IB Analysis & Topology

Ishan Nath, Michaelmas 2022

Based on Lectures by Dr. Paul Russell

February 10, 2023

Contents

Part I

Generalizing Continuity and Convergence

1   Three Examples of Convergence

1.1   Convergence in $\mathbb{R}$

Let (x_n) be a sequence in $\mathbb{R}$ and $x \in \mathbb{R}$. We say (x_n) *converges* to x, and write $x_n \to x$, if for all $\varepsilon > 0$, there exists N such that for all $n \geq N$, $|x_n - x| < \varepsilon$.

In $\mathbb{R}$, one useful fact is the *triangle inequality*: $|a + b| \leq |a| + |b|$. We also have two key theorems:

Theorem 1.1 (Bolzano-Weierstrass Theorem)**.**

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

Recall that a sequence (x_n) in $\mathbb{R}$ is *Cauchy* if for all $\varepsilon > 0$, there exists N, such that for all $m, n \geq N$, $|x_m - x_n| < \varepsilon$. It is easy to show every convergent sequence is Cauchy. We also have the following:

Theorem 1.2 (General Principle of Convergence)**.**

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

This can be proven by Bolzano-Weierstrass theorem.

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

Let (z_n) be a sequence in $\mathbb{R}^2$, and $z \in \mathbb{R}^2$. We wish to define $(z_n) \to z$.

In $\mathbb{R}$, we used the norm $|x|$. In $\mathbb{R}^2$, if we have $z = (x, y)$, then we can say $\|z\| = \sqrt{x^2 + y^2}$. This also satisfies the triangle inequality – $\|a + b\| \leq \|a\| + \|b\|$.

Definition 1.1. Let (z_n) be a sequence in $\mathbb{R}^2$, and $z \in \mathbb{R}^2$. We say that (z_n) *converges* to z, and write $z_n \to z$, if for all $\varepsilon > 0$, there exists N such that for all $n \geq N$, $\|z_n - z\| < \varepsilon$.

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

Lemma 1.1. *If (z_n), (w_n) are sequences in $\mathbb{R}^2$ with $z_n \to z$, $w_n \to w$. Then $z_n + w_n \to z + w$.*

Proof:

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

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

Proposition 1.1. *Let (z_n) be a sequence in $\mathbb{R}^2$ and let $z \in \mathbb{R}^2$. Write $z_n = (x_n, y_n)$ and $z = (x, y)$. Then $z_n \to z$ if and only if $x_n \to x$ and $y_n \to y$.*

Proof: First, note $|x_n - x|, |y_n - y| \leq \|z_n - z\|$, so $\|z_n - z\| \to 0$ implies $|x_n - x|, |y_n - y| \to 0$.

Now, if $|x_n - x|, |y_n - y| \to 0$, then $\|z_n - z\| = \sqrt{|x_n - x|^2 + |y_n - y|^2} \to 0$.

Definition 1.2. A sequence (z_n) in $\mathbb{R}^2$ is *bounded* if there exists $M \in \mathbb{R}$ such that for all n, $\|z_n\| \leq M$.

Theorem 1.3 (Bolzano-Weierstrass in $\mathbb{R}^2$).

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

Proof: Let (z_n) be a bounded subsequence in $\mathbb{R}^2$. Write $z_n = (x_n, y_n)$. Now $|x_n|, |y_n| \leq \|z_n\|$, so x_n, y_n are bounded in $\mathbb{R}$.

By Bolzano-Weierstrass, x_n has a convergent subsequence, say $x_{n_j} \to x \in \mathbb{R}$. Similarly (y_{n_j}) is bounded, so it has a convergent subsequence $y_{n_{j_k}} \to y$. Since we know $x_{n_{j_k}} \to x$, $z_{n_{j_k}} \to z = (x, y)$.

Definition 1.3. A sequence $(z_n) \in \mathbb{R}^2$ is *Cauchy* if for all $\varepsilon > 0$, there exists N such that for all $m, n \geq N$, $\|z_m - z_n\| < \varepsilon$.

It is easy to show a convergent sequence in $\mathbb{R}^2$ is Cauchy.

Theorem 1.4 (General Principle of Convergence for $\mathbb{R}^2$).

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

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

1.3 Convergence of Functions

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

Definition 1.4. Say (f_n) *converges pointwise* to f, and we write $f_n \to f$ pointwise, if for all $x \in X$, $f_n(x) \to f(x)$ as $n \to \infty$.

Although this is simple and easy to check, it doesn't preserve some 'nice' properties that we want.

Example 1.1.

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

1. We will construct f_n continuous, but f not. Take

$$f_n(x) = \begin{cases} nx & x \leq \frac{1}{n}, \\ 1 & x \geq \frac{1}{n}. \end{cases}, f = \begin{cases} 0 & x = 0, \\ 1 & x > 0. \end{cases}$$

Then $(f_n) \to f$ pointwise, but f is not continuous.

2. We will construct f_n Riemann integrable, but f not. Take the function

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

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

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

3. We will construct f_n Riemann integrable, f Riemann integrable, but the integrals do not converge. Take $f(x) = 0$ for all x. We construct f_n with integral 1, such as

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

We consider another definition of convergence.

Definition 1.5 (Uniform Convergence)**.** Let $X \subset \mathbb{R}$, $f_n : X \to \mathbb{R}$, $f : X \to \mathbb{R}$. We say (f_n) *converges uniformly* to f, and write $f_n \to f$ uniformly, if for all $\varepsilon > 0$,

there exists N, such that for all $x \in X$ and all $n \geq N$, $|f_n(x) - f(x)| < \varepsilon$.

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

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

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

Theorem 1.5. *Let $X \subset \mathbb{R}$, $f_n : X \to \mathbb{R}$ be continuous, and let $f_n \to f : X \to \mathbb{R}$ uniformly. Then f is continuous.*

Proof: Let $x \in X$, and pick $\varepsilon > 0$. As $f_n \to f$ uniformly, we can find N such that for all $n \geq N$ and $\in X$,

$$|f_n(y) - f(y)| < \varepsilon.$$

In particular, we may take $n = N$. As f_N is continuous, we can find $\delta > 0$ such that for all $y \in X$,

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

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

$$|f(y) - f(x)| \leq |f(y) - f_N(y)| + |f_N(y) - f_N(x)| + |f_N(x) - f(x)|$$
$$< \varepsilon + \varepsilon + \varepsilon = 3\varepsilon.$$

But 3ε can be made arbitrarily small, so f is continuous.

Remark. This is often called a '3ε proof' (or a '$\varepsilon/3$ proof').

Theorem 1.6. *Let $f_n : [a, b] \to \mathbb{R}$ be integrable and let $f_n \to f : [a, b] \to \mathbb{R}$ uniformly. Then f is integrable and*

$$\int_a^b f_n \to \int_a^b f$$

as $n \to \infty$.

Proof: As $f_n \to f$ uniformly, we can pick n sufficiently large such that $f_n - f$ is bounded. Also f_n is bounded, so by the triangle inequality $f = (f - f_n) + f_n$ is bounded.

Let $\varepsilon > 0$. As $f_n \to f$ uniformly, there is some N such that for all $n \geq N$

and $x \in [a, b]$, we have $|f_n(x) - f(x)| < \varepsilon$. By Riemann's criterion, there is some dissection $\mathcal{D}$ of $[a, b]$ for which

$$S(f_N, \mathcal{D}) - s(f_N, \mathcal{D}) < \varepsilon.$$

Let $\mathcal{D} = \{x_0, x_1, \ldots, x_k\}$, where $a = x_0 < x_1 < \cdots < x_k = b$. Now,

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

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

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

But this can be made arbitrarily small, so by Riemann's criterion, f is integrable over $[a, b]$.

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

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

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

Unfortunately, uniform convergence cannot preserve all properties.

Example 1.2.

Take $f_n : [-1, 1] \to \mathbb{R}$, where each f_n is differentiable and $f_n \to f$ uniformly,

but f is not differentiable. Take

$$f_n = \sqrt{\left(\frac{1}{n} + x^2\right)}.$$

Then f_n is differentiable, and also uniformly converges to $f(x) = |x|$. But f is not differentiable.

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

Theorem 1.7. *Let $f_n : (u, v) \mapsto \mathbb{R}$ with $f_n \to f : (u, v) \to \mathbb{R}$ pointwise. Suppose further that each f_n is continuously differentiable and that $f'_n \to g : (u, v) \to \mathbb{R}$ uniformly. Then f is differentiable with $f' = g$.*

Proof: Fix $a \in (u, v)$. Let $x \in (u, v)$. By FTC, we have each f'_n is integrable over $[a, x]$ and

$$\int_a^x f'_n = f_n(x) - f_n(a).$$

But $f'_n \to g$ uniformly, so by theorem 5, g is integrable over $[a, x]$ and

$$\int_a^x g = \lim_{n \to \infty} \int_a^x f'_n(x) = f(x) - f(a).$$

So we have shown that for all $x \in (u, v)$,

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

By theorem 4, g is continuous so by FTC, f is differentiable with $f' = g$.

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

Definition 1.6. Let $X \subset \mathbb{R}$ and let $f_n : X \to \mathbb{R}$ for each $n \geq 1$. We say (f_n) is *uniformly Cauchy* if for all $\varepsilon > 0$, there exists N such that for all $m, n \geq N$ and for all $x \in X$,

$$|f_m(x) - f_n(x)| < \varepsilon.$$

It is easy to show that a uniformly convergent sequence is uniformly Cauchy.

Theorem 1.8 (General Principle of Uniform Convergence)**.** *Let (f_n) be a uniformly Cauchy sequence of functions $X \to \mathbb{R}$. Then (f_n) is uniformly convergent.*

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

Let $\varepsilon > 0$. Then we can find an N such that for all $m, n \geq N$ and all $y \in X$, $|f_m(y) - f_n(y)| < \varepsilon$. Fixing y, m and letting $n \to \infty$, $|f_m(y) - f(y)| \leq \varepsilon$. But since y is arbitrary, this shows $f_n \to f$ uniformly.

We will also try to take Bolzano-Weierstrass over to the space of functions.

Definition 1.7. Let $X \subset \mathbb{R}$ and let $f_n : X \to \mathbb{R}$ for each $n \geq 1$. We say (f_n) is *pointwise bounded* if for all x, there exists M such that for all n, $|f_n(x)| \leq M$.

We say (f_n) is *uniformly bounded* if there exists M, such that for all x and n, $|f_n(x)| \leq M$.

We would like a uniform Bolzano-Weierstrass, saying if (f_n) is a uniformly bounded sequence of functions, then it has a uniformly convergent subsequence. But this is not true.

> **Example 1.3.**
>
> Take $f_n : \mathbb{R} \to \mathbb{R}$,
> $$f_n(x) = \begin{cases} 1 & x = n, \\ 0 & x \neq n \end{cases}.$$
>
> Then (f_n) is uniformly bounded, but if $m \neq n$, then $f_m(m) = 1$ and $f_n(m) = 0$, so $|f_m(m) - f_n(m)| = 1$, hence (f_n) are not uniformly Cauchy, so cannot be uniformly convergent.

1.4   Application to Power Series

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

Definition 1.8. Let $f_n : X \to \mathbb{R}$ for each $n \geq 0$. We say that the series

$$\sum_{n=0}^{\infty} f_n$$

converges uniformly if the sequence of partial sums (F_n) does, where $F_n = f_0 + f_1 + \cdots + f_n$.

If we can prove that $\sum a_n x^n$ is uniformly convergent, then we can apply earlier theorems to show differentiability. However this is not quite true, for example take

$$\sum_{n=0}^{\infty} x^n.$$

However, we do have another approach. We can show that if $0 < r < R$, then we do have uniform convergence on $(-r, r)$, and then given $x \in (-R, R)$, we can choose $|x| < r < R$ and use the above to show all the properties we want. This is known as the *local uniform convergence of power series*.

Lemma 1.2. *Let $\sum a_n x^n$ be a real power series with radius of convergence $R > 0$. Let $0 < r < R$. Then $\sum a_n x^n$ converges uniformly on $(-r, r)$.*

Proof: Define $f, f_m : (-r, r) \to \mathbb{R}$ by

$$f(x) = \sum_{n=0}^{\infty} a_n x^n, \quad f_m(x) = \sum_{n=0}^{m} a_n x^n.$$

Recall that $\sum a_n x^n$ converges absolutely for all x with $|x| < R$. Let $x \in (-r, r)$. Then

$$|f(x) - f_m(x)| = \left| \sum_{n=m+1}^{\infty} a_n x^n \right|$$

$$\leq \sum_{n=m+1}^{\infty} |a_n||x|^n \leq \sum_{n=m+1}^{\infty} |a_n| r^n,$$

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

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

as $m \to \infty$.

Theorem 1.9. *Let $\sum a_n x^n$ be a real power series with radius of convergence $R > 0$. Define $f : (-R, R) \to \mathbb{R}$ by*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Then,

(i) *f is continuous;*

(ii) *For any $x \in (-R, R)$, f is integrable over $[0, x]$ with*

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

Proof: Let $x \in (-R, R)$. Pick r such that $|x| < r < R$. By the above lemma, $\sum a_n y^n$ converges uniformly on $(-r, r)$. But the partial sum functions are all continuous on $(-r, r)$, hence $f|_{(-r,r)}$ is continuous. Thus f is a continuous function on $(-R, R)$.

Moreover, $[0, x] \subset (-r, r)$ so we also have $\sum a_n y^n$ converges uniformly on $[0, x]$. Each partial sum on $[0, x]$ is a polynomial, so can be integrated with

$$\int_0^x \sum_{n=0}^{m} a_n y^n \, dy = \sum_{n=0}^{m} \int_0^x a_n y^n \, dy = \sum_{n=0}^{m} \frac{a_n}{n+1} x^{n+1}.$$

Thus, f is integrable over $[0, x]$ with

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

For differentiation, we need the following lemma:

Lemma 1.3. *Let $\sum a_n x^n$ be a real power series with radius of convergence $R > 0$. Then the power series $\sum n a_n x^{n-1}$ has radius of convergence at least R.*

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

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

Fix n, let $\alpha = |x/w| < 1$, and let $c = M/|x|$, a constant. Then $|n a_n x^{n-1}| \leq c n \alpha^n$. By comparison test, it suffices to show $\sum n \alpha^n$ converges. Note

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

as $n \to \infty$, so this converges by the ratio test.

Theorem 1.10. *Let $\sum a_n x^n$ be a real power series with radius of convergence $R > 0$. Let $f : (-R, R) \to \mathbb{R}$ be defined by*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Then f is differentiable and for all $x \in (-R, R)$,

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

Proof: Let $x \in (-R, R)$. Pick r with $|x| < r < R$. Then $\sum a_n y^n$ converges uniformly on $(-r, r)$. Moreover, the power series $\sum n a_n y^{n-1}$ had radius of convergence at least R, and so also converges uniformly on $(-r, r)$.

The partial sum functions $f_m(y)$ are polynomials, so are differentiable with

$$f'_m(y) = \sum_{n=1}^{m} n a_n y^{n-1}.$$

we now have f'_m converging uniformly on $(-r, r)$ to the function

$$g(y) = \sum_{n=1}^{\infty} n a_n y^{n-1}.$$

Hence, $f|_{(-r,r)}$ is differentiable and for all $y \in (-r, r)$, $f'(y) = g(y)$. In particular, f is differentiable at x with $f'(x) = g(x)$. This gives f is a differentiable function on $(-R, R)$ with derivative g as desired.

1.5   Uniform Continuity

Let $X \subset \mathbb{R}$. Let $f : X \mapsto \mathbb{R}$. Recall that f is *continuous* if for all $\varepsilon > 0$ and for all $x \in X$, there exists $\delta > 0$, such that for all $y \in X$ with $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$.

Definition 1.9. We say f is *uniformly continuous* if for all $\varepsilon > 0$, there exists $\delta > 0$, such that for all $x, y \in X$ with $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$.

Remark. Clearly if f is uniformly continuous, then f is continuous. The converse is not true.

Example 1.4.

Consider $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$. Then f is continuous as it is a polynomial. Suppose $\delta > 0$. Then,

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

as $x \to \infty$. So the condition fails for $\varepsilon = 1$.

Even on the bounded interval $(0, 1)$, take $f(x) = 1/x$. This is clearly continuous, but cannot be uniformly continuous as it approaches infinity as x approaches 0.

Theorem 1.11. *A continuous real-valued function on a closed bounded interval is uniformly continuous.*

Proof: Let $f : [a, b] \to \mathbb{R}$, and suppose f is continuous but not uniformly continuous. Then we can find an $\varepsilon > 0$ such that, for all $\delta > 0$, there exist $x, y \in [a, b]$ with $|x - y| < \delta$ but $|f(x) - f(y)| \geq \varepsilon$. In particular, take $\delta = 1/n$.

Thus, we can find sequence $(x_n), (y_n)$ in $[a, b]$ with $|x_n - y_n| < 1/n$ but $|f(x_n) - f(y_n)| \geq \varepsilon$. The sequence (x_n) is bounded, so by Bolzano-Weierstrass it has a convergent subsequence $x_{n_j} \to x$. Since $[a, b]$ is closed, $x \in [a, b]$.

Then $x_{n_j} - y_{n_j} \to 0$, so also $y_{n_j} \to x$. But f is continuous at x, so there exists $\delta > 0$ such that for all $y \in [a, b]$, $|y - x| < \delta$ implies $|f(y) - f(x)| < \varepsilon/2$. Take such a δ. As $x_{n_j} \to x$, we can find J_1 such that $j \geq J_1$ implies $|x_{n_j} - x| < \delta$. Similarly, we can find J_2 such that for $j \geq J_2$, $|y_{n_j} - x| < \delta$. Let $j = \max\{J_1, J_2\}$. Then we have $|f(x_{n_j}) - f(x)|, |f(y_{n_j}) - f(x)| < \varepsilon/2$. But by triangle inequality,

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

a contradiction.

Corollary 1.1. *A continuous real-valued function on a closed bounded interval is bounded.*

Proof: Let $f : [a, b] \mapsto \mathbb{R}$ be continuous, and so uniformly continuous. Then we can find $\delta > 0$ such that for all $x, y \in [a, b]$, $|x - y| < \delta$ implies $|f(x) - f(y)| < 1$. Let $M = \lceil (b - a)/\delta \rceil$. Then for any $x \in [a, b]$, we can find

$a = x_0 \leq x_1 \leq \ldots \leq x_M = x$, with $|x_i - x_{i-1} < \delta$. Then we have

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

Corollary 1.2. *A continuous real-valued function on a closed bounded interval is integrable.*

Proof: Let $f : [a, b] \to \mathbb{R}$ be continuous, and so uniformly continuous. Let $\varepsilon > 0$. Then we can find $\delta > 0$ such that for all $x, y \in [a, b]$, $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$. Let $\mathcal{D} = \{x_0 < x_1 < \ldots < x_n\}$ be a dissection such that $x_i - x_{i-1} < \delta$, and $i \in \{1, \ldots, n\}$. Then for any $u, v \in [x_{i-1}, x_i]$, we have $|u - v| < \delta$, so $|f(u) - f(v)| < \varepsilon$. Hence

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

This gives

$$S(f, \mathcal{D}) - s(f, \mathcal{D}) \leq \sum_{i=1}^{n} (x_i - x_{i-1})\varepsilon = \varepsilon(b - a).$$

But this can be made arbitrarily small, so by Riemann's criterion, f is integrable over $[a, b]$.

2   Metric Spaces

2.1   Definitions and Examples

Our goal is to generalize the idea of convergence. This requires a notion of distance.

We have seen in $\mathbb{R}$, we have a norm $|x - y|$, in $\mathbb{R}^2$ we have $\|x - y\|$, and in function space, we can take

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

We have seen that the triangle inequality is very useful, so we wish to preserve this property.

Definition 2.1. A *metric space* is a set X endowed with a *metric d*, i.e. a function $d : X^2 \to \mathbb{R}$, satisfying:

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

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

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

We could define a metric space as an ordered pair (X, d), but usually it is obvious what d is, so we often refer to the metric space as the set X.

> **Example 2.1.**
>
> (i) If $X = \mathbb{R}$, we have the usual metric $d(x, y) = |x - y|$.
>
> (ii) If $X = \mathbb{R}^n$, we can take the Euclidean metric
>
> $$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$
>
> (iii) Uniform convergence might not work. We wish to take $d(f, g) = \sup |f - g|$, but this might not exist if $f - g$ is unbounded. However, with the appropriate subspace of functions, we can take this metric. Let $Y \subset \mathbb{R}$, and take
>
> $$X = B(Y) = \{f : Y \to \mathbb{R} \mid f \text{ bounded}\},$$
>
> with the *uniform metric*
>
> $$d(f, g) = \sup_{x \in Y} |f(x) - g(x)|.$$

We can check the triangle inequality: if $f, g, h \in B(Y)$, then for all $x \in Y$,

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

Taking the sup over all $x \in Y$, we get

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

Remark. Suppose (X, d) is a metric space and $Y \subset X$. Then $d|_Y$ is a metric on Y. We say Y with this metric is a *subspace* of X.

Example 2.2.

(i) We can take $\mathbb{Q}, \mathbb{Z}, \mathbb{N}, [0, 1]$ as subspaces of $\mathbb{R}$.

(ii) A continuous function on a bounded interval is bounded, so $\mathcal{C}([a, b])$ is a subspace of $B([a, b])$, with the uniform metric.

(iii) We can take the empty metric space $X = \emptyset$ with the empty metric.

Moreover, we can define different metrics on the same set.

Example 2.3.

(i) We can take the l_1 metric on $\mathbb{R}^n$:

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

(ii) We can also take the l_∞ metric on $\mathbb{R}^n$:

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

(iii) On $\mathcal{C}([a, b])$, we can define the L_1 metric

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

(iv) If $X = \mathbb{C}$, we can define a metric

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

We can check that the triangle inequalitys. This is known as the British Rail metric or SNCF metric.

(v) Let X be any set. Define a metric d on X by

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

This is called the discrete metric on X.

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

$$d(x, y) = \begin{cases} 0 & x = y, \\ p^{-a} & p^a \| x - y. \end{cases}$$

We show the triangle quality holds. If any of x, y, z are the same, this is easy, so assume all of x, y, z are distinct. Let $x - y = p^a m$, $y - z = p^b n$. Then if $a \leq b$, we have

$$x - z = (x - y) + (y - z) = p^a(m + p^{b-a}n).$$

Hence $p^a \mid x - z$, so $d(x, z) \leq p^{-a}$.

Definition 2.2. Let (X, d) be a metric space. Let (x_n) be a sequence in X and let $x \in X$. We say (x_n) *converges* to x, and write $x_n \to x$, if for all $\varepsilon > 0$, there exists N such that for all $n \geq N$,
$$d(x_n, x) < \varepsilon.$$
Equivalently $x_n \to x$ if and only if $d(x_n, x) \to 0$ in $\mathbb{R}$.

Proposition 2.1. *Limits are unique. That is, if (X, d) is a metric space, (x_n) is a sequence in X, $x, y \in X$ with $x_n \to x$ and $x_n \to y$, then $x = y$.*

Proof: For each n,

$$d(x, y) \leq d(x, x_n) + d(x_n, y) \leq d(x_n, x) + d(x_n, y)$$
$$\to 0 + 0 = 0.$$

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

Remark.

1. This justifies talking about the limit of a convergent sequence in a metric space, and writing $x = \lim x_n$.

2. Constant sequences and eventually constant sequences converge.

3. Suppose (X, d) is a metric space and Y is a subspace of X. Suppose (x_n) is a sequence in Y which converges in Y to x. Then (x_n) also converges in X to x.

   However the converse is false: take the reals, then $1/n \to 0$. But if we consider the subspace $\mathbb{R} \setminus \{0\}$, then $(1/n)$ is a sequence, but does not converge in $\mathbb{R} \setminus \{0\}$.

Example 2.4.

Let d be the Euclidean metric on $\mathbb{R}^n$. Then we have $x_n \to x$ if and only if the sequence converges in each coordinate in the usual way in $\mathbb{R}$. Let's consider other metrics, such as the uniform metric

$$d_\infty(x, y) = \max_i |x_i - y_i|, \text{ then}$$

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

So $d(x, y) \leq \sqrt{n} d_\infty(x, y)$. But also $d_\infty(x, y) \leq d(x, y)$. So for (x_n) in $\mathbb{R}^n$,

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

So the same sequences converge in $(\mathbb{R}^n, d)$ and $(\mathbb{R}^n, d_\infty)$. Similarly, for

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

$$d_\infty(x, y) \leq d_1(x, y) \leq n d_\infty(x, y).$$

Consider $X = \mathcal{C}([0, 1])$. Let d_∞ be the uniform metric on X, so

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

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

Similarly we can take the L_1 metric

$$d_1(f, g) = \int_0^1 |f - g|, \text{ then}$$

$$d_1(f, g) = \int_0^1 |f - g| \le \int_0^1 d_\infty(f, g) = d_\infty(f, g).$$

So we can prove $f_n \to f$ in (X, d_∞) implies $f_n \to f$ in (X, d_1). But the converse is not true, from our previous examples on uniform convergence.

We can also take (X, d) a discrete metric. Consider a convergence sequence $x_n \to x$. Then letting $\varepsilon = 1$, the definition of convergence says for all $n \ge N$, $d(x_n, x) < 1$, so $x_n = x$. Thus (x_n) is eventually constant. So in a discrete metric, (x_n) converges if and only if (x_n) is eventually constant.

Definition 2.3. Let (X, d) and (Y, e) be metric spaces, and let $f : X \to Y$.

(i) Let $a \in X$ and $b \in Y$. We say $f(x) \to b$ as $x \to a$ if for all $\varepsilon > 0$, there exists $\delta > 0$, such that for all $x \in X$,

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

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

(iii) If for all $a \in X$, f is continuous, then we say f is a continuous function.

(iv) We say f is uniformly continuous if for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in X$,
$$d(x, y) < \delta \implies e(f(x), f(y)) < \varepsilon.$$

(v) Suppose $W \subset X$. We say f is continuous on W (resp. uniformly continuous on W) if the function $f|_W$ is continuous (resp. uniformly continuous), as a function $W \to Y$.

Remark.

1. We don't have a nice rephrasing of (i) in terms of concepts in the reals: we want something like

$$e(f(x), b) \to 0 \text{ as } d(x, a) \to 0,$$

but this is meaningless.

2. (i) says nothing about what happens at the point a itself. For example, $f : \mathbb{R} \to \mathbb{R}$ as $f(x) = \delta_{0x}$ tends to 0 as $x \to 0$. If we have f continuous, then $d(x, a) = 0$ implies $e(f(x), f(a)) = 0$, so we may take $0 \leq d(x, a) < \delta$.

3. We can rewrite (v): f is continuous on W if and only if $f|_W$ is a continuous function if and only if for all $a \in W$ and $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in W$,
$$d(x, a) < \delta \implies e(f(x), f(a)) < \varepsilon.$$

In particular, note that this only considers points in W. For example,
$$f(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & x \notin [0, 1] \end{cases}$$

is continuous on $[0, 1]$, but not continuous at 0 or 1.

Proposition 2.2. *Let (X, d), (Y, e) be metric spaces. Let $f : X \to Y$ and $a \in X$. Then f is continuous at a if and only if whenever (x_n) is a sequence in X with $x_n \to a$, then $f(x_n) \to f(a)$.*

Proof: Suppose f is continuous at a. Let (x_n) be a sequence in X with $x_n \to a$. Let $\varepsilon > 0$. As f is continuous at a we can find $\delta > 0$ such that for all $x \in X$, $d(x, a) < \delta$ implies $e(f(x), f(a)) < \varepsilon$.

As $x_n \to x$, we can find N such that for all $n \geq N$, $d(x_n, a) < \delta$. Hence, for $n \geq N$, $e(f(x), f(a)) < \varepsilon$. This gives $f(x_n) \to f(a)$.

Now suppose f is not continuous at a. Then there is some $\varepsilon > 0$ such that for all $\delta > 0$, there exists $x \in X$ with $d(x, a) < \delta$ but $e(f(x), f(a0) \geq \varepsilon$. Take $\delta_n = 1/n$, to obtain a sequence (x_n) with $d(x_n, a) < 1/n$ but $(f(x_n), f(a)) \geq \varepsilon$. Hence $x_n \to a$ but $f(x_n) \not\to f(a)$.

Proposition 2.3. *Let (W, c), (X, d), (Y, e) be metric spaces. Let $f : W \to X$, $g : X \to Y$ and let $a \in W$. Suppose f is continuous at a and g is continuous at $f(a)$. Then $g \circ f$ is continuous at a.*

Proof: Let (x_n) be a sequence in W with $x_n \to a$. Then, $f(x_n) \to f(a)$, and so also $g(f(x_n)) \to g(f(a))$. So $g \circ f$ is continuous at a.

Example 2.5.

1. Consider $\mathbb{R} \to \mathbb{R}$ with the usual metric. This is the same metric as defined for $\mathbb{R}$ only. We know many continuous function on $f : \mathbb{R} \to \mathbb{R}$, such as polynomials, trigonometric functions, exponential functions, etc.

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

3. Consider $\mathbb{R}^n$ with the EUclidean metric and $\mathbb{R}$ with the usual metric. The projection maps $\pi_i : \mathbb{R}^n \to \mathbb{R}$ given by $\pi_i(x) = x_i$ are continuous. Let's denote a sequence in $\mathbb{R}^n$ by $(x^{(m)})_{m \geq 1}$. We known that $x^{(m)} \to x$ if and only if for each i, $x_i^{(m)} \to x_i$. Hence π_i is continuous.

   Similarly, suppose $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$, and let $f : \mathbb{R} \to \mathbb{R}^n$ by $f(x) = (f_1(x), \ldots, f_n(x))$. Then f is continuous at a point if and only if all of $f_1, \ldots, f_n$ are.

   Using these facts, we can show that $f(x, y, z) = (e^{-x} \sin y, 2x \cos z)$ is continuous.

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

5. Let (X, d) be a discrete metric space and let (Y, e) be any metric space. Then all functions $f : X \to Y$ are continuous. Suppose $a \in X$ and (x_n) is a sequence in X with $x_n \to a$. Then (x_n) is eventually constant, so $f(x_n) \to f(a)$.

2.2   Completeness

We saw a version of the general principle of convergence held in each of the three examples we considered. We try to extend this to all metric spaces:

Definition 2.4. Let (X, d) be a metric space and let (x_n) be a sequence in X. We say (x_n) is *Cauchy* if for all $\varepsilon > 0$, then there exists N such that for all $m, n \geq N$, $d(x_m, d_n) < \varepsilon$.

It is easy to show that if (x_n) is convergent, then (x_n) is Cauchy, but the converse

is not true in general.

For example, let $X = \mathbb{R} \setminus \{0\}$, and let $x_n = 1/n$. Then the (x_n) do not converge, but are Cauchy as they are Cauchy in $\mathbb{R}$.

Similarly, we can consider $\mathbb{Q}$, then this does not satisfy the general principle of convergence.

Definition 2.5. Let (X, d) be a metric space. We say X is *complete* if every Cauchy sequence in X converges.

> ### Example 2.6.
>
> 1. $\mathbb{R} \setminus \{0\}$ and $\mathbb{Q}$ are not complete.
>
> 2. $\mathbb{R}$ with the usual metric is complete.
>
> 3. The general principle of convergence for $\mathbb{R}^n$ say $\mathbb{R}^n$ with the Euclidean metric is complete.
>
> 4. If $x \subset \mathbb{R}$ and $B(X) = \{f : X \to \mathbb{R} \mid f \text{ is bounded}\}$ with the uniform norm, then $B(X)$ is complete. Indeed, let (f_n) be a Cauchy sequence in $B(X)$. Then (f_n) is uniformly Cauchy, so by the general principle of uniform convergence, it is uniformly convergent, so $f_n \to f$ uniformly for some $f : X \to \mathbb{R}$.
>
>   This gives $f_n - f$ is bounded for n sufficiently large. Take such an n, and then since f_n bounded, $f = f_n - (f_n - f)$ implies f is bounded, so $f \in B(X)$. Finally, $f_n \to f$ uniformly implies $d(f_n, f) \to 0$, so $f_n \to f$ in $(B(X), d)$.

Remark. This is a typical example of a proof that a given space (X, d) is complete:

  (i) Take (x_n) Cauchy in X.

 (ii) Construct a limit object where it seems $(x_n) \to x$.

(iii) Show $x \in X$.

(iv) Show $x_n \to x$ in the metric space (X, d).

It is important to do things in this order, as we cannot talk about $d(x_n, x)$ until we known $x \in X$.

> ### Example 2.7.
>
> If $[a, b]$ is a closed interval, then $\mathcal{C}([a, b])$ with the uniform norm d is complete. Indeed, let (f_n) be Cauchy in $\mathcal{C}([a, b])$. Then since $\mathcal{C}([a, b]) \subset B([a, b])$, and $B([a, b])$ is complete, then $f_n \to f$ for some $f \in B([a, b])$.
>
> Each function is continuous, and $f_n \to f$ uniformly, so f is continuous, giving $f \in \mathcal{C}([a, b])$. Finally, $f_n \to f$ uniformly gives $d(f_n, f) \to 0$.

Definition 2.6. Let (X, d) be a metric space and $Y \subset X$. We say Y is *closed* if whenever (x_n) is a sequence in Y with $x_n \to x \in X$, then $x \in Y$.

Proposition 2.4. *A closed subset of a complete metric space is complete.*

Remark. This makes sense: if $Y \subset X$, then Y itself is a metric space as a subspace of X, so we can say Y is complete.

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

> ### Example 2.8.
>
> Define
> $$\ell_1 = \left\{ (x_n)_{n \geq 1} \in \mathbb{R}^{\mathbb{N}} \mid \sum_{n=1}^{\infty} |x_n| \text{ converges} \right\}.$$
> We can define a metric d on ℓ_1 by
> $$d((x_n), (y_n)) = \sum_{n=1}^{\infty} |x_n - y_n|.$$
> Then since $\sum |x_n|$, $\sum |y_n|$ converge and $|x_n - y_n| \leq |x_n| + |y_n|$, by comparison test $\sum |x_n - y_n|$ converges, so d is well-defined. It is easy to check that d is a metric. Then (l_1, d) is complete. Indeed, let $(x^{(n)})$ be a Cauchy sequence in l_1, so $(x_i^{(n)})$ is a sequence in $\mathbb{R}$.

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

We next show that $x \in l_1$. Given $y \in l_1$, define

$$\sigma(y) \sum_{i=1}^{\infty} |y_i|,$$

i.e. $\sigma(y) = d(y, \bar{0})$, where $\bar{0}$ is the constant zero sequence. Now we have, for any m, n,

$$\sigma(x^{(m)}) = d(x^{(m)}, z) \leq d(x^{(m)}, x^{(n)}) + d(x^{(n)}, \bar{0}) = d(x^{(m)}, x^{(n)}) + \sigma(x^{(n)}).$$

This gives $\sigma(x^{(m)}) - \sigma(x^{(n)}) \leq d(x^{(m)}, x^{(n)})$. Similarly, we can swap around m, n, to give $|\sigma(x^{(m)}) - \sigma(x^{(n)})| \leq d(x^{(m)}, x^{(n)})$, so $(\sigma(x^{(m)}))$ is a Cauchy sequence in $\mathbb{R}$, and so converges to K. Now we claim for any $I \in \mathbb{N}$,

$$\sum_{i=1}^{I} |x_i| \leq K + 2.$$

Indeed, as $\sigma(x^{(n)}) \to K$ as $n \to \infty$, we can find N_1 such that for all $n \geq N_1$,

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

For all $i \in \{1, 2, \ldots, I\}$, we have $x_i^{(n)} \to x$, so we can find N_2 such that for all $n \geq N_2$ and $i \in \{1, 2, \ldots, I\}$ we have $|x_i^{(n)} - x_i| < 1/I$. Letting $n = \max N_1, N_2$, we have

$$\sum_{i=1}^{I} |x_i| \leq \sum_{i=1}^{I} |x_i^{(n)}| + \sum_{i=1}^{I} |x_i^{(n)} - x_i| \leq K + 1 + 1 = K + 2.$$

Since the partial sums $\sum |x_i|$ are increasing and bounded above, they converge.

Finally, we need to check $x^{(n)} \to x$ as $n \to \infty$ in l_1, i.e. $d(x^{(n)}, x) \to 0$. Note that

$$d(x^{(n)}, x) = \sum_{i=1}^{\infty} |x_i^{(n)} - x_i| \leq \sum_{i=1}^{I} |x_i^{(n)} - x_i| + \sum_{i=I+1}^{\infty} |x_i^{(n)}| + \sum_{i=I+1}^{\infty} |x_i|.$$

Let $\varepsilon > 0$. We know that $\sum |x_i|$ is convergent, so we can pick I_1 such that

$$\sum_{i=I_1+1}^{\infty} |x_i| < \varepsilon.$$

As $(x^{(n)})$ is Cauchy, we can find N_1 such that for all $m, n \geq N_1$, $d(x^{(m)}, x^{(n)}) < \varepsilon$. As $\sum |x_i^{(N_1)}|$ converges, we can find I_2 such that

$$\sum_{i=I_2+1}^{\infty} |X_i^{(N_1)}| < \varepsilon.$$

Then for $n \geq N_1$, we have

$$\sum_{i=I_2+1}^{\infty} |x_i^{(n)}| \leq \sum_{i=I_2+1}^{\infty} |x_i^{(N_1)}| + \sum_{i=I_2+1}^{\infty} |x_i^{(n)} - x_i^{(N_1)}| < \varepsilon + d(x^{(n)}, x^{(N_1)}) < 2\varepsilon.$$

Let $I = \max\{I_1, I_2\}$. For each $i = 1, 2, \ldots, I$, we have $|x_i^{(n)} - i| \to 0$ as $n \to \infty$, so

$$\sum_{i=1}^{I} |x_i^{(n)} - x_i| \to 0$$

as $n \to \infty$. Hence we can find N_2 such that for $n \geq N_2$,

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

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

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

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

Definition 2.7. Let (X, d) be a metric space and $f : X \to X$. We say f is a *contraction* if there exists $\lambda \in [0, 1)$ such that for all $x, y \in X$, $d(f(x), f(y)) \leq$

$\lambda d(x, y)$.

Theorem 2.1 (Contraction Mapping Theorem)**.** *Let (X, d) be a complete, non-empty metric space and $f : X \to X$ a contraction. Then f has a unique fixed point.*

Proof: We show the existence of a fixed point. Let $\lambda \in [0, 1]$ satisfy $d(f(x), f(y)) \leq \lambda d(x, y)$. Let $x_0 \in X$, and define $x_n = f(x_{n-1})$. Let $\Delta = d(x_0, x_1)$. Then by induction, $d(x_n, x_{n+1}) \leq \lambda^n \Delta$. Suppose $N \leq m < n$. Then

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

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

Now suppose $f(y) = y$ for some $y \in X$. Then $d(x, y) = d(f(x), f(y)) \leq \lambda d(x, y)$, with $\lambda < 1$. So $d(x, y) = 0$, i.e. $x = y$.

Remark.

1. f is continuous as for all $x, y \in X$, $d(f(x), f(y)) \leq d(x, y)$. So for all $\varepsilon > 0$, if $d(x, y) < \varepsilon$, then $d(f(x), f(y)) < \varepsilon$. Indeed, this implies f is uniformly continuous.

2. We have proven more than claimed. Not only does f have a unique fixed point, but starting from any point of the space and repeatedly applying f, the resulting sequence converges to the fixed point. In fact, the speed of convergence is exponential.

Example 2.9.

Consider $f(x) = \cos x$ on $\mathbb{R}$. Suppose we want to find a solution to $f(x) = x$. Any root must lie in $[-1, 1]$, so consider $X = [-1, 1]$ with the usual metric. Then X is closed and bounded, hence complete. Thinking of $f : X \to X$, then suppose $x, y \in [-1, 1]$. We have

$$|\cos x - \cos y| = |x - y| |\cos' z| = |x - y| |-\sin z| \leq |x - y| \sin 1.$$

But $0 \leq \sin 1 < 1$, so cos is a contraction of $[-1, 1]$. Thus by the contraction

mapping theorem, cos has a unique fixed point in $[-1, 1]$. Thus $\cos x = x$ has a unique solution. To find this solution, we can apply f repeatedly to a point, which exponentially converges to the root.

2.3   Sequential Compactness

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

Definition 2.8. Let (X, d) be a metric space. We say X is bounded if there exists $M \in \mathbb{R}$, such that for all $x, y \in X$, $d(x, y) \leq M$.

Remark. It is easy to check (from the triangle inequality) that X bounded is equivalent to either $X = \emptyset$ or there existing $M \in \mathbb{R}, x \in X$ such that $d(x, y) \leq M$ for all $y \in X$.

Recall the definition of a closed set. We say $Y \subset X$ is closed in X if whenever (x_n) is a sequence in Y with $x_n \to x \in X$, then $x \in Y$.

Definition 2.9. A metric space is *sequentially compact* if every sequence has a compact subsequence.

Theorem 2.2. *Let $X \subset \mathbb{R}^n$ with the Euclidean metric. Then X is sequentially compact if and only if X is closed and bounded.*

> **Proof:**   Suppose X is closed and bounded. Let (x_n) be a sequence in X. Then (x_n) is a bounded sequence in $\mathbb{R}^n$ so by Bolzano-Weierstrass, in $\mathbb{R}^n$, $x_{n_j} \to x$ for some $x \in \mathbb{R}^n$ and some subsequence (x_{n_j}) of (x_n).
>
> As X is closed, $x \in X$. Hence the subsequence (x_{n_j}) converges in X. So X is sequentially compact.
>
> Suppose X is not closed. Then we can find a sequence $(x - n)$ in X such that in $\mathbb{R}^n$, $x_n \to x \in \mathbb{R}^n$ with $x \notin X$. Now any subsequence $x_{n_j} \to x$ in $\mathbb{R}^n$. But $x \notin X$, so (x_{n_j}) does not converge in X. Hence X is not sequentially compact.
>
> Suppose instead X is not bounded. Then we can find a sequence (x_n) in X with $\|x_n\| \geq n$. Suppose we have a subsequence $x_{n_j} \to x \in X$. Then $\|x_{n_j}\| \to \|x\|$, but $\|x_{n_j}\| \to \infty$. So X is not sequentially compact.

Consider the same problem in a general metric space. Then this does not work: consider $\mathbb{R} \setminus \{0\}$, and $X = (0, 1]$, which is closed and bounded. Then the sequence

$(1/n)_{n \geq 1}$ has no convergent subsequence.

This does not work as the space is not complete, so we consider the same problem in a complete space. However, this still doesn't work. Let

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

with the uniform metric. Then X is complete and bounded, but consider

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

Then (f_n) is a sequence in X, but for all $m \neq n$, $d(f_m, f_n) = 1$. So (f_n) cannot have a convergent subsequence.

In this case, our space X is too big, so we can find infinitely many objects which are far apart. This calls for a stronger concept of boundedness.

Definition 2.10. Let (X, d) be a metric space. We say X is *totally bounded* if for all $\delta > 0$, we can find a finite set $A \subset X$ such that for all $x \in X$, there exists $a \in A$ such that $d(x, a) < \delta$.

This is the relevant condition needed to extend our theorem from $\mathbb{R}^n$ to a general metric space.

Theorem 2.3. *A metric space is sequentially compact if and only if it is complete and totally bounded.*

> **Proof:**   Suppose (X, d) is complete and totally bounded. Let (x_n) be a sequence in X.
>
> As X is totally bounded, we can find a finite sequence $A_1 \subset X$ such that for all $x \in X$, there exists $a \in A_1$ such that $d(x, a) < 1$. In particular, there is an infinite set $N_1 \subset \mathbb{N}$ and a point $a_i \in A_i$ such that for all $n \in N_1$, $d(x_n, a_1) < 1$. Hence for all $m, n \in N_1$, $d(x_m, x_n) < 2$.
>
> Similarly, we can find finite $A_2 \subset X$ such that for all $x \in X$, there exists $a \in A_2$ with $d(x, a) < 1/2$. In particular, there is an infinite $N_2 \subset N_1$ such that for all $n \in N_2$, $d(x_n, a_2) < 1/2$, and then for all $m, n \in N_2$, $d(x_m, x_n) < 1$.
>
> Continue to get a sequence $N_1 \supset N_2 \supset N_3 \supset \cdots$ of infinite subset of $\mathbb{N}$ such that for all $m, n \in N_1$, we have $d(x_m, x_n) < 2/i$.
>
> Pick $n_1 \in N_1$, $n_2 \in N_2$ with $n_2 > n_1$, and successively pick $n_k \in N_k$ with

$n_k > n_{k-1}$. We obtain a subsequence (x_{n_j}) of (x_n) such that for all j, $x_{n_j} \in N_j$. Thus if $i \leq j$, then $x_{n_i}, x_{n_j} \in N_1$, and so $d(x_{n_i}, x_{n_j}) < 2/i$. Hence (x_{n_j}) is a Cauchy sequence, and hence by completeness converges. Thus X is sequentially compact.

Suppose now that X is not complete. Then X has a Cauchy sequence (x_n) which doesn't converge. Suppose we have a convergent subsequence, say $x_{n_j} \to x$. Then we can prove $x_n \to x$.

Suppose instead X is not totally bounded. Then there is some $\delta > 0$ such that whenever $A \subset X$ is finite, there exists $x \in X$ such that for all $a \in A$, $d(x, a) \geq \delta$.

Thus, pick $x_1 \in X$. Pick $x_2 \in X$ such that $d(x_1, x_2) \geq \delta$, pick $x_3 \in X$ such that $d(x_1, x_3), d(x_2, x_3) \geq \delta$, and continue to get a sequence (x_n) in X such that for all $i \neq j$, $d(x_i, x_j) \geq \delta$. Hence (x_n) has no convergent subsequence.

Corollary 2.1. *A continuous function on a sequentially compact metric space is uniformly continuous. If the function is real-valued, the it is bounded and attains its bounds.*

2.4   Topology of Metric Spaces

In $\mathbb{R}^n$, we have different concepts of distance given by Euclidean distance ℓ_1 and ℓ_∞. But these all give the same concept of convergence and continuity. We want some way to visualise why these are all the same.

Definition 2.11. Let (X, d) and (Y, e) be metric spaces. Let $f : X \to Y$. We say f is a *homeomorphism* and that X and Y are homeomorphic if f is a continuous bijection with continuous inverse.

Remark. Homeomorphism is an equivalence relation.

Example 2.10.

1. If $x, y \in \mathbb{R}^n$, then
$$d_\infty(x, y) \leq d_1(x, y) \leq n d_\infty(x, y).$$

   So the identity map $\mathbb{R}^n \to \mathbb{R}^n$ is continuous from $(\mathbb{R}^n, d_1) \to (\mathbb{R}^n, d_\infty)$, and inverse is simply itself. So this is a homeomorphism.

   Similarly, $\mathbb{R}^n$ with the Euclidean metric is homeomorphic to both these spaces.

2. If $(X, d) \to (Y, e)$ are metric spaces and $f : X \to Y$ is a bijection satisfying

   (i) There exists A such that for all $x, y \in X$, $e(f(x), f(y)) \leq d(x, y)$;

   (ii) There exists B such that for all $x, y \in X$, $d(x, y) \leq Be(f(x), f(y))$;

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

3. Define
$$f : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$$
   by $f(x) = \tan x$. Then f is a homeomorphism, but is not of type 2.

Proposition 2.5. *Let $(V, b), (W, c), (X, d), (Y, e)$ be metric spaces and $f : X \to V$, $g : Y \to W$ be homemorphisms.*

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

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

Proof:

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

(ii) h is continuous implies $g \circ h \circ f^{-1}$ is continuous. But $g \circ h \circ f^{-1}$ continuous implies $h = g^{-1} \circ (g \circ h \circ f^{-1}) \circ f$ is continuous.

We now have examples of metric spaces that look different but behave identically with respect to convergence and continuity. This begs the question: can we dispense with the notion of distance altogether?

We begin by looking at continuity in a different way.

Definition 2.12. Let (X, d) be a metric space. Let $a \in X$ and let $\varepsilon > 0$. The *open ball* of radius ε about a is the set

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

Remark. Suppose $f : X \to Y$, where d is the metric on X, e is the metric on Y Let $a \in X$. Then f is continuous at a if and only if for all $\varepsilon > 0$, there exists $\delta > 0$ such that $d(x, a) < \delta \implies d(f(x), f(y)) < \varepsilon$.

But this is equivalent to, for all $\varepsilon > 0$, there existing $\delta > 0$ such that $x \in B_\delta(a) \implies$ $f(x) \in B_\varepsilon(f(a))$, or $f(B_\delta(a)) \subset B_\varepsilon(f(a))$, or $B_\delta(a) \subset f^{-1}(B_\varepsilon(f(a)))$.

So we can redefine continuity in terms of open balls. But open balls still refer to distance with their radii, so we shall extend this notion further.

Definition 2.13. Let X be a metric space. A subset $G \subset X$ is *open* if for all $x \in G$, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset G$.

A subset $N \subset X$ is a *neighbourhood* of a point $a \in X$ if there exists an open set $G \subset X$ such that $a \in G \subset N$.

Remark.

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

2. The open ball $B_\varepsilon(x)$ is open. If $x \in B_{\varepsilon(a)}$, then $d(x, a) = \delta < \varepsilon$. So by the triangle inequality, $B_{\varepsilon-\delta}(x) \subset B_\varepsilon(a)$.

3. If $\mathcal{N}$ is an open set and $a \in \mathcal{N}$, then certainly $\mathcal{N}$ is a neighbourhood of a, as $a \in \mathcal{N} \subset \mathcal{N}$.

   However a neighbourhood of a need not be open, for example $[-1, 1]$ is a neighbourhood of 0 in $\mathbb{R}$, as

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

4. $\mathcal{N}$ is a neighbourhood of a if and only if there exists $\varepsilon > 0$ such that $B_\varepsilon(a) \subset \mathcal{N}$.

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

Proposition 2.6. *Let $(X, d), (Y, e)$ be metric spaces and $f : X \to Y$.*

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

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

Proof:

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

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

(ii) Suppose f is continuous. Let $G \subset Y$ be open, and let $a \in f^{-1}(G)$. Then $f(a) \in G$, and G is open so G is a neighbourhood of $f(a)$. Moreover, f is continuous at a, so $f^{-1}(G)$ is a neighbourhood of a. Hence there exists $\delta > 0$ such that $B_\delta(a) \subset f^{-1}(G)$. So $f^{-1}(G)$ is open.

Let $a \in X$. Let $\mathcal{N} \subset Y$ be a neighbourhood of $f(a)$. Then there exists $G \subset Y$ open such that $f(a) \in G \subset \mathcal{N}$. By our assumption, $f^{-1}(G) \subset X$ is open. Now $a \in f^{-1}(G) \subset f^{-1}(\mathcal{N})$ with $f^{-1}(G)$ open, so $f^{-1}(\mathcal{N})$ is a neighbourhood of a. So by (i), f is continuous at a, hence is continuous everywhere.

Remark.

1. This says that we can define continuity entirely in terms of open sets without mentioning the metric.

2. We say previously that homeomorphisms preserve convergence and continuity. This proposition says homeomorphisms also preserve open sets.

   Precisely, if $f : X \to Y$ is a homeomorphism, then $G \subset X$ is open is open if and only if $f(G) \subset Y$ is open. This is by using the fact that f and f^{-1} are continuous.

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

$$y_{n_j} = f(f^{-1})(y_{n_j}) \to f(x) \in Y.$$

So Y is sequentially compact. Thus if X, Y are homeomorphic, then X is sequentially compact if and only if Y is sequentially compact.

Thus we can say sequential compactness is a topological property. However, completeness is not the same.

> **Example 2.11.**
>
> We see $(0, 1)$ and $\mathbb{R}$ with the usual metric in each case are homeomorphic. But $\mathbb{R}$ is complete and $(0, 1)$ is not. So completeness is not a topological property.

In this case, the property of being a Cauchy sequence is not preserved by homeomorphisms.

Remark. Suppose (x_n) is a sequence in a metric space X and $x \in X$. Then $x_n \to x$ if and only if, for all $\varepsilon > 0$, there exists N such that for all $n \geq N$, $d(x_n, x) < \varepsilon$. This is equivalent to, for all neighbourhood $\mathcal{N}$ of x, there exists N such that for all $n \geq N$, $x_n \in \mathcal{N}$.

This defines convergence solely in terms of neighbourhoods. We can't do something similar for Cauchy sequences, however.

We have just seen that sequential compactness is a topological property. We can define sequential compactness just in terms of neighbourhoods and open sets, as we can express convergence of sequence in terms of neighbourhoods.

We may ask whether there is a better way to do this.

Definition 2.14. Let X be a metric space. An open cover of X is a collection $\mathcal{C}$ of open subsets of X such that

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

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

Proposition 2.7 (Heine-Borel Theorem). *$[0, 1]$ with the usual metric is compact.*

> **Proof:**   Let $\mathcal{C}$ be an open cover of $[0, 1]$. Let
>
> $$A = \{x \in [0, 1] \mid \exists \mathcal{B} \subset \mathcal{C}, [0, x] \subset \bigcup_{G \in \mathcal{B}} G, |\mathcal{B}| < \infty\}.$$
>
> We know there exists $G \in \mathcal{C}$ with $0 \in G$. So $0 \in A$. Clearly A is bounded above by 1, so A has a supremum $\sigma = \sup A$.
>
> As G is open, there exists $\varepsilon > 0$ such that $[0, \varepsilon) = B_\varepsilon(0) \subset G$. So $\sigma > 0$.
>
> Suppose $\sigma < 1$. We can find $G' \in \mathcal{C}$ with $\sigma \in G'$. As $\sigma = \sup A$, we can find

$x \in A$ with $x \in G'$. So we have $\mathcal{B} \subset \mathcal{C}$ finite with

$$[0, x] = \bigcup_{G' \in \mathcal{B}} G'.$$

But there exists $\varepsilon > 0$ such that $(\sigma - \varepsilon, \sigma + \varepsilon) = [B_\varepsilon(\sigma)]$. So

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

This is a contradiction, so $\sigma = 1$. Hence we can find $G'' \in \mathcal{C}$ such that $1 \in G''$. As G'' is open, we can find $\varepsilon > 0$ such that $(1 - \varepsilon, 1] \subset G''$. As $1 = \sup A$, we can find $x \in A \cap (1 - \varepsilon, 1]$. This says that we have a finite $\mathcal{B} \subset \mathcal{C}$ with

$$[0, x] \subset \bigcup_{G \in \mathcal{B}} G.$$

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

Theorem 2.4. *Let X be a metric space. Then the following are equivalent:*

(i) *X is compact;*

(ii) *X is sequentially compact;*

(iii) *X is complete and totally bounded;*

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

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

Proof: We have already shown (ii) $\iff$ (iii) (and possibly $\iff$ (iv)). So it suffices to show (i) $\iff$ (ii).

Suppose X is not sequentially compact. Then there is some sequence (x_n) in X with no convergent subsequence. Hence for every point $a \in X$, we can find a neighbourhood of a, and hence an open set G_a containing a, with only finitely many x_n.

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

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

Suppose note, then for all $\delta > 0$, there exist $a \in X$ such that for all $G \in \mathcal{C}$, $B_\delta(a) \not\subset G$. Taking $\delta = 1/n$, we obtain a sequence $(x_n) \in X$. By sequential compactness, we can find a convergent subsequence $x_{n_j} \to a \in X$.

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

Take δ as in the claim. As X is sequentially compact, it is totally bounded, so we can find a finite set $A \subset X$, such that for all $x \in X$, there exists $a \in A$ with $d(x, a) < \delta$. That is, $X = \bigcup_{a \in A} B_\delta(a)$.

By choice of δ, for each $a \in A$, we can pick $G_a \in \mathcal{C}$ such that $B_\delta(a) \subset G_a$. So $\{G_a \mid a \in A\}$ is a finite subcover, and X is compact.

Finally, we prove two important properties of open sets. First is a relationship between open and closed sets.

Proposition 2.8. *Let X be a metric space and $G \subset X$. Then G is open if and only if $F = X \setminus G$ is closed.*

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

Now suppose G is not open. Then there is some $x \in G$ such that for all $\varepsilon > 0$, $B_\varepsilon(x) \not\subset G$. That is, $B_\varepsilon(x) \cap F \neq \emptyset$. So for $n \in \mathbb{N}$, we can pick $x_n \in B_{1/n}(x) \cap F$. Then (x_n) is a sequence in F with $x_n \to x \in G$. So F is not closed.

Secondly, if X is a metric space, we can say a few things about the structure of the collection of all open subset of X.

Proposition 2.9. *Let X be a metric space and let $\tau = \{G \subset X \mid G \text{ open}\}$. Then:*

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

(ii) *If $\sigma \subset \tau$, then $\bigcup_{G \in \sigma} G \in \tau$;*

(iii) *If $G_1, G_2, \ldots, G_n \in \tau$, then $\bigcap_{i=1}^{n} G_i \in \tau$.*

Remark. We need finiteness in (iii). Indeed, take $(-1/n, 1/n)$, which is open in $\mathbb{R}$. Then the intersection is $\{0\}$, which is not open.

Proof:

(i) True by definition.

(ii) Suppose $\sigma \subset \tau$. Let $H = \bigcup_{G \in \sigma} G$, and let $a \in H$. Then $a \in G$ for some $G \in \sigma$, so G is a neighbourhood of a. Thus H is a neighbourhood of a as $G \subset H$, and so H is open.

(iii) Suppose $G_1, \ldots, G_n \in \tau$, and let $J = \bigcap_{i=1}^{n} G_i$. Suppose $a \in J$. For each i, $a \in G_i$, and since G_i open, there exists $\delta_i > 0$ such that $B_{\delta_i}(a) \subset G_i$.

Let $\delta = \min\{\delta_1, \ldots, \delta_n\}$. Then $\delta = 0$ and

$$B_\delta(a) = \bigcap_{i=1}^{n} B_{\delta_i}(a) \subset \bigcap_{i=1}^{n} G_i = J.$$

So J is open.

3   Topological Spaces

Currently, we have extended our notion of continuity from $\mathbb{R}$ to general metric spaces, which are sets equipped with a distance. However, we can push this idea further, and talk about continuity without mentioning distances at all.

3.1   Definitions and Examples

Definition 3.1. A topological space is a set X endowed with a topology τ that is a subset $\tau \subset \mathcal{P}(X)$ satisfying

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

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

(iii) If $G_1, G_2, \dots, G_n \in \tau$, then $\bigcap_{i=1}^{n} G_i \in \tau$.

Remark. We could replace (iii) by $G, H \in \tau \implies G \cap H \in \tau$, which is equivalent to (iii) by induction.

If it is clear what the topology on a set X is, we may just write X is a topological space.

Example 3.1.

Let (X, d) be a metric space, and let $\tau = \{G \subset X \mid G \text{ open}\}$. Then τ is a topology on X.

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

We want to define the notion of being open, closed and continuous for topological spaces. As metric spaces are topological spaces, we have to ensure that our definitions for topological spaces don't contradict our definitions for metric spaces.

Definition 3.2. Let (X, τ) be a topological space. We say $G \subset X$ is *open* if $G \in \tau$. We say F is *closed* if $X \setminus F \in \tau$. We say $\mathcal{N} \subset X$ is a *neighbourhood* of $a \in X$ if there exists $G \subset X$ open with $a \in G \subset \mathcal{N}$.

Definition 3.3. Let (X, τ) and (Y, σ) be another topological space, and let $f : X \to Y$. We say f is *continuous* if, whenever $G \subset Y$ is open, then $f^{-1}(G) \subset X$ is open.

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

We say f is a *homeomorphism* and X, Y are *homeomorphic* if f is a bijection and both f, f^{-1} are continuous.

A property if *topological* if it is preserved by homeomorphisms.

Remark.

1. If τ is induced by a metric, then this is all consistent with the metric space definitions of these concepts.

2. Given our definition, then G open $\iff$ $G \in \tau$. We often don' need to explicitly name the topology.

3. Homeomorphism is an equivalence relation.

4. If $a \in G$ and G open, then G is a neighbourhood of a, however neighbourhoods need not be open in general. A set $G \subset X$ is open if and only if G is a neighbourhood of each of its points (for all $a \in \mathcal{N}$, take the union of $a \in G_a \subset \mathcal{N}$).

Proposition 3.1. *Let X, Y be topological spaces and let $f : X \to Y$. Then f is continuous if and only if for all $a \in X$, f is continuous at a.*

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

Suppose for all $a \in X$, we have f continuous at a. Let $G \subset Y$ be open. Let $a \in f^{-1}(G)$. Then $f(a) \in G$, but G is open, so G is a neighbourhood of $f(a)$ Since f is continuous at a, $f^{-1}(G)$ is a neighbourhood of a in X.

But a was arbitrary, so $f^{-1}(G)$ is a neighbourhood of each of its points. Hence it is open, and f is continuous.

Proposition 3.2. *Let $(X, \tau), (Y, \sigma), (Z, \rho)$ be topological spaces, and let $f : X \to Y$ and $g : Y \to Z$ be continuous. Then $g \circ f : X \to Z$ is continuous.*

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

Example 3.2.

1. We can take the discrete topology, by letting X be an arbitrary set and $\tau = \mathcal{P}(X)$, so every set is open.

   However, this is simply the topology induced by the discrete metric

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

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

2. Let X be any set and let $\tau = \{\emptyset, X\}$. Then this is the indiscrete topology, and is not induced by a metric if $|X| \geq 2$.

   Indeed, suppose $|X| \geq 2$ and τ is induced by a metric d. Let $x, y \in X$ with $x \neq y$, so $d(x, y) = \delta > 0$. Then $B_\delta(x)$ is open with $x \in B_\delta(x)$ and $y \notin B(\delta(x))$.

3. Let X be any infinite set and let $\tau = \{G \subset X \mid X \setminus G \text{ finite}\} \cup \{\emptyset\}$. This is the cofinite topology, and we can check this is a topology:

   (i) It contains $\emptyset$ and $X \setminus X = \emptyset$ is finite so $X \in \tau$.

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

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

   is finite.

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

   So the cofinite topology is indeed a topology. Suppose that it is induced by a metric d. Then if G, H are open an non-empty, then $G \cap H \neq \emptyset$. Now suppose $x, y \in X$ with $x \neq y$. Then $(x, y) = \delta > 0$ so $B_{\delta/2}(x), B_{\delta/2}(y)$ are non-empty disjoint open sets. So d doesn't induce τ.

4. Let X be any uncountable set and let $\tau = \{G \subset X \mid X \setminus G \text{ countable}\} \cup \{\emptyset\}$. This is known as the cocountable topology. Similarly to the cofinite topology, this is not induced by any metric.

3.2   Sequences and Hausdorff Spaces

Definition 3.4. Let X be a topological space, and (x_n) be a sequence in X, with $x \in X$. We say (x_n) converges to x and write $x_n \to x$ if whenever $\mathcal{N} \subset X$ is a neighbourhood of x, then there exists N such that for all $n \geq N$, $x_n \in \mathcal{N}$.

> **Example 3.3.**
>
> 1. Let X be an uncountable set with the cocountable topology. Suppose $x_n \to x$. Then let $\mathcal{N} = (X \setminus \{x_n \mid n \in \mathbb{N}\}) \cup \{x\}$. Then $\mathcal{N}$ is open and $x \in \mathcal{N}$, so $\mathcal{N}$ is a neighbourhood of x. Then there exist N such that for all $n \geq N$, $x_n \in \mathcal{N}$. This implies $x_n = x$. Thus the only convergent sequences in this space are eventually constant.
>
> 2. Let $X = \{1, 2, 3\}$ with the indiscrete topology. Let $x_n = i \in X$ with $i \equiv n \pmod 3$, so x_n is the sequence $(1, 2, 3, 1, 2, 3, \ldots)$. Then $x_n \to 2$.
>
>   Indeed, let $\mathcal{N}$ be a neighbourhood of 2. Then there exists G open such that $2 \in G \subset \mathcal{N}$. But the only open sets are $\emptyset$ and $\{1, 2, 3\}$, so $G = \{1, 2, 3\}$. Hence $\mathcal{N} = \{1, 2, 3\}$, so for all n, $x_n \in \mathcal{N}$. Similarly, $x_n \to 1$ and $x_n \to 3$.

Importantly, this says that the **limit of convergent subsequences may not be unique**, so we cannot write $\lim x_n$. In fact, the above proof shows that in any indiscrete space, every sequence converges to every point of the space.

Definition 3.5. A topological space X is *Hausdorff* if whenever $x, y \in X$ with $x \neq y$, then there are disjoint open $G, H \subset X$ with $x \in G$ and $y \in H$.

> **Example 3.4.**
>
> 1. Metric spaces are Hausdorff. Indeed, if (X, d) is metric and $x, y \in X$ with $x \neq y$, then let $\delta = d(x, y) > 0$ and take $G = B_{\delta/2}(x)$ and $H = B_{\delta/2}(y)$.
>
> 2. Indiscrete spaces are not Hausdorff (assuming $|X| \geq 2$).
>
> 3. The cofinite topology is not Hausdorff. Let X be an infinite set with the cofinite topology and let $x, y \in X$ with $x \neq y$. Let $G, H \subset X$ be open with $x \in G, y \in H$. Clearly $G, H \neq \emptyset$ so $X \setminus G, X \setminus H$ is finite and so $X \setminus (G \cap H) = (X \setminus G) \cup (X \setminus H)$ is finite. In particular, $G \cap H \neq \emptyset$.
>
>   Similarly, the cocountable topology is not Hausdorff.

Proposition 3.3. *Limits of convergent sequences in Hausdorff spaces are unique.*

Proof: Let X be Hausdorff, and let $a, b \in X$ and (x_n) be a sequence in X with $x_n \to a$ and $x_n \to b$.

Suppose $a \neq b$. Take open G, H with $a \in G$, $b \in H$ and $G \cap H = \emptyset$.

Now G is a neighbourhood of a, so there is some N_1 such that $x_n \in G$ for all $n \geq N_1$. Similarly, there is some N_2 such that $x_n \ni H$ for all $n \geq N_2$. $Taken = \max\{N_1, N_2\}$. Then $x_n \in G \cap H = \emptyset$, which is a contradiction. So $a = b$.

Proposition 3.4. *Let X, Y be topological spaces and let $f : X \to Y$ be continuous at $a \in X$. Let (x_n) be a sequence in X with $x_n \to a$. Then $f(x_n) \to f(a)$.*

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

However, **the converse is not true in general**.

> **Example 3.5.**
>
> Let $X = Y = \mathbb{R}$, where X has the cocountable topology and Y has the usual topology. Define $f : X \to Y$ as the identity.
>
> Suppose $x_n \to 0$ in X. Then for sufficiently large n, $x_n = 0$, and so for sufficiently large n, $f(x_n) = x_n = 0 = f(0)$, so $f(x_n) \to f(0)$ in Y.
>
> However $(-1, 1) \subset Y$ is open and $0 \in (-1, 1)$, so $(-1, 1)$ is a neighbourhood of 0 in Y. But $f^{-1}((-1, 1)) = (-1, 1) \subset X$ is not a neighbourhood of 0 in X, so f is not continuous at 0.
>
> Moreover, this is not true even if both spaces are Hausdorff. Take the above example but replace the topology on X by
>
> $$\sigma = \{G \subset \mathbb{R} \mid (X \setminus G) \text{ countable or } 0 \notin G\}.$$
>
> This is a topology and is Hausdorff. as for $x, y \neq 0$, $\{x\}, \{y\} \in \sigma$, and for $x = 0$, $X \setminus \{y\}, \{y\} \in \sigma$.
>
> Notice the neighbourhoods of 0 in σ are exactly the same in the cocountable topology. So exactly as before, $x_n \to 0$ in X implies $x_n \to 0$ in Y, but f is not continuous at 0.

Remark. In a metric space, the topology is completely determined by convergence of sequences. This is not true for a general topological space. Hence we tend to focus on continuity more than convergence of sequences.

3.3   Subspaces

Definition 3.6. Let (X, τ) be a topological space and let $Y \subset X$. The *subspace topology* on Y is

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

It is easy to see this is a topology.

Since we have define a subspace of a metric space, we need to check backwards compatibility.

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

Proof: Let σ be the topology on Y induced by the metric $d \mid_Y$.

Suppose $G \in \tau$. Let $y \in G \cap Y$. As $y \in G$ and G open in X, we can find $\delta > 0$ such that for all $x \in X$, $d(x, y) < \delta \implies x \in G$. Then, for all $x \in Y$, $d(x, y) < \delta \implies x \in G \cap Y$. So $G \cap Y$ is a neighbourhood of y.

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

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

Each ball is open, and for each $y \in H$, $y \in B_{\delta_y}(y)$ and $B\delta_y(y) \cap Y \subset H$. Let G be the union of these balls. Then G is open and $G \cap Y = H$, so we can find $G \in \tau$ such that $G \cap Y = H$.

Proposition 3.6. *A subspace of a Hausdorff space is Hausdorff.*

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

3.4   Compactness

Definition 3.7. Let (X, τ) be a topological space. An *open cover* of X is a subset $\mathcal{C} \subset \tau$ such that $X = \bigcup_{G \in \mathcal{C}} G$. A *subcover* of $\mathcal{C}$ is a $\mathcal{D} \subset \mathcal{C}$ which is itself an open cover.

We say X is *compact* if every open cover of X has a finite subcover. We say X is *sequentially compact* if every sequence in X has a convergent subsequence.

It can be shown a continuous real-valued function on a sequentially compact topological space is bounded and attains its bounds.

Remark. Here and elsewhere, if no topology is specified on $\mathbb{R}$, it is generally assumed to have the usual topology.

We've seen for a metric space that compactness and sequential compactness are equivalent. This is **not true for general topological spaces**. There exist compact spaces which are not sequentially compact, and sequentially compact spaces which are not compact.

Observe that compactness and sequential compactness are both topological properties. Given we want to focus on continuity rather than convergence, we will be focusing on compactness rather than sequential compactness.

Remark. If X is a topological space, we might want to say 'K is compact'. This is meaningful, as K is a topological space with the subspace topology.

More clearly, if τ is the topology on X, then K is compact if and only if whenever $\mathcal{C} \subset \tau$ with $K = \bigcup_{G \in \mathcal{C}} G \cap K$, then there is a finite $\mathcal{D} \subset \mathcal{C}$ such that $K = \bigcup_{G \in \mathcal{D}} G \cap K$.

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

> ### Example 3.6.
>
> 1. $[0, 1]$ with the usual topology is compact (from the Heine-Borel theorem). More generally, $S \subset \mathbb{R}^n$ is compact if and only if S is closed and bounded.
>
> 2. A metric space is compact if and only if it is complete and totally bounded.
>
> 3. Suppose X is a discrete topological space. Then $\{\{x\} \mid x \in X\}$ is an

open cover. Hence X is compact if and only if X is finite (note any finite space is compact).

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

Theorem 3.1. *A continuous real-valued function on a compact topological space is bounded and attains its bounds.*

Proof: Let X be compact and $f : X \to \mathbb{R}$ be continuous. Let $G_n = f^{-1}((-n, n))$, for $n \in \mathbb{N}$. Then $\{G_n \mid n \in \mathbb{N}\}$ is an open cover of X, and so, as X compact, there is a finite subcover $\{G_{n_1}, \ldots, G_{n_k}\}$. Note for all $x \in G_{n_i}$, $|f(x)| < n_i$. Hence for all $x \in X$, $|f(x)| < \max\{n_1, \ldots, n_k\}$, giving f bounded.

Let $\sigma = \sup f(x)$, and suppose σ is not attained by f. Then we can define $g : X \to \mathbb{R}$, by $g(x) = (\sigma - f(x))^{-1}$, which is well-defined and continuous. Hence g is bounded. But as $\sigma = \sup f(x)$, given $\varepsilon > 0$ we can find x such that $\sigma - f(x) < \varepsilon$, hence $g(x) > \varepsilon^{-1}$. So σ is attained, and similarly $\inf f(x)$ is attained.

Remark. We can think of compactness as a 'smallness' condition, perhaps the next best thing to being finite.

For example, a real-valued function on a finite space is bounded.

Analogously, for a continuous function on a compact space, we show boundedness by using compactness to show that the space is not 'too big'.

We can extend the above theorem to settings outside of $\mathbb{R}$.

Theorem 3.2. *A continuous image of a compact space is compact.*

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

Then $\mathcal{D}$ is an open cover of K in Y. So K is compact.

Remark. This implies the previous theorem as compact subsets of $\mathbb{R}$ are closed and bounded.

Lemma 3.1.

(i) *A closed subset of a compact space is compact.*

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

> **Proof:**
>
> (i) Let X be a compact topological space and let $F \subset X$ be closed. Let $\mathcal{C}$ be an open cover of F in X. Then $X \setminus F$ is open, so $\mathcal{C}' = \mathcal{C} \cup \{X \setminus F\}$, then $\mathcal{C}'$ is an open cover of X.
>
> Since X is compact, $\mathcal{C}'$ has a finite subcover $\mathcal{D}'$. Let $\mathcal{D} = \mathcal{D}' \setminus \{X \setminus F\}$ if $X \setminus F \in \mathcal{D}'$, otherwise $\mathcal{D} = \mathcal{D}'$. Then $\mathcal{D}$ is a finite subcover of $\mathcal{C}$, so F is compact.
>
> (ii) Let X be a Hausdorff space, and let $K \subset X$ be compact. We want to show K is closed, i.e. $X \setminus K$ is open, so $X \setminus K$ is a neighbourhood of each of its points.
>
> Let $y \in X \setminus K$. Given $x \in K$, $x \neq y$ so as X is Hausdorff, we can find open $U_x, V_x \subset X$ with $x \in U_x$ and $y \in V_x$. Then $\{U_x \mid x \in X\}$ is an open cover of K in X, so it has a finite subcover $\{U_{x_1}, \ldots, U_{x_n}\}$. Define $U = \bigcup_{i=1}^n U_{x_i}$ and $V = \bigcap_{i=1}^n V_{x_i}$. Now we have U, V are open, $K \subset U$, $y \in V$ and $U \cap V = \emptyset$.
>
> In particular, we have found an open set V such that $y \in V \subset X \setminus K$. So $X \setminus K$ is a neighbourhood of each of its points, hence K is closed.

Theorem 3.3. *A continuous bijection from a compact space to a Hausdorff space is a homeomorphism.*

> **Proof:** Let $f : X \to Y$ be a continuous bijection, X compact and Y Hausdorff. To show f is a homeomorphism, we want to show $f^{-1} : Y \to X$ is continuous.
>
> Let $G \subset X$ be open. Then $X \setminus G$ is closed, so since X is compact, $X \setminus G$ is compact. Hence $f(X \setminus G)$ is compact, and so $f(X \setminus G)$ is closed. That is, $Y \setminus f(G)$ is closed as f is bijective, or $f(G)$ is open. But this gives $(f^{-1})^{-1}(G) = f(G)$ is open, so f^{-1} is continuous.

3.5   Products

We have $\mathbb{R}$ with the usual topology, and we have $\mathbb{R}^2$ with the usual topology. We would like a way to define $\mathbb{R} \times \mathbb{R}$ to be $\mathbb{R}^2$ with the Euclidean topology.

In general, given (X, τ) and (Y, σ), two topological spaces, we would like a sensible way to define a topology on $X \times Y$.

In general $\tau \times \sigma$ will not be a topology, as it won't be closed under union. However, by including arbitrary unions of $\tau \times \sigma$, we can turn this into a topology. To do this, we will use the following definitions:

Definition 3.8. A π-system on a set X is a non-empty subset $\Pi \subset \mathcal{P}(X)$ such that if $A, B \in \Pi$, then $A \cap B \in \Pi$.

Proposition 3.7. *Let Π be a π-system on a set X. Then,*

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

is a topology on X.

> **Proof:** Clearly $\emptyset, X \in \tau$, and τ is closed under arbitrary unions.
>
> Now suppose $G, H \in \tau$. If $G = \emptyset, X$ or $H = \emptyset, X$, then $G \cap H \in \tau$ trivially. Otherwise, we can write
>
> $$G = \bigcup_{A \in \Phi} A, \quad H = \bigcup_{B \in \Theta} B,$$
>
> for some $\Phi, \Theta \subset \Pi$. Then,
>
> $$G \cap H = \bigcup_{\substack{A \in \Phi \\ B \in \Theta}} (A \cap B) = \bigcup_{C \in \Sigma} C,$$
>
> where $\Sigma = \{A \cap B \mid A \in \Phi, B \in \Theta\} \subset \Pi$. So $G \cap H \in \tau$.

We call τ the topology generated by Π.

Proposition 3.8. *Let (X, τ), (Y, σ) be topological spaces. Then $\tau \times \sigma$ is a π-system on $X \times Y$.*

> **Proof:** Note $\emptyset = \emptyset \times \emptyset \in \tau \times \sigma$, and $X \times Y \in \tau \times \sigma$.
>
> Now suppose $A, B \in \tau \times \sigma$. We can write $A = G \times H$, $B = K \times L$ for some $G, K \in \tau$ and $H, L \in \sigma$. So,
>
> $$A \cap B = (G \cap K) \times (H \cap L) \in \tau \times \sigma.$$

Definition 3.9. Let (X, τ), (Y, σ) be topological spaces. The *product topology* on $X \times Y$ is the topology generated by the π-system $\tau \times \sigma$.

> **Example 3.7.**
>
> If $X = Y = \mathbb{R}$, then the product topology on $\mathbb{R} \times \mathbb{R}$ is the Euclidean topology on $\mathbb{R}^2$.

Theorem 3.4.

(i) *A product of Hausdorff spaces is Hausdorff.*

(ii) *A product of compact spaces is compact.*

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

(i) Suppose X, Y are Hausdorff. Let (x, y), $(z, w) \in X \times Y$ with $(x, y) \neq (z, w)$. Say $x \neq z$. As X is Hausdorff, we can find $G, H \in \tau$ with $G \cap H = \emptyset$, $x \in G$, $z \in H$.

Then $G \times Y$, $H \times Y \in \rho$ with $(G \times Y) \cap (H \times Y) = \emptyset$ and $(x, y) \in G \times Y$, $(z, w) \in H \times Y$. So $X \times Y$ is Hausdorff.

(ii) Suppose X, Y are compact. Let $\mathcal{C} \subset \rho$ be an open cover of $X \times Y$.

Fix each $x \in X$. For each $y \in Y$, there is some $G_y \in \mathcal{C}$ such that $(x, y) \in G_y$. Hence we can find $U_y \in \sigma$ and $V_y \in \tau$ such that $(x, y) \in U_y \times V_y \subset G_y$. In particular, $x \in U_y$ and $y \in V_y$.

Then $\{V_y \mid y \in Y\} \subset \sigma$ is an open cover of Y. As Y is compact, it has a finite subcover $\{V_{y_1}, \ldots, V_{y_n}\}$.

Let $W = \bigcap_{i=1}^{n} U_{y_i}$. Then W is open in X and $x \in W$. Moreover, $W \times X \subset \bigcup_{i=1}^{n} G_{y_i}$. Doing this for all $x \in X$, we obtain $W_x = W$, $n_x = n$ and $G_{y_i}^{(x)} = G_{y_i}$, as above.

Since $\{W_x \mid x \in X\} \subset \tau$ is an open cover, it has a finite subcover $\{W_{x_1}, \ldots, W_{x_m}\}$. Now $X = \bigcup_{j=1}^{m} W_{x_j}$, and for each j, $W_{x_j} \times Y \subset \bigcup_{i=1}^{n_{x_j}} G_{y_i}^{(x_j)}$. Thus,

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

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

3.6   Quotients

Consider the surface of a torus in $\mathbb{R}^3$. We may be interested in continuous function $T \to T$. However, such analysis is likely to be unpleasant if we do not know much about T.

However, since we only care about continuity and convergence, replacing T with a space homeomorphic to T, For example, taking the closed unit square $[0,1] \times [0,1]$ with the Euclidean topology and joining opposite sides seems to give us T.

More formally, we can define an equivalence relation on $[0,1] \times [0,1]$, $\sim$, say, with equivalence classes:

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

- $\{(x,0), (x,1)\}$, $0 < x < 1$;

- $\{(0,y), (1,y)\}$, $0 < y < 1$;

- $\{(0,0), (0,1), (1,0), (1,1)\}$.

This is a step in the right direction, but we wish to define $T = [0,1]^2/\sim$, hence maybe there's a better way to do this.

Instead, we could define an equivalence relation $\sim$ on $\mathbb{R}^2$ by $(x,y) \sim (z,w) \iff x - z, y - w \in \mathbb{Z}$. Again, hopefully we could define this to be T. However, we need a way to define the topology on such a space.

Definition 3.10. Let (X, τ) be a topological space and let $\sim$ be an equivalence relation on X. The *quotient topology* on $X/\sim$ is

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

where $q : X \to X/\sim$ is the quotient map, i.e. $q(x) = [x]_\sim$.

Remark.

1. ρ is indeed a topology, using $q^{-1}(\bigcup G) = \bigcup q^{-1}(G)$ and $q^{-1}(G \cap H) = q^{-1}(G) \cap q^{-1}(H)$.

2. ρ is the smallest topology on $X/\sim$ making the quotient map q continuous.
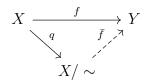
> **Example 3.8.**
>
> 1. Take $\mathbb{R}$ with the usual topology and let $x \sim y$ if and only if $x - y \in \mathbb{Z}$. Then $\mathbb{R}/\sim$ is homeomorphic to S^1, the unit circle.
>
> 2. Take $\mathbb{R}$ with the usual metric and let $x \sim y$ if and only if $x - y \in \mathbb{Q}$.

Suppose $G \subset \mathbb{R}/\sim$ is open and $G \neq \emptyset$. Then $q^{-1}(G)$ is open and non-empty, so contains some interval $(a, b) \subset q^{-1}(G)$ with $a \neq b$.

Now take any $x \in \mathbb{R}$. Then there exists $y \in (a, b)$ with $x - y \in \mathbb{Q}$, and so $q(x) = [x]_\sim = [y]_\sim \in G$. Hence $G = \mathbb{R}/\sim$.

So the quotient topology on $\mathbb{R}/\sim$ is the indiscrete topology. Hence quotient of metrizable spaces are not necessarily metrizable, and quotients of Hausdorff spaces are not necessarily Hausdorff.

Suppose X is a set and $\sim$ is an equivalence relation on X. Then the quotient map $q : X \to X/\sim$ is surjective. Suppose now Y is also a set and $f : X \to Y$. Assume f respects $\sim$, so if $x \sim y$ then $f(x) = f(y)$.

$$
\begin{array}{ccc}
X & \xrightarrow{\quad f \quad} & Y \\
& \searrow{\scriptstyle q} \quad {\scriptstyle \bar{f}} \nearrow & \\
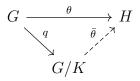& X/\sim &
\end{array}
$$

Then there is a unique function $\bar{f} : X/\sim \to Y$ such that $f = \bar{f} \circ q$.

Indeed, we must have $\bar{f}([x]_\sim) = \bar{f}(q(x)) = f(x)$, so as f respects $\sim$, this is well defined.

Example 3.9.

Suppose G, H are groups and $\theta : G \to H$ is a homomorphism. Let $K = \operatorname{Ker} \theta$ and define $\sim$ on G by $g \sim h \iff g^{-1}h \in K$. Then $G/K = G/\sim$, and we can define a map $\bar{\theta}$ that commutes with the projection $q : G \to G/K$:

$$
\begin{array}{ccc}
G & \xrightarrow{\quad \theta \quad} & H \\
& \searrow{\scriptstyle q} \quad {\scriptstyle \bar{\theta}} \nearrow & \\
& G/K &
\end{array}
$$

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

Proposition 3.9. *Let (X, τ) be a topological space and $\sim$ an equivalence relation on X. Let ρ be the quotient topology on $X/\sim$. Suppose $f : X \to Y$ is a continuous function respecting $\sim$, where (Y, σ) is a topological space.*

Then there is a unique continuous function $\bar{f} : X/\sim \to Y$ such that $f = \bar{f} \circ q$, where q is the quotient map.

Proof: Define $\bar{f} : X/\sim \to Y$ by $\bar{f}([x]_\sim) = f(x)$. This is well-defined as f respect $\sim$, and clearly $\bar{f} \circ q = f$.

Let $G \in \sigma$, then

$$q^{-1}(\bar{f}^{-1}(G)) = (\bar{f} \circ q)^{-1}(G) = f^{-1}(G) \in \tau,$$

as f is continuous. So by the definition of a quotient topology, $\bar{f}^{-1}(G) \in \rho$, hence $\bar{f}$ is continuous.

Finally, if $f = h \circ q$ for some $h : X/\sim \to Y$, then for all $x \in X$,

$$h([x]_\sim) = h(q(x)) = f(x) = \bar{f}([x]_\sim).$$

So $h = \bar{f}$.

Remark. This is ultimately what makes quotients useful. Recall we can define a torus as $T = \mathbb{R}^2/\sim$, for an appropriate relation $\sim$. Then it is easier to work in $\mathbb{R}^2$ rather than a subspace of $\mathbb{R}^3$.

So instead of thinking about continuous functions on T, we can think about an appropriate continuous function on $\mathbb{R}^2$ respecting $\sim$.

Example 3.10.

Recall that we had $\mathbb{R}$ with the usual topology, and $x \sim y$ if and only if $x - y \in \mathbb{Z}$. We also have the circle $S^1 = \{x \in \mathbb{R}^2 \mid \|x\| = 1\}$, with subspace topology induced by the Euclidean topology on $\mathbb{R}^2$.

We claimed $\mathbb{R}/\sim$ is homeomorphic to S^1. Indeed, define $f : \mathbb{R} \to S^1$ by

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

Clearly f is a continuous surjection, and it respects $\sim$. Hence there is a unique continuous $\bar{f} : \mathbb{R}/\sim \to S^1$ with $\bar{f} \circ q = f$. Then $\bar{f}$ is a continuous bijection.

Since $\mathbb{R}/\sim = q([0, 1])$ is a continuous image of a compact set it is compact. Moreover, S^1 is Hausdorff. Since any continuous bijection from a compact space to a Hausdorff space is a homeomorphism, $\bar{f}$ is a homeomorphism from $\mathbb{R}/\sim$ to S^1.

3.7   Connectedness

Recall the *intermediate value theorem*:

If $f : [a, b] \to \mathbb{R}$ is continuous, and $f(x) < f(b)$, then

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

Moreover, if $c, d \in f([a, b])$ with $c < d$, then

$$[c, d] \subset f([a, b]).$$

This doesn't work in general. Replace $[a, b]$ with $[-1, 0) \cup (0, 1] = X$. Then define $f : X \to \mathbb{R}$ by

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

Then f is continuous on X, $0, 1 \in f(X)$ but $[0, 1] \not\subset f(X)$. The problem here is $[-1, 0) \cup (0, 1]$ is 'disconnected' in some sense.

Definition 3.11. A topological space X is *disconnected* if there exist disjoint, non-empty open sets U, V with $X = U \cup V$. We say X is *connected* if X is not disconnected.

Remark. Recall that $U \subset X$ is closed if and only if its complement is open. Hence X is disconnected if and only if there exist disjoint non-empty closed sets U, V with $X = U \cup V$.

Analogously, X is connected if and only if the only subsets of X that are both open and closed are $\emptyset$ and X.

Another equivalent condition is that X is connected if and only if, whenever $U, V \subset X$ are open and disjoint with $X = U \cup V$, then either $U = \emptyset$ or $V = \emptyset$. Again, we could replace 'open' with 'closed'.

Now if X is disconnected, we say that the sets U and V which partition X *disconnect* X.

Note that connectedness is a topological property, and if $S \subset X$ and X is a topological space, then S is disconnected if and only if there exist open sets $U, V \subset X$ such that $S \cap U \cap V = \emptyset$, $S \subset U \cup V$, and $S \cap U \neq \emptyset$, $S \cap V \neq \emptyset$.

Example 3.1. We don't necessarily have $U \cap V = \emptyset$ for a set $S \subset X$ to be disconnected. Indeed, consider $\mathbb{N}$ with the cofinite topology, and consider the set $\{1, 2\}$.

Then it is disconnected in $\mathbb{N}$ with sets $\mathbb{N} \setminus \{1\}$ and $\mathbb{N} \setminus \{2\}$, but these two sets intersect in X.

A natural question is to classify the connected subsets of $\mathbb{R}$.

Definition 3.12. A subset $I \subset \mathbb{R}$ is an *interval* if whenever $a < b < c$ with $a, c \in I$, then $b \in I$.

Proposition 3.10. *Let $I \subset \mathbb{R}$ with the usual topology. Then I is connected if and only if I is an interval.*

> **Proof:** Suppose I is not an interval. Then we can find $a < b < c$ with $a, c \in I$ but $b \notin I$. Then $(-\infty, b)$ and (b, ∞) disconnect I in $\mathbb{R}$.
>
> Now suppose I is an interval. We work in the subspace topology on I. Let $S \subset I$ be open, closed an non-empty. Let $a \in S$, and suppose we have $b \in I \setminus S$. We can let $b > a$.
>
> Let $c = \sup([a, b] \cap S)$. Then this exists, so we can find a sequence (x_n) in S with $x_n \to c$. But S is closed in I, so $c \in S$. In particular $c \neq b$, so $c < b$.
>
> But also S is open in I, so there exists $\delta > 0$ such that $(c - \delta, c + \delta) \subset S$. Let $\delta < b - c$. Then $c + \delta/2 \in S \cap [a, b]$, which is a contradiction to our definition of c.
>
> So in fact $S = I$, and I is connected.

We have another equivalent version of connectedness as follows:

Theorem 3.5. *Let X be a topological space. Then X is connected if and only if every continuous function $f : X \to \mathbb{Z}$ is constant.*

> **Proof:** First, suppose X is connected and $f : X \to \mathbb{Z}$ is continuous. Then for any $n \in \mathbb{Z}$, $\{n\} \subset \mathbb{Z}$ is open and closed, so $f^{-1}(\{n\}) \subset X$ is open and closed, hence either $\emptyset$ or X. This proves f is constant.
>
> Now suppose U, V disconnect X. Define $f : X \to \mathbb{Z}$ by
>
> $$f(x) = \begin{cases} 0 & x \in U, \\ 1 & x \in V. \end{cases}$$
>
> Then for any $A \subset \mathbb{Z}$, $f^{-1}(A) = \emptyset, X, U$ or V, which is open. So f is continuous and non-constant.

Remark. This proposition, and the intermediate value theorem give an alternative proof of the previous proposition. Moreover, the same result holds if we replace $\mathbb{Z}$ with a non-trivial discrete space Y.

Proposition 3.11. *A continuous image of a connected space is connected.*

> **Proof:** Let X be a connected topological space, and Y be a topological space. Let $f : X \to Y$ be continuous.
>
> Suppose $U, V \subset Y$ are open with $f(x) \subset U \cup V$ and $U \cap V \cap f(X) = \emptyset$. As f is continuous, $f^{-1}(U), f^{-1}(V) \subset X$ are open. Also $X = f^{-1}(U) \cup f^{-1}(V)$ and $f^{-1}(U) \cap f^{-1}(V) = \emptyset$. As X is connected, one of these sets, say $f^{-1}(U)$, is equal to $\emptyset$. Then $U \cap f(X) = \emptyset$, so $f(X)$ is connected.

Proposition 3.12. *A product of connected spaces is connected.*

> **Proof:** Let (X, τ) and (Y, σ) be connected topological spaces, and let ρ be the product topology on $X \times Y$.
>
> Suppose $U, V \in \rho$ with $U \cup V = X \times Y$ and $U \cap V = \emptyset$. Fix $x \in X$. Then $\{x\} \times Y$ is homeomorphic to Y, in particular, it is connected. So $\{x\} \times Y \subset Y$ or $\{x\} \times Y \subset V$, otherwise they would disconnect $\{x\} \times Y$.
>
> Let $A = \{x \in X \mid \{x\} \times Y \subset U\}$ and $B = \{x \in X \mid \{x\} \times Y \subset V\}$. Clearly $A \cap B = \emptyset$ as $U \cap V = \emptyset$, and we have proven $X = A \cup B$.
>
> Now suppose $x \in A$, so $\{x\} \times Y \subset Y$. Then for any $y \in Y$, $(x, y) \in U$. Since U is open, we can find $T \in \tau$, $S \in \sigma$ such that $(x, y) \in T \times S \subset U$. In particular, for all $w \in T$, $(w, y) \in U$, and so $\{w\} \times Y \subset U$, i.e. $w \in A$. So we have $T \in \tau$ with $x \in T \subset A$, so A is a neighbourhood of $x \in X$. Hence A is open.
>
> Similarly, B is open. But X is connected, so either $A = \emptyset$, giving $U = \emptyset$, or $B = \emptyset$, giving $V = \emptyset$. Hence $X \times Y$ is connected.

Example 3.11.

Take $[-1, 0) \cup (0, 1]$. We have shown it is not connected. However it is a disjoint union of connected sets (by how we defined it). Moreover, if we have a proper superset of $[-1, 0)$ or $(0, 1]$ in $[-1, 0) \cup (0, 1]$, then it is disconnected.

Definition 3.13. Let X be a topological space. A *connected component* of X is a maximal connected subset A of X, that is to say, A is connected but if $A \subset B \subset X$ with B connected, then $A = B$.

Theorem 3.6. *The connected components of a topological space X form a partition of X.*

Proof: Define an equivalence relation $\sim$ on X by $x \sim y$ if and only if there exists $A \subset X$ connected with $x, y \in A$. Clearly $\sim$ is reflexive (as $\{x\}$ is connected) and symmetric, so we want to show that $\sim$ is transitive.

Suppose $x, y \in X$ with $x \sim y$ and $y \sim z$. Then there exist $A, B \subset X$ connected with $x, y \in A$ and $y, z \in B$. Note $x, z \in A \cup B$.

Suppose U, V disconnect $A \cup B$ in X, and say $y \in U$. Pick $w \in V \cap (A \cup B)$, and say $w \in A$. Then since $y \in A$, U, V disconnect A. So $A \cup B$ is connected, giving $x \sim z$. Hence $\sim$ is an equivalence relation.

Now we show the equivalence classes are connected themselves. Suppose S is an equivalence class of $\sim$, and that U and V disconnect S. Then, we can find $x \in U \cap S$, $y \in V \cap S$. Since $x \sim y$, there is a connected $A \subset X$ with $x, y \in A$. Then for all $z \in A$, x, z are connected so $x \sim z$ and $z \in S$.

So A is a subset of S, and U, V disconnect A, which is a contradiction. So S is connected. Now suppose S is not maximal, that is $S \subset T \subset X$ with T connected. Let $x \in S$. Then for all $y \in T$, $x, y \in T$ with T connected so $x \sim y$. Thus $T \subset S$, hence $S = T$, and S is maximal.

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

Remark. This tells us that connected components do actually exist.

Instead of defining connectedness in this way, we may choose to define it in a way that is more intuitive to understand.

Definition 3.14. A *path* from x to y in a topological space X is a continuous function $\phi : [0, 1] \to X$ with $\phi(0) = x$ and $\phi(1) = y$. X is *path-connected* if for all $x, y \in X$, there is a path from $x \to y$.

Proposition 3.13. *A path-connected space X is connected.*

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

However, the converse is not true.

Example 3.12. (Closed Topologist's Sine Curve)

Take the following subspace of $\mathbb{R}^2$: Let

$$A = \{(0, y) \mid -1 \leq y \leq 1\},$$
$$B = \{(x, \sin \frac{1}{x}) \mid 0 < x \leq 1\}.$$

Let $X = A \cup B \subset \mathbb{R}^2$. We prove X is connected.(insert picture)

Indeed, A and B are path-connected, so are connected. Suppose U, V disconnect X in $\mathbb{R}^2$. Then we may assume $A \subset U$, $B \subset V$. So $(0, 0) \in A \subset U$. Since U is open, we have some $\delta > 0$ such that $B_\delta((0, 0)) \subset U$. But then we can pick n such that $\frac{1}{2n\pi} < \delta$, which gives $(\frac{1}{2n\pi}, 0) \in U \cap B$.

However, X is not path-connected. Suppose $\phi = (\phi_1, \phi_2)$ is a path from $(0, 0)$ to $(1, \sin 1)$, in X. Let $\sigma = \sup\{t \in [0, 1] \mid \phi_1(t) = 0\}$, and let $y = \phi_2(\sigma)$. As ϕ is continuous, $\phi(\sigma) = (0, y)$. Choose $\delta > 0$ such that $|\sigma - t| < \delta \implies |\phi(\sigma) - \phi(t)| < 1$, and $\delta < 1 - \sigma$.

By the definition of σ, there exists ε such that $\phi_1(\sigma + \varepsilon) = x > 0$. Choose $w \in (0, x)$ such that $|\sin \frac{1}{w} - y| \geq 1$. Then, by intermediate value theorem, there is some $t \in (\sigma, \sigma + \varepsilon)$ such that $\sigma_1(t) = w$.

Then $|\sigma - t| < \delta$, but

$$|\phi(\sigma) - \phi(t)| \geq |\phi_2(\sigma) - \phi_2(t)| = \left|\sin \frac{1}{w} - y\right| \geq 1,$$

a contradiction.

However, we can prove the following nice proposition.

Proposition 3.14. *An open, connected subset of Euclidean space is path-connected.*

Proof: Let $X \subset \mathbb{R}^n$ be open and connected. If X is empty, we are done, so assume $X \neq \emptyset$.

Fix $a \in X$. Let

$$U = \{x \in X \mid \text{there is a path in } X \text{ from } a \text{ to } x\}.$$

Now $U \neq \emptyset$, as $a \in U$. Moreover, U is open in X: Suppose $b \in U$. Then X is open, so we can pick $\delta > 0$ such that $B_\delta(b) \subset X$. Let ϕ be a path from a

to b in X, and let $x \in B_\delta(b)$. Then, define a path θ from a to x, given by

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

So this gives $x \in X$.

Moreover, U is closed in X. Let $b \in X \setminus U$. Choose $\delta > 0$ such that $B_\delta(b) \subset X$. Suppose $x \in B_\delta(b) \cap U$. Then, if ϕ is a path from a to x, we may define a path θ from a to b by

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

This is a contradiction to $b \in X \setminus U$, so $B_\delta(b) \subset X \setminus U$.

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

Part II

Differentiation

4 Higher Dimensional Derivatives

4.1 Extending the Derivative

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

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

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

As we have done before, we wish to generalise this definition to functions $f : \mathbb{R}^n \to \mathbb{R}^m$.

For $n = 1$, we can easily modify the above definition and get a satisfactory result. However if $n \geq 2$, then dividing by $h \in \mathbb{R}^n$ makes no sense. Hence we introduce the following definition.

Definition 4.1. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a function, the *i-th partial derivative* of f at $a \in \mathbb{R}^n$ is

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

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

Example 4.1.

Take a function $f : \mathbb{R}^2 \to \mathbb{R}$ by

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

Then both partial derivatives exist at $(0,0)$: Indeed, $D_1 f(0,0) = D_2 f(0,0) = 0$. But f is not continuous at $(0,0)$.

Hence we may want to look for a better definition of a derivative. Returning to

functions from $\mathbb{R}$ to $\mathbb{R}$, note

$$f'(a) = A \iff \frac{f(a+h) - f(a)}{h} \to A \qquad\qquad \text{as } h \to 0,$$

$$\iff \frac{f(a+h) - f(a)}{h} = A + \varepsilon(h) \qquad \text{where } \varepsilon(h) \to 0 \text{ as } h \to 0,$$

$$\iff f(a+h) = f(a) + Ah + \varepsilon(h)h \qquad \text{where } \varepsilon(h) \to 0 \text{ as } h \to 0.$$

We can now generalise this notion to higher dimensions.

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

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

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

Remark. In this definition, $h \in \mathbb{R}^n$.

We say that α is the *derivative* of f at a, and write $Df|_a = \alpha$. So if f is differentiable at a,

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

where $\varepsilon(h) \to 0$ as $h \to 0$. But if this definition is to make any sense, then α must be unique.

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

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

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

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

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

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

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

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

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

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

Remark.

1. To consider differentiability of f at a, it only matters what happens on some neighbourhood of a. Hence, our definition works if instead of $f : \mathbb{R}^n \to \mathbb{R}^m$, we have $f : \mathcal{N} \to \mathbb{R}^m$, where $\mathcal{N} \subset \mathbb{R}^n$ is a neighbourhood of a, or in particular, $f : B_\delta(a) \to \mathbb{R}^m$, where $\delta > 0$.

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

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

$$
\|x\|_\infty = d_\infty(0, x) = \max_i |x_i|.
$$

Note that $\|x\|_1 \geq 0$ with equality of and only if $x = 0$, and

$$
\|\lambda x\|_1 = |\lambda|\|x\|_1, \qquad \|x + y\|_1 \leq \|x\|_1 + \|y\|_1.
$$

This is the same for $\|\cdot\|_\infty$.

We have seen that for all $x \in \mathbb{R}^n$,

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

$$
\|x\|_\infty \leq \|x\|_1 \leq n\|x\|_\infty.
$$

So we can replace our norm $\|\cdot\|$ in the definition of the derivative with either $\|\cdot\|_1$ or $\|\cdots\|_\infty$, and the definition will be the same. This can be sometimes useful for computation.

Consider the vector space $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ of linear maps $\mathbb{R}^n \to \mathbb{R}^m$. Then $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{mn}$ with a standard isomorphism.

So we could define the Euclidean norm of a linear map, although it is a bit unnatural. Instead, we will use the following definition.

Definition 4.3. The *operator norm* of $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is defined by

$$\|\alpha\| = \sup\{\|\alpha x\| \mid \|x\| = 1\}.$$

Proposition 4.2. *Let $\|\cdot\|$ be the operator norm of $V = \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Let $\alpha, \beta \in V$, and $\gamma \in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^n)$. Then,*

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

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

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

  (iv) *For all $x \in \mathbb{R}^n$, $\|\alpha x\| \leq \|\alpha\|\|x\|$;*

   (v) $\|\alpha\gamma\| \leq \|\alpha\|\|\gamma\|$, *with appropriate operator norms;*

  (vi) *If $\|\cdot\|'$ is the Euclidean norm of $V \cong \mathbb{R}^{mn}$ with the standard isomorphism, then there are constants c, d such that*

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

Remark.

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

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

Proof: (i) to (iii) are easy to prove.

To prove (iv), let $x \in \mathbb{R}^n$. If $x = 0$, then we are done. Otherwise

$$\alpha(x) = \|x\|\alpha\left(\frac{x}{\|x\|}\right) \leq \|x\|\|\alpha\|,$$

as $\|\frac{x}{\|x\|}\| = 1$.

Now for (v), let $x \in \mathbb{R}^k$ with $\|x\| = 1$. Then, by applying (iv) multiple times,

$$\|\alpha\gamma x\| \leq \|\alpha\|\|\gamma x\| \leq \|\alpha\|\|\gamma\|\|x\| = \|\alpha\|\|\gamma\|.$$

So $\|\alpha\gamma\| \leq \|\alpha\|\|\gamma\|$.

For (vi), let $x \in \mathbb{R}^n$ with $\|x\| = 1$. Then,

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

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

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

Hence $\|\alpha\| \leq n\sqrt{m}\|\alpha\|'$. On the other hand, pick i, j that maximize $|A_{ij}|$. Then $\|\alpha e_j\| \geq \|(\alpha e_j)_i\| = |A_{ij}|$. But then,

$$\|\alpha\|' \leq \sqrt{mn}|A_{ij}| \leq \sqrt{mn}\|\alpha e_j\| \leq \sqrt{mn}\|\alpha\|.$$

This prove (vi) with $d = \sqrt{mn}$, $c = \frac{1}{n\sqrt{m}}$.

Proposition 4.3. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$, which is differentiable at $a \in \mathbb{R}^n$. Then, f is continuous at a.*

Proof: Write

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

where $\varepsilon(h) \to 0$ as $h \to 0$. Also, $\|h\| \to 0$ as $h \to 0$, and $Df|_a$ is linear, so continuous, so $Df|_a(h) \to Df|_a(0) = 0$ as $h \to 0$.

Hence, $f(a + h) \to f(a)$ as $h \to 0$.

Proposition 4.4. *Let $f, g : \mathbb{R}^n \to \mathbb{R}^m$, and $\lambda : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $a \in \mathbb{R}^n$. Then $f + g$ and λf are differentiable at a with*

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

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

Proof: We have

$$f(a+h) = f(a) + Df|_a(h) + \varepsilon(h)\|h\|,$$
$$g(a+h) = g(a) + Dg|_a(h) + \eta(h)\|h\|,$$
$$\lambda(a+h) = \lambda(a) + D\lambda|_a(h) + \zeta(h)\|h\|,$$

where $\varepsilon(h), \eta(h), \zeta(h) \to 0$ as $h \to 0$. Now,

$$(f+g)(a+h) = (f+g)(a) + (Df|_a + Dg|_a)(h) + (\varepsilon(h) + \eta(h))\|h\|,$$

where $Df|_a + Dg|_a$ is linear and $\varepsilon(h) + \eta(h) \to 0$ as $h \to 0$. Similarly, we get

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

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

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

as $h \to 0$, since $\varepsilon(h), \zeta(h) \to 0$ as $h \to 0$, $\|h\| \to 0$ as $h \to 0$, and $Df|_a, D\lambda|_a$ are linear so continuous, so they tend to 0 as $h \to 0$. So,

$$\left\|Df|_a(h)Dg|_a(h)\frac{1}{\|h\|}\right\| \leq \|Df|_a(h)\|\|D\lambda|_a(h)\|\frac{1}{\|h\|}$$
$$\leq \|Df|_a\|\|h\|\|D\lambda|_a\|\|h\|\frac{1}{\|h\|}$$
$$= \|Df|_a\|\|D\lambda|_a\|\|h\| \to 0,$$

as $h \to 0$.

Although we have gotten rid of partial derivatives, they can still be useful for computation.

Proposition 4.5. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $a \in \mathbb{R}^n$. Write $f = \begin{pmatrix} f_1 & \cdots & f_m \end{pmatrix}^T$ where for each i, $f_i : \mathbb{R}^n \to \mathbb{R}$. Then,*

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

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

and

(b) *If f is differentiable at a, and A is the matrix of $Df|_a$ in terms of the standard bases, then $A_{ij} = D_j f_i(a)$.*

Proof: For (a), if f is differentiable, write

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

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

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

where $(Df|_a)_i : \mathbb{R}^n \to \mathbb{R}$ is linear, and $|\varepsilon_i(h)| \leq \|\varepsilon(h)\| \to 0$ as $h \to 0$. So the f_i are differentiable.

Now suppose f_i is differentiable for each i. Then we can write

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

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

where $\alpha = \begin{pmatrix} Df_1|_a & \cdots & Df_m|_a \end{pmatrix}^T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, and

$$\|\varepsilon(h)\| = \left\| \begin{pmatrix} \varepsilon_1(h) \\ \vdots \\ \varepsilon_m(h) \end{pmatrix} \right\| = \sqrt{\sum_{i=1}^{m} \varepsilon_i(h)^2} \to 0$$

as $h \to 0$.

Now for (b), again write

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

where $\varepsilon(h) \to 0$ as $h \to 0$. Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{R}^n$. Then,

$$\frac{f(a + he_j) - f(a)}{h} = \frac{Df|_a(he_j) + \varepsilon(he_j)\|he_j\|}{h}$$
$$= Df|_a(e_j) \pm \varepsilon(he_j) \to Df|_a(e_j),$$

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

Definition 4.4. The matrix A in (b) is called the *Jacobian* matrix of f at a.

Theorem 4.1 (Chain Rule). *Let $f : \mathbb{R}^p \to \mathbb{R}^n$ be differentiable at $a \in \mathbb{R}^p$, and let $g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $f(a) \in \mathbb{R}^n$. Then, $g \circ f$ is differentiable at a with*

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

Remark. In principle, this makes sense. If f is approximately linear near a, and g is approximately linear near $f(a)$, then $g \circ f$ should be approximately linear near a, and the linear approximation to $g \circ f$ near a should be the obvious thing.

However the proof is a bit messy, we need to do a calculation to make sure that the error terms behave.

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

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

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

Now, $\varepsilon(h) \to 0$ as $h \to 0$ and β is linear, so continuous, hence $\beta(\varepsilon(h)) \to \beta(0) = 0$ as $h \to 0$.

Next, a is linear, so continuous, so $\alpha(h) \to \alpha(0) = 0$ as $h \to 0$, and $\varepsilon(h)\|h\| \to 0$ as $h \to 0$. So $\alpha(h) + \varepsilon(h)\|h\| \to 0$ as $h \to 0$.

Now let $\eta(0) = 0$, so η is continuous at 0. Then $\eta(\alpha(h) + \varepsilon(h)\|h\|) \to 0$ as $h \to 0$. Finally,

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

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

Example 4.2.

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

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

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

4. Using this together with the chain rule, we get many differentiable functions, such as $f : \mathbb{R}^2 \to \mathbb{R}^2$, defined by

$$f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} e^{x+y} \\ \cos(xy) \end{pmatrix},$$

which is differentiable. This is because the projection maps are differentiable, so by the chain rule

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

are differentiable. Hence f is differentiable.

The derivative of f at z is a linear map $\mathbb{R}^2 \to \mathbb{R}^2$. Then the matrix of the derivative is given by the partial derivatives:

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

5. Let $\mathcal{M}_n$ be the vector space of $n \times n$ real matrices. Then $\mathcal{M}_n \cong \mathbb{R}^{n^2}$, so we can consider the differentiability of $f : \mathcal{M}_n \to \mathcal{M}_n$.

   Recall that the definition is still the same if we replace the Euclidean norm by the operator norm, so write $\| \cdot \|$ for the operator norm on $\mathcal{M}_n$. Define $f : \mathcal{M}_n \to \mathcal{M}_n$ by $f(A) = A^2$. Then,

$$f(A + H)^2 = (A + H)^2 = A^2 + (AH + HA) + H^2,$$

where $\|\frac{H^2}{\|H\|}\| \leq \frac{\|H\|^2}{\|H\|} = \|H\| \to 0$ as $H \to 0$. So f is everywhere differentiable with derivative

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

6. Consider $\det : \mathcal{M}_n \to \mathbb{R}$. Then we have

$$\det(I + H) = \begin{vmatrix} 1 + H_{11} & & H_{ij} \\ & \ddots & \\ H_{ij} & & 1 + H_{nn} \end{vmatrix}$$
$$= 1 + \operatorname{tr}(H) + \varepsilon(H)\|H\|,$$

where $\varepsilon(H)\|H\|$ is sum of terms with two or more H_{ij}'s multiplied together. Then since

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

as $H \to 0$ (where $\|\cdot\|_2$ is the Euclidean norm), we have $\varepsilon(H) \to 0$. So det is differentiable at I, with derivative $D\det|_I(H) = \operatorname{tr}(H)$.

Now suppose $A \in \mathcal{M}_n$ is invertible. Then,

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

Then, since $\varepsilon(K) \to 0$ as $K \to 0$ and

$$\left| \frac{(\det A)\varepsilon(A^{-1}H)\|A^{-1}H\|}{\|H\|} \right| \leq |(\det A)\varepsilon(A^{-1}H)\|A^{-1}\|| \to 0$$

as $H \to 0$, we have det is differentiable at A with

$$D\det|_A(H) = (\operatorname{tr} A^{-1}H)(\det A).$$

4.2   The Mean Value Inequality

Recall that in one dimension, if $f : \mathbb{R} \to \mathbb{R}$ is differentiable with zero derivative everywhere, then f is constant. To prove this, we used the mean value theorem, so if we want to replicate this result in higher dimensions, we need an analogue of the mean value theorem.

Theorem 4.2 (Mean Value Inequality)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$. Suppose f is differentiable on an open set $X \subset \mathbb{R}^n$ with $a, b \in X$. Suppose further that*

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

Then

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

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

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

$$\phi(t) = f(a + t(b - a)) \cdot (f(b) - f(a)).$$

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

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

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

Therefore,

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

So by the mean value theorem,

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

Hence,

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

Corollary 4.1. *Let $X \subset \mathbb{R}^n$ be open and connected, and let $f : X \to \mathbb{R}^m$ be differentiable with $Df|_x$ the zero map for all $x \in X$. Then f is constant on X.*

Proof: By the mean value inequality, f is locally constant, so for each $x \in X$, there is some $\delta > 0$ such that $B_\delta(x) \subset X$, and so f is constant of $B_\delta(x)$.

Note that as X is open, if $U \subset X$, then U is open in X if and only if U is open in $\mathbb{R}^n$. If $X = \emptyset$, we are done. So suppose otherwise. Fix $a \in X$, and let
$$U = \{x \in X \mid f(x) = f(a)\}.$$

Then U is non-empty, as $a \in U$, and U is open, as if $b \in U$, then there is some $\delta > 0$ such that $B_\delta(b) \subset X$. Since f is constant on $B_\delta(b)$, $D_\delta(b) \subset U$. Hence U is open.

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

However X is connected, so $U = X$.

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

However there are some cases when the converse does hold true.

Theorem 4.3. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and let $a \in \mathbb{R}^n$. Suppose there is some neighbourhood of a, such that the partial derivatives $D_i f$ all exist and are continuous on this neighbourhood. Then f is differentiable at a.*

For simplicity, we prove this for that case $n = 2$, $m = 1$. So $f : \mathbb{R}^2 \to \mathbb{R}$.

Write $a = (x, y)$, then we want to think about $f(x + h, y + k)$ for small h, k. Now by the definition of the partial derivatives,
$$f(x + h, y + k) = f(x + h, y) + k D_2 f(x + h, y) + o(k),$$
and
$$f(x + h, y) = f(x, y) + h D_1 f(x, y) + o(h).$$
Hence, putting these together,
$$\begin{aligned}
f(x + h, y + k) &= f(x, y) + h D_1 f(x, y) + k D_2 f(x + h, y) + o(h) + o(k) \\
&= f(x, y) + h D_1 f(x, y) + k(D_2 f(x, y) + o(1)) + o(h) + o(k) \\
&= f(x, y) + h D_1 f(x, y) + k D_2 f(x, y) + o(h) + o(k).
\end{aligned}$$

So this works. Or does it?

Our claim that $f(x + h, y + k) = f + kD_2f + o(h)$, is not true, as the $o(h)$ term is a function of both h and k. If we call it $\eta(h, k)$, then we need

$$\frac{\eta(h, k)}{k} \to 0 \text{ as } (h, k) \to 0.$$

But we only know that for all h,

$$\frac{\eta(h, k)}{k} \to 0 \text{ as } k \to 0,$$

which is weaker. In fact, for our proof we need the mean value theorem.

Proof: We again take the case $n = 2$, $m = 1$, so $a = (x, y)$. Take (h, k) small. Then by mean value theorem,

$$f(x + h, y + k) - f(x + h, y) = kD_2f(x + h, y + \theta_{h,k}k),$$

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

$$f(x + h, y) - f(x, y) = hD_1f(x + \phi_h h, y),$$

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

$$f(x + h, y + k) - f(x, y) = kD_2f(x + h, y + \theta_{h,k}k) + hD_1f(x + \phi_h h, y).$$

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

$$D_2f(x + h, y + \theta_{h,k}k) \to D_2f(x, y),$$

$$D_1f(x + \phi_h h, y) \to D_1f(x, y).$$

Then write
$$D_2f(x + h, y + \theta_{h,k}k) = D_2f(x, y) + \eta(h, k),$$
$$D_1f(x + \phi_h h, y) = D_1f(x, y) + \zeta(h, k),$$
where $\eta(h, k), \zeta(h, k) \to 0$ as $(h, k) \to (0, 0)$. Then,

$$f(x + h, y + k) = f(x, y) + hD_1f(x, y) + kD_2f(x, y) + h\zeta(h, k) + k\eta(h, k).$$

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

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

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

Remark.

1. We only need $D_i f$ to be continuous at a for this proof to work.

2. The same proof basically works for $f : \mathbb{R}^n \to \mathbb{R}$ for general n. Then we can get $f : \mathbb{R}^n \to \mathbb{R}^m$ by looking at each $f_i : \mathbb{R}^n \to \mathbb{R}$.

3. In general, all proofs of this use mean value theorem at some point.

4.3   The Second Derivative

We will start with a result on partial derivatives that is used in many courses, but not proven until now.

Theorem 4.4. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$, $a \in \mathbb{R}^n$ and $\varepsilon > 0$. Suppose $D_i D_j f$ and $D_j D_i f$ exist on $B_\varepsilon(a)$ and are continuous at a. Then $D_i D_j f(a) = D_j D_i f(a)$.*

Proof: Without loss of generality, say $m = 1$ and $n = 2$. Then let $a = (x, y)$. Define

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

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

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

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

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

Letting $h \to 0$ and using the continuity of $D_1 D_2 f$ and $D_2 D_1 f$, we get

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

Before we look formally at the second derivative, we need to think about what the second derivative really means.

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be everywhere differentiable. Then for each $x \in \mathbb{R}^n$, we have $Df|_x \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Hence we can define $F : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{nm}$ by $F(x) = Df|_x$.

If F is differentiable at $a \in \mathbb{R}^n$, then we say f is *twice-differentiable* at a and the

second derivative of f at a is

$$D^2 f|_a = DF|_a.$$

If we look at $D^2 f|_a$, then it is

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

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

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

or analogously,

$$Df|_{a+h}(k) = Df|_a(k) + D^2 f|_a(h, k) + o_k(h).$$

Example 4.3.

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

$$f(A + K) = (A + K)^3 = A^3 + (A^2 K + AKA + KA^2) + o(K),$$

so f is everywhere differentiable with

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

Now, we can compute the second derivative:

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

So f is twice differentiable at A and

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

Remark. As with differentiability, for this definition to work, it is enough to have f twice-differentiable on some neighbourhood of a.

We have seen that $Df|_a$ is related to $D_j f(a)$. Now we would like to see how $D^2 f|_a$ is related to $D_i D_j f(a)$.

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

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

So $D_i D_j f(a) = D^2 f|_a(e_i, e_j)$. Hence if H is the $n \times n$ matrix representing the bilinear form $D^2 f|_a$, we have

$$
H_{ij} = D_i D_j f(a).
$$

We call H the *Hessian* matrix of f.

Now if we want to generalize this to $f : \mathbb{R}^n \to \mathbb{R}^m$, we could do this for each $f_i : \mathbb{R}^n \to \mathbb{R}$, or we could think about the matrices whose entries are elements of $\mathbb{R}^m$.

Definition 4.5. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $a \in \mathbb{R}^n$. We say f is *continuously differentiable* at a if $Df|_x$ exists for all x is some ball $B_\delta(a)$, and the function $x \mapsto Df|_x$ is continuous at a.

If f is twice differentiable at a, then due to the commutation of mixed derivatives, we get that H is a symmetric matrix. Hence, under this condition, $D^2 f|_a$ is a symmetric bilinear form.

Definition 4.6. Let $f : \mathbb{R}^n \to \mathbb{R}$, $a \in \mathbb{R}^n$. We say a is a *local maximum* (resp. *minimum*) for f if there is some $\delta > 0$ such that for all $x \in B_\delta(a)$, we have $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$).

Proposition 4.6. *Let $f : \mathbb{R}^n \to \mathbb{R}$, and let a be a local maximum/minimum for f. Then $Df|_a$ is the zero map.*

> **Proof:**   Let $u \in \mathbb{R}^n$. For $\lambda \neq 0 \in \mathbb{R}$,
>
> $$
> \frac{f(a + \lambda u) - f(a)}{\lambda} = \frac{Df|_a(\lambda u) + o(\lambda)}{\lambda} \to Df|_a(u),
> $$
>
> as $\lambda \to 0$. Assume that a is a local maximum (otherwise consider $-f$). Then,
>
> $$
> \frac{f(a + \lambda u) - f(a)}{\lambda} \text{ is } \begin{cases} \geq 0 & \text{if } \lambda < 0, \\ \leq 0 & \text{if } \lambda > 0. \end{cases}
> $$
>
> Hence $Df|_a(u) = 0$.

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

Lemma 4.1 (Second-order Taylor Theorem)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be twice-differentiable at* $a \in \mathbb{R}^n$. *Then,*

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

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

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

Then g is continuous on $[0, 1]$, $g(0) = 0$ and g is differentiable on $(0, 1)$ with derivative
$$g'(t) = Df|_{a+th}(h) - Df|_a(h) - tD^2f_a(h, h).$$

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

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

as $h \to 0$.

Theorem 4.5. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $a \in \mathbb{R}^n$. *Suppose* f *is twice continuously differentiable at* a *(in particular,* $D^2f|_a$ *is a symmetric bilinear form) and* $Df|_a = 0$. *Then,*

- *If* $D^2f|_a$ *is positive definite, then* a *is a local minimum.*

- *If* $D^2f|_a$ *is negative definite, then* a *is a local maximum.*

Proof: Suppose that $D^2f|_a$ is positive definite (otherwise consider $-f$). Then with respect to some basis, $D^2f|_a$ is a diagonal matrix (as it is symmetric, hence orthogonally diagonalizable) with strictly positive elements on the leading diagonal (as it is positive definite).

Hence, for all $x \in \mathbb{R}^n$, $D^2f|_a(x, x) \geq \mu\|x\|^2$, where $\mu > 0$ is the least

eigenvalue of $D^2 f|_a$. Then, by our previous lemma,

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

$$\geq \frac{1}{2}\mu + o(1) \to \frac{1}{2}\mu$$

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

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

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

4.4   Ordinary Differential Equations

Lemma 4.2. *Let $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$ with A compact and B closed. Let $X = \mathcal{C}(A, B) = \{f : A \to B \mid f \text{ continuous}\}$, with the uniform metric*

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

Then X is a complete metric space.

Proof: As A is compact, d is well-defined. Let (f_n) be a Cauchy sequence in X. Then (f_n) is uniformly Cauchy, so uniformly convergent. Hence $f_n \to f$ uniformly for some $f : A \to \mathbb{R}^m$.

Then the uniform limit of continuous functions is continuous, so f is continuous, and for all $x \in A$, $f_n(x) \to f(x)$, so, as B is closed, $f \in X$ and $d(f_n, f) \to 0$.

Often, we want to solve an ODE, however there is no closed-form solution. To solve, we can do one of two things:

- Use numerical methods to approximate the solution.

- Think about what the solution looks like, using phase-plane portraits.

However, this only works if the ODE has a solution. So we want a result telling us that under appropriate conditions, ODEs have unique solutions.

A typical example of an ODE is

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \phi(x, y),$$

subject to $y = y_0$ when $x = x_0$. In the case $\mathbb{R} \to \mathbb{R}^n$, we want to solve the initial value problem

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

In the sequel, we take the *closed ball* of radius δ about a as

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

Theorem 4.6 (Lindelöf-Picard). *Let $a, b \in \mathbb{R}$, with $a < b$, $y_0 \in \mathbb{R}^n$, $\delta > 0$ and $t_0 \in (a, b)$.*

Let $\phi : [a, b] \times \overline{B_\delta(y_0)} \to \mathbb{R}^n$ be continuous, and suppose there is some $K > 0$ such that for all $t \in [a, b]$ and for all $y, z \in \overline{B_\delta(y_0)}$,

$$\|\phi(t, y) - \phi(t, z)\| \leq K\|y - z\|.$$

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

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

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

Proof: As ϕ is a continuous function on a compact set, we can find M such that for all $t \in [a, b]$, and for all $y \in \overline{B_\delta(y_0)}$, $\|\phi(t, y)\| \leq M$.

Take $\varepsilon > 0$ such that $[t_0 - \varepsilon, t_0 + \varepsilon] \subset [a, b]$, and let $X = \mathcal{C}([t_0 - \varepsilon, t_0 + \varepsilon], \overline{B_\delta(y_0)})$. Then X is complete with the uniform metric d, and $X \neq \emptyset$.

For $g \in X$, define $Tg : [t_0 - \varepsilon, t_0 + \varepsilon] \to \mathbb{R}^n$ by

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

By the fundamental theorem of calculus, $Tf = f$ if and only if f is a solution to the initial value problem. To do this, we will use the contraction mapping theorem.

If $g \in X$ and $t \in [t_0 - \varepsilon, t_0 + \varepsilon]$, then

$$\|Tg(t) - y_0\| = \left\| \int_{t_0}^t \phi(x, g(x)) \, \mathrm{d}x \right\|$$

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

$$\leq M\varepsilon.$$

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

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

Thus, $d(Tg, Th) \leq K\varepsilon d(g, h)$. So, by taking $\varepsilon = \min\{\frac{\delta}{M}, \frac{1}{2K}\}$, we have that T is a contraction of X, and so has a unique fixed point by the contraction mapping theorem, as desired.

Remark. As stated, this doesn't have not much use. This does not provide a global solution, and indeed, there may or may not be a global solution. In practice, given appropriate conditions on ϕ, we can often patch together local solutions, however this is beyond the scope of this course.

4.5   The Inverse Function Theorem

Theorem 4.7 (Inverse Function Theorem). *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable at $a \in \mathbb{R}^n$, with $\alpha = Df|_a$ being non-singular. Then there exist open neighbourhoods U of a and V of $f(a)$ such that $f|_U$ is a homeomorphism of U onto V.*

Moreover, if $g : V \to U$ is the inverse of $f|_U$, then it is differentiable at $f(a)$ with

$$Dg|_{f(a)} = \alpha^{-1}.$$

Proof: Write
$$f(a + h) = f(a) + \alpha(h) + \varepsilon(h)\|h\|,$$
where $\varepsilon(h) \to 0$ as $h \to 0$. Let $\delta, \eta > 0$ such that f is differentiable on $\overline{B_\delta(a)}$. Let $W = \overline{B_\delta(a)}$, $V = B_\eta(f(a))$.

Define $\phi : \mathbb{R}^n \to \mathbb{R}^n$ by $\phi(x) = f(x) - \alpha(x)$. Then for $x \in W$, ϕ is differentiable at x with

$$D\phi|_x = Df|_x - \alpha \to 0 \text{ as } x \to a.$$

Note W is a complete, non-empty metric space. Fix $y \in V$, and define $T_y : W \to \mathbb{R}^n$ by
$$T_y(x) = x - \alpha^{-1}(f(x) - y).$$
Then, $f(x) = y \iff T_y(x) = x$. We will now show T_y is a contraction. Given $x \in W$,

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

For given $w, x \in W$, we have

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

by the mean value inequality. Pick $\delta > 0$ sufficiently small such that for all $x \in \overline{B_\delta(a)}$, we have

$$\|\varepsilon(x-a)\| \leq \frac{1}{2\|\alpha^{-1}\|}, \qquad \sup_{z \in W} \|D\phi|_z\| < \frac{1}{\|\alpha^{-1}\|}.$$

We can do this as $\varepsilon(x-a) \to 0, D\phi|_x \to 0$ as $x \to a$. Moreover, we can take $\eta = \frac{\delta}{2}$.

Then, for each $y \in V$, we have, for all $x \in W$, $\|T_y x - a\| \leq \delta$, and for all $x, w \in W$, $\|T_y x - T_y w\| \leq K\|w - x\|$, where $K < 1$ is a constant.

So T_y is a contraction of W, so by the contraction mapping theorem, it has a unique fixed point $x_y \in T_y(W) \subset B_\delta(a)$. That is, for each $y \in V$, there is a unique $x \in W$ with $f(x) = y$, and $x \in B_\delta(a)$.

Let U be the set of all such x, and let $h = f|_{B_\delta(a)}$. Then $U = h^{-1}(V)$, so U is open in $B_\delta(a)$. But $B_\delta(a)$ is open in $\mathbb{R}^n$, so U is open in $\mathbb{R}^n$.

So now we have open neighbourhoods U of a and V of $f(a)$ such that f maps U bijectively onto V. Now it remains to show that the inverse function is continuous.

Let $X = \mathcal{C}(V, W)$. As W is bounded, we have X is a complete, non-empty metric space with the uniform metric. Define $S : X \to X$ by

$$(Sg)(y) = g(y) - \alpha^{-1}(f(g(y)) - y)$$
$$= T_y(g(y)).$$

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

$$\|(Sg)(y) - (Sh)(y)\| = \|T_y(g(y)) - T_y(h(y))\|$$
$$\leq K\|g(y) - h(y)\| \leq Kd(g, h),$$

so $d(Sg, Sh) \leq Kd(g, h)$. So S is a contraction of X, and has a unique fixed point g. By the definition of S, for each $y \in V$, we have $g(y)$ is the unique $x \in W$ with $f(x) = y$.

Hence $g = (f|_U)^{-1}$. But $g \in X$, so $(f|_U)^{-1}$ is continuous, and then $f|_U$ is a homeomorphism from U onto V.

Index