary.

Definition. We say that a topological space (X, τ) (or that the topology τ) is *second countable* if there is a countable base for τ.

Example: $\mathbb{R}$ is second uncountable: the family of open intervals (a, b) with $a, b \in \mathbb{Q}$ is a countable base. Similarly $\mathbb{R}^n$ is second countable.

Definition. A function $f : X \to Y$ between topological spaces is said to be *continuous* if $f^{-1}(V)$ is open in X for every open subset V of Y.

Note: For functions between metric spaces, this agrees with the ε-δ definition by Proposition 2.8.

Examples:

1. Constant functions are continuous: if $f : X \to Y$ is constant, then for any $V \subseteq Y$, we have that $f^{-1}(V)$ is either X or $\emptyset$ depending on whether the constant value that f takes is in V or not.
2. The identity function Id: $X \to X$ on a space X is continuous.
3. If Y is a subspace of a topological space X, then the inclusion map $\iota : Y \to X$ is continuous. Indeed, if V is open in X, then $\iota^{-1}(V) = V \cap Y$, which is open in Y by the definition of the subspace topology.

   It follows that for any continuous function $f : X \to Z$, the restriction $f|_Y : Y \to Z$ is also continuous since $f|_Y = f \circ \iota$ (See the next proposition).

Proposition 4.8. *Let $f : X \to Y$ be a function between topological spaces.*

(i) f is continuous $\iff$ $f^{-1}(V)$ is closed in X for all closed sets V in Y

(ii) If $\mathcal{B}$ is a base for Y, then we have

$$f \text{ is continuous} \iff f^{-1}(B) \text{ is open in } X \text{ for every } B \in \mathcal{B}$$

(iii) If f is continuous and if $g : Y \to Z$ is another continuous function between topological spaces, then the composite $g \circ f : X \to Z$ is also continuous.

Proof.

(i) If f is continuous and V is a closed subset of Y, then $Y \setminus V$ is open in Y, and hence $f^{-1}(Y \setminus V) = X \setminus f^{-1}(V)$ is open in X. Thus, $f^{-1}(X)$ is closed in X.

If $f^{-1}(V)$ is closed in X for all closed sets V in Y, take some open set U in Y. Then $Y \setminus U$ is closed in Y. Hence $f^{-1}(Y \setminus U) = X \setminus f^{-1}(U)$ is closed in X, so $f^{-1}(U)$ is open in X.

(ii) Since members of $\mathcal{B}$ are open in Y, the given condition is necessary.

Conversely, assuming the condition, if V is open in Y, then there is a subfamily $\mathcal{C} \subseteq \mathcal{B}$ with $V = \bigcup_{B \in \mathcal{C}} B$. It follows that $f^{-1}(V) = \bigcup_{B \in \mathcal{C}} f^{-1}(B)$, which is a union of open sets in X, and hence open in X.

(iii) Given an open set W in Z, we have $(g \circ f)^{-1}(W) = f^{-1}(g^{-1}(W))$. This is open in X, since $V = g^{-1}(W)$ is open in Y by continuity of g, and in turn $f^{-1}(V)$ is open in X by continuity of f.

$\square$

Remarks: There is a notion of continuity at a point. Let $f : X \to Y$ be a function between topological spaces, and let $x \in X$. We say f is *continuous* at x if for every ngbd V of $f(x)$ in Y, there is a ngbd U of x in X such that $f(U) \subseteq V$. Equivalently, $f^{-1}(V)$ is a ngbd of x in X for every ngbd V of $f(x)$ in Y. Then f is continuous iff f is continuous at every $x \in X$. If f is continuous at x and $x_n \to x$ in X, then $f(x_n) \to f(x)$. Then converse is false in general! If $f(x_n) \to f(x)$ whenever $x_n \to x$ in X, then it does not follow in general that f is continuous at x.

Definition. A function $f : X \to Y$ between topological spaces is a *homeomorphism* if f is a bijection and both f and f^{-1} are continuous.
Topological spaces X and Y are said to be *homeomorphic* if there is a homeomorphism $f : X \to Y$. This is an equivalence relation on the class of topological spaces.
A property $\mathfrak{B}$ of topological spaces is called a *topological property* (or a *topological invariant*) if for any pair X, Y of homeomorphic topological spaces, X has $\mathfrak{B} \iff Y$ has $\mathfrak{B}$.

Examples:

1. Being metrizable is a topological property.

   Indeed, assume $f : (X, d) \to (Y, \tau)$ is a homeomorphism, where d is a metric on X and τ is a topology on Y. We define $d'(y, z) = d(f^{-1}(y), f^{-1}(z))$ for $y, z \in Y$. A routine check shows that d' is a metric. We now let $g = f$ but thought of as a function $g : (X, d) \to (Y, d')$. Note that g is isometric, and so g is a homeomorphism. It follows that $\mathrm{Id} = g \circ f^{-1} : (Y, \tau) \to (Y, d')$ is a homeomorphism. Thus if $U \subseteq Y$ then U is τ-open iff U is d'-open.

2. Being a complete metric space is not a topological property. There are examples where a complete metric on a set is equivalent to another metric on the same set that is not complete.

Note: Let $f : X \to Y$ be a continuous bijection between topological spaces. Such an f need not be a homeomorphism. In fact, such an f is a homeomorphism iff for every open set U in X, the set $(f^{-1})^{-1}(U) = f(U)$ is open in Y.

Definition. A function $f : X \to Y$ between topological spaces is called *open* (or an *open map*) if $f(U)$ is open in Y for every open subset U of X.

Remark: A homeomorphism is a continuous and open bijection.

4.2 Product topology

Let X and Y be topological spaces. Consider the family

$$\mathcal{B} = \{U \times V : U \text{ is open in } X \text{ and } V \text{ is open in } Y\}$$

of subsets of $X \times Y$. Note that $X \times Y \in \mathcal{B}$. Also, for open sets U, U' in X and V, V' in Y, we have $U \times V \cap U' \times V' = (U \cap U') \times (V \times V')$. It follows that $\mathcal{B}$ is closed under taking intersections.

By Lemma 4.7, there is a unique topology on $X \times Y$ for which $\mathcal{B}$ is a base, called the *product topology* of $X \times Y$. The product of topological spaces will always be assumed to have the product topology unless otherwise stated.

Note: For a subset $W \subseteq X \times Y$, we have

$$W \text{ is open} \iff \forall z \in W\ \exists \text{open } U \text{ in } X \text{ and } V \text{ in } Y\ z \in U \times V \subseteq W$$

Example: Let (M, d) and (M', d') be metric spaces. We consider the metric topology on M and M', which induces a product topology on $M \times M'$. Let $W \subseteq M \times M'$. Then W is open in the product topology iff

$$\forall (x, x') \in W\ (x, x') \in U \times U' \subseteq W \text{ for some open sets } U \text{ in } M \text{ and } U' \text{ in } M'$$

$$\iff \forall (x, x') \in W\ (x, x') \in D_r(x) \times D_r(x') \subseteq W \text{ for some } r > 0$$

Observe that $D_r(x) \times D_r(x') = D_r(x, x')$ in $M \oplus_\infty M'$. it follows that W is open in the product topology iff W is d_∞-open. Thus, the product topology on the product of metrizable spaces is metrizable: it is induces by d_∞, and hence also by d_p for $p = 1$ or 2.

E.g. taking the product of two copies of $\mathbb{R}$ with its usual topology, we deduce that the product topology on $\mathbb{R}^2$ is the standard euclidean topology.

Proposition 4.9. *Let X and Y be topological spaces. The coordinate projections $q_X : X \times Y \to X$, $(x, y) \mapsto x$, and $q_Y : X \times Y \to Y$, $(x, y) \mapsto y$, satisfy*

(i) q_X and q_Y are continuous

(ii) Given a topological space Z and a function $f : Z \to X \times Y$, we have

$$f \text{ is continuous} \iff q_X \circ f : Z \to X \text{ and } q_Y \circ f : Z \to Y \text{ are continuous}$$

Proof.

(i) If U is an open subset of X, then $q_X^{-1}(U) = U \times Y$, which is open in $X \times Y$. Thus, q_X is continuous, as is q_Y by a similar argument.

(ii) If f is continuous, then so are the composites $q_X \circ f$ and $q_Y \circ f$ by part (i).

For the converse, set $g = q_X \circ f$ and $h = q_Y \circ f$, and assume that g and h are continuous. Note that $f(z) = (g(z), h(z))$ for $z \in Z$. Let W be a member of the defining base of the product topology on $X \times Y$, i.e $W = U \times V$ where

U is open in X and V is open in Y. Then for $z \in Z$, we have $f(z) \in W$ iff $g(z) \in U$ and $h(z) \in V$. It follows that $f^{-1}(W) = g^{-1}(U) \cap h^{-1}(V)$, which is open in Z since g and h are assumed continuous. Finally, it follows by Proposition 4.8(ii) that f is continuous.

$\square$

Remarks: All of the above extends to products of any finite number of topological spaces. Given $n \in \mathbb{N}$ and spaces $X_1, \ldots, X_n$, the family

$$\mathcal{B} = \{U_1 \times \ldots \times U_n : U_i \text{ is open in } x_i \text{ for } 1 \leq i \leq n\}$$

of subsets of $X = X_1 \times \ldots \times X_n$ is a base for a unique topology on X called the product topology. A subset $W \subseteq X$ is open in X iff

$$\forall (x_1, \ldots, x_n) \in W \ \exists \text{open } U_i \subseteq X_i \ (x_1, \ldots, x_n) \in U_1 \times \ldots \times U_n \subseteq W$$

As before the product of metrizable spaces is metrizable using any of the metrics d_1, d_2, d_∞ on the product. E.g. the product of n copies of $\mathbb{R}$ with the usual topology yields $\mathbb{R}^n$ with the standard topology.

The obvious analogue of Proposition 4.9 holds.

Finally, it is possible to define the product topology on arbitrary (not necessarily finite) products. This is beyond the scope of the course.

4.3 Quotient topology

Let X be a topological space and let R be an equivalence relation on X. This means that $R \subseteq X \times X$ (as usual we write $x \sim y$ instead of $(x, y) \in R$), and R is reflexive, symmetric and transitive.

For $x \in X$ we let $q(x)$ be the equivalence class of x: $q(x) = \{y \in X : y \sim x\}$.

The set X/R of all equivalence classes is called the *quotient set* of X by R.

The map $q : X \to X/R$, $x \mapsto q(x)$, is called the *quotient map*. We define the *quotient topology* on X/R (induced by the topology of X) as the family

$$\{V \subseteq X/R : q^{-1}(V) \text{ is open in } X\}$$

We check this is a topology:

(i) $q^{-1}(\emptyset) = \emptyset$ and $q^{-1}(X/R) = X$, so $\emptyset$ and X/R are open.

(ii) $q^{-1}\left(\bigcup_{i\in I} V_i\right) = \bigcup_{i\in I} q^{-1}(V_i)$, so union of open sets is open.

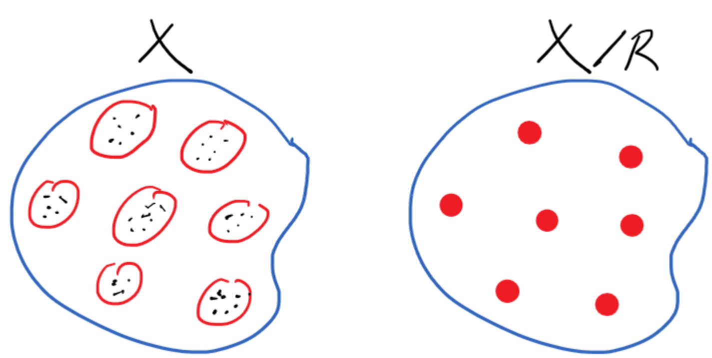
(iii) $q^{-1}(U \cap V) = q^{-1}(U) \cap q^{-1}(V)$, so $U \cap V$ is open if U and V are.

Remarks:

1. The quotient map $q : X \to X/R$ is continuous. Indeed, if V is open in X/R, then $q^{-1}(V)$ is open in X by definition.

2. For $x \in X$ and $t \in X/R$, we have $x \in t \iff t = q(x)$. This is because the equivalence classes partition X. It follows that for $V \subseteq X/R$ we have

$$q^{-1}(V) = \{x \in X : q(x) = V\} = \{x \in X : \exists t \in V \; x \in t\} = \bigcup \{t : t \in V\}$$



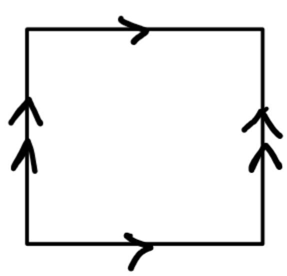
Examples:

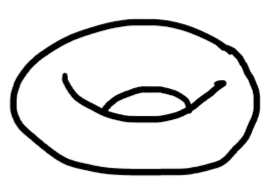
1. $\mathbb{R}$ in the usual topology is also an abelian group under addition. Then $\mathbb{Z} \leq \mathbb{R}$ and we can form the quotient group $\mathbb{R}/\mathbb{Z}$, which is an example of a quotient space with relation $x \sim y \iff x - y \in \mathbb{Z}$. Elements of $[0, 1]$ represent different cosets except for $0 + \mathbb{Z} = 1 + \mathbb{Z}$: in the quotient space, 0 and 1 are "glued together". There are no other cosets, so the quotient space $\mathbb{R}/\mathbb{Z}$ is a circle. More precisely, $\mathbb{R}/\mathbb{Z}$ in the quotient topology is homeomorphic to a circle. This is not obvious.

2. This time consider the quotient group $\mathbb{R}/\mathbb{Q}$. The quotient map $q : \mathbb{R} \to \mathbb{R}/\mathbb{Q}$ is a group homeomorphism. Assume that $V \subseteq \mathbb{R}/\mathbb{Q}$ is open and non-empty. Then $q^{-1}(V)$ is a non-empty open subset of $\mathbb{R}$, and hence $(a, b) \subseteq q^{-1}(V)$ for some $a < b$ in $\mathbb{R}$. Given $x \in \mathbb{R}$, there exists $r \in \mathbb{Q} \cap (a-x, b-x)$. It follows that $x+r \in (a, b)$, and hence $q(x) = q(x+r) \in V$ and $x \in q^{-1}(V)$. We have shown that $q^{-1}(V) = \mathbb{R}$, and thus $V = q(q^{-1}(V)) = \mathbb{R}/\mathbb{Q}$. Thje quotient topology is the indiscrete topology! So a quotient of a metrizable space need not be metrizable.

3. Consider the unit square $Q = [0, 1] \times [0, 1]$ in $\mathbb{R}^2$ with the following equivalence relation:

$$(x, y) \sim (x', y') \iff \begin{cases} (x, y) = (x', y') \text{ or} \\ x = x', \{y, y'\} = \{0, 1\} \text{ or} \\ \{x, x'\} = \{0, 1\}, y = y' \end{cases}$$





From IA Numbers & Sets

Let X be a set and R be an equivalence relation on X. Let $q : X \to X/R$ be the quotient map.

Suppose Y is another set and a function $f : X \to Y$ *respects* the relation R:

$$\forall x, y \in X x \sim y \implies f(x) = f(y)$$

Then there is a unique function $\tilde{f} : X/R \to Y$ such that $f = \tilde{f} \circ q$. So the following diagram

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

commutes.

Proof: given $z \in X/R$, write $z = q(x)$ for some $x \in X$, then set $\tilde{f}(z) = f(x)$ (easy to check this is well defined).

Note:

1. $\mathrm{Im}(f) = \mathrm{Im}(\tilde{f})$ so f surjective $\implies$ $\tilde{f}$ surjective.
2. Supppose f *fully* respects R:

   $$\forall x, y \in X \; x \sim y \iff f(x) = f(y)$$

   Then $\tilde{f}$ is injective.

Proposition 4.10. *Let X, Y be topological spaces, let R be an equivalence relation on X and let $q : X \to X/R$ be the quotient map. Let $f : X \to Y$ be a function that respects the relation R. Then the following hold*

(i) If f is continuous, then $\tilde{f}$ is continuous.

(ii) If f is an open map, then $\tilde{f}$ is an open map.

(Here $\tilde{f} : X/R \to Y$ is the unique map with $\tilde{f} \circ q = f$) In particular if f is continuous, surjective and fully respects R, then $\tilde{f}$ is a continuous bijection. If in addition, f is an open map, then $\tilde{f}$ is a homeomorphism.

Proof.

(i) Let V be an open subset of Y. Is $\tilde{f}^{-1}(V)$ open in X/R? We have $q^{-1}(\tilde{f}^{-1}(V)) = (\tilde{f} \circ q)^{-1}(V) = f^{-1}(V)$, which is open in X since f is continuous. Hence $\tilde{f}^{-1}(V)$ is open in X/R, and so $\tilde{f}$ is continuous.

(ii) Let V be an open subset of X/R. Then $U = q^{-1}(V)$ is open in X and $q(U) = q(q^{-1}(V)) = V$. Hence $\tilde{f}(V) = \tilde{f}(q(U)) = f(U)$, which is open in Y. Thus, $\tilde{f}$ is open.

□

Example: $\mathbb{R}/\mathbb{Z}$ is homeomorphic to $S^1 = \{x \in \mathbb{R}^2 : ||x|| = 1\}$.

Define $f : \mathbb{R} \to S^1$ by $f(t) = (\cos(2\pi t), \sin(2\pi t))$. Then f is continuous, surjective andfor $s, t \in \mathbb{R}$, we have $s - t \in \mathbb{Z} \iff f(s) = f(t)$. So by Proposition 10, there is a unique map $\tilde{f} : \mathbb{R}/\mathbb{Z} \to S^1$ with $f = \tilde{f} \circ q$, and moreover, $\tilde{f}$ is a continuous bijection.

It remains to check that f is an open map. So let U be an open subset of $\mathbb{R}$, and assume for contradiction that $f(U)$ is not open in S^1. Then its complement is not closed, so it contains a sequence (z_n) that converges to some $z \in f(U)$. For each $n \in \mathbb{N}$ choose $x_n \in [0, 1]$ such that $z_n = f(x_n)$. Note that $x_n \notin U$ since $z_n \notin f(U)$. After passing to a subsequence, we may assume that $x_n \to x$ for some $x \in [0, 1]$ (Bolzano-Weierstrass). Then $f(x_n) \to f(x)$ by continuity, and so $f(x) = z$. Since $z \in f(U)$, we have $z = f(y)$ for some $y \in U$. Then $k = y - x \in Z$. Now $f(x_m + k) = f(x_n) = z_n \notin f(U)$, and thus $x_n + k \notin U$. On the other hand, $x_n + k \to x + k = y$. Since $\mathbb{R} \setminus U$ is closed, we have $y \notin U$ - a contradiction.

Later we will have a much better way of proving f is open.

Proposition 4.11. *Let X be a topological space and R an equivalence relation on X.*

(a) If X/R is Hausdorff, then R is closed in $X \times X$.

(b) If R is closed in $X \times X$ and $q : X \to X/R$ is open, then X/R is Hausdorff.

Proof.

(a) Given $(x, y) \in X \times X \setminus R$, we have $x \not\sim y$, so $q(x) \neq q(y)$. Thus, there exist disjoint open sets S, T in X/R such that $q(x) \in S$ and $q(y) \in T$. Then $U = q^{-1}(S)$ and $V = q^{-1}(T)$ are disjoin open sets in X with $x \in U$ and $y \in V$. Now for all $a \in U$ and $b \in V$, we have $q(a) \in S$ and $q(b) \in T$, and so $a \not\sim b$, i.e $(a, b) \notin R$. It follows that $(x, y) \in U \times V \subseteq X \times X \setminus R$. This proves that R has open complement, and thus R is closed.

(b) Given $z \neq w$ in X/R, choose $x, y \in X$ with $z = q(x)$ and $w = q(y)$. Then $(x, y) \notin R$, i.e. the open set $X \times X \setminus R$ contains (x, y). Thus, there exist open sets U, V in X such that $(x, y) \in U \times V \subseteq X \times X \setminus R$. Since q is an open map, $q(U)$ and $q(V)$ are open sets in X/R with $z = q(x) \in q(U)$ and $w = q(y) \in q(V)$. Finally, $q(U) \cap q(V) = \emptyset$ since otherwise, there exist $a \in U$ and $b \in V$ with $q(a) = q(b)$, and thus $(a, b) \in R \cap U \times V$, contradicting the choice of U, V.

$\square$

5 Connectedness

Recall the Intermediate Value Theorem (IVT): Let $f : I \to \mathbb{R}$ be a continuous function on the interval I. Given $x < y$ in I and $c \in \mathbb{R}$ strictly between $f(x)$ and $f(y)$, there exists $z \in (x, y)$ such that $f(z) = c$.

Note: A subset I of $\mathbb{R}$ is an interval if the following holds: for all $x, y, z \in \mathbb{R}$ if $x < y < z$ and $x, z \in I$, then $y \in I$. So the IVT says that the continuous image of an interval is an interval.

Question: For what topological spaces does the IVT hold?

Example: The function

$$f : X = [0,1) \cup (1,2] \to \mathbb{R},\ x \mapsto \begin{cases} 0 \text{ if } x \in [0,1) \\ 1 \text{ if } x \in (1,2] \end{cases}$$

is continuous but its image is not an interval.

Definitions. A topological space X is *disconnected* if there exist subsets U and V of X such that

- $U \neq \emptyset$ and $V \neq \emptyset$
- U and V are open in X
- $U \cap V = \emptyset$
- $U \cup V = X$

In this case, we say that the subsets U and V of X *disconnect* X.

We say X is *connected* if X is not disconnected.

Theorem 5.1. *For a topological space X the following are equivalent*

(i) X is connected

(ii) $f : X \to \mathbb{R}$ continuous $\implies$ $f(X)$ is an interval

(iii) $f : X \to \mathbb{Z}$ continuous $\implies$ f is constant

Proof.
To show (i) $\implies$ (ii): suppose $f : X \to \mathbb{R}$ is continuous but $f(X)$ is not an interval. So there exists $a < b < c$ in $\mathbb{R}$ with $a, c \in f(X)$ and $b \notin f(X)$. Choose $x, z \in X$ with $f(x) = a$ and $f(z) = c$. We show that $U = f^{-1}(-\infty, b)$ and $V = f^{-1}(b, \infty)$ disconnect X, which will contradict (i). Indeed, U and V are non-empty ($x \in U$ and $z \in V$), open (as f is continuous), disjoint and $U \cup V = X$

since $b \notin f(X)$.

To show (ii) $\implies$ (iii): The inclusion map $\iota : \mathbb{Z} \to \mathbb{R}$ is continuous. So if $f : X \to \mathbb{Z}$ is continuous, then so is $g = \iota \circ f : X \to \mathbb{R}$. Then $g(X)$ is an interval by (ii) but at the same time $g(X) = f(X) \subseteq \mathbb{Z}$. So f must have been constant.

To show (iii) $\implies$ (i): Assume that U and V disconnect X. Define

$$f : X \to \mathbb{Z},\ x \mapsto \begin{cases} 0 \text{ if } x \in U \\ 1 \text{ if } x \in V \end{cases}$$

Then for any $A \subseteq \mathbb{Z}$, we have $f^{-1}(A)$ is one of $\emptyset, U, V$ or X depending on whether $A \cap \{0,1\}$ is $\emptyset, \{0\}, \{1\}$ or $\{0,1\}$, respectively. Thus, f is continuous but not constant as U and V are non-empty. □

Corollary 5.2. *Let $X \subseteq \mathbb{R}$. Then X is connected $\iff$ X is an interval.*

Proof. To show "$\implies$": Then inclusion map $\iota : X \to \mathbb{R}$ is continuous, so by Theorem 5.1 (ii) its image X is an interval.

Alternatively, if X is not an interval, then there exist $a < b < c$ in $\mathbb{R}$ with $a, c \in X$ and $b \notin X$. It follows that $U = (-\infty, b) \cap X$ and $V = (b, \infty) \cap X$ disconnect X.

To show "$\impliedby$": By the IVT, (ii) in Theorem 5.1 holds, and hence X is connected.

We now give a direct proof. Assume U and V disconnect X. Fix $x \in U$ and $y \in V$. Assume wlog $x < y$. Let $z = \sup(U \cap [x, y])$. Then $z \in [x, y] \subseteq X$. We show that $z \in U \cap V$. For each $n \in \mathbb{N}$, there exists $x_n \in U \cap [x, y]$ with $z - \frac{1}{n} < x_n \leq z$. Then $x_n \to z$ and hence $z \in U$ as $U = X \setminus V$ is closed in X. It follows that $z < y$ (otherwise $z \in V$ and we are done). Hence for large n, we have $z + \frac{1}{n} \in [x, y]$ and so $z + \frac{1}{n} \in V$. Since $V = X \setminus U$ is closed in X, it follows that $z \in V$. □

More examples:

1. Any indiscrete space is connected.

2. The cofinite topology on an infinite set is connected.

3. The discrete topology on a set X of size at least 2 is disconnected. Indeed, for any $x \in X$, the subsets $\{x\}$ and $X \setminus \{x\}$ disconnect X.

Lemma 5.3. *Let Y be a subspace of a topological space X. Then Y is disconnected iff there exist open subsets U and V of X such that $U \cap Y \neq \emptyset$, $V \cap Y \neq \emptyset$, $U \cap V \cap Y = \emptyset$ and $Y \subseteq U \cup V$.*

Proof. To show " $\Longrightarrow$ ": Let subsets U' and V' of Y disconnect Y. By definition of the subspace topology, there are open sets U and V in X such that $U' = U \cap Y$ and $V' = V \cap Y$. Then $U \cap Y = U' \neq \emptyset, V \cap Y = V' \neq \emptyset, U \cap V \cap Y = U' \cap V' = \emptyset$ and $Y = U' \cup V' \subseteq U \cup V$.

To show " $\Longleftarrow$ ": Now assume that U and V are open subsets of X such that $U \cap Y \neq \emptyset$, $V \cap Y \neq \emptyset$, $U \cap V \cap Y = \emptyset$ and $Y \subseteq U \cup V$. Then $U' = U \cap Y$ and $V' = V \cap Y$ disconnect Y. $\square$

Note: If Y is a subspace of X, and U and V are open subsets of X such that $U \cap Y \neq \emptyset$, $V \cap Y \neq \emptyset$, $U \cap V \cap Y = \emptyset$ and $Y \subseteq U \cup V$, then we will say that the open subsets U, V of X disconnect Y.

Proposition 5.4. *Let Y be a subspace of a topological space X. If Y is connected, then its closure $\overline{Y}$ in X is also connected.*

Proof. Assume $\overline{Y}$ is disconnected. Then by Lemma 5.3, then there are open subsets U, V in X such that $U \cap \overline{Y} \neq \emptyset$, $V \cap \overline{Y} \neq \emptyset$, $U \cap V \cap \overline{Y} = \emptyset$ and $\overline{Y} \subseteq U \cup V$. Then $U \cap V \cap Y = \emptyset$ and $Y \subseteq U \cup V$. If in addition, we also have $U \cap Y \neq \emptyset$ and $V \cap Y \neq \emptyset$, then Y is disconnected by Lemma 5.3. So one of $U \cap Y$ and $V \cap Y$ must be empty. Wlog $U \cap Y = \emptyset$, and so Y is a subset of the closed set $X \setminus U$. It follows that $\overline{Y} \subseteq X \setminus U$ contradicting $U \cap \overline{Y} \neq \emptyset$. $\square$

Remarks:

1. More generally, if Y is connected and $Y \subseteq Z \subseteq \overline{Y}$, then Z is connected. Indeed, the closure of Y in Z is $\overline{Y} \cap Z$ (proposition 4.6).

2. An alternative proof of proposition 5.4 is via Theorem 1 (iii).

Theorem 5.5. *Let $f : X \to Y$ be a continuous function. If X is connected, then $f(X)$ is also connected (the continuous image of a connected space is connected).*

Proof. Assume that U, V are open subsets of Y that disconnect $f(X)$. Then $f^{-1}(U)$ and $f^{-1}(V)$ are open in X since f is continuous.

$f^{-1}(U)$ is non-empty since $U \cap f(X) \neq \emptyset$, and similarly $f^{-1}(V)$ is non-empty. If $x \in f^{-1}(U) \cap f^{-1}(V)$, then $f(x) \in U \cap V \cap f(X) = \emptyset$. This contradiction shows that $f^{-1}(U) \cap f^{-1}(V) = \emptyset$.

Finally, $f^{-1}(U) \cup f^{-1}(V) = f^{-1}(U \cup V) = X$ since $U \cup V \supseteq f(X)$. So $f^{-1}(U)$ and $f^{-1}(V)$ disconnect X contradicting the assumption that X is connected. $\square$

Remarks:

1. Connectedness is a topological property.

2. If $f : X \to Y$ is a continuous function and $A \subseteq X$ is a connected subset of X, then $f(A)$ is connected (apply Theorem 5.5 to $f|_A : A \to Y$).

Corollary 5.6. *The quotient space of a connected space is connected.*

Proof. Trivial by the theorem. $\square$

Example: Consider the subset $Y = \{(x, \sin\frac{1}{x}) : x > 0\}$ of $\mathbb{R}^2$. The function $f : (0, \infty) \to \mathbb{R}^2$, $x \mapsto (x, \sin\frac{1}{x})$, is continuous, and hence its image Y is connected. It follows that $\overline{Y}$ is also connected by proposition 5.4. Note that

$$\overline{Y} = Y \cup \{(0, y) : -1 \leq y \leq 1\}$$

Indeed, let $-1 \leq y \leq 1$. For $n \in \mathbb{N}$, the function $x \mapsto \frac{1}{x}$ maps $(0, \frac{1}{n})$ onto (n, ∞), and hence $\sin\frac{1}{x_n} = y$ for some $x_n \in (0, \frac{1}{n})$. Then $(x_n, \sin\frac{1}{x_n}) \to (0, y)$, and hence $(0, y) \in \overline{Y}$. This shows that $\tilde{Y} = Y \cup \{(0, y) : -1 \leq y \leq 1\}$ is contained in $\overline{Y}$.

Conversely, $\overline{Y} \subseteq \tilde{Y}$ follows if we show that $\tilde{Y}$ is closed. So assume $(x_n, y_n) \to (x, y)$ in $\mathbb{R}^2$ with $(x_n, y_n) \in \tilde{Y}$ for all $n \in \mathbb{N}$. Since $x_n \geq 0$ and $-1 \leq y_n \leq 1$ for all n, we have $x \geq 0$ and $-1 \leq y \leq 1$. So if $x = 0$, then $(x, y) \in \tilde{Y}$. Otherwise, $x > 0$, so $x_n > 0$ for all large n. Now, if $x_n > 0$, then $y_n = \sin\frac{1}{x_n}$. Thus, $y_n \to \sin\frac{1}{x}$ and $(x, y) = (x, \sin\frac{1}{x}) \in \tilde{Y}$.

Lemma 5.7. *Let X be a topological space and let $\mathcal{A}$ be a family of connected subsets of X. Assume that $A \cap B \neq \emptyset$ for all $A, B \in \mathcal{A}$. Then $\bigcup_{A \in \mathcal{A}} A$ is connected.*

Proof. Set $Y = \bigcup_{A \in \mathcal{A}} A$ and assume that $f : Y \to \mathbb{Z}$ is continuous. We show that f must be constant. Then by Theorem 5.1 it will follow that Y is connected.

For each $A \in \mathcal{A}$, the restriction $f|_A : A \to \mathbb{Z}$ is continuous, and thus constant by Theorem 5.1. For any $A, B \in \mathcal{A}$, the functions $f|_A$ and $f|_B$ must take the same constant value since $A \cap B \neq \emptyset$. It follows that f itself must be a constant function. □

Theorem 5.8. *Let X and Y be connected topological spaces. Then $X \times Y$ is connected in the product topology.*

Proof. Wlog X and Y are non-empty.

Fix $x_0 \in X$ and consider the map $f : Y \to X \times Y$, $y \mapsto (x_0, y)$. We show that f is continuous. Given open sets U in X and V in Y, if $x_0 \in U$, then $f^{-1}(U \times V) = V$ and if $x_0 \notin U$, then $f^{-1}(U \times V) = \emptyset$. It follows by Proposition 4.8 (ii) that f is continuous.

Alternatively, recall that $q_X : X \times Y \to X$ and $q_Y : X \times Y \to Y$ denote coordinate projections. Then $q_X \circ f : Y \to X$ is the constant function $y \mapsto x_0$ and $q_Y \circ f : Y \to Y$ is the identity function $y \mapsto y$, and thus both are continuous. It follows by Proposition 4.9 (ii) that f is continuous.

Since Y is connected and f is continuous, it follows by Theorem 5.5 that $\{x_0\} \times Y$, the image of f, is connected. Similarly, for any $y_0 \in Y$, the subspace $X \times \{y_0\}$ of $X \times Y$ is connected.

From here, we give two ways of completing the proof. In the first method, we assume that U and V disconnect $X \times Y$. Fix $x_0 \in X$. Since $\{x_0\} \times Y$ is connected, U and V cannot disconnect it. Hence either U or V has empty intersection with $\{x_0\} \times Y$, and so wlog we may assume that $\{x_0\} \times Y \subseteq U$. Similarly, for any $y \in Y$, the set $X \times \{y\}$ is contained either in U or in V, and since $(x_0, y) \in U$, we must in fact have $X \times \{y\} \subseteq U$. Thus $X \times Y \subseteq U$ and $V = \emptyset$, a contradiction.

In the second method, we first observe that for $x_0 \in X$ and $y_0 \in Y$, the connected sets $\{x_0\} \times Y$ and $X \times \{y_0\}$ have non-empty intersection $\{(x_0, y_0)\}$. It follows by Lemma 5.7 that $\{x_0\} \times Y \cup X \times \{y_0\}$ is connected. Now fix $x_0 \in X$ and for each $y \in Y$ set $A_y = \{x_0\} \times Y \cup X \times \{y\}$. For $y, y' \in Y$, we have $A_y \cap A_{y'} \supseteq \{x_0\} \times Y$, and hence $A_y \cap A_{y'}$ is non-empty. Hence by Lemma 5.7, it follows that $\bigcup_{y \in Y} A_y = X \times Y$ is connected. □

Example: $\mathbb{R}^n$ is connected for any $n \in \mathbb{N}$.

Remarks:

1. In the proof we showed that $y \mapsto (x, y)$ is a continuous map $Y \to X \times Y$ with image $\{x\} \times Y$. In fact, this map is injective with inverse $\{x\} \times Y \to Y$ being the second coordinate projection, which is continuous. It follows that $\{x\} \times Y$ is homeomorphic to Y for any $x \in X$. Similarly, $X \times \{y\}$ is homeomorphic to X for any $y \in Y$.

2. The converse of Theorem 5.8 also holds provided X and Y are non-empty: if $X \times Y$ is connected, then so are X and Y. Simply apply Theorem 5.5 to the coordinate projections.

5.1 Components

We define a relation $\sim$ on a topological space X as follows:

$$\forall x, y \in X \ x \sim y \iff \exists \text{connected subset } A \text{ of } X \text{ such that } x, y \in A$$

We check that this is an equivalence relation. For any $x \in X$, the set $\{x\}$ is connected, and so $x \sim x$. Symmetry is immediate from the definition. Finally, if $x \sim y$ and $y \sim z$, then there exist connected subsets A, B of X with $x, y \in A$ and $y, z \in B$. Since $A \cap B \neq \emptyset$, it follows from Lemma 5.7 that $A \cup B$ is connected, and hence $x \sim z$. For $x \in X$, we shall write C_x for the equivalence class containing x. Equivalence classes are called *connected components* of X.

Proposition 5.9. *Connected components of a topological space X are non-empty, maximal (with respect to inclusion) connected subsets of X. Connected components are closed and they partition X (i.e X is the disjoint union of its connected components).*

Proof. Let C be a connected component of X. Then $C = C_x$ for some $x \in X$. Thus $C \neq \emptyset$ since $x \in C$.

If A is connected subset of X with $x \in A$, then $y \sim x$ for all $y \in A$, and hence $A \subseteq C$. It follows that if A is a connected subset of X with $A \supseteq C$, then $A = C$.

For each $y \in C$ there is a connected subset A_y of X such that $x, y \in A_y$. Then $A_y \cap A_{y'} \neq \emptyset$ for all $y, y' \in C$, and hence $A = \bigcup_{y \in C} A - y$ is connected by Lemma 5.7 and $A \supseteq C$. By above, $A = C$, and thus C is connected.

By Proposition 5.4, the closure C° of C in X is also connected. Hence by the maximality of C, we have $C = C^\circ$ and C is closed.

Finally, X is the disjoint union of components, since in general, equivalence classes of an equivalence relation on a set partition the set. $\square$

Definitions. Let X be a topological space. For $x, y \in X$, a *path* from x to y in X is a continuous function $\gamma : [0, 1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = y$. We say that a topological space X is *path-connected* if for all $x, y \in X$ there is a path from x to y in X.

Note: what we called a path is sometimes called a *continuous path*

Theorem 5.10. *Every path-connected topological space is connected.*

Proof. Assume X is path connected but not connected. Let U, V disconnect X. Fix $x \in U, y \in V$ and a continuous function $\gamma : [0,1] \to X$ with $\gamma(0) = x$ and $\gamma(1) = y$. It is straightforward to check that $\gamma^{-1}(U)$ and $\gamma^{-1}(V)$ disconnect $[0,1]$ contradicting Corollary 5.2. □

Example: We show that the converse of Theorem 5.10 is false. Recall that the subset $X = \{(x, \sin\frac{1}{x}) : x > 0\} \cup \{(0,y) : -1 \leq y \leq 1\}$ of $\mathbb{R}^2$ is connected. We now show that X is not path-connected.

Assume $\gamma : [0,1] \to X$ is continuous with $\gamma(0) = (0,0)$ and $\gamma(1) = (1, \sin(1))$. Let γ_1 and γ_2 be the components of γ. So $\gamma(t) = (\gamma_1(t), \gamma_2(t))$ for all $t \in [0,1]$. Since γ is continuous, so are γ_1 and γ_2. Assume that $x = \gamma_1(t) > 0$ for some t (e.g $t = 1$). By the IVT the image of the interval $(0,t)$ under γ_1 contains $(0,x)$. Choose $n \in \mathbb{N}$ such that $\frac{1}{2\pi n} \in (0,x)$ and $s \in (0,t)$ such that $\gamma_1(s) = \frac{1}{2\pi n}$. Then $\gamma_2(s) = \sin\frac{1}{\gamma_1(s)} = 0$. Similarly, there exists $s \in (0,t)$ with $\gamma_1(s) = \frac{1}{2\pi n + \frac{\pi}{2}}$ and $\gamma_2(s) = 1$.

We now inductively construct $t_1 > t_2 > t_3 > \ldots > 0$ in $[0,1]$ such that $\gamma_2(t_n) = 0$ if n is even and $\gamma_2(t_n) = 1$ otherwise. Let $t_n \to t$ in $[0,1]$. Then $\gamma_2(t_n) \to \gamma_2(t)$. However, $(\gamma_2(t_n))$, is not convergent, a contradiction.

Lemma 5.11. *Let X be a topological space and let A, B be closed subsets with $X = A \cup B$. If $f : X \to Y$ is a function such that $f|_A : A \to Y$ and $f|_B : B \to Y$ are continuous, then f is continuous.*

Proof. Let V be a closed subset of Y. Since $f|_A$ is continuous, it follows that $(f|_A)^{-1}(V) = A \cap f^{-1}(V)$ is closed in A, and hence also in X (Proposition 4.6). Similarly, $B \cap f^{-1}(V)$ is closed in X. It follows that

$$f^{-1}(V) = (A \cap f^{-1}(V)) \cup (B \cap f^{-1}(V))$$

is also closed in X. Finally, f is continuous by Proposition 4.8 (i). □

Note: This result is sometimes called the gluing lemma.

5.2 Path-connected components

Let X be a topological space. We define a relation on X as follows

$$\forall x, y \in X \ x \sim y \iff \exists \text{path from } x \text{ to } y \text{ in } X$$

For $x \in X$, the constant function with value x is continuous, which shows that $x \sim x$. If γ is a path from x to y, then $t \mapsto \gamma(1-t)$ is a path from y to x, which shows symmetry. Finally, if γ is a path from x to y and δ is a path from y to z, then

$$\eta(t) = \begin{cases} \gamma(2t) & \text{if } t \in [p, \frac{1}{2}] \\ \delta(2t-1) & \text{if } t \in [\frac{1}{2}, 1] \end{cases}$$

defines a path from x to z. Indeed, this is well-defined at $\frac{1}{2}$ since $\gamma(1) = \delta(0) = y$. Secondly, the sets $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$ are closed in $[0,1]$, $[0,1] = [0, \frac{1}{2}] \cup [\frac{1}{2}, 1]$ and the functions $\eta|_{[0,\frac{1}{2}]}$, $\eta|_{[\frac{1}{2},1]}$ are continuous. So η is continuous by Lemma 5.11. This shows that $\sim$ is transitive.

Equivalence classes of this relation are called *path-connected components* of X. It is immediate from the fact that $\sim$ is an equivalence relation that path-connected components are path-connected.

Theorem 5.12. *An open subset U of $\mathbb{R}^n$ is connected $\iff$ path-connected.*

Proof. To show "$\Leftarrow$": This is true in general by Theorem 5.10.

To show "$\Rightarrow$": Fix $x_0 \in U$ and set $P = \{x \in U : \exists \text{path from } x_0 \text{ to } x\}$. In other words, P is the path-connected component of x_0 in U. We show that P is both open and closed in U. Since P and $U \setminus P$ cannot disconnect U, one of them has to be empty. Since $x_0 \in P$, it follows that $P = U$, and thus U is path-connected.

Fix $x \in P$. Since U is open, there exists $r > 0$ such that $D_r(x) \subseteq U$. Now, for any $y \in D_r(x)$, there is a path from x to y inside $D_r(x)$. For example, the straight line segment $t \mapsto (1-t)x + ty$ is continuous and takes values in $D_r(x)$ since

$$||((1-t)x + ty) - x|| = ||t(y-x)|| = t||y-x|| < r$$

This shows that $D_r(x) \subseteq P$, and thus P is open in U.

Now fix $x \in U \setminus P$ and again choose $r > 0$ such that $D_r(x) \subseteq U$. As before, for any $y \in D_r(x)$, we have $y \sim x$. Hence, if there exists $y \in D_r(x) \cap P$, then $y \sim x$ and $y \sim x_0$, which gives the contraction $x \in P$. It follows that $D_r(x) \subseteq U \setminus P$, and thus $U \setminus P$ is open in U. □

5.3 An application of connectedness

For $n \geq 2$, $\mathbb{R}$ and $\mathbb{R}^n$ are not homeomorphic.

Proof. Assume that $f : \mathbb{R} \to \mathbb{R}^n$ is a homeomorphism with inverse $g : \mathbb{R}^n \to \mathbb{R}$. Then $f|_{\mathbb{R}\setminus\{0\}}$ is a continuous bijection from $\mathbb{R} \setminus \{0\}$ onto $\mathbb{R}^n \setminus \{f(0)\}$ whose inverse is $g|_{\mathbb{R}^n\setminus\{f(0)\}}$ which is also continuous. Thus, $\mathbb{R} \setminus \{0\}$ and $\mathbb{R}^n \setminus \{f(0)\}$ are homeomorphic. Now, $\mathbb{R} \setminus \{0\}$ is not connected since it is not an interval. However, it is easy to see that $\mathbb{R}^n \setminus \{f(0)\}$ is path-connected, and hence connected. Since connectedness is a topological property, we have reached the desired contradiction. □

6 Compactness

Recall that a continuous real-valued function on a closed bounded interval is bounded and attains its bounds.

Question: For which topological spaces X is it true that every continuous function $X \to \mathbb{R}$ is bounded?

Some answers:

1. For finite spaces
2. For spaces X with the following property:

   For every continuous function $f : X \to \mathbb{R}$, there exist $n \in \mathbb{N}$ and subsets $A_1, \ldots, A_n$ of X such that $X = \bigcup_{k=1}^{n} A_k$ and f is bounded on each A_k.

Note: Let X be a topological space and $f : X \to \mathbb{R}$ continuous. For $x \in X$ let $U_x = f^{-1}(f(x) - 1, f(x) + 1)$. Then U_x is open, $x \in U_x$ and for every $y \in U_x$ we have
$$|f(y)| \leq |f(y) - f(x)| + |f(x)| \leq |f(x)| + 1$$
So $X = \bigcup_{x \in X} U_x$ and f is bounded on each U_x. If there exists a finite set $F \subseteq X$ such that $X = \bigcup_{x \in F} U_x$, then f is bounded.

Definitions. Let X be a topological space.

An *open cover* for X is a family $\mathcal{U}$ of open subsets of X such that $\bigcup_{U \in \mathcal{U}} U = X$. A *subcover* of an open cover $\mathcal{U}$ for X is a subfamily $\mathcal{V} \subseteq \mathcal{U}$ such that $\bigcup_{U \in \mathcal{V}} U = X$. If $\mathcal{V}$ is finite, then we call this a *finite subcover* of $\mathcal{U}$.

A topological space is *compact* if every open cover for X has a finite subcover.

Theorem 6.1. *Let X be a compact topological space and $f : X \to \mathbb{R}$ a continuous function. Then f is bounded, and moreover, if $X \neq \emptyset$, then f attains its bounds.*

Proof. The proof of boundedness was essentially given in the Note above. Here is another proof.

For $n \in \mathbb{N}$ let $U_n = \{x \in X : |f(x)| < n\}$. Then $U_n = f^{-1}(-n, n)$ is open in X. For every $x \in X$ there exists $n \in \mathbb{N}$ with $|f(x)| < n$. It follows that $X = \bigcup_{n \in \mathbb{N}} U_n$. Since X is compact, there is a finite $F \subseteq \mathbb{N}$ with $X = \bigcup_{n \in F} U_n = U_N$ where $N = \max F$. So f is bounded by N.

Next, we let $m = \inf_X f$ and $M = \sup_X f$. Assume that there is no $y \in X$ with $f(y) = m$. Then for every $y \in X$ we have $f(y) > m$. Let $y \in X$. Choose $\alpha_y \in \mathbb{R}$ with $m < \alpha_y < f(y)$ and set $U_y = f^{-1}(\alpha_y, \infty)$. Then U_y is open, it contains y and f is bounded below by α_y on U_y. In particular, $\{U_y : y \in X\}$ is an open cover for X. Hence there is a finite $F \subseteq X$ such that $\bigcup_{y \in F} U_y$. Let $\alpha = \min\{\alpha_y : y \in F\}$. Then $\alpha > m$ and f is bounded below by α on X. This contradiction shows that $m = f(y)$ for some $y \in X$. A similar argument shows M is also attained. $\square$

Remark: "Compactness is the next best thing after finiteness"

Lemma 6.2. *Let Y be a subspace of a topological space X. Then Y is compact $\iff$ for any family $\mathcal{U}$ of open subsets of X satisfying $Y \subseteq \bigcup_{U \in \mathcal{U}} U$, there is a finite subfamily $\mathcal{V} \subseteq \mathcal{U}$ such that $Y \subseteq \bigcup_{U \in \mathcal{V}} U$.*

Proof. "$\Rightarrow$" Let $\mathcal{U}$ be a family of open subsets of X satisfying $Y \subseteq \bigcup_{U \in \mathcal{U}} U$. Then $\{U \cap Y : U \in \mathcal{U}\}$ is an open cover for Y. Since Y is compact, there is a finite subcover: there is a finite $\mathcal{V} \subseteq \mathcal{U}$ such that $Y = \bigcup_{U \in \mathcal{V}} U \cap Y$. It follows that $Y \subseteq \bigcup_{U \in \mathcal{V}} U$.

"$\Leftarrow$" Let $\mathcal{W}$ be an open cover for Y. For each $W \in \mathcal{W}$ fix an open set $\tilde{W}$ in X such that $W = \tilde{W} \cap Y$. Then $\mathcal{U} = \{\tilde{W} : W \in \mathcal{W}\}$ is a family of open subsets of X satisfying $Y \subseteq \bigcup_{U \in \mathcal{U}} U$. By assumption, there is a finite subfamily of $\mathcal{U}$ whose union still contains Y. So there is a finite $\mathcal{V} \subseteq \mathcal{W}$ such that $\bigcup_{W \in \mathcal{V}} \tilde{W} \supseteq Y$. It follows that $\bigcup_{W \in \mathcal{V}} W = Y$, i.e $\mathcal{V}$ is a finite subcover of $\mathcal{W}$. $\square$

Theorem 6.3. *The unit interval $[0,1]$ is compact.*

Proof. Let $\mathcal{U}$ be a family of open subsets of $\mathbb{R}$ such that $[0,1] \subseteq \bigcup_{U \in \mathcal{U}} U$. For $I \subseteq [0,1]$ we will say that $\mathcal{U}$ *finitely covers* I if there is a finite subfamily $\mathcal{V} \subseteq \mathcal{U}$ satisfying $I \subseteq \bigcup_{U \in \mathcal{V}} U$.

We now make an observation. Suppose $I = J \cap K$ for subsets $I, J, K of [0,1]$. If $\mathcal{U}$ finitely covers J and K, then it finitely covers I.

Now asusme that $\mathcal{U}$ does not finitely cover $[0,1]$. Then at least one of $[0,1/2]$ and $[1/2,1]$ cannot be finitely covered by $\mathcal{U}$. Call that interval $[a_1,b_1]$. It follows that, putting $c = \frac{1}{2}(a_1+b_1)$, at least one of the intervals $[a_1,c]$ and $[c,b_1]$ cannot be finitely covered by $\mathcal{U}$. Call that interval $[a_2,b_2]$. Continue inductively and obtain a nested sequence

$$[0,1] \supseteq [a_1,b_1] \supseteq [a_2,b_2] \supseteq \dots$$

such that for each $n \in \mathbb{N}$ the interval $[a_n,b_n]$ cannot be finitely covered by $\mathcal{U}$ and $b_n - a_n = 2^{-n}$.

Since (a_n) is an increasing, bounded above sequence, it converges to some $x \in [0,1]$. Then $b_n = a^n + 2^{-n} \to x$ as well.

Since $[0,1] \subseteq \bigcup_{U \in \mathcal{U}} U$, there exists $U \in \mathcal{U}$ such that $x \in U$. Since U is open, there exists $\varepsilon > 0$ such that $(x-\varepsilon, x+\varepsilon)$. Since $a_n, b_n \to x$ there exists an

$n \in \mathbb{N}$ with $a_n, b_n \in (x-\varepsilon, x+\varepsilon)$. It follows that $[a_n, b_n] \subseteq (x-\varepsilon, x+\varepsilon) \subseteq U$ contradicting that $[a_n, b_n]$ cannot be finitely covered by $\mathcal{U}$, So $[0,1]$ can be finitely covered by $\mathcal{U}$, and thus $[0,1]$ is compact by Lemma 2. $\square$

More examples:

1. Finite spaces are compact.
2. Any set X with the cofinite topology is compact. Indeed, given an open cover $\mathcal{U}$ for X, fix a non-empty $U \in \mathcal{U}$. Then $F = X \setminus U$ is finite. For each $x \in F$, fix $U_x \in \mathcal{U}$ such that $x \in U_x$. Then $\{U\} \cup \{U_x : x \in F\}$ is a finite subcover of $\mathcal{U}$.
3. If $x_n \to x$ in a topological space X, then the subspace $Y = \{x\} \cup \{x_n : n \in \mathbb{N}$ is compact. To see this, assume $\mathcal{U}$ is a family of open subsets of X covering Y. Then $x \in U$ for some $U \in \mathcal{U}$. Since $x_n \to x$, there exists $N \in \mathbb{N}$ with $x_n \in U$ for $n > N$. Now we simply choose $U_k \in \mathcal{U}$ with $x_k \in U_k$ for each $k = 1, 2, \ldots, N$. Then $\{U\} \cup \{U_k : 1 \leq k \leq N\}$ is a finite subcover.
4. An infinite set X with the discrete topology is not compact. Indeed, the open cover $\{\{x\} : x \in X\}$ has no finite subcover.
5. $\mathbb{R}$ is not compact. Indeed, $\{(-n, n) : n \in \mathbb{N}$ is an open cover for $\mathbb{R}$ without a finite subcover.

Proposition 6.4. *Let Y be a subspace of a topological space X.*

(a) If X is compact and Y is closed in X, then Y is compact.

(b) If X is Hausdorff and Y is compact, then Y is closed in X.

Proof.

(a) Let $\mathcal{U}$ be a family of open subsets of X covering Y (i.e $Y \subseteq \bigcup_{U \in \mathcal{U}} U$).

Then $\mathcal{U} \cup \{X \setminus Y\}$ is an open cover for X. Since X is compact, this has a finite subcover. So there is a finite subfamily $\mathcal{V}$ of $\mathcal{U}$ such that $\mathcal{V} \cup \{X \setminus Y\}$ is still an open cover for X. It follows that $\mathcal{V}$ covers Y. By Lemma 2, it follows that Y is compact.

(b) Fix $x \in X \setminus Y$. For each $y \in Y$, since $x \neq y$, there exist some disjoint open sets U_y, V_y in X such that $x \in U_y$ and $y \in V_y$. Then $\{V_y : y \in Y\}$ is a family of open sets in X covering Y. Since Y is compact, there is a finite $F \subseteq Y$ such thta $Y \subseteq \bigcup_{y \in F} V_y$. It follows that $U = \cap_{y \in F} U_y$ is an open set containing x and disjoint from Y. This shows that $X \setminus Y$ is a neighbourhood of x, and thus $X \setminus Y$ is open and Y is closed.

$\square$

Proposition 6.5. *Let $f : X \to Y$ be a continuous function between topological spaces. If X is compact, then $f(X)$ is compact.*

Proof. Let $\mathcal{U}$ be a family of open subsets of Y covering $f(X)$. Since f is continuous, $f^{-1}(U)$ is open in X for every $U \in \mathcal{U}$ and moreover $X = f^{-1}\left(\bigcup_{U\in\mathcal{U}} U\right) = \bigcup_{U\in\mathcal{U}} f^{-1}(U)$. Thus $\{f^{-1}(U) : U \in \mathcal{U}\}$ is an open cover for X. Since X is compactm this has a finite subcover. So there is a finite subfamily $\mathcal{V} \subseteq \mathcal{U}$ such that $X = \bigcup_{U\in\mathcal{V}} f^{-1}(U)$. It follows that $f(X) \subseteq \bigcup_{U\in\mathcal{V}} U$. Now from Lemma 2 again, we deduce that $f(X)$ is compact. □

Remarks:

1. Compactness is a topological property.
2. If $f : X \to Y$ is continuous and A is a compact subset of X, then $f(A)$ is a compact: simply apply the above to $f|_A : A \to Y$.

Example: For $a < b$ in $\mathbb{R}$, the unit interval $[0, 1]$ is homeomorphic to the interval $[a, b]$ via the map $t \mapsto (1 - t)a + tb$. Hence $[a, b]$ is compact.

Corollary 6.6. *Let R be an equivalence relation on a compact topological space X. Then $X \setminus R$ is compact in the quotient topology.*

Theorem 6.7 (The topological inverse function theorem (TIVT))**.** *Let $f : X \to Y$ be a continuous bijection between topological spaces. If X is compact and Y is Hausdorff, then f is an open map, and hence a homeomorphism.*

Proof. Let U be an open subset of X. Set $K = X \setminus U$. Since f is a bijection, it follows that $f(U) = Y \setminus f(K)$. So it is enough to show that $f(K)$ is closed in Y. Since K is a closed subset of the compact space X, it is compact (Proposition 4(a)).

Since f is continuous, $f(K)$ is the continuous image of the compact set K, and hence compact (Proposition 5).

Since $f(K)$ is a compact subspace of the Hausdorff space Y, it is closed (Proposition 4(b)). □

An application:

The quotient space $\mathbb{R}/\mathbb{Z}$ is homeomorphic to the unit circle $S^1 = \{x \in \mathbb{R}^2 : ||x|| = 1\}$.

Proof. Recall that the map $f : \mathbb{R} \to S^1$, $f(t) = (\cos(2\pi t), \sin(2\pi t))$ fully respects the coset relation ($s \sim t$, i.e $s - t \in \mathbb{Z}$ iff $f(s) = f(t)$), is continuous and surjective. It follows that there is a unique map $\tilde{f} : \mathbb{R}/\mathbb{Z} \to S^1$ such that $f = \tilde{f} \circ q$ (where $q : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ is the quotient map), and moreover $\tilde{f}$ is a continuous bijection (See Proposition 4.10).

Note that for any $x \in \mathbb{R}$, we have $q(x) = q(x - \lfloor x \rfloor)$. Thus $\mathbb{R}/\mathbb{Z} = q(\mathbb{R}) = q([0, 1])$. Since $[0, 1]$ is compact (Theorem 3) and since q is continuous, it follows that $\mathbb{R}/\mathbb{Z}$ is compact (Proposition 5). On the other hand, S^1 is a metric space, and hence Hausdorff. Hence $\tilde{f}$ is a homeomorphism by Theorem 7. □

Theorem 6.8 (Tychonov's Theorem)**.** *The product of compact topological spaces is compact in the product topology.*

Proof. Let X and Y be compact topological spaces. We will show that $X \times Y$ is compact in the product topology. It then follows by induction that the product of any finite number of compact spaces is compact. Theorem 8 remains true for arbitrary products, however this is not covered in this course.

Let $\mathcal{W}$ be an open cover for $X \times Y$. We need to show that $\mathcal{W}$ has a finite subcover. Define

$$\mathcal{U} = \{U \times V : U \text{ open in } X, V \text{ open in } Y \text{ and } \exists W \in \mathcal{W}\ U \times V \subseteq W\}$$

Note that $\mathcal{U}$ is an open cover for $X \times Y$. Indeed, given $z \in X \times Y$, there exists $W \in \mathcal{W}$ with $z \in W$, and by the definition of the product topology, there exist open sets U, V in X, Y respectively, such that $z \in U \times V \subseteq W$.

We next claim that it is enough to show $\mathcal{U}$ has a finite subcover. Indeed, assume that for some $n \in \mathbb{N}$ there exist open sets $U_1, \dots, U_n$ in X and open sets $V_1, \dots, V_n$ in Y such that $U_i \times V_i \in \mathcal{U}$ for $1 \le i \le n$ and $X \times Y = \bigcup_{i=1}^n U_i \times V_i$. For each $i = 1, \dots, n$ we can choose $W_i \in \mathcal{W}$ with $U_i \times V_i \subseteq W_i$. Then $X \times Y = \bigcup_{i=1}^n W_i$, and thus $\{W_1, \dots, W_n\}$ is a finite subcover of $\mathcal{W}/$

Fix $x \in X$. From the proof of Theorem 5.8 we know that $\{x\} \times Y$ is a continuous image of Y, and thus compact by Proposition 5. Since $\mathcal{U}$ covers $\{x\} \times Y$, there is a finite subfamily of $\mathcal{U}$ that covers $\{x\} \times Y$ (Lemma 2). So there exist $n_x \in \mathbb{N}$ and open sets $U_{x,1}, U_{x,2}, \dots, U_{x,n_x}$ in X and open sets $V_{x,1}, V_{x,2}, \dots, V_{x,n_x}$ in Y such that $U_{x,i} \times V_{x,i} \in \mathcal{U}$ for $1 \le i \le n_x$ and $\{x\} \times Y \subseteq \bigcup_{i=1}^{n_x} U_{x,i} \times V_{x,i}$. Wlog we may assume that $x \in U_{x,i}$ for all $i = 1, 2, \dots, n_x$. (If $x \notin U_{x,i}$, then $U_{x,i} \times V_{x,i} \cap \{x\} \times Y = \emptyset$, and so $U_{x,i} \times V_{x,i}$ can be removed from the finite subcover). It follows that $U_x = \bigcap_{i=1}^{n_x} U_{x,i}$ is an open set in X containing x. Morevover, $\bigcup_{i=1}^{n_x} U_{x,i} \times V_{x,i} \supseteq U_x \times Y$ (indeed, given $z \in U_x$ and $y \in Y$, we have $(x, y) \in U_{x,i} \times V_{x,i}$ for some $1 \le i \le n_x$ and hence $(z, y) \in U_{x,i} \times V_{x,i}$).

We carry out the above process for each $x \in X$, and obtain the open cover $\{U_x : x \in X\}$ for X. Since X is compact, there is a finite subset $F \subseteq X$ such that $X = \bigcup_{x \in F} U_x$. It follows that

$$X \times Y = \bigcup_{x \in F} U_x \times Y \subseteq \bigcup_{x \in F} \bigcup_{i=1}^{n_x} U_{x,i} \times V_{x,i}$$

and hence

$$\{U_{x,i} \times V_{x,i} : x \in F,\ 1 \le i \le n_x\}$$

is a finite subcover of $\mathcal{U}$. □

Remark: The coverse of Theorem 8 is true and is easy to prove. If $X \times Y$ is compact and X, Y are non-empty, then X and Y are compact. This is because X, Y are continuous images of $X \times Y$ under the coordinate projections q_X, q_Y respectively.

Theorem 6.9 (Heine-Borel Theorem). *A subset K of $\mathbb{R}^n$ is compact $\iff$ it is closed and bounded.*

Proof. To show "$\Rightarrow$": the function $\mathbb{R}^n \to \mathbb{R}$, $x \mapsto ||x||$, is continuous (indeed, $|||x|| - ||y||| \le ||x - y||$ for all $x, y \in \mathbb{R}^n$). Hence the image of the compact set K is bounded: there exists $M \ge 0$ such that $||x|| \le M$ for all $x \in K$. Thus $K \subseteq B_M(0)$ and K is bounded.

As a compact subset of the Hausdorff space $\mathbb{R}^n$, the set K is closed in $\mathbb{R}^n$ (Proposition 4(b)).

To show "$\Leftarrow$": Fix $M \ge 0$ such that $||x|| \le M$ for all $x \in K$. Then $K \subseteq [-M, M]^n$, By Tychonov's theorem, $[-M, M]^n$ is compact in the product topology. So K is a closed subset of the compact space $[-M, M]^n$, and hence compact by Proposition 4(a). □

Another application of the TIFT:

Let $Q = [0,1]^2$ (the unit square), and let R be the equivalence relation defined as

$$(x,y) \sim (x',y') \iff \begin{cases} (x,y) = (x',y') \text{ or} \\ x = x', \{y,y'\} = \{0,1\} \text{ or} \\ \{x,x'\} = \{0,1\}, y = y' \end{cases}$$

Then the quotient space Q/R is homeomorphic to the torus

$$T^2 = \{((2+\cos\theta)\cos\varphi, (2+\cos\theta)\sin\varphi, \sin\theta) : \theta, \varphi \in [0,2\pi]\} \subseteq \mathbb{R}^3$$

Before proving this, note that $\{(2+\cos\theta, 0, \sin\theta) : \theta \in [0,2\pi]\}$ is the circle with centre $(2,0,0)$ and radius 1 in the xz plane. Applying the matrix

$$\begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

rotates the circle about the z-axis, which sweeps out the surface of a torus.

We now turn to the proof. Define

$f : Q \to T^2$ by $f(s,t) = ((2+\cos(2\pi t))\cos(2\pi s), (2+\cos(2\pi t))\sin(2\pi s), \sin(2\pi t))$

It is easy to check that f fully respects the relation R, Moreover, f is continuous and surjective, It follows that there is a unique map $\tilde{f} : Q/R \to T^2$ such that $f = \tilde{f} \circ q$ (where $q : Q \to Q/R$ is the quotient map) and $\tilde{f}$ is a continuous bijection.

Finally, $Q = [0,1]^2$ is compact by the Heine-Borel theorem. Hence Q/R, the continuous image of Q under the quotient map, is also compact. On the other hand, T^2 is Hausdorff being a subset of the metric space $\mathbb{R}^3$. By Theorem 7 it follows that $\tilde{f}$ is a homeomorphism.

Definition. Let f_n, for each $n \in \mathbb{N}$, and f be scalar functions on a topological space X. We say (f_n) converges *locally uniformly* on X if for every $x \in X$ there is a neighbourhood U of x such that $f_n \to f$ uniformly on U.

Remark: Let f_n for each $n \in \mathbb{N}$, and f be scalar functions on an open subset U of $\mathbb{R}^d$. Then $f_n \to f$ locally uniformly on U iff $f_n \to f$ uniformly on every compact set $K \subseteq U$.

Proof. To show "$\Rightarrow$": assume $K \subseteq U$ and K is compact. For each $x \in L$, fix an open neighbourhood U_x of x such that $U_x \subseteq U$ and $f_n \to f$ uniformly on U_x. Then $\{U_x : x \in K\}$ is a family of open subsets of U that covers K. Since K is compact, there is a finite subset $F \subseteq K$ such that $K \subseteq \bigcup_{x\in F} U_x$. It follows that $f_n \to f$ uniformly on K.

To show "$\Leftarrow$": since U is open, for each $x \in U$ there exists $r > 0$ such that $B_r(x) \subseteq U$. The set $B_r(x)$ is closed and bounded, and hence compact by the Heine-Borel theorem. By assumption, $f_n \to f$ uniformly on $B_r(x)$, which is a neighbourhood of x. $\square$

Sequential Compactness

A topological space X is *sequentially compact* if every sequence in X has a convergent subsequence. I.e, for every sequence (x_n) in X, there is a subsequence (x_{k_n}) of (x_n) and $x \in X$ such that $x_{k_n} \to x$.
Notation: For a sequence $(x_n)_{n=1}^{\infty}$ (in some set S) and for an infinite $M \subseteq \mathbb{N}$, we write $(x_m)_{m\in M}$ for the subsequence $(x_{m_n})_{n=1}^{\infty}$, where $m_1 < m_2 \ldots$ is an enumerate of M in strictly increasing order.

Note that for infinite subsets L, M of $\mathbb{N}$, if $L \subseteq M$, then $(x_n)_{n\in L}$ is a subsequence of $(x_n)_{n\in M}$.
Examples:

1. Every closed and bounded subset of $\mathbb{R}$ is sequentially compact. This follows from the Bolzano-Weierstrass theorem.

2. More generally, every closed and bounded subset X of $\mathbb{R}^n$ is sequentially compact. Indeed, given a sequence $(\mathbf{x}_m)$ in X, write each $\mathbf{x}_m = (x_{m,1}, \ldots, x_{m,n})$ in terms of its coordinates. Since X is bounded, $(x_{m,1})_{m\in\mathbb{N}}$ is a bounded sequence in $\mathbb{R}$, so by Bolzano-Weierstrass there is an infinite $M_1 \subseteq \mathbb{N}$ such that $(x_{m,1})_{m\in M_1}$ is convergent. Now, $(x_{m,2})_{m\in M_1}$ is also bounded, so there is an infinite $M_2 \subseteq M_1$ such that $(x_{m,2})_{m\in M_2}$ is convergent. Continuing this way we find infinite subsets $M_1 \supseteq M_2 \supseteq \ldots \supseteq M_n$ of $\mathbb{N}$ such that $(x_{m,k})_{m\in M_k}$ is convergent for each $k = 1, \ldots, n$. It follows that $(x_{m,k})_{m\in M_n}$ is convergent for each $k = 1, \ldots, n$. Hence $(\mathbf{x}_m)_{m\in M_n}$ is convergent in $\mathbb{R}^n$. Moreover, the limit is in X since X is closed in $\mathbb{R}^n$.

Remark: It follows from the above example and from Heine-Borel that a subset X of $\mathbb{R}^n$ is compact iff it is sequentially compact. This turns out to be true if X is any metric space. Our aim is to prove this and various other things.

Definitions. We fix a metric space (M, d) for the rest of this chapter. For $\varepsilon > 0$, a subset $F \subseteq M$ is called an *ε-net* for M if

$$\forall x \in M \ \exists y \in F \ d(x, y) \leq \varepsilon$$

Or equivalently

$$M = \bigcup_{y\in F} B_\varepsilon(y)$$

If in addition F is a finite set, we call it a *finite ε-net* for M. We say M is *totally bounded* if for all $\varepsilon > 0$, there is a finite ε-net for M.

Example: Given $\varepsilon > 0$, choose $n \in \mathbb{N}$ with $\frac{1}{n} < \varepsilon$. Then $\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\}$ is a finite ε-net for $(0, 1)$. Thus $(0, 1)$ is totally bounded.

Definition. For a non-empty subset $A \subseteq M$, define the *diameter* of A, denoted diamA, by

$$\text{diam} A = \sup\{d(x, y) : x, y \in A\}$$

Note: diam$A < \infty$ iff A is bounded.

Lemma 6.10. *Assume that M is totally bounded and A is a non-empty closed subset of M. Then for any $\varepsilon > 0$, there exist $K \in \mathbb{N}$ and non-empty closed subsets $B_1, B_2, \ldots, B_K$ for A such that $A = \bigcup_{k=1}^{K} B_k$ and diam$B_k \le \varepsilon$ for $1 \le k \le K$.*

Proof. Let $F \subseteq M$ be a finite $\frac{\varepsilon}{2}$-net for M. Then $M = \bigcup_{x \in F} B_{\frac{\varepsilon}{2}}(x)$, and hence $A = \bigcup_{x \in F} A \cap B_{\frac{\varepsilon}{2}}(x)$. For $x \in F$ set $B_x = A \cap B_{\frac{\varepsilon}{2}}(x)$ and let $G = \{x \in F : B_x \neq \emptyset\}$. Then $A = \bigcup_{x \in G} B_x$, and for each $x \in G$, the set B_x is non-empty, closed and has diam$B_x \le$ diam$B_{\frac{\varepsilon}{2}}(x) \le \varepsilon$. $\square$

Theorem 6.11. *For a metric space (M, d) TFAE*

(a) M is compact

(b) M is sequentially compact

(c) M is complete and totally bounded

Proof. To show $(a) \Rightarrow (b)$: Let (x_n) be a sequence in M. For each $n \in \mathbb{N}$ set $T_n = \{x_k : k > n\}$. Note that if a subsequence of (x_n) converges to some $x \in M$, then $x \in \bigcap_{n \in \mathbb{N}} \mathrm{Cl}(T_n)$.

We first show that $\bigcap_{n \in \mathbb{N}} \mathrm{Cl}(T_n)$ is not empty. Assume otherwise. Then $\bigcup_{n \in \mathbb{N}} M \setminus \mathrm{Cl}(T_n) = M$, and so $\{M \setminus \mathrm{Cl}(T_n) : n \in \mathbb{N}\}$ is an open cover for M. Since M is compact, there is a finite subcover. Since $\mathrm{Cl}(T_m) \supseteq \mathrm{Cl}(T_n)$ for all $m \le n$, it follows that $M = M \setminus \mathrm{Cl}(T_n)$ for some $n \in \mathbb{N}$, which is absurd.

Fix $x \in \bigcap_{n \in \mathbb{N}} \mathrm{Cl}(T_n)$. We show that some subsequence of (x_n) converges to x. Since $x \in \mathrm{Cl}(T_1)$, have $D_1(x) \cap T_1 \neq \emptyset$, so $\exists k_1 > 1$ such that $x_{k_1} \in D_1(x)$. Since $x \in \mathrm{Cl}(T_{k_1})$, we have $D_{\frac{1}{2}}(x) \cap T_{k_1} \neq \emptyset$, so $\exists k_2 > k_1$ such that $x_{k_2} \in D_{\frac{1}{2}}(x)$. In general, assume we have found $k_1 < k_2 < \ldots < k_n$ such that $x_{k_j} \in D_{\frac{1}{j}}(x)$ for $j = 1, 2, \ldots, n$.

Since $x \in \mathrm{Cl}(T_{k_n})$, have $D_{\frac{1}{n+1}}(x) \cap T_{k_n} \neq \emptyset$, so $\exists k_{n+1} > k_n$ such that $x_{k_{n+1}} \in D_{\frac{1}{n+1}}(x)$. By induction, we have constructed a subsequence (x_{k_n}) of (x_n) such that $d(x_n, x) < \frac{1}{n}$ for all n. It follows that $x_{k_n} \to x$.

To show $(b) \Rightarrow (c)$: We first show that M is complete. Let (x_n) be a Cauchy sequence in M. Since M is sequentially compact, there is a subsequence (x_{k_n}) of (x_n) that converges to some $x \in M$. We show that $x_n \to x$. Given $\varepsilon > 0$, choose $N \in \mathbb{N}$ such that $d(x_m, x_n) < \varepsilon$ for all $m, n \ge N$. Since $x_{k_n} \to x$, we have $d(x_{k_n}, x) < \varepsilon$ for all sufficiently large n. In particular, we can fix $n_0 \in \mathbb{N}$ such that $n_0 \ge N$ and $d(x_{k_{n_0}}, x) < \varepsilon$. Then $k_{n_0} \ge n_0 \ge N$, and hence for all $n \ge N$, we have

$$d(x_n, x) \le d(x_n, x_{k_{n_0}}) + d(x_{k_{n_0}}, x) < 2\varepsilon$$

We next show that M is totally bounded. If not, then for some $\varepsilon > 0$, there is no finite ε-net for M. Then we can construct a sequence (x_n) inductively as follows. We pick $x_1 \in M$ arbitrarily, and for $n \geq 2$, we choose $x_n \in M \setminus \bigcup_{k=1}^{n-1} B_\varepsilon(x_k)$ (this set must be non-empty or else we have an ε-net). The sequence (x_n) satisfies $d(x_m, x_n) > \varepsilon$ for all $m \neq n$, and thus it has no Cauchy subsequence, contradicting sequential compactness.

To show $(c) \Rightarrow (a)$: We follow the proof of the compactness of $[0, 1]$. Let $\mathcal{U}$ be an open cover for M and assume M is not finitely covered by $\mathcal{U}$. We now inductively construct a nested sequence $A_0 \supseteq A_1 \supseteq \ldots$ of non-empty closed subsets of M such that $\text{diam}A_n \to 0$ and A_n cannot be finitely covered by $\mathcal{U}$ for any n. We let $A_0 = M$. Assume $n \geq 1$ and A_{n-1} has already been constructed. By Lemma 10, there exist $K \in \mathcal{N}$ and non-empty closed subsets $B_1, B_2, \ldots, B_K$ of A_{n-1} such that $A_{n-1} = \bigcup_{k=1}^{K} B_k$ and $\text{diam}B_k \leq \frac{1}{n}$ for $1 \leq k \leq K$. Since A_{n-1} cannot be finitely covered by $\mathcal{U}$, there must exist k, such that B_k cannot be finitely covered by $\mathcal{U}$. We then set $A_n = B_k$. This completes the inductive construction.

For each $n \in \mathbb{N}$, choose $x_n \in A_n$. For all $N \in \mathbb{N}$ and for all $m, n \geq N$, we have $x_m, x_n \in A_N$, and thus $d(x_m, x_n) \leq \text{diam}(A_N)$. It follows that (x_n) is a Cauchy sequence, and thus converges to some $x \in M$, since M is complete. Since $x_m \in A_n$ for $m \geq n$, and since A_n is closed, it follows that $x \in A_n$ for all $n \in \mathbb{N}$. Since $\mathcal{U}$ is an open cover for M, there exists $U \in \mathcal{U}$ with $x \in U$. As U is open, there is an $r > 0$ such that $D_r(x) \subseteq U$. Choose $n \in \mathbb{N}$ with $\text{diam}A_n < r$. Then for all $y \in A_n$, we have $d(y, x) < r$, and thus $A_n \subseteq D_r(x) \subseteq U$. This contradicts that A_n cannot be finitely covered by $\mathcal{U}$. $\square$

Remarks:

1. In $\mathbb{R}^n$, the Heine-Borel and Bolzano-Weierstrass theorems can be deducted from each other.

2. We now have another proof that the product of compact metric spaces is compact in the product topology. Just replace compactness by sequential compactness, for which the proof is straightforward.

3. There are topological spaces that are compact but not sequentially compact and vice versa.

7 Differentiation and the Inverse Function Theorem

Preliminaries

Let $m, n \in \mathbb{N}$. Then $L(\mathbb{R}^m, \mathbb{R}^n) = \{T : \mathbb{R}^m \to \mathbb{R}^n : T \text{ is linear}\} \cong \mathcal{M}_{n,m} \cong \mathbb{R}^{mn}$.

Let $e_1, e_2, \ldots, e_m$ be the standard basis vectors in $\mathbb{R}^m$.

Let $e'_1, e'_2, \ldots, e'_n$ be the standard basis vectors in $\mathbb{R}^n$.

Recall that for $x = \sum_{j=1}^n x_j e'_j$ and $y = \sum_{j=1}^n y_j e'_j$ in $\mathbb{R}^n$, their inner product is

$$\langle x, y\rangle = \sum_{j=1}^n x_j y_j$$

The Cauchy-Schwarz inequality states that

$$|\langle x, y\rangle| \le (\sum_{j=1}^n x_j^2)^{1/2} (\sum_{j=1}^n y_j^2)^{1/2} = ||x|| ||y||$$

A linear map $T \in L(\mathbb{R}^m, \mathbb{R}^n)$ is identified with the $n \times m$ matrix

$$(T_{j,i})_{1\le j\le n,\ 1\le i\le m} \in \mathcal{M}_{n,m}$$

where $T_{j,i} = \langle Te_i, e'_j\rangle$ (the jth coordinate of the ith column).

We can arrange the entries $(T_{j,i})_{1\le j\le n,\ 1\le i\le m}$ into a column vector of size mn, i.e into an element of $\mathbb{R}^{mn}$. So we can view $L(\mathbb{R}^m, \mathbb{R}^n)$ as a real mn-dimensional euclidean space with euclidean norm

$$||t|| = (\sum_{1\le j\le n,\ 1\le i\le m} T_{j,i}^2)^{1/2} = (\sum_{i=1}^m ||Te_i||^2)^{1/2}$$

Thus $L(\mathbb{R}^m, \mathbb{R}^n)$ becomes a metric space with the euclidean distance $d(S, T) = ||S - T||$ for $S, T \in L(\mathbb{R}^m, \mathbb{R}^n)$.

Lemma 7.1. *(a) For $T \in L(\mathbb{R}^m, \mathbb{R}^n)$ and $x \in \mathbb{R}^m$, we have $||Tx|| \le ||T|| ||x||$. It follows that T is Lipschitz, and hence continuous.*

(b) For $S \in L(\mathbb{R}^n, \mathbb{R}^p)$ and $T \in L(\mathbb{R}^m, \mathbb{R}^n)$, we have $||ST|| \le ||S|| ||T||$.

Proof.

(a) Writing $x = \sum_{i=1}^m x_i e_i$, we have

$$\begin{aligned}
||Tx|| &= ||\sum_{i=1}^m x_i Te_i|| \\
&\leq \sum_{i=1}^m |x_i|||Te_i|| \\
&\leq (\sum_{i=1}^m |x_i|^2)^{1/2}(\sum_{i=1}^m ||Te_i||^2)^{1/2} \\
&= ||T||||x||
\end{aligned}$$

It follows that for any $x, y \in \mathbb{R}^m$ we have

$$d(Tx, Ty) = ||Tx - Ty|| = ||T(x - y)|| \leq ||T||||x - y|| = ||T||d(x, y)$$

(b) Using part (a) we obtain

$$||ST|| = (\sum_{i=1}^m ||STe_i||^2)^{1/2} \leq (\sum_{i=1}^m ||S||^2||Te_i||^2)^{1/2} = ||S||||T||$$

$\square$

From Analysis I

Given a function $f : \mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$, we say f is differentiable at a if the limit

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

exists. We then denote this limit by $f'(a)$ and call it the derivative of f at a. Assuming this, define

$$\varepsilon : \mathbb{R} \to \mathbb{R} \quad \varepsilon(h) = \begin{cases} \frac{f(a+h)-f(a)}{h} - f'(a) & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

Then $\varepsilon(h) \to 0$ as $h \to 0$, and moreover

$$f(a + h) = f(a) + f'(a)h + \varepsilon(h)h$$

for all $h \in \mathbb{R}$. We proved one half of the following equivalence

1. f is differentiable at a
2. $\exists \lambda \in \mathbb{R}\ \exists \varepsilon : \mathbb{R} \to \mathbb{R}$ with $\varepsilon(0) = 0$ and ε continuous at 0 such that

   $$f(a + h) = f(a) + \lambda h + \varepsilon(h)h$$

   for all $h \in \mathbb{R}$.

So f is approximated by a linear function and the error of approximation is $o(h)$.

Compare this to continuity. f is continuous at a iff there is a function $\eta : \mathbb{R} \to \mathbb{R}$ with $\eta(0) = 0$ and η continuous at 0 such that

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

for all $h \in \mathbb{R}$. Here f is approximated by a constant function and the error of approximation is $o(1)$.

More generally, if f is n-times differentiable at a, then f is approximated by a polynomial of degree n and the error of approximation is $o(h^n)$.

Definition. We are given a function $f : \mathbb{R}^m \to \mathbb{R}^n$ and a point $a \in \mathbb{R}^m$.

We say f is *differentiable* at a if there exists $T \in L(\mathbb{R}^m, \mathbb{R}^n)$ and a function $\varepsilon : \mathbb{R}^m \to \mathbb{R}^n$ with $\varepsilon(0) = 0$ and ε continuous at 0 such that

$$f(a+h) = f(a) + T(h) + ||h||\varepsilon(h) \qquad (\dagger)$$

for all $h \in \mathbb{R}^m$. So ε is given by

$$\varepsilon(h) = \begin{cases} \frac{f(a+h)-f(a)-T(h)}{||h||} & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

Note that if there exists $T \in L(\mathbb{R}^m, \mathbb{R}^n)$ such that $\frac{f(a+h)-f(a)-T(h)}{||h||} \to 0$ as $h \to 0$, then the function ε above is continuous at 0 and ($\dagger$) holds for all $h \in \mathbb{R}^m$. It follows that f is differentiable at a.

Note that ($\dagger$) can also be written as

$$f(a+h) = f(a) + T(h) + o(||h||)$$

Note: suppose $S, T \in L(\mathbb{R}^m, \mathbb{R}^n)$ satisfy

$$\frac{f(a+h)-f(a)-S(h)}{||h||} \to 0 \text{ and } \frac{f(a+h)-f(a)-T(h)}{||h||} \to 0 \text{ as } h \to 0$$

Then $\frac{S(h)-T(h)}{0}$ as $h \to 0$. It follows that for any non-zero $x \in \mathbb{R}^m$, we have

$$\frac{S(x)-T(x)}{||x||} = \frac{S(x-k)-T(x/k)}{||x/k||} \to 0 \text{ as } k \to \infty$$

and hence $S(x) = T(x)$. Thus $S = T$.

Definition. If f is differentiable at a then the unique linear map $T \in L(\mathbb{R}^m, \mathbb{R}^n)$ satisfying $\frac{f(a+h)-f(a)-T(h)}{||h||} \to 0$, is called the *derivative* of f at a and is denoted by $f'(a)$ (or $Df(a)$ or $Df|_a$).

Note that if f is differentiable at a, then

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

where $\varepsilon : \mathbb{R}^m \to \mathbb{R}^n$ is continuous at 0 and $\varepsilon(0) = 0$.

Remark: for the case $m = 1$:

First consider $T = L(\mathbb{R}, \mathbb{R}^n)$. Set $v = T(1)$. Then $T(h) = T(h \cdot 1) = h \cdot v$ for all $h \in \mathbb{R}$. So $L(L(\mathbb{R}, \mathbb{R}^n)) \cong \mathbb{R}^n$ with the isomorphism being given by $T \mapsto T(1)$.

Now we are given a function $f : \mathbb{R} \to \mathbb{R}^n$ and a point $a \in \mathbb{R}$. Then f is differentiable at a iff there exists $v \in \mathbb{R}^n$ and a function $\varepsilon : \mathbb{R} \to \mathbb{R}^n$ with $\varepsilon(0) = 0$ and ε continuous at 0 such that

$$f(a+h) = f(a) + hv + h\varepsilon(h)$$

for all $h \in \mathbb{R}$. This $v \in \mathbb{R}^n$ is then unique and is called the derivative of f at a denoted $f'(a)$. Note that

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

and so

$$f(a+h) = f(a) + hf'(a) + o(h)$$

Definitions. A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is *differentiable* on $\mathbb{R}^m$ is f is differentiable at a for every $a \in \mathbb{R}^m$. In this case, the *derivative* of f on $\mathbb{R}^m$ is the function

$$f' : \mathbb{R}^m \to L(\mathbb{R}^m, \mathbb{R}^n) \quad a \mapsto f'(a)$$

Examples:

1. Constant functions. Let $b \in \mathbb{R}^n$ and $f : \mathbb{R}^m \to \mathbb{R}^n$ be given by $f(x) = b$ for all $x \in \mathbb{R}^m$. Then for any $a \in \mathbb{R}^m$, we have

   $$f(a+h) = b = f(a) + 0 + 0$$

   for all $h \in \mathbb{R}^m$. So f is differentiable at a with $f'(a) = 0$. So f is differentiable on $\mathbb{R}^m$ and its derivative is the constant zero function $\mathbb{R}^m \to L(\mathbb{R}^m, \mathbb{R}^n)$.

2. Linear functions. Let $f \in L(\mathbb{R}^m, \mathbb{R}^n)$. Then for any $a \in \mathbb{R}^m$, we have

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

   for all $h \in \mathbb{R}^m$. Thus f is differentiable at a with $f'(a) = f$. Thus, f is differentiable on $\mathbb{R}^m$ and its derivative is the constant function $\mathbb{R}^m \to L(\mathbb{R}^m, \mathbb{R}^n)$ with value f.

3. Consider $f : \mathbb{R}^m \to \mathbb{R}$, $f(x) = ||x||^2$. Fix $a \in \mathbb{R}^m$. Then

$$f(a+h) = ||a+h||^2 = ||a|| + \langle h, a\rangle + ||h|^2 = f(a) + 2\langle h, a\rangle + ||h||^2$$

Since $||h||^2 = o(||h||)$, it follows f is differentiable at a with $f'(a)(h) = 2\langle h, a\rangle$.

4. Let $\mathcal{M}_n$ denote the space of $n \times n$ real matrices (which we can think of as $\mathbb{R}^{n^2}$). Consider $f : \mathcal{M}_n \to \mathcal{M}_n$, $f(A) = A^2$. For fixed $A \in \mathcal{M}_n$, we have

$$f(A+H) = (A+H)^2 = f(A) + (AH + HA) + H^2$$

for all $H \in \mathcal{M}_n$. Since $||H^2|| \leq ||H||^2$ (Lemma 1), it follows that f is differentiable at A with $f'(A)(H) = AH + HA$.

Examples (continued):

5 Let $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p$ be a bilinear map. This means that $f(x, y)$ is linear in each variable with the other variable fixed. Fix $(a, b) \in \mathbb{R}^m \times \mathbb{R}^n$. Then

$$f((a, b) + (h, k)) = f(a + h, b + k) = f(a, b) + f(a, k) + f(h, b) + f(h, k)$$

Note that the map $T : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p$, $T(h, k) = f(a, k) + f(h, b)$ is linear. We show that $f(h, k) = o(||(h, k)||)$. Write $h = \sum_{i=1}^m h_i e_i$ and $k = \sum_{j=1}^n k_j e'_j$. We then have

$$||f(h, k)|| = \left|\left|\sum_{i=1}^m \sum_{j=1}^n h_i k_k f(e_i, e'_j)\right|\right| \leq \sum_{i=1}^m \sum_{j=1}^n |h_i||k_j|||f(e_i, e'_j)|| \leq C||(h, k)||^2$$

where $C = \sum_{i=1}^m \sum_{j=1}^n ||f(e_i, e'_j)||$ and we used $|h_i| \leq ||(h, k)||$, $|k_j| \leq ||(h, k)||$. It follows that f is differentiable at (a, b) with derivative $f'(a, b)(h, k) = f(a, k) + f(h, b)$. Note that $f' : \mathbb{R}^m \times \mathbb{R}^n \to L(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^p)$ is linear.

So far we considered functions with domain the whole of $\mathbb{R}^m$. We now repeat the definitions in a general setting

Definition. We are given an open set $U \subseteq \mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \in U$.

We say f is *differentiable* at a if there exists $T \in L(\mathbb{R}^m, \mathbb{R}^n)$ and a function $\varepsilon : \{h \in \mathbb{R}^m : a + h \in U\} \to \mathbb{R}^n$ with $\varepsilon(0) = 0$ and ε continuous at 0 such that

$$f(a + h) = f(a) + T(h) + ||h||\varepsilon(h) \qquad (\dagger)$$

for all $h \in \mathbb{R}^m$ such that $a + h \in U$. It follows that

$$\varepsilon(h) = \begin{cases} \frac{f(a+h)-f(a)-T(h)}{||h||} & \text{if } h \neq 0 \text{ and } a + h \in U \\ 0 & \text{if } h = 0 \end{cases}$$

Since U is open, $D_r(a) \subseteq U$ for some $r > 0$ and hence ε is defined on $D_r(0)$.

Note: f is differentiable at a iff

$$\exists T \in L(\mathbb{R}^m, \mathbb{R}^n) \text{ such that } \frac{f(a + h) - f(a) - T(h)}{||h||} \to 0 \text{ as } h \to 0$$

Definition. We are given an open set $U \subseteq \mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \in U$.

If f is differentiable at a then (as we saw earlier) there is a unique $T \in L(\mathbb{R}^m, \mathbb{R}^n)$ satisfying

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

The unique linear map T is called the *derivative* of f at a and is denoted $f'(a)$.

Note: If f is differentiable at a then

$$f(a+h) = f(a) + f'(a)(h) + o(||h||)$$

If $m = 1$, then we identify $\mathbb{R}^n$ with $L(\mathbb{R}, \mathbb{R}^n)$ and we have

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

Proposition 7.2 (Differentiable implies continuous)**.** *We are given an open set $U \subseteq \mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \in U$. If f is differentiable at a, then f is continuous at a.*

Proof. By assumption we have

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

where $\varepsilon(0) = 0$ and ε is continuous at 0. Given $x \in U$, putting $h = x - a$ we get

$$f(x) = f(a) + f'(a)(x-a) + ||x-a||\varepsilon(x-a)$$

By Lemma 1 the linear map $f'(a)$ is continuous; $|| \cdot ||$ is continuous by the triangle inequality; and ε is continuous at 0. Using results about continuity of sums, products and composites (section 2), we see f is continuous at a. $\square$

Proposition 7.3 (Chain Rule)**.** *Let U be an open subset of $\mathbb{R}^m$ and V be an open subset of $\mathbb{R}^n$. We are given functions $f : U \to \mathbb{R}^n$ with $f(U) \subseteq V$ and $g : V \to \mathbb{R}^p$. Let $a \in U$ and $b = f(a)$. If f is differentiable at a, and if g is differentiable at $b = f(a)$, then $g \circ f$ is differentiable at a, and moreover $(g \circ f)'(a) = g'(f(a)) \circ f'(a)$.*

Proof. Set $S = f'(a)$ and $T = g'(b)$. Then

$$f(a+h) = f(a) + S(h) + ||h||\varepsilon(h) \text{ and } g(b+k) = g(b) + T(k) + ||k||\zeta(k)$$

for suitable error functions ε and ζ.

$$\begin{aligned}
(g \circ f)(a+h) = g(f(a+h)) &= g(f(a) + \underbrace{S(h) + ||h||\varepsilon(h)}_{k=k(h)}) \\
&= g(f(a)) + T(S(h) + ||h||\varepsilon(h)) + ||k||\zeta(k) \\
&= (g \circ f)(a) + (T \circ S)(h) + ||h||T(\varepsilon(h)) + ||k||\zeta(k) \\
&= (g \circ f)(a) + g'(f(a)) \circ f'(a)(h) + ||h||\eta(h)
\end{aligned}$$

where

$$\eta(h) = \begin{cases} T(\varepsilon(h)) + \frac{||k||}{||h||}\zeta(k) & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

it remains to show that η is continuous at 0. Recall $k = S(h) + ||h||\varepsilon(h)$. Using Lemma 1, note that

$$||k|| \leq ||S(h) + ||h||\varepsilon(h)|| \leq ||S(h)|| + ||h||||\varepsilon(h)||)$$

Choose $\delta > 0$ such that if $||h|| < \delta$, then $||\varepsilon(h)|| < 1$, and hence $\frac{||k||}{||h||} < ||S|| + 1$.

Note that $k(0) = 0$ and k is continuous at 0 since S and $|| \cdot ||$ are continuous everywhere and ε is continuous at 0. It follows that $\zeta(k(0)) = 0$ and $\zeta(k)$ is also continuous at 0. Since T is also continuous everywhere, $T(\varepsilon)$ is also continuous at 0. It follows that η is continuous at 0. $\square$

Proposition 7.4. *We are given an open subset U of $\mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \in U$. For each $j = 1, 2, \dots, n$, let $f_j : U \to \mathbb{R}$ be the jth component of f. Then*

$$f \text{ is differentiable at } a \iff \text{each } f_j \text{ is differentiable at } a$$

in which case we have $f'(a)(h) = \sum_{j=1}^n f_j'(a)(h)e_j'$ for all $h \in \mathbb{R}^m$.

Proof. Let $q_j : \mathbb{R}^n \to \mathbb{R}$ be the jth coordinate projection $q_j(y) = \langle y, e_j' \rangle$ for $y \in \mathbb{R}^n$. Then $f_j = q_j \circ f$ for each j, and so $f(x) = (f_1(x), \dots, f_n(x))$ for $x \in \mathbb{R}^n$. Now, if f is differentiable at a, then so is f_j by the chain rule: $f_j'(a) = q_j'(f(a)) \circ f'(a) = q_j \circ f'(a)$ by linearity of q_j.

Conversely, assume each f_j is differentiable. Then for suitable functions ε_j

$$\begin{aligned} f(a+h) &= \sum_{j=1}^n f_j(a+h)e_j' = \sum_{j=1}^n [f_j(a) + f_j'(a)(h) + ||h||\varepsilon_j(h)]e_j' \\ &= f(a) + \sum_{j=1}^n f_j'(a)(h)e_j' + ||h|| \sum_{j=1}^n \varepsilon_j(h)e_j' \end{aligned}$$

Since $h \mapsto \sum_{j=1}^n f_j'(a)(h)e_j'$ is linear, and since $\varepsilon(h) = \sum_{j=1}^n \varepsilon_j(h)e_j'$ is continuous at 0 with $\varepsilon(0) = 0$, the result follows. $\square$

Corollary 7.5. *We are given an open subset U of $\mathbb{R}^m$, functions $f, g : U \to \mathbb{R}^n$ and $\lambda : U \to \mathbb{R}$ and a point $a \in U$. If f and g are differentiable at a, then so is $f + g$ with $(f+g)'(a) = f'(a) + g'(a)$. If f and λ are differentiable at a, then so is λf with*

$$(\lambda f)'(a)(h) = \lambda'(a)(h)f(a) + \lambda(a)f'(a)(h) \text{ for all } h \in \mathbb{R}^m$$

Proof. We define functions

$$H : U \to \mathbb{R}^n \times \mathbb{R}^n,\ H(x) = (f(x), g(x)) \quad A : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n,\ A(y,z) = y + z$$

$$K : U \to \mathbb{R} \times \mathbb{R}^n,\ K(x) = (\lambda(x), f(x)) \quad S : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n,\ S(t,y) = ty$$

Observe that $f + g = A \circ H$ and $\lambda f = S \circ K$. It follows from Proposition 4 that H and K are differentiable at a with

$$H'(a)(h) = (f'(a)(h), g'(a)(h)) \text{ and } K'(a)(h) = (\lambda'(a)(h), f'(a)(h))$$

As A is linear and S is bilinear, they are differentiable with

$$A'(y,z) = A \text{ and } S'(t,y)(u,h) = S(t,h) + S(u,y) = th + uy$$

for all $y, z, h \in \mathbb{R}^n$ and $t, u \in \mathbb{R}$. Finally, it follows by the chain rule that $f + g = A \circ H$ and $\lambda f = S \circ K$ are differentiable at a with

$$\begin{aligned}(f+g)'(a)(h) &= [A'(H(a)) \circ H'(a)](h) = A'(H(a))[H'(a)(h)] \\ &= A[(f'(a)(h), g'(a)(h))] = f'(a)(h) + g'(a)(h)\end{aligned}$$

and similarly,

$$\begin{aligned}(\lambda f)'(a)(h) &= [S'(K(a)) \circ K'(a)](h) = S'(K(a))[K'(a)(h)] \\ &= S'(\lambda(a), f(a))[(\lambda'(a)(h), f'(a)(h))] \\ &= \lambda(a)f'(a)(h) + \lambda'(a)(h)f(a)\end{aligned}$$

$\square$

Remark: There is an alternative direct proof, which we will give next.

Proof. We begin with λf. By definition of differentiability, we have

$$\begin{aligned}(\lambda f)(a+h) &= \lambda(a+h)f(a+h)\\ &= (\lambda(a)+\lambda'(a)(h)+||h||\varepsilon(h))\,(f(a)+f'(a)(h)+||h||\zeta(h))\\ &= (\lambda f)(a)+\lambda(a)f'(a)(h)+\lambda'(a)(h)f(a)+||h||\eta(h)\end{aligned}$$

Where

$$\eta(h) = \lambda(a)\zeta(h)+\frac{\lambda'(a)(h)}{||h||}f'(a)(h)+\lambda'(a)(h)\eta(h)+\varepsilon(h)[f(a)+f'(a)(h)+||h||\zeta(h)]$$

if $h \neq 0$ and $\eta(0)=0$.

By Lemma 1, we have $||\lambda'(a)(h)|| \leq ||\lambda'(a)||||h||$, and hence $\frac{\lambda'(a)(h)}{||h||}$ is bounded. It follows easily that $\eta$is continuous at 0. A similar argument shows $(f+g)'(a) = f'(a)+g'(a)$. $\square$

Partial Derivatives

We are given an open set U of $\mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \in U$. We fix a direction u in $\mathbb{R}^m$, which means that $u \in \mathbb{R}^m$ and $u \neq 0$. It the limit

$$\lim t \to 0 \frac{f(a+tu)-f(a)}{t}$$

exists, we call it the *directional derivative* of f at a in the direction u and we denote it by $D_u f(a)$.

Note:

1. $D_u f(a) \in \mathbb{R}^n$ and $f(a+tu) = f(a)+tD_u f(a)+o(t)$
2. Define $\gamma : \mathbb{R} \to \mathbb{R}^m$ by $\gamma(t) = a+tu$. Then $f \circ \gamma$ is defined on $\gamma^{-1}(U)$, which is open and contains 0. Then

$$\frac{f(a+tu)-f(a)}{t} = \frac{(f\circ\gamma)(t)-(f\circ\gamma)(0)}{t}$$

and hence $D_u f(a)$ exists iff $f \circ \gamma$ is differentiable at 0, and in that case $D_u f(a) = (f\circ\gamma)'(0)$.

Special case: $u = e_i$ for some $1 \leq i \leq m$. When $D_{e_i} f(a)$ exists, it is called the ith partial derivative of f at a and is denoted by $D_i f(a)$.

Proposition 7.6. *Let U, f, a be as above. If f is differentiable at a, then $D_u f(a)$ exists for all $u \in \mathbb{R}^m \setminus \{0\}$, and moreover $D_u f(a) = f'(a)(u)$. It follows that*

$$f'(a)(h) = \sum_{i=1}^{m} h_i D_i f(a)$$

for all $h = \sum_{i=1}^m h_i e_i \in \mathbb{R}^m$.

Proof. Fix $u \in \mathbb{R}^m \setminus \{0\}$. By assumption, for a suitable error function ε, we have

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

Put $h = tu$ and use linearity of $f'(a)$ to get

$$f(a+tu) = f(a) + tf'(a)(u) + |t|||u||\varepsilon(tu)$$

Hence

$$\frac{f(a+tu) - f(a)}{t} = f'(a)(u) + \frac{|t|}{t}||u||\varepsilon(tu) \to f'(a)(u) \text{ as } t \to 0$$

For the last part, observe

$$f'(a)(h) = \sum_{i=1}^{m} h_i f'(a)(e_i) = \sum_{i=1}^{m} h_i D_i f(a)$$

□

The Jacobian Matrix

We are given an open subset U of $\mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \mathrm{int} U$. Assume f is differentiable at a. Then *Jacobian matrix* of f at a, denoted $Jf(a)$, is the $n \times m$ matrix representing $f'(a)$ with respect to the standard bases of $\mathbb{R}^m$ and $\mathbb{R}^n$. Thus for $1 \leq i \leq m$, the ith column of $Jf(a)$ is

$$f'(a)(e_i) = D_i f(a)$$

and for $1 \leq j \leq n$, the (j,i)-entry of $Jf(a)$ is

$$[Jf(a)]_{j,i} = \langle D_i f(a), e'_j \rangle = q_j(f'(a)(e_i)) = f'_j(a)(e_i) = D_i f_j(a) = \frac{\partial f_j}{\partial x_i}(a)$$

Here $q_j : \mathbb{R}^n \to \mathbb{R}$ is the jth coordinate projection. Moreover, $f_j = q_j \circ f$ is the jth component of f, and we are using Proposition 4 and Proposition 6.

Remarks: We are given an open subset U of $\mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \mathrm{int} U$. We also fix a direction $u \in \mathbb{R}^m \setminus \{0\}$.

1. Alternative proof of Proposition 6: Let $\gamma(t) = a + tu$ for $t \in \mathbb{R}$. Restricting γ to the open neighbourhood $\gamma^{-1}(U)$ of 0, the composite $f \circ \gamma$ is defined. Note that γ is differentiable with $\gamma'(t) = u$ for all $t \in \mathbb{R}$. It follows that if f is differentiable at $a = \gamma(0)$, then by the chain rule $f \circ \gamma$ is differentiable at 0. Hence $D_u f(a)$ exists and

$$D_u f(a) = (f \circ \gamma)'(0) = f'(a)(\gamma'(0)) = f'(a)(u)$$

2. If $D_u f(a)$ exists, then $D_u f_j(a)$ exists for all $1 \leq j \leq n$. Indeed,

$$\frac{f_j(a+tu) - f_j(a)}{t} = q_j\left(\frac{f(a+tu) - f(a)}{t}\right) \to q_j(D_u f(a))$$

   using linearity and continuity of q_j. Note that we do note assume f is differentiable at a.

3. The converse of Proposition 6 is false in general.

Theorem 7.7. *We are given an open set U of $\mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \in U$. Assume that there exists $r > 0$ such that $D_r(a) \subseteq U$ and for each $1 \leq i \leq m$, the partial derivative $D_i f(x)$ exists for all $x \in D_r(a)$ and the function $x \mapsto D_i f(x) : D_r(a) \to \mathbb{R}^n$ is continuous at a. Then f is differentiable at a.*

Proof. By Proposition 4 and by Remark 2 above, we can assume wlog that $n = 1$. We will do the proof for $m = 2$. The proof for general m is completely analogous.

Let $a = (p, q)$. If f is differentiable at a, then $f'(a)(h, k) = hD_1 f(a) + kD_2 f(a)$. So it is necessary (and sufficient) to show that

$$f((p, q) + (h, k)) = f(p, q) + hD_1 f(p, q) + kD_2 f(p, q) + o(||(h, k)||)$$

So for all $(h, k) \in D_r(0)$, we have

$$\begin{aligned}
&f((p, q) + (h, k)) - f(p, q) - hD_1(p, q) - kD_2(p, q) \\
&= f(p + h, q + k) - f(p + h, q) - kD_2 f(p, q) + f(p + h, q) - f(p, q) - hD_1 f(p, q)
\end{aligned}$$

The expression $f(p+h, q) - f(p, q) - hD_1 f(p, q)$ is $o(h)$ by definition of $D_1 f(p, q)$, and hence $f(p + h, q) - f(p, q) - hD_1 f(p, q) = o(||(h, k)||)$. To deal with the other terms, fix $(h, k) \in D_r(0)$ and define

$$\varphi : [0, 1] \to \mathbb{R} \quad \varphi(t) = f(p + h, q + tk)$$

Note that φ is continuous on $[0, 1]$ and differentiable on $(0, 1)$. Indeed, for $t \in (0, 1)$ we have $\varphi'(t) = kD_2 f(p + h, q + tk)$. Then by the Mean Value Theorem, for some $t = t(h, k) \in (0, 1)$, we have

$$\begin{aligned}
f(p + h, q + k) - f(p + h, q) - kD_2 f(p, q) &= \varphi(1) - \varphi(0) - kD_2 f(p, q) \\
&= \varphi'(t) - kD_2 f(p, q) \\
&= k[D_2 f(p + h, q + tk) - D_2 f(p, q)]
\end{aligned}$$

Since $D_2 f$ is continuous at $a = (p, q)$, it follows that

$$\frac{|f(p + h, q + k) - f(p + h, q) - kD_2 f(p, q)|}{||(h, k)||} \leq |D_2 f(p+h, q+tk) - D_2 f(p, q)| \to 0$$

as $(h, k) \to (0, 0)$. $\square$

Remark: For $D_1 f$, we only needed existence at a. For general m, it is enough to assume that for all but one value of i, the partial derivative $D_i f(x)$ exists for all x in a neighbourhood of a and it is continuous at a, whereas for the remaining value of u, it is enough to assume that existence of $D_i f$ at a.

Remarks (continued):

2 The Mean Value Theorem (MVT) works for functions defined on intervals with values in $\mathbb{R}$. There is also a version for functions with values in $\mathbb{R}^2$, but not with values in $\mathbb{R}^n$ for $n \geq 3$. We saw one remedy in the previous proof: we condisered components f_j of f, and restricted f_j to a line: for fixed $a \in U$ and for a fixed direction u, we considered $t \mapsto f_j(a + tu)$. We also have the following result.

Theorem 7.8 (The Mean Value Inequality)**.** *Let U be an open subset of $\mathbb{R}^m$, let $f : U \to \mathbb{R}^n$ be a function that is differentiable at every $z \in U$ and let $a, b \in U$. Assume that the line segment $[a, b]$ joining a and b is contained in U and that for some M we have*

$$\forall z \in [a, b] \ ||f'(z)|| \leq M$$

Then

$$||f(b) - f(a)|| \leq M||b - a||$$

Proof. We can assume that $a \neq b$. Set $u = b - a$ and $v = f(b) - f(a)$. For $t \in \mathbb{R}$ define $\gamma(t) = a + tu$. Then $f \circ \gamma$ is defined on the open set $\gamma^{-1}(U)$, which contains the closed interval $[0, 1]$ since $\gamma([0, 1]) = [a, b]$. On $\gamma^{-1}(U)$, by the Chain Rule we have

$$(f \circ \gamma)'(t) = f'(\gamma(t))(\gamma'(t)) = f'(a + tu)(u) = D_u f(a + tu)$$

For $t \in \gamma^{-1}(U)$ define $\varphi(t) = \langle f(a + tu), v\rangle = \langle (f \circ \gamma)(t), v\langle$. Then

$$\varphi(1) - \varphi(0) = \langle f(b) - f(a), v\rangle = ||f(b) - f(a)||^2$$

Moreover, since the map $\mathbb{R}^n \to \mathbb{R}$, $y \mapsto \langle y, v\rangle$, is linear, and hence differentiable, it follows by the Chain Rule that φ is differentiable with

$$\varphi'(t) = \langle (f \circ \gamma)'(t), v\rangle = \langle f'(a + tu)(u), v\rangle$$

Now, by the MVT, there exists $\theta \in (0, 1)$ such that

$$\begin{aligned}
\varphi(1) - \varphi(0) = \varphi'(\theta) &= \langle f'(a + \theta u)(u), v\rangle \\
&\leq ||f'(a + \theta u)(u)||||v|| \\
&\leq ||f'(a + \theta u)||||u||||v|| \\
&\leq M||b - a||||f(b) - f(a)||
\end{aligned}$$

Hence, $||f(b) - f(a)|| \leq M||b - a||$. □

Corollary 7.9. *Let U be an open connected subset of $\mathbb{R}^m$ and $f : U \to \mathbb{R}^n$ be a function that is differentiable on U with $f'(a) = 0$ for every $a \in U$. Then f is constant on U.*

Proof. For $a, b \in U$, if the line segment $[a, b] \subseteq U$, then $f(a) = f(b)$ by the MVI. Given $a \in U$, there exists $r > 0$ such that $D_r(a) \subseteq U$. Then for all $x \in D_r(a)$, we have $[a, x] \subseteq U$, and hence $f(a) = f(x)$. Thus f is locally constant. Since U is connected, it follows that f is constant. □

Remark: Suppose $V \subseteq \mathbb{R}^m$ and $W \subseteq \mathbb{R}^n$ are open sets and $f : V \to W$ is a bijection. Let $a \in V$ and assume that f is differentiable at a and f^{-1} is differentiable at $f(a)$. Put $S = f'(a)$ and $T = (f^{-1})'(f(a))$. Then by the Chain Rule

$$TS = (f^{-1} \circ f)'(a) = I_m \text{ and } ST = (f \circ f^{-1})'(f(a)) = I_n$$

It follows that $m = \text{tr}(TS) = \text{tr}(ST) = n$.

Definition. We are given an open set $U \subseteq \mathbb{R}^m$ and a function $f : U \to \mathbb{R}^n$. We say that f is *differentiable* on U if f is differentiable at a for every $a \in U$.

If f is differentiable on U, then the *derivative* of f on U is the function $f' : U \to L(\mathbb{R}^m, \mathbb{R}^n)$ mapping $a \mapsto f'(a)$.

We say f is a C^1*-function* on U if it is continuously differentiable on U, i.e. if f is differentiable on U and its derivative $f' : U \to L(\mathbb{R}^m, \mathbb{R}^n)$ is continuous.

Theorem 7.10 (The Inverse Function Theorem (IFT))**.** *Let U be an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^n$ be a C^1-function on U. Let $a \in U$ and assume that $f'(a)$ is invertible. Then there are open subsets V, W of $\mathbb{R}^n$ such that $a \in V \subseteq U$ and $f|_V : V \to W$ is a bijection whose inverse $g : W \to V$ is a C^1-function with $g'(y) = f'(g(y))^{-1}$ for every $y \in W$.*

Proof. Step 1: We show that we can assume that $a = f(a) = 0$ and $f'(a) = I$. To see this, let $T = f'(a)$ and define $h(x) = T^{-1}(f(x+a) - f(a))$. Then domain of h is $U - a$ and h is a C^1-function with $h'(x) = T^{-1} \circ f'(x+a)$ by the Chain Rule. Note that by Lemma 1 we have

$$||h'(x) - h'(y)|| = ||T^{-1} \circ (f'(x+a) - f'(y+a))|| \leq ||T^{-1}|| ||f'(x+a) - f'(y+a)||$$

Note also that $h(0) = 0$ and $h'(0) = T^{-1} \circ f'(a) = I$. Now, if we can prove the result for h, the result follows for f since $f(x) = T(h(x-a)) + f(a)$ for all $x \in U$.

Step 2: We now assume $f(0) = 0$ and $f'(0) = I$. Since f' is continuous, we can choose $r > 0$ such that $B_r(0) \subseteq U$ and $||f'(x) - I|| \leq 1/2$ for all $x \in B_r(0)$. We show that for all $x, y \in B_r(0)$, we have

$$||f(x) - f(y)|| \geq \frac{1}{2} ||x - y||$$

To see this, define $p : U \to \mathbb{R}^n$ by $p(x) = f(x) - x$. Then p is differentiable with $p'(x) = f'(x) - I$ for all $x \in U$. It follows that $||p'(x)|| < 1/2$ for all $x \in B_r(0)$. Now, given $x, y \in B_r(0)$, the line segment $[x, y] \subseteq B_r(0)$:

$$||(1-t)x + ty|| \leq (1-t)||x|| + t||y|| \leq (1-t)r + tr = r$$

for any $0 \leq t \leq 1$. Hence by the MVI (Theorem 8) $||p(x) - p(y)|| \leq \frac{1}{2}||x - y||$. It follows that

$$||f(x) - f(y)|| = ||p(x) + x - (p(y) + y))|| \geq ||x - y|| - ||p(x) - p(y)|| \geq \frac{1}{2}||x - y||$$

Step 3: Put $s = \frac{r}{2}$. We show that $f(D_r(0)) \supseteq D_s(0)$. Fix $w \in D_s(0)$. For $x \in B_r(0)$, let $q(x) = w - p(x) = w - f(x) + x$. Since $p(0) = f(0) - 0 = 0$, it follows that

$$||q(x)|| \leq ||w|| + ||p(x) - p(0)|| \leq ||w|| + \frac{1}{2}||x - 0|| < 2s = r$$

Thus, q maps $B_r(0)$ into $B_r(0)$. Next, for $x, y \in B_r(0)$, we have

$$||q(x) - q(y)|| = ||p(x) - p(y)|| \leq \frac{1}{2}||x - y||$$

Hence q is a contraction mapping on the non-empty complete metric space $B_r(0)$. By the contraction mapping theorem (3.20) q has a unique fixed point, i.e there is a unique $x \in B_r(0)$ such that $f(x) = w$. Note that $||x|| = ||q(x)|| < r$ from above.

Step 4: Set $W = D_s(0)$ and $V = f^{-1}(D_s(0)) \cap D_r(0)$. We show that V and W satisfy the conclusions of the theorem.

W is open and $f(0) \in W$. Since f is continuous, V is open with $0 \in V \subseteq U$. It follows from Step 3 that $f|_V : V \to W$ is a bijection. Let $g : W \to V$ be its inverse. Given $u, v \in W$, set $x = g(u)$ and $y = g(v)$. It follows from Step 2 that

$$||g(u) - g(v)|| = ||x - y|| \leq 2||f(x) - f(y)|| = 2||u - v||$$

Thus g is a Lipschitz map and hence continuous.

The proof of the final 5th Step is **non-examinable**. As a preliminary step of Step 5, overserve that for $x \in B_r(0)$, we have $||f'(x) - I|| < 1/2$ and hence

$$||f'(x)(h)|| \geq ||h|| - ||h - f'(x)(h)|| \geq \frac{1}{2}||h||$$

It follows that $f'(x)$ is injective, and hence invertible.

Step 5 $g : W \to V$ is a C^1-function and $g'(y) = f'(g(y))^{-1}$ for every $y \in W$. Fix $y \in W$. Set $x = g(y)$ and $T = f'(x)$. Then for a suitable error function ε, we have

$$f(x + h) = f(x) + T(h) + ||h||\varepsilon(h)$$

Choose $\delta > 0$ such that $D_\delta(y) \subseteq W$. For $k \in D_\delta(0)$, define $h = h(k) = g(y+k) - g(y)$. Then $g(y+k) = g(y) + h = x + h$, and hence $y + k = -f(x+h)$ and $k = f(x + h) - f(x)$. From above, it follows that $k = T(h) + ||h||\varepsilon(h)$, and so $h = T^{-1}(k) - ||h||T^{-1}(\varepsilon(h))$. We then obtain

$$g(y + k) = g(y) + h = g(y) + T^{-1}(k) - ||h||T^{-1}(\varepsilon(h))$$

The composite $T{-}1 \circ \varepsilon \circ h$ is 0 at 0 and continuous at 0, whereas $||h|| = ||g(y + k) - g(y)||| \leq 2||k||$ is by Step 4. Hence, we get

$$g(y + k) = g(y) + T^{-1}(k) + o(||k||)$$

which shows that g is differentiable at y and $g'(y) = T^{-1} = f'(g(y))^{-1}$. $\square$

Second derivative

We are given an open set $U \subseteq \mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \in U$. We say f is *twice differentiable* at a if f is differentiable on some open set V with $a \in V \subseteq U$ and $f' : V \to L(\mathbb{R}^m, \mathbb{R}^n)$ is differentiable at a. We let $f''(a) = (f')'(a)$ and call $f''(a)$ the *second derivative* of f at a.

Note: $f''(a) \in L(\mathbb{R}^m, L(\mathbb{R}^m, \mathbb{R}^n))$.

Second derivative is a linear map

Given $T \in L(\mathbb{R}^m, L(\mathbb{R}^m, \mathbb{R}^n))$, the map $\mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n$ given by $(h, k) \mapsto T(h)(k)$ is bilinear. Conversely, given a bilinear map $T : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n$, for each $h \in \mathbb{R}^m$, the map $T(h) : \mathbb{R}^m \to \mathbb{R}^n$ defined by $T(h)(k) = T(h, k)$ is linear, and moreover, the map $\mathbb{R}^m \to L(\mathbb{R}^m, \mathbb{R}^n)$ given by $h \mapsto T(h)$ is linear. This proves that $L(\mathbb{R}^m, L(\mathbb{R}^m, \mathbb{R}^n)) \cong \text{Bil}(\mathbb{R}^m \times \mathbb{R}^m, \mathbb{R}^n)$, the space of all bilinear maps $\mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n$.

Under this identification, for a member T of either of these two spaces we write $T(h)(k) = T(h, k)$ for $h, k \in \mathbb{R}^m$.

Proposition 7.11. *We are given an open set $U \subseteq \mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \in U$. Assume that f is differentiable on some open set V with $a \in V \subseteq U$. Then f is twice differentiable at a $\iff$ there is a bilinear map $T \in Bil(\mathbb{R}^m \times \mathbb{R}^m, \mathbb{R}^n)$ such that for every fixed $k \in \mathbb{R}^m$, we have*

$$f'(a + h)(k) = f'(a)(k) + T(h, k) + o(||h||)$$

Proof.
To show "$\Rightarrow$": Let $T = f''(a)$. Then

$$f'(a + h) = f'(a) + T(h) + ||h||\varepsilon(h)$$

where $\varepsilon : V - a \to L(\mathbb{R}^m, \mathbb{R}^n)$ is continuous at 0 and $\varepsilon(0) = 0$. Fix $k \in \mathbb{R}^m$ and evaluate the above expression at k:

$$f'(a + h)(k) = f'(a)(k) + T(h, k) + ||h||\varepsilon(h)(k)$$

Here we think of T as a bilinear map. By Lemma 1

$$||\varepsilon(h)(k)|| \leq |\varepsilon(h)||||k|| \to 0 \text{ as } h \to 0$$

and hence $||h||\varepsilon(h)(k) = o(||h||)$.

To show "$\Leftarrow$": Thinking of the given $T \in \text{Bil}(\mathbb{R}^m \times \mathbb{R}^m, \mathbb{R}^n)$ as a member of $L(\mathbb{R}^m, L(\mathbb{R}^m, \mathbb{R}^n))$, define

$$\varepsilon : V - a \to L(\mathbb{R}^m, \mathbb{R}^n) \quad \varepsilon(h) = \begin{cases} \frac{f'(a+h) - f'(a) - T(h)}{||h||} & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

Then $f'(a+h) = f'(a) + T(h) + ||h||\varepsilon(h)$ for all h. We just need to show that ε is continuous at 0. By assumption, for fixed $k \in \mathbb{R}^m$ we have

$$\varepsilon(h)(k) = \frac{f'(a+h)(k) - f'(a)(k) - T(h,k)}{||h||} \to 0 \text{ as } h \to 0$$

It follows that

$$||\varepsilon(h)|| = \left(\sum_{i=1}^{m} ||\varepsilon(h)(e_i)||^2\right)^{1/2} \to 0 \text{ as } h \to 0$$

So we are done. □

Examples:

1. Suppose $f : \mathbb{R}^m \to \mathbb{R}^n$ is linear. Then f is differentiable on $\mathbb{R}^m$ and $f'(x) = f$ for all $x \in \mathbb{R}^m$. Thus, $f' : \mathbb{R}^m \to L(\mathbb{R}^m, \mathbb{R}^n)$ is the constant function with value f. It follows that f is twice differentiable on $\mathbb{R}^m$ and $f''(x) = 0$ for all $x \in \mathbb{R}^m$.

2. Now let $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p$ be a bilinear map. Then f is differentiable on $\mathbb{R}^m \times \mathbb{R}^n$ and $f' : \mathbb{R}^m \times \mathbb{R}^n \to L(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^p)$ is given by

$$f'(a,b)(h,k) = f(a,k) + f(h,b)$$

   This expression is linear in (a,b) with (h,k) fixed. In other words, f' is a linear map, and hence differentiable with $f''(a,b) = f'$ for all $(a,b) \in \mathbb{R}^m \times \mathbb{R}^n$.

3. Consider $f : \mathcal{M}_n \to \mathcal{M}_n$ given by $f(A) = A^3$. Then

$$\begin{aligned} f(A+H) = (A+H)^3 &= A^3 + A^2H + AHA + HA^2 + AH^2 + HAH + H^2A + H^3 \\ &= f(A) + (A^2H + AHA + HA^2) + o(||H||) \end{aligned}$$

   Indeed, for example $||AH^2|| \leq ||A||||H||^2$ by Lemma 1, and thus $AH^2 = o(||H||)$. The other terms can be dealt with in a similar way. This proves that f is differentiable at A with $f'(A)(H) = A^2H + AHA + HA^2$. We now consider

$$\begin{aligned} f'(A+H)(K) &= (A+H)^2K + (A+H)K(A+H) + K(A+H)^2 \\ &= f'(A)(K) + T(H,K) + H^2K + KHK + KH^2 \end{aligned}$$

   The map $T : \mathcal{M}_n \times \mathcal{M}_n \to \mathcal{M}_n$ given by $T(H,K) = AHK + HAK + AKH + HKA + KAH + KHA$ is bilinear. The remiander $H^2K + KHK + HK^2$ us $o(||H||)$ (with K fixed). It follows by Proposition 11 that f is twice differentiable at A with $f''(A) = T$.

Second derivatives and partial derivatives

We are given an open set $U \subseteq \mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \in U$. Assume f is twice differentiable at a. Then for each fixed $k \in \mathbb{R}^m$, we have

$$f'(a+h)(k) = f'(a)(k) + f''(a)(h,k) + o(||h||)$$

Now fix directions $u, v \in \mathbb{R}^m \setminus \{0\}$. Putting $k = v$ above yields

$$D_v f(a+h) = D_v f(a) + f''(a)(h,v) + o(||h||)$$

It follows that $D_v f : V \to \mathbb{R}^n$ is differentiable at a and

$$(D_v f)'(a)(h) = f''(a)(h,v)$$

Hence

$$D_u D_v f(a) = D_u(D_v f)(a) = (D_v f)'(a)(u) = f''(a)(u,v)$$

In particular, $D_i D_j f(a) = f''(a)(e_i, e_j)$.

Theorem 7.12 (Symmetry of mixed partial derivatives)**.** *We are given an open set $U \subseteq \mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \in U$. Assume f is twice differentiable on an open set V with $a \in V \subseteq U$. Assume that $f'' : V \to Bil(\mathbb{R}^m \times \mathbb{R}^m, \mathbb{R}^n)$ is continuous at a. Then*

$$D_u D_v f(a) = D_v D_u f(a) \text{ or equivalently } f''(a)(u,v) = f''(a)(v,u)$$

for all directions $u, v \in \mathbb{R}^m \setminus \{0\}$. Thus $f''(a)$ is a symmetric bilinear map.

Proof. We first reduce to the case $n = 1$. For $x \in V$ we have

$$(D_u f)_j(x) = [D_u f(x)]_j = f_j'(x)(u) = D_u f_j(x)$$

(In the third equality, we used Proposition 4). Hence $(D_u f)_j = D_u f_j$, which in turn implies $(D_u D_v f)_j = D_u(D_v f)_j = D_u D_v f_j$. Thus, it is enough to show that $D_u D_v f_j(a) = D_v D_u f_j(a)$ for each $1 \leq j \leq n$. So wlog $n = 1$.

We define a function φ on a suitable neighbourhood of $(0,0)$ in $\mathbb{R}^2$ as follows.

$$\varphi(s,t) = f(a+su+tv) - f(a+tv) - f(a+su) + f(a)$$

Fix s and t and consider $\psi(y) = f(a+yu+tv) - f(a+yu)$. Note that $\varphi(s,t) = \psi(s) - \psi(0)$. By the MVT we find $\alpha = \alpha(s,t) \in (0,1)$ such that

$$\varphi(s,t) = \psi(s) - \psi(0) = s\psi'(\alpha s) = s(D_u f(a+\alpha su+tv) - D_u f(a+\alpha su))$$

Applying the MVT to $D_u f(a+\alpha su+yv)$, we find $\beta = \beta(s,t) \in (0,1)$ such that

$$\varphi(s,t) = st D_v D_u f(a+\alpha su+\beta tv) = st f''(a+\alpha su+\beta tv)(v,u)$$

We do this for every (s,t) and use the continuity of f'' at a to get

$$\frac{\varphi(s,t)}{st} = f''(a+\alpha su+\beta tv)(v,u) \to f''(a)(v,u) = D_v D)uf(a) \text{ as } (s,t) \to (0,0)$$

Now repeat the above, starting with $\psi(y) = f(a+su+yv) - f(a+yv)$ ending up with

$$\frac{\varphi(s,t)}{st} \to f''(a)(u,v) = D_u D_v f(a) \text{ as } (s,t) \to (0,0)$$

$\square$

Definitions. We are given an open set $U \subseteq \mathbb{R}^m$, a function $f : U \to \mathbb{R}$ and a point $a \in U$.

We say f has a *local maximum* at a if there is an $r > 0$ such that $D_r(A) \subseteq U$ and $f(x) \leq f(a)$ for all $x \in D_r(a)$.

We say f has a *local minimum* at a if there is an $r > 0$ such that $D_r(a) \subseteq U$ and $f(x) \geq f(a)$ for all $x \in D_r(a)$.

We say f has a *local extremum* at a if f has a local maximum or local minimum at a.

Proposition 7.13. *We are given an open set $U \subseteq \mathbb{R}^m$, a function $f : U \to \mathbb{R}$ and a point $a \in U$. If f is differentiable at a and f has a local extremum at a, then $f'(a) = 0$.*

Definition. We say a is a *stationary point* of f if f is differentiable at a and $f'(a) = 0$.

Proof. Replacing f with $-f$ if neccessary, we may assume f has a local maximum. Now assume that $f'(a) \neq 0$. Then there exists $u \in \mathbb{R}^m$ such that $f'(a)(u) \neq 0$. By rescaling, we may assume that $f'(a)(u) > 0$ and $||u|| = 1$. Now, by the definition of differentiability we have

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

where $\varepsilon(0) = 0$ and ε is continuous at 0. Choose $\delta > 0$ such that if $||h|| \leq \delta$, then $|\varepsilon(h)| < f'(a)(u)$ and $f(a+h) \leq f(a)$. Putting $h = \delta u$, we have

$$0 \geq f(a+\delta u) - f(a) = \delta(f'(a)(u) + \varepsilon(\delta u)) > 0$$

which is a contradiction. □

Remark: The converse of Proposition 13 is false in general. E.g 0 is a stationary point of $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$, but f has no local extremum at 0.

Lemma 7.14. *We are given an open set $U \subseteq \mathbb{R}^m$, a function $f : U \to \mathbb{R}^n$ and a point $a \in U$. If f is twice differentiable at a, then*

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

Proof. By considering components of f, we may assume that $n = 1$. By definition of the second derivative, we have

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

where $\varepsilon(0) = 0$ and ε is continuous at 0. We next define

$$g(h) = f(a+h) - f(a) - f'(a)(h) - \frac{1}{2}f''(a)(h,h)$$

which is defined on some open neighbourhood of 0. We need to show that $g(h) = o(||h||^2)$. Fix h and define $\varphi : [0,1] \to \mathbb{R}$ by $\varphi(t) = g(th)$. Note that

$$\varphi(t) = f(a+th) - f(a) - tf'(a)(h) - \frac{t^2}{2} f''(a)(h,h)$$

So φ is continuous on $[0,1]$ and differentiable on $(0,1)$ with

$$\begin{aligned}
\varphi'(t) &= f'(a+th)(h) - f'(a)(h) - tf''(a)(h,h) \\
&= f'(a+th)(h) - f'(a)(h) - f''(a)(th,h) \\
&= [f'(a+th) - f'(a) - f''(a)(th)](h) \\
&= ||th||\varepsilon(th)(h)
\end{aligned}$$

By the MVT, there exists $t = t(h) \in (0,1)$ such that

$$g(h) = \varphi(1) - \varphi(0) = \varphi'(t) = ||th||\varepsilon(th)(h)$$

It follows by Lemma 1 that $|\varepsilon(th)(h)| \leq ||\varepsilon(th)||||h||$, and hence $|g(h)| \leq ||h||^2||\varepsilon(th)||$, from which the result follows. □

Recall from Linear Algebra: A symmetric bilinear map $T : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is positive definite if $T(x,x) > 0$ for all $x \in \mathbb{R}^m \setminus \{0\}$ and negative definite if $T(x,x) < 0$ for all $x \in \mathbb{R}^m \setminus \{0\}$.

Theorem 7.15. *We are given an open set $U \subseteq \mathbb{R}^m$, a function $f : U \to \mathbb{R}$ and a point $a \in U$. Assume that f is twice differentiable on U and f'' is continuous at a. If a is a stationary point of f and $f''(a)$ is positive definite, then f has a local minimum at a. If a is a stationary point of f and $f''(a)$ is negative definite, then f has a local maximum at a.*

Note: By Theorem 12 if f'' exists on a neighbourhood of a and is continuous at a, then $f''(a)$ is a symmetric bilinear map. The $m \times m$ matrix H defined by $H_{i,j} = f''(a)(e_i, e_j) = D_i D_j f(a)$ is called the Hessian of f at a.

Proof. Assume that $f'(a) = 0$ and $f''(a)$ is positive definite. Let $u_1, \ldots, u_m$ be an orthonormal basis of $\mathbb{R}^m$ such that $f''(a)(u_i, u_j) = 0$ if $i \neq j$. Set

$$\mu = \min\{f''(a)(u_i, u_j) : 1 \leq i \leq m\}$$

Then $\mu > 0$ and for $h = \sum_{i=1}^m h_i u_i \in \mathbb{R}^m$, we have

$$f''(a)(h,h) = \sum_{i,j=1}^m h_i h_j f''(a)(u_i, u_j) = \sum_{i=1}^m h_i^2 f''(a)(u_i, u_i) \geq \mu \sum_{i=1}^m h_i^2 = \mu||h||^2$$

Now, by Lemma 14 we have

$$f(a+h) = f(a) + \frac{1}{2} f''(a)(h,h) + ||h||^2\varepsilon(h)$$

where $\varepsilon(0) = 0$ and ε is continuous at 0. Choose $\delta > 0$ such that $|\varepsilon(h)| < \mu/2$ whenever $||h|| < \delta$. Then

$$f(a+h) - f(a) \geq ||h||^2(\mu/2 + \varepsilon(h)) \geq 0$$

whenever $h \in D_\delta(0)$. Thus, f has a local minimum at a. The proof of the second statement is similar. $\square$