# Analysis II+ Met & Top Review Sheet

Michael Li

April 28, 2016

**Uniform convergence**
The general principle of uniform convergence. A uniform limit of continuous functions is continuous. Uniform convergence and termwise integration and differentiation of series of real-valued functions. Local uniform convergence of power series. [3]

**Uniform continuity and integration**
Continuous functions on closed bounded intervals are uniformly continuous. Review of basic facts on Riemann integration (from Analysis I). Informal discussion of integration of complex-valued and $\mathbb{R}^n$-valued functions of one variable; proof that $\| \int_a^b f(x) \, \mathrm{d}x\| \leq \int_a^b \|f(x)\| \, \mathrm{d}x$. [2]

**$\mathbb{R}^n$ as a normed space**
Definition of a normed space. Examples, including the Euclidean norm on $\mathbb{R}v^n$ and the uniform norm on $C[a, b]$. Lipschitz mappings and Lipschitz equivalence of norms. The Bolzano-Weierstrass theorem in $\mathbb{R}^n$. Completeness. Open and closed sets. Continuity for functions between normed spaces. A continuous function on a closed bounded set in $\mathbb{R}^n$ is uniformly continuous and has closed bounded image. All norms on a finite-dimensional space are Lipschitz equivalent. [5]

**Differentiation from $\mathbb{R}^m$ to $\mathbb{R}^n$**
Definition of derivative as a linear map; elementary properties, the chain rule. Partial derivatives; continuous partial derivatives imply differentiability. Higher-order derivatives; symmetry of mixed partial derivatives (assumed continuous). Taylor's theorem. The mean value inequality. Path-connectedness for subsets of $\mathbb{R}^n$; a function having zero derivative on a path-connected open subset is constant. [6]

**Metric spaces**
Definition and examples. *Metrics used in Geometry*. Limits, continuity, balls, neighbourhoods, open and closed sets. [4]

**The Contraction Mapping Theorem**
The contraction mapping theorem. Applications including the inverse function theorem (proof of continuity of inverse function, statement of differentiability). Picard's solution of differential equations. [4]

# Contents

# 1 Uniform Convergence

From Analysis I we know how real numbers converge. But how does functions converge? Well we can try setting:

**Definition** (Pointwise convergence)**.** The sequence $f_n$ converges *pointwise* to $f$ if for all $x$

$$f(x) = \lim_{n\to\infty} f(x)$$

The only problem is that $f$ does not necessarily inherit any of the properties of $f_n$ (like $f(x, n) = x^{1/(2n+1)}$ converges to the Heaviside Step function as $n \to \infty$. Not differentiable. Not continuous. Not Nice). So we need a new definition:

**Definition** (Uniform convergence)**.** A sequence of functions $f_n : E \to \mathbb{R}$ converges *uniformly* to $f$ if

$$(\forall\varepsilon)(\exists N)(\forall x)(\forall n > N)\ |f_n(x) - f(x)| < \varepsilon.$$

Notice the only difference is that we have changed the order of $\exists N$ and $\forall x$ from pointwise convergence. This means $N$ is not a function of $x$.

This basically says that after a certain $N$, $f_n(x)$ is at most $\epsilon$ away from $f(x)$ at *every* given point. This makes all the difference. Now we can roll out the definitions and theorems similar to ones in Analysis I:

**Definition** (Uniformly Cauchy sequence)**.** A sequence $f_n : E \to \mathbb{R}$ of functions is *uniformly Cauchy* if

$$(\forall\varepsilon > 0)(\exists N)(\forall m, n > N)\ \sup_{x\in E} |f_n(x) - f_m(x)| < \varepsilon.$$

You sensed what was coming...

**Theorem.** Let $f_n : E \to \mathbb{R}$ be a sequence of functions. Then $(f_n)$ converges uniformly if and only if $(f_n)$ is uniformly Cauchy.

*Proof.* First suppose that $f_n \to f$ uniformly. Given $\varepsilon$, we know that there is some $N$ such that

$$(\forall n > N)\ \sup_{x\in E} |f_n(x) - f(x)| < \frac{\varepsilon}{2}.$$

Then if $n, m > N$, $x \in E$ we have

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

So done.

Now suppose $(f_n)$ is uniformly Cauchy. Then $(f_n(x))$ is Cauchy for all $x$. So it converges. Let $f(x) = \lim_{n\to\infty} f_n(x)$. Given $\varepsilon > 0$, choose $N$ such that whenever $n, m > N$, $x \in E$, we have $|f_n(x) - f_m(x)| < \frac{\varepsilon}{2}$. Letting $m \to \infty$, $f_m(x) \to f(x)$. So we have $|f_n(x) - f(x)| \le \frac{\varepsilon}{2} < \varepsilon$. So done. □

## 1.1 Properties Inherited Under Uniform Convergence

**Note.** These properties can come in very handy for solving questions involving uniform convergence. Here are some notes:

- If you are trying to prove convergence for a general function, prove Cauchy instead.

**Theorem** (Uniform convergence and continuity)**.** Let $E \subseteq \mathbb{R}$, $x \in E$ and $f_n, f : E \to \mathbb{R}$. Suppose $f_n \to f$ uniformly, and $f_n$ are continuous at $x$ for all $n$. Then $f$ is also continuous at $x$.

*Proof.* Let $\varepsilon > 0$. Choose $N$ such that for all $n \geq N$, we have

$$\sup_{y \in E} |f_n(y) - f(y)| < \varepsilon.$$

Since $f_N$ is continuous at $x$, there is some $\delta$ such that

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

Then for each $y$ such that $|x - y| < \delta$, we have

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

□

**Theorem** (Uniform convergence and integrals)**.** Let $f_n, f : [a, b] \to \mathbb{R}$ be Riemann integrable, with $f_n \to f$ uniformly. Then

$$\int_a^b f_n(t) \, \mathrm{d}t \to \int_a^b f(t) \, \mathrm{d}t.$$

*Proof.* We have

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

□

**Theorem.** Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of functions differentiable on $[a, b]$ (at the end points $a$, $b$, this means that the one-sided derivatives exist). Suppose the following holds:

(i) For some $c \in [a, b]$, $f_n(c)$ converges.

(ii) The sequence of derivatives $(f'_n)$ converges uniformly on $[a, b]$.

Then $(f_n)$ converges uniformly on $[a, b]$, and if $f = \lim f_n$, then $f$ is differentiable with derivative $f'(x) = \lim f'_n(x)$.

*Proof.* Fix $x \in [a, b]$. We apply the mean value theorem to $f_n - f_m$ to get

$$(f_n - f_m)(x) - (f_n - f_m)(c) = (x - c)(f'_n - f'_m)(t)$$

for some $t \in (x, c)$.

Taking the supremum and rearranging terms, we obtain

$$\sup_{x\in[a,b]} |f_n(x) - f_m(x)| \leq |f_n(c) - f_m(c)| + (b - a) \sup_{t\in[a,b]} |f'_n(t) - f'_m(t)|.$$

So given any $\varepsilon$, since $f'_n$ and $f_n(c)$ converge and are hence Cauchy, there is some $N$ such that for any $n, m \geq N$,

$$\sup_{t\in[a,b]} |f'_n(t) - f'_m(t)| < \varepsilon, \quad |f_n(c) - f_m(c)| < \varepsilon.$$

Hence we obtain

$$n, m \geq N \Rightarrow \sup_{x\in[a,b]} |f_n(x) - f_m(x)| < (1 + b - a)\varepsilon.$$

So by the Cauchy criterion, we know that $f_n$ converges uniformly. Let $f = \lim f_n$.

Now we have to check differentiability. Let $f'_n \to h$. For any fixed $y \in [a, b]$, define

$$g_n(x) = \begin{cases} \frac{f_n(x)-f_n(y)}{x-y} & x \neq y \\ f'_n(y) & x = y \end{cases}$$

Then by definition, $f_n$ is differentiable at $y$ iff $g_n$ is continuous at $y$. Also, define

$$g(x) = \begin{cases} \frac{f(x)-f(y)}{x-y} & x \neq y \\ h(y) & x = y \end{cases}$$

Then $f$ is differentiable with derivative $h$ at $y$ iff $g$ is continuous at $y$. The only thing left to do is to show that the convergence is uniform, so that $g$ is continuous.

For $x \neq y$, we know that

$$g_n(x) - g_m(y) = \frac{(f_n - f_m)(x) - (f_n - f_m)(y)}{x - y} = (f'_n - f'_m)(t)$$

for some $t \in [x, y]$. This also

Let $\varepsilon > 0$. Since $f'$ converges uniformly, there is some $N$ such that for all $x \neq y$, $n, m > N$, we have

$$|g_n(x) - g_m(y)| \leq \sup |(f'_n - f'_m)(t)| < \varepsilon.$$

For $x = y$, this is still true since $|g_n(y) - g_m(y)| = |f'_n(y) - f'_m(y)|$ by definition. So

$$n, m \geq N \Rightarrow \sup_{[a,b]} |g_n - g_m| < \varepsilon,$$

ie. $g_n$ converges uniformly. Hence the limit function $g$ is continuous, in particular at $x = y$. So $f$ is differentiable at $y$ and $f'(y) = h(y) = \lim f'_n(y)$. □

**Proposition.** If $(f_n) \to f$ and $(g_n) \to g$ uniformly, and $h$ is bounded, then $af_n + bg_n + h(f_n) \to af + bg + hf$ uniformly.

This is just the extension of the normal limits and the proof is literally expanding this statement.

## 2 Series of Functions

After sequences, it is natural to look at: Series!

**Definition** (Convergence of series)**.** Let $g_n; E \to \mathbb{R}$ be a sequence of functions. Then we say the series $\sum_{n=1}^{\infty} g_n$ converges at a point $x \in E$ if the sequence of partial sums

$$f_n = \sum_{j=1}^{n} g_j$$

converges at $x$. The series converges uniformly if $f_n$ converges uniformly. Similarly to Analysis I, we define that $\sum g_n$ converges *absolutely (uniformly)* at a point $x \in E$ if $\sum |g_n|$ converges (uniformly) at $x$.

Now let's rerun the proofs from Analysis I:

**Proposition.** Let $g_n : E \to \mathbb{R}$. If $\sum g_n$ converges absolutely uniformly, then $\sum g_n$ converges uniformly.

*Proof.* Let $f_n = \sum_{j=1}^{n} g_j$ and $h_n(x) = \sum_{j=1}^{n} |g_j|$ be the partial sums. Then for $n > m$, we have

$$|f_n(x) - f_m(x)| = \left| \sum_{j=m+1}^{n} g_j(x) \right| \leq \sum_{j=m+1}^{n} |g_j(x)| = |h_n(x) - h_m(x)|.$$

By hypothesis, we have

$$\sup_{x \in E} |h_n(x) - h_m(x)| \to 0 \text{ as } n, m \to \infty.$$

So we get

$$\sup_{x \in E} |f_n(x) - f_m(x)| \to 0 \text{ as } n, m \to \infty.$$

So the result follows from the Cauchy criteria. □

The following is a handy test for uniform convergence:

**Theorem** (Weierstrass M-test)**.** Let $g_n : E \to \mathbb{R}$ be a sequence of functions. Suppose there is some sequence $M_n$ such that for all $n$, we have

$$\sup_{x \in E} |g_n(x)| \leq M_n.$$

If $\sum M_n$ converges, then $\sum g_n$ converges absolutely uniformly.

*Proof.* Let $f_n = \sum_{j=1}^{n} |g_j|$ be the partial sums. Then for $n > m$, we have

$$\sup |f_n(x) - f_m(x)| = \sup \sum_{j=m+1}^{n} |g_j(x)| \leq \sum_{j=m+1}^{n} M_j \to 0 \text{ as } n, m \to \infty.$$

So done by the Cauchy criterion. □

## 2.1 Power Series

Remember the results we proved in Analysis I? Now we are going to prove a slightly stronger result:

**Theorem.** Let $\sum_{n=0}^{\infty} c_n(x-a)^n$ be a real power series. Then there exists a unique number $R \in [0, +\infty]$ (called the radius of convergence) such that

(i) If $|x-a| < R$, then $\sum c_n(x-a)^n$ converges absolutely. (Analysis I)

(ii) If $|x-a| > R$, then $\sum c_n(x-a)^n$ diverges. (Analysis I)

(iii) If $R > 0$ and $0 < r < R$, then $\sum c_n(x-a)^n$ converges absolutely uniformly on $[a-r, a+r]$.

*Proof.* For (i) and (ii), see Analysis I and note that $R = \frac{1}{\limsup \sqrt[n]{c_n}}$.

For (iii), note that from (i), taking $x = a-r$, we know that $\sum |c_n| r^n$ is convergent. But we know that if $x \in [a-r, a+r]$, then

$$|c_n(x-a)^n| \le |c_n| r^n.$$

So the result follows from the Weierstrass M-test by taking $M_n = |c_n| r^n$. □

The following theorem is proved in Analysis I, but it is important, so we do it again:

**Theorem** (Termwise differentiation of power series)**.** Suppose $\sum c_n(x-n)^n$ is a real power series with radius of convergence $R > 0$. Then

(i) The "derived series" $\sum_{n=1}^{\infty} nc_n(x-a)^{n-1}$ has radius of convergence $R$.

(ii) The function defined by $f(x) = \sum c_n(x-a)^n$, $x \in (a-R, a+R)$ is differentiable with derivative $f'(x) = \sum nc_n(x-a)^{n-1}$ within the (open) circle of convergence.

*Proof.*

(i) Let $R_1$ be the radius of convergence of the derived series. We know

$$|c_n(x-a)^n| = |c_n||x-a|^{n-1}|x-a| \le |nc_n(x-a)^{n-1}||x-a|.$$

Hence if the derived series $\sum nc_n(x-a)^{n-1}$ converges absolutely for some $x$, then so does $\sum c_n(x-a)^n$. So $R_1 \le R$.

Suppose that the inequality is strict, ie. $R_1 < R$, then there are $r_1, r$ such that $R_1 < r_1 < r < R$, where $\sum n|c_n|r_1^{n-1}$ diverges while $\sum |c_n| r^n$ converges. But this cannot be true since $n|c_n|r_1^{n-1} \le |c_n| r^n$ for sufficiently large $n$. So we must have $R_1 = R$.

(ii) Let $f_n(x) = \sum_{j=0}^{n} c_j(x-a)^j$. Then $f_n'(x) = \sum_{j=1}^{n} jc_j(x-a)^{j-1}$. We want to use the result that the derivative of limit is limit of derivative. This requires that $f_n$ converges at a point, and that $f_n'$ converges uniformly. The first is obviously true, and we know that $f_n'$ converges uniformly on $[a-r, a+r]$ for any $r < R$.

So for each $x_0$, there is some interval containing $x_0$ on which $f_n'$ is convergent. So on this interval, we know that

$$f(x) = \lim_{n\to\infty} f_n(x)$$

is differentiable with

$$f'(x) = \lim_{n\to\infty} f_n'(x) = \sum_{j=1}^{\infty} jc_j(x-a)^j.$$

In particular,

$$f'(x_0) = \sum_{j=1}^{\infty} jc_j(x_0-a)^j.$$

Since this is true for all $x_0$, the result follows.

□

# 3 Uniform Continuity and Integration

Remember how we boosted convergence to uniform convergence? We are going to do the exact same thing to continuity.

**Definition** (Uniform continuity)**.** Let $E \subseteq \mathbb{R}$ and $f : E \to \mathbb{R}$. We say that $f$ is *uniformly continuous* on $E$ if

$$(\forall\varepsilon)(\exists\delta > 0)(\forall x)(\forall y)\ |x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon.$$

**Theorem.** Any continuous function on a closed, bounded interval is uniformly continuous.

**Note.** Note TWO conditions. Both conditions are there to prevent $f$ from blowing up. This would be important on tests.

*Proof.* We are going to prove by contradiction. Suppose $f : [a,b] \to \mathbb{R}$ is not uniformly continuous. Since $f$ is not uniformly continuous, there is some $\varepsilon > 0$ such that for all $\delta = \frac{1}{n}$, there is some $x_n, y_n$ such that $|x_n - y_n| < \frac{1}{n}$ but $|f(x_n) - f(y_n)| > \varepsilon$.

Since we are on a closed, bounded interval, by Bolzano-Weierstrass, $(x_n)$ has a convergent subsequence $(x_{n_i}) \to x$. Then we also have $y_{n_i} \to x$. So by continuity, we must have $f(x_{n_i}) \to f(x)$ and $f(y_{n_i}) \to f(x)$. But $|f(x_{n_i}) - f(y_{n_i})| > \varepsilon$ for all $n_i$. This is a contradiction. □

## 3.1 Integrability

We first review the definition of Riemann Integrability:

**Definition** (Riemann integrability)**.** A bounded function $f : [a,b] \to \mathbb{R}$ is *Riemann integrable* on $[a,b]$ if $I^*(f) = \inf U(P,f) = I_*(f) = \sup L(P,f)$, where $U(P,f)$ is an upper sum of the function $f$ and partition $P$, and $L(P,f)$ a lower sum. We write

$$\int_a^b f(x)\ \mathrm{d}x = I^*(f) = I_*(f).$$

Also we knew the following equivalence:

**Theorem** (Riemann criterion for integrability)**.** A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable iff for every $\varepsilon$, there is a partition $P$ such that

$$U(P, f) - L(P, f) < \varepsilon.$$

### 3.1.1 Uniform Continuity & Integrability

**Theorem.** If $f : [a, b] \to [A, B]$ is integrable and $g : [A, B] \to \mathbb{R}$ is continuous, then $g \circ f : [a, b] \to \mathbb{R}$ is integrable. Therefore, *any* continuous function is integrable.

*Proof.* Let $\varepsilon > 0$. Since $g$ is continuous, $g$ is uniformly continuous. Let's find $\delta = \delta(\varepsilon) > 0$ such that for any $x, y \in [A, B]$, we have $|x-y| < \delta$ implies $|g(x)-g(y)| < \varepsilon$.

Since $f$ is integrable, for arbitrary $\varepsilon'$, we can find a partition $P = \{a = a_0 < a_1 < \cdots < a_n = b\}$ such that

$$U(P, f) - L(P, f) = \sum_{j=0}^{n-1} (a_{j+1} - a_j) \left( \sup_{I_j} f - \inf_{I_j} f \right) < \varepsilon'. \qquad (*)$$

Our objective is to make $U(P, g \circ f) - L(P, g \circ f)$ small. By uniform continuity of $g$, if $\sup_{I_j} f - \inf_{I_j} f$ is less than $\delta$, then $\sup_{I_j} g \circ f - \inf_{I_j} g \circ f$ will be less that $\varepsilon$. We like these sorts of intervals. So we let

$$J = \left\{ j : \sup_{I_j} f - \inf_{I_j} f < \delta \right\},$$

We now show properly that these intervals are indeed "nice". For any $j \in J$, for all $x, y \in I_j$, we must have

$$|f(x) - f(y)| \leq \sup_{z_1, z_2 \in I_j} (f(z_1) - f(z_2)) = \sup_{I_j} f - \inf_{I_j} f < \delta.$$

Hence, for each $j \in J$ and all $x, y \in I_j$, we know that

$$|g \circ f(x) - g \circ f(y)| < \varepsilon.$$

Hence, we must have

$$\sup_{I_j} \Big( g \circ f(x) - g \circ f(y) \Big) \leq \varepsilon. \Rightarrow \sup_{I_j} g \circ f - \inf_{I_j} g \circ f \leq \varepsilon.$$

Hence we know that

$$\begin{aligned} U(P, g \circ f) - L(P, g \circ f) &= \sum_{j=0}^{n} (a_{j+1} - a_j) \left( \sup_{I_j} g \circ f - \inf_{I_j} g \circ f \right) \\ &= \sum_{j \in J} (a_{j+1} - a_j) \left( \sup_{I_j} g \circ f - \inf_{I_j} g \circ f \right) \\ &\qquad + \sum_{j \notin J} (a_{j+1} - a_j) \left( \sup_{I_j} g \circ f - \inf_{I_j} g \circ f \right) \\ &\leq \varepsilon(b - a) + 2 \sup_{[A,B]} |g| \sum_{j \notin J} (a_{j+1} - a_j) \end{aligned}$$

Hence, it suffices here to make $\sum\limits_{j\notin J}(a_{j+1}-a_j)$ small. From $(*)$, we know that we must have

$$\sum_{j\notin J}(a_{j+1}-a_j) < \frac{\varepsilon'}{\delta},$$

or else $U(P,f)-L(P,f) > \varepsilon'$. So we can bound

$$U(P,g\circ f)-L(P,g\circ f) \le \varepsilon(b-a) + 2\sup_{[A,B]}|g|\frac{\varepsilon'}{\delta}.$$

So if we are given an $\varepsilon$ at the beginning, we can get a $\delta$ by uniform continuity. Afterwards, we pick $\varepsilon'$ such that $\varepsilon' = \varepsilon\delta$. Then we have shown that given any $\varepsilon$, there exists a partition such that

$$U(P,g\circ f)-L(P,g\circ f) < \left((b-a)+2\sup_{[A,B]}|g|\right)\varepsilon.$$

Then the claim follows from the Riemann criterion. □

**Note.** Main idea of the proof above: Find the difference of the partitions. Use uniform continuity to first take out the good intervals (those smaller than $\delta$): they are easy to bound. The bad ones? We know $f$ is integrable, so we can make the $f$ partition as small as we like. So we can bound the "bad" intervals. Done.

**Theorem.** Let $f_n : [a,b] \to \mathbb{R}$ be bounded and integrable for all $n$. Then if $(f_n)$ converges uniformly to a function $f : [a,b]\to\mathbb{R}$, then $f$ is bounded and integrable.

*Proof.* Let

$$c_n = \sup_{[a,b]}|f_n - f|.$$

Then uniform convergence says that $c_n \to 0$. By definition, for each $x$, we have

$$f_n(x) - c_n \le f(x) \le f_n(x) + c_n.$$

Since $f_n$ is bounded, this implies that $f$ is bounded by $\sup|f_n| + c_n$. Also, for any $x, y \in [a,b]$, we know

$$f(x) - f(y) \le (f_n(x) - f_n(y)) + 2c_n.$$

Hence for any partition $P$,

$$U(P,f) - L(L,f) \le U(P,f_n) - L(P,f_n) + 2(b-a)c_n.$$

So given $\varepsilon > 0$, first choose $n$ such that $2(b-a)c_n < \frac{\varepsilon}{2}$. Then choose $P$ such that $U(P,f_n) - L(P,f_n) < \frac{\varepsilon}{2}$. Then for this partition, $U(P,f) - L(P,f) < \varepsilon$. □

If we extend Riemann integrability trivially to a vector-valued function (by defining the integrability on each component), we will have the following theorem: (Hint: If this looks random, it is because this is used in the Norms section)

**Proposition.** If $\mathbf{f} : [a,b] \to \mathbb{R}^n$ is integrable, then the function $\|\mathbf{f}\| : [a,b] \to \mathbb{R}$ defined by

$$\|\mathbf{f}\|(x) = \|\mathbf{f}(x)\| = \sqrt{\sum_{j=1}^{n} f_j^2(x)}.$$

is integrable, and

$$\left\| \int_a^b \mathbf{f}(x)\ \mathrm{d}x \right\| \leq \int_a^b \|\mathbf{f}\|(x)\ \mathrm{d}x.$$

*Proof.* The integrability of $\|\mathbf{f}\|$ is clear since squaring and taking square roots are continuous, and a finite sum of integrable functions is integrable. To show the inequality, we let

$$\mathbf{v} = (v_1, \cdots, v_n) = \int_a^b \mathbf{f}(x)\ \mathrm{d}x.$$

Then by definition,

$$v_j = \int_a^b f_j(x)\ \mathrm{d}x.$$

If $\mathbf{v} = \mathbf{0}$, then we are done. Otherwise, we have

$$\|\mathbf{v}\|^2 = \sum_{j=1}^{n} v_j^2 = \sum_{j=1}^{n} v_j \int_a^b f_j(x)\ \mathrm{d}x = \int_a^b \sum_{j=1}^{n} (v_j f_j(x))\ \mathrm{d}x = \int_a^b \mathbf{v} \cdot \mathbf{f}(x)\ \mathrm{d}x$$

Using the Cauchy-Schwarz inequality, we get

$$\leq \int_a^b \|\mathbf{v}\| \|\mathbf{f}\|(x)\ \mathrm{d}x = \|\mathbf{v}\| \int_a^b \|\mathbf{f}\|\ \mathrm{d}x.$$

Divide by $\|\mathbf{v}\|$ and we are done. □

# 4 $\mathbb{R}^n$ as a Normed Space, aka Met and Top 0.5

## 4.1 Terminology and Introduction

A normed space is a space equipped with a norm, an abstracted version of a distance measure. Well, we like to abstract everything:

**Definition** (Normed space)**.** Let $V$ be a real vector space. A *norm* on $V$ is a function $\|\cdot\| : V \to \mathbb{R}$ satisfying

(i) $\|\mathbf{x}\| \geq 0$ with equality iff $\mathbf{x} = \mathbf{0}$ (non-negativity)

(ii) $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$ (linearity in scalar multiplication)

(iii) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (triangle inequality)

A *normed space* is a pair $(V, \|\cdot\|)$. If the norm is understood, we just say $V$ is a normed space. We do have to be slightly careful since there can be multiple norms on a vector space.

**Example** (Important Norms for the Test)**.**

$p$-Norms for $\mathbb{R}^n$

$$\|\mathbf{x}\|_p = \left(\sum |x_i|^p\right)^{1/p}.$$

The triangle inequality is NOT trivial to check for this norm. You shouldn't be doing this outside of cases $p = 1, 2, \infty$, where the infinity norm is defined as the limit of this, which is simply $\max\{|x_i|\}$.

$l$-Norms for $l$-spaces Notice this is infinite dimensional. We have the following set of norms defined on $l^p$ space, similar to $p$-norms:

$$\ell^p = \left\{(x_k) \in \mathbb{R}^{\mathbb{N}} : \sum |x_k|^p < \infty\right\}$$

with the norm

$$\|(x_k)\|_p = \|(x_k)\|_{\ell^p} = \left(\sum |x_k|^p\right)^{1/p}.$$

The "space" condition is there to keep the sum converging. Infinity norm is defined the same way, with $\max$ replaced by $\sup$.

$L$-Norms for $C(a, b)$

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

Similarly, the infinity norm...Yeah, you understand.

So we have a lot of norms now. How should we compare them?

**Definition** (Lipschitz equivalence of norms). Let $V$ be a (real) vector space. Two norms $\|\cdot\|, \|\cdot\|'$ on $V$ are *Lipschitz equivalent* if there are real constants $0 < a < b$ such that

$$a\|\mathbf{x}\| \leq \|\mathbf{x}\|' \leq b\|\mathbf{x}\|$$

for all $\mathbf{x} \in V$.

It is easy to show this is indeed an equivalence relation on the set of all norms on $V$.

Now to make this a bit abstract, we introduce the equivalent notion:

**Definition** (Open ball). Let $(V, \|\cdot\|)$ be a normed space, $\mathbf{a} \in V$, $r > 0$. The *open ball* centered at $\mathbf{a}$ with radius $r$ is

$$B_r(\mathbf{a}) = \{\mathbf{x} \in V : \|\mathbf{x} - \mathbf{a}\| < r\}.$$

Then the requirement that $a\|\mathbf{x}\| \leq \|\mathbf{x}\|' \leq b\|\mathbf{x}\|$ for all $\mathbf{x} \in V$ is equivalent to saying

$$B_{1/b}(\mathbf{0}) \subseteq B'_1(\mathbf{0}) \subseteq B_{1/a}(\mathbf{0}),$$

Next up is our good old friend Cauchy-Schwarz:

**Lemma** (Cauchy-Schwarz inequality). If $f, g \in C([a, b])$, $f, g \geq 0$, then

$$\int_a^b fg \, \mathrm{d}x \leq \left(\int_a^b f^2 \, \mathrm{d}x)\right)^{1/2} \left(\int_a^b g^2 \, \mathrm{d}x\right)^{1/2}.$$

*Proof.* If $\int_a^b f^2 \, \mathrm{d}x = 0$, then the inequality holds trivially.

Otherwise, let $A^2 = \int_a^b f^2 \, \mathrm{d}x \neq 0$, $B^2 = \int_a^b g^2 \, \mathrm{d}x$. Consider the function

$$\phi(t) = \int_a^b (g - tf)^2 \, \mathrm{d}t = t^2 A^2 - 2t \int_a^b gf \, \mathrm{d}x + B^2 \geq 0.$$

The conditions for a quadratic in $t$ to be non-negative is exactly

$$\left( \int_a^b gf \, \mathrm{d}x \right)^2 - A^2 B^2 \leq 0.$$

□

More terminology:

**Definition** (Bounded subset)**.** Let $(V, \|\cdot\|)$ be a normed space. A subset $E \subseteq V$ is *bounded* if there is some $R > 0$ such that

$$E \subseteq B_R(\mathbf{0}).$$

Two equivalent norms agree on what is bounded.

**Definition** (Convergence of sequence)**.** Let $(V, \|\cdot\|)$ be a normed space. A sequence $(x_k)$ in $V$ *converges to* $\mathbf{x} \in V$ if $\|\mathbf{x}_k - \mathbf{x}\| \to 0$ (as a sequence in $\mathbb{R}$), ie.

$$(\forall \varepsilon > 0)(\exists N)(\forall k \geq N) \; \|\mathbf{x}_k - \mathbf{x}\| < \varepsilon.$$

Two equivalent norms agree on what converges.

## 4.2 Convergence in Normed Spaces

Some basic results extend trivially in *finite* dimensional spaces: (Remember, all norms in $\mathbb{R}^n$ are actually equivalent (proof later), so we don't need to care which one we are talking about)

**Proposition.** Convergence in $\mathbb{R}^n$ is equivalent to coordinate-wise convergence, ie. $\mathbf{x}^{(k)} \to \mathbf{x}$ if and only if $x_j^{(k)} \to x_j$ for all $j$.

*Proof.* Fix $\varepsilon > 0$. Suppose $\mathbf{x}^{(k)} \to \mathbf{x}$. Then there is some $N$ such that for any $k \geq N$ such that

$$\|\mathbf{x}^{(k)} - \mathbf{x}\|_2^2 = \sum_{j=1}^n (x_j^{(k)} - x_j)^2 < \varepsilon.$$

Hence $|x_j^{(k)} - x_j| < \varepsilon$ for all $k \leq N$.

On the other hand, for any fixed $j$, there is some $N_j$ such that $k \geq N_j$ implies $|x_j^{(k)} - x_j| < \frac{\varepsilon}{\sqrt{n}}$. So if $k \geq \max\{N_j : j = 1, \cdots, n\}$, then

$$\|\mathbf{x}^{(k)} - \mathbf{x}\|_2 = \left( \sum_{j=1}^n (x_j^{(k)} - x_j)^2 \right)^{\frac{1}{2}} < \varepsilon.$$

□

**Theorem** (Bolzano-Weierstrass theorem in $\mathbb{R}^n$)**.** Any bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

*Proof.* We induct on $n$. The $n = 1$ case is the usual Bolzano-Weierstrass on the real line, which was proved in IA Analysis I.

Assume the theorem holds in $\mathbb{R}^{n-1}$, and let $\mathbf{x}^{(k)} = (x_1^{(k)}, \cdots, x_n^{(k)})$ be a bounded sequence in $\mathbb{R}^n$. Then let $\mathbf{y}^{(k)} = (x_1^{(k)}, \cdots, x_{n-1}^{(k)})$. Since for any $k$, we know that

$$\|\mathbf{y}^{(k)}\|^2 + |x_n^{(k)}|^2 = \|\mathbf{x}^{(k)}\|^2,$$

it follows that both $(\mathbf{y}^{(k)})$ and $(x_n^{(k)})$ are bounded. So by the induction hypothesis, there is a subsequence $(k_j)$ of $(k)$ and some $\mathbf{y} \in \mathbb{R}^{n-1}$ such that $\mathbf{y}^{(k_j)} \to \mathbf{y}$. Also, by Bolzano-Weierstrass in $\mathbb{R}$, there is a further subsequence $(x_n^{(k_{j_\ell})})$ of $(x_n^{(k_j)})$ that converges to, say, $y_n \in \mathbb{R}$. Then we know that

$$\mathbf{x}^{(k_{j_\ell})} \to (\mathbf{y}, y_n).$$

So done. □

**Note.** Remember. This only works in *finite* dimensional spaces. In infinite dimensional spaces neither work, since you can send non-zero components down to infinity, so they pointwise converge to 0, but the norm never changes.

**Definition** (Cauchy sequence)**.** Let $(V, \|\cdot\|)$ be a normed space. A sequence $(\mathbf{x}^{(k)})$ in $V$ is a *Cauchy sequence* if

$$(\forall \varepsilon)(\exists N)(\forall n, m \geq N)\ \|\mathbf{x}^{(n)} - \mathbf{x}^{(m)}\| < \varepsilon.$$

Clearly,a Cauchy sequence is bounded and every convergent sequence is Cauchy. Standard results such as if a subsequence of a Cauchy sequence converges, then the whole sequence converges also extends trivially.

**Definition** (Complete normed space)**.** A normed space $(V, \|\cdot\|)$ is *complete* if every Cauchy sequence converges to an element in $V$.

We can have the following proposition based on definition:

**Proposition.** If $\|\cdot\|'$ is Lipschitz equivalent to $\|\cdot\|$ on $V$, then $(\mathbf{x}_k)$ is Cauchy with respect to $\|\cdot\|$ if and only if $(\mathbf{x}_k)$ is Cauchy with respect to $\|\cdot\|'$. Also, $(V, \|\cdot\|)$ is complete if and only if $(V, \|\cdot\|')$ is complete.

Here comes an important theorem. Remember this for the test!

**Theorem.** $\mathbb{R}^n$ (with the Euclidean norm, say) is complete.

*Proof.* The important thing is to know this is true for $n = 1$, which we have proved from Analysis I.

If $(\mathbf{x}^k)$ is Cauchy in $\mathbb{R}^n$, then $(x_j^{(k)})$ is a Cauchy sequence of real numbers for each $j \in \{1, \cdots, n\}$. By the completeness of the reals, we know that $x_j^k \to x_j \in \mathbb{R}$ for some $x$. So $x^k \to x = (x_1, \cdots, x_n)$ since convergence in $\mathbb{R}^n$ is equivalent to componentwise convergence. □

**Example.** Some useful examples: $l^p$ are complete with respect to $l_p$ norms. $C([0, 1])$ is complete with respect to the infinity norm, but NOT the $L^1$ or $L^2$ norm.

## 4.3 Open/Closed Sets, Sequential Compactness

**Definition** (Open set)**.** Let $(V, \|\cdot\|)$ be a normed space. A subspace $E \subseteq V$ is *open* in $V$ if for any $\mathbf{y} \in E$, there is some $r > 0$ such that

$$B_r(\mathbf{y}) = \{\mathbf{x} \in V : \|\mathbf{x} - \mathbf{y}\| < r\} \subseteq E.$$

This is just generalizing the open ball. We now add some more definitions:

**Definition** (Closed set)**.** Let $(V, \|\cdot\|)$ be a normed space. Then $E \subseteq V$ is *closed* if $V \setminus E$ is open, ie. $E$ contains all its limit points.

**Definition** (Limit point)**.** Let $(V, \|\cdot\|)$ be a normed space, $E \subseteq V$. A point $\mathbf{y} \in V$ is a *limit point* of $E$ if there is a sequence $(\mathbf{x}_k)$ in $E$ with $\mathbf{x}_k \neq \mathbf{y}$ for all $k$ and $\mathbf{x}_k \to \mathbf{y}$.

Now we will connect limit points and open sets:

**Lemma.** Let $(V, \|\cdot\|)$ be a normed space, $E$ any subset of $V$. Then a point $\mathbf{y} \in V$ is a limit point of $E$ if and only if

$$(B_r(\mathbf{y}) \setminus \{\mathbf{y}\}) \cap E \neq \emptyset$$

for every $r$.

*Proof.* ($\Rightarrow$) If $\mathbf{y}$ is a limit point of $E$, then there exists a sequence $(\mathbf{x}_k) \in E$ with $\mathbf{x}_k \neq \mathbf{y}$ for all $k$ and $\mathbf{x}_k \to \mathbf{y}$. Then for every $r$, for sufficiently large $k$, $\mathbf{x}_k \in B_r(\mathbf{y})$. Since $\mathbf{x}_k \neq \{\mathbf{y}\}$ and $\mathbf{x}_k \in E$, the result follows.

($\Leftarrow$) For each $k$, let $r = \frac{1}{k}$. By assumption, we have some $\mathbf{x}_k \in (B_{\frac{1}{k}}(\mathbf{y}) \setminus \{\mathbf{y}\}) \cap E$. Then $\mathbf{x}_k \to \mathbf{y}$, $\mathbf{x}_k \neq \mathbf{y}$ and $\mathbf{x}_k \in E$. So $\mathbf{y}$ is a limit point of $E$. ☐

Now we prove the following:

**Proposition.** Let $E \subseteq V$. Then $E$ contains all of its limit points iff $V \setminus E$ is open in $V$.

*Proof.* ($\Rightarrow$) Suppose $E$ contains all its limit points. Let $\mathbf{y} \in V \setminus E$. So $\mathbf{y}$ is not a limit point of $E$. So for some $r$, we have $(B_r(\mathbf{y}) \setminus \{\mathbf{y}\}) \cap E = \emptyset$. Hence it follows that $B_r(\mathbf{y}) \subseteq V \setminus E$ (since $\mathbf{y} \notin E$).

($\Leftarrow$) Suppose $V \setminus E$ is open. Let $\mathbf{y} \in V \setminus E$. Since $V \setminus E$ is open, there is some $r$ such that $B_r(\mathbf{y}) \subseteq V \setminus E$. By the lemma, $\mathbf{y}$ is not a limit point of $E$. So all limit points of $E$ are in $E$. ☐

Now we introduce the concept of Sequentially compact:

**Definition** ((Sequentially) compact set)**.** Let $V$ be a normed vector space. A subset $K \subseteq V$ is said to be *compact* (or *sequentially compact*) if every sequence in $K$ has a subsequence that converges to a point in $K$.

There are things we can immediately know about the spaces:

**Theorem.** Let $(V, \|\cdot\|)$ be a normed vector space, $K \subseteq V$ a subset. Then

(i) If $K$ is compact, then $K$ is closed and bounded.

(ii) If $V$ is $\mathbb{R}^n$, then if $K$ is closed and bounded, then $K$ is compact.

*Proof.*

(i) Let $K$ be compact. If $K$ is unbounded, then we can generate a sequence $\mathbf{x}_k$ such that $\|\mathbf{x}_k\| \to \infty$. Then this cannot have a convergent subsequence, since any subsequence will also be unbounded, and convergent sequences are bounded. So $K$ must be bounded.

   To show $K$ is closed, let $\mathbf{y}$ be a limit point of $K$. Then there is some $\mathbf{y}_k \in K$ such that $\mathbf{y}_k \to \mathbf{y}$. Then by compactness, there is a subsequence of $\mathbf{y}_k$ converging to some point in $K$. But the subsequence must converge to $\mathbf{y}$. So $\mathbf{y} \in K$.

(ii) Let $K$ be closed and bounded. Let $\mathbf{x}_k$ be a sequence in $K$. Since $V = \mathbb{R}^n$ and $K$ is bounded, $(\mathbf{x}_k)$ is a bounded sequence in $\mathbb{R}^n$. So by Bolzano-Weierstrass, this has a convergent subsequence $\mathbf{x}_{k_j}$. By closedness of $V$, we know that the limit is in $K$. So $K$ is compact.

□

## 4.4 Mappings between Normed Spaces

Let $(V, \|\cdot\|)$, $(V', \|\cdot\|')$ be normed spaces, and let $E \subseteq K$ be a subset, and $f : E \to V'$ a mapping (which is just a function, although we reserve the terminology "function" or "functional" for when $V' = \mathbb{R}$).

**Definition** (Continuity of mapping)**.** Let $\mathbf{y} \in E$. We say $f : E \to V'$ is *continuous at* $\mathbf{y}$ if for all $\varepsilon > 0$, there is $\delta > 0$ such that the following holds:

$$(\forall \mathbf{x} \in E)\ \|\mathbf{x} - \mathbf{y}\|_V < \delta \Rightarrow \|f(\mathbf{x}) - f(\mathbf{y})\|_{V'} < \varepsilon.$$

Note this is really just extending the definition of continuity in $\mathbb{R}$.

**Theorem.** Let $(V, \|\cdot\|)$, $(V', \|\cdot\|')$ be normed spaces, $E \subseteq V$, $f : E \to \mathbf{V}'$. Then $f$ is continuous at $\mathbf{y} \in E$ iff for any sequence $\mathbf{y}_k \to \mathbf{y}$ in $E$, we have $f(\mathbf{y}_k) \to f(\mathbf{y})$.

*Proof.* ($\Rightarrow$) Suppose $f$ is continuous at $\mathbf{y} \in E$, and that $\mathbf{y}_k \to \mathbf{y}$. Given $\varepsilon > 0$, by continuity, there is some $\delta > 0$ such that

$$B_\delta(\mathbf{y}) \cap E \subseteq f^{-1}(B_\varepsilon(f(\mathbf{y}))).$$

For sufficiently large $k$, $\mathbf{y}_k \in B_\delta(\mathbf{y}) \cap E$. So $f(\mathbf{y}_k) \in B_\varepsilon(f(\mathbf{y}))$, or equivalently,

$$|f(\mathbf{y}_k) - f(\mathbf{y})| < \varepsilon.$$

So done.

($\Leftarrow$) If $f$ is not continuous at $y$, then there is some $\varepsilon > 0$ such that for any $k$, we have

$$B_{\frac{1}{k}}(\mathbf{y}) \not\subseteq f^{-1}(B_\varepsilon(f(\mathbf{y}))).$$

Choose $\mathbf{y}_k \in B_{\frac{1}{k}}(\mathbf{y}) \setminus f^{-1}(B_\varepsilon(f(\mathbf{y})))$. Then $\mathbf{y}_k \to \mathbf{y}$, $\mathbf{y}_k \in F$, but $\|f(\mathbf{y}_k) - f(\mathbf{y})\| \geq \varepsilon$, contrary to the hypothesis. □

**Definition** (Continuous function)**.** $f : E \to V'$ is *continuous* if $f$ is continuous at every point $\mathbf{y} \in E$.

**Theorem.** Let $(V, \|\cdot\|)$ and $(V', \|\cdot\|')$ be normed spaces, and $K$ a compact subset of $V$, and $f : V \to V'$ a continuous function. Then $f(K)$ is compact in $V'$, thus closed and bounded.

*Proof.* Let $(\mathbf{x}_k)$ be a sequence in $f(K)$ with $\mathbf{x}_k = f(\mathbf{y}_k)$ for some $\mathbf{y}_k \in K$. By compactness of $K$, there is a subsequence $(\mathbf{y}_{k_j})$ such that $\mathbf{y}_{k_j} \to \mathbf{y}$. By the previous theorem, we know that $f(\mathbf{y}_{j_k}) \to f(\mathbf{y})$. So $\mathbf{x}_{k_j} \to f(\mathbf{y}) \in f(K)$. So $f(K)$ is compact.

□

Finally, we will prove what we said is true: Any norm in $\mathbb{R}^n$ are Lipschitz equivalent. We first introduce a lemma:

**Lemma.** Let $V$ be an $n$-dimensional vector space with a basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. Then for any $\mathbf{x} \in V$, write $\mathbf{x} = \sum_{j=1}^n x_j \mathbf{v}_j$, with $x_j \in \mathbb{R}$. We define the Euclidean norm by

$$\|\mathbf{x}\|_2 = \left(\sum x_j^2\right)^{\frac{1}{2}}.$$

Then this is a norm, and $S = \{\mathbf{x} \in V : \|\mathbf{x}\|_2 = 1\}$ is compact in $(V, \|\cdot\|_2)$.

After we show this, we can easily show that every other norm is equivalent to this norm.

This is not hard to prove, since we know that the unit sphere in $\mathbb{R}^n$ is compact, and we can just pass our things on to $\mathbb{R}^n$.

*Proof.* $\|\cdot\|_2$ is well-defined since $x_1, \cdots, x_n$ are uniquely determined by $\mathbf{x}$ (by definition of basis). It is easy to check that $\|\cdot\|_2$ is a norm.

Given a sequence $\mathbf{x}^{(k)}$ in $S$, if we write $\mathbf{x}^{(k)} = \sum_{j=1}^n x_j^{(k)} \mathbf{v}_j$. We define the following sequence in $\mathbb{R}^n$:

$$\tilde{\mathbf{x}}^{(k)} = (x_1^{(k)}, \cdots, x_n^{(k)}) \in \tilde{S} = \{\tilde{\mathbf{x}} \in \mathbb{R}^n : \|\tilde{\mathbf{x}}\|_{\text{Euclid}} = 1\}.$$

As $\tilde{S}$ is closed and bounded in $\mathbb{R}^n$ under the Euclidean norm, it is compact. Hence there exists a subsequence $\tilde{x}^{(k_j)}$ and $\tilde{x} \in \tilde{S}$ such that $\|\tilde{\mathbf{x}}^{(k_j)} - \tilde{\mathbf{x}}\|_{\text{Euclid}} \to 0$. This says that $\mathbf{x} = \sum_{j=1}^n x_j \mathbf{v}_j \in S$, and $\|\mathbf{x}^{k_j} - \mathbf{x}\|_2 \to 0$. So done. □

**Theorem.** Any two norms on a finite dimensional vector space are Lipschitz equivalent.

The idea is to pick a basis, and prove that any norm is equivalent to $\|\cdot\|_2$.

To show that an arbitrary norm $\|\cdot\|$ is equivalent to $\|\cdot\|_2$, we have to show that for any $\|\mathbf{x}\|$, we have

$$a\|\mathbf{x}\|_2 \leq \|\mathbf{x}\| \leq b\|\mathbf{x}\|_2.$$

We can divide by $\|\mathbf{x}\|_2$ and obtain an equivalent requirement:

$$a \leq \left\|\frac{\mathbf{x}}{\|\mathbf{x}\|_2}\right\| \leq b.$$

We know that any $\mathbf{x}/\|\mathbf{x}\|_2$ lies in the unit sphere $S = \{\mathbf{x} \in V : \|\mathbf{x}\|_2 = 1\}$. So we want to show that the image of $\|\cdot\|$ is bounded. But we know that $S$ is compact. So it suffices to show that $\|\cdot\|$ is continuous.

*Proof.* Fix a basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ for $V$, and define $\|\cdot\|_2$ as in the lemma above. Then $\|\cdot\|_2$ is a norm on $V$, and $S = \{\mathbf{x} \in V : \|\mathbf{x}\|_2 = 1\}$, the unit sphere, is compact by above.

To show that any two norms are equivalent, it suffices to show that if $\|\cdot\|$ is any other norm, then it is equivalent to $\|\cdot\|_2$, since equivalence is transitive.

For any

$$\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{v}_j,$$

we have

$$\|\mathbf{x}\| = \left\|\sum_{j=1}^{n} x_j \mathbf{v}_j\right\| \leq \sum |x_j|\|\mathbf{v}_j\| \leq \|\mathbf{x}\|_2 \left(\sum_{j=1}^{n} \|\mathbf{v}_j\|^2\right)^{\frac{1}{2}}$$

by the Cauchy-Schwarz inequality. So $\|\mathbf{x}\| \leq b\|\mathbf{x}\|_2$ for $b = \left(\sum \|\mathbf{v}_j\|^2\right)^{\frac{1}{2}}$.

To find $a$ such that $\|\mathbf{x}\| \geq a\|\mathbf{x}\|_2$, consider $\|\cdot\| : (S, \|\cdot\|_2) \to \mathbb{R}$. By above, we know that

$$\|\mathbf{x} - \mathbf{y}\| \leq b\|\mathbf{x} - \mathbf{y}\|_2$$

By the triangle inequality, we know that $\big|\|\mathbf{x}\| - \|\mathbf{y}\|\big| \leq \|\mathbf{x} - \mathbf{y}\|$. So when $\mathbf{x}$ is close to $\mathbf{y}$ under $\|\cdot\|_2$, then $\|\mathbf{x}\|$ and $\|\mathbf{y}\|$ are close. So $\|\cdot\| : (S, \|\cdot\|_2) \to \mathbb{R}$ is continuous. So there is some $\mathbf{x}_0 \in S$ such that $\|\mathbf{x}_0\| = \inf_{x \in S} \|\mathbf{x}\| = a$, say. Since $\|\mathbf{x}_0\| > 0$, we know that $\|\mathbf{x}_0\| > 0$. So $\|\mathbf{x}\| \geq a\|\mathbf{x}\|_2$ for all $\mathbf{x} \in V$. □

**Corollary.** Let $(V, \|\cdot\|)$ be a finite-dimensional normed space.

(i) The Bolzano-Weierstrass theorem holds for $V$, ie. any bounded sequence sequence in $V$ has a convergent subsequence.

(ii) A subset of $V$ is compact if and only if it is closed and bounded.

*Proof.* If a subset is bounded in one norm, then it is bounded in any Lipschitz equivalent norm. Similarly, if it converges to $\mathbf{x}$ in one norm, then it converges to $\mathbf{x}$ in any Lipschitz equivalent norm.

Since these results hold for the Euclidean norm $\|\cdot\|_2$, it follows that they hold for arbitrary finite-dimensional vector spaces. □

**Corollary.** Any finite-dimensional normed vector space $(V, \|\cdot\|)$ is complete.

*Proof.* This is true since if a space is complete in one norm, then it is complete in any Lipschitz equivalent norm, and we know that $\mathbb{R}^n$ under the Euclidean norm is complete. □

# 5 Metric Spaces, or Met and Top 0.8

If you look at the section above, you will realize that we didn't really use the concept of a norm. We kind of used more the concept of distance. And that is how we will continue to abstract our spaces. A normed space naturally induces a metric, but NOT the other way around.

**Note.** Note there is a LOT of similarities between results in normed spaces and metric spaces. Most things just pass over from normed spaces to metric spaces so you do not have to remember twice.

## 5.1 Definitions and Terms

**Definition** (Metric space)**.** Let $X$ be any set. A *metric* on $X$ is a function $d : X \times X \to \mathbb{R}$ that satisfies

- $d(x, y) \geq 0$ with equality iff $x = y$ (non-negativity)
- $d(x, y) = d(y, x)$ (symmetry)
- $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality)

The pair $(X, d)$ is called a *metric space.*

Besides the "usual" metrics, this one is worth mentioning:

**Example.** Given a metric space $(X, d)$, we define

$$g(x, y) = \min\{1, d(x, y)\}.$$

Then this is a metric on $X$. Similarly, if we define

$$h(x, y) = \frac{d(x, y)}{1 + d(x, y)}$$

is also a metric on $X$. In both cases, we obtain a *bounded* metric.

Similarly we define all the concepts, including equivalent metrics, subspaces, and convergence.

**Definition** (Lipschitz equivalent metrics)**.** Metrics $d, d'$ on a set $X$ are said to be *Lipschitz equivalent* if there are (positive) constants $A, B$ such that

$$Ad(x, y) \leq d'(x, y) \leq Bd(x, y)$$

for all $x, y \in X$.

**Definition** (Metric subspace)**.** Given a metric space $(X, d)$ and a subset $Y \subseteq X$, the restriction $d|_{Y\times Y} \to \mathbb{R}$ is a metric on $Y$. This is called the *induced metric* or *subspace metric.*

Note that unlike vector subspaces, we do not require our subsets to have any structure. We can take *any* subset of $X$ and get a metric subspace.

**Definition** (Convergence)**.** Let $(X, d)$ be a metric space. A sequence $x_n \in X$ is said to *converge* to $x$ if $d(x_n, x) \to 0$ as a real sequence. In other words,

$$(\forall \varepsilon)(\exists K)(\forall k > K)\, d(x_k, x) < \varepsilon.$$

**Proposition.** The limit of a convergent sequence is unique.

The proof is this follows from the same proof we did in normed spaces.

Now, we can also define the *inner product* on a real vector space, which is the extension of the dot product:

**Definition** (Inner product)**.** Let $V$ is a real vector space. An *inner product* on $V$ is a function $\langle \cdot,$
$cdot\rangle : V \times V \to \mathbb{R}$ such that

(i) $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ for all $\mathbf{v} \in V$

(ii) $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

(iii) $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$.

(iv) $\langle \mathbf{v}_1 + \lambda \mathbf{v}_2, \mathbf{w} \rangle = \langle \mathbf{v}_1, \mathbf{w} \rangle + \lambda \langle \mathbf{v}_2, \mathbf{w} \rangle$.

This is really just FYI. We would not do anything about it.

## 5.2 Topology of Metric Spaces

Now before we enter the "topological spaces", which concerns only the open and closed sets, we need some definitions:

**Definition** (Open subset)**.** Let $(X, d)$ be a metric space. A subset $U \subseteq X$ is *open* if for every $y \in U$, there is some $r > 0$ such that $B_r(y) \subseteq U$.

**Definition** (Topology)**.** Let $(X, d)$ be a metric space. The *topology* on $(X, d)$ is the collection of open subsets of $X$. We say it is the topology induced by the metric.

Note that non-Lipschitz equivalent metrics can induce the same open sets. (See the "interesting" metric example above: a bounded metric $\min\{1, |x - y\}$ induces the same set of open sets as $|x - y|$)

**Definition** (Topological notion)**.** A notion or property is said to be a *topological* notion or property if it only depends on the topology, and not the metric.

**Definition** (Neighbourhood)**.** Given a metric space $X$ and a point $x \in X$, a *neighbourhood* of $x$ is an open set containing $x$.

**Definition** (Limit point)**.** Let $(X, d)$ be a metric space and $E \subseteq X$. A point $y \in X$ is a *limit point* of $E$ if there exists a sequence $x_k \in E$, $x_k \neq y$ such that $x_k \to y$.

**Definition** (Closed subset)**.** A subset $E \subseteq X$ is *closed* if $E$ contains all its limit points.

**Theorem.** Let $(X, d)$ be a metric space. Then

(i) The union/intersection of *any* collection of open/closed sets is open/closed

(ii) The intersection/union of finitely many open/closed sets is open/closed.

(iii) $\emptyset$ and $X$ are open/closed.

**Note.** This theorem is important for Met and Top but not really for Analysis II.

*Proof.*

(i) Let $U = \bigcup_\alpha V_\alpha$, where each $V_\alpha$ is open. If $x \in U$, then $x \in V_\alpha$ for some $\alpha$. Since $V_\alpha$ is open, there exists $\delta > 0$ such that $B_\delta(x) \subseteq V_\alpha$. So $B_\delta(x) \subseteq \bigcup_\alpha V_\alpha = U$. So $U$ is open.

(ii) Let $U = \bigcap_{i=1}^n V_\alpha$, where each $V_\alpha$ is open. If $x \in V$, then $x \in V_i$ for all $i = 1, \cdots, n$. So $\exists \delta_i > 0$ with $B_{\delta_i}(x) \subseteq V_i$. Take $\delta = \min\{\delta_1, \cdots, \delta_n\}$. So $B_\delta(x) \subseteq V_i$ for all $i$. So $B_\delta(x) \subseteq V$. So $V$ is open.

(iii) $\emptyset$ satisfies the definition of an open subset vacuously. $X$ is open since for any $x$, $B_1(x) \subseteq X$.

For the closed cases, just take the complement of the result for the open ones. □

## 5.3 Cauchy Sequences and Completeness

**Definition** (Cauchy sequence)**.** Let $(X, d)$ be a metric space. A sequence $(x_n)$ in $X$ is *Cauchy* if

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

**Definition** (Complete metric space)**.** A metric space is complete if all Cauchy sequences converge.

Once AGAIN, we define Cauchy sequences. Every theorem that follows from normed spaces still follow. So we would only list the one that is "kind of" special to metric spaces.

**Theorem.** Let $(X, d)$ be a metric space, $Y \subseteq X$ any subset. Then

(i) If $(Y, d|_{Y\times Y})$ is complete, then $Y$ is closed in $X$.

(ii) If $(X, d)$ is complete, then $(Y, d|_{Y\times Y})$ is complete if and only if it is closed.

*Proof.*

(i) Let $x \in X$ be a limit point of $Y$. Then there is some sequence $x_k \to x$, where each $x_k \in Y$. Since $(x_k)$ is convergent, it is a Cauchy sequence. Hence it is Cauchy in $Y$. By completeness of $Y$, $(x_k)$ has to converge to some point in $Y$. By uniqueness of limits, this limit must be $x$. So $x \in Y$. So $Y$ contains all its limit points.

(ii) We have just showed that if $Y$ is complete, then it is closed. Now suppose $Y$ is closed. Let $(x_k)$ be a Cauchy sequence in $Y$. Then $(x_k)$ is Cauchy in $X$. Since $X$ is complete, $x_k \to x$ for some $x \in X$. Since $x$ is a limit point of $Y$, we must have $x \in Y$. So $x_k$ converges in $Y$.

$\square$

## 5.4 Compactness

It has the exact same definition as the one in normed spaces. And we also have that all compact spaces are complete and bounded, which also follows over from normed spaces. Thus, we omit the statement and the proof. This would help you realize Analysis II is not that long. It is just repetitive. Seriously.

## 5.5 Continuous Functions

Now we would also NOT redefine continuity/Uniform continuity in the metric space context, because you should really know this by now. Like you have seen it 3 times already. Here is a new term:

**Definition** (Lipschitz function and Lipschitz constant)**.** $f$ is said to be *Lipschitz* on $X$ if there is some $K \in [0, \infty)$ such that for all $x, y \in X$,

$$d'(f(x), f(y)) \leq K d(x, y)$$

*Any* such $K$ is called a Lipschitz constant.

Here is an example of how one theorem in $\mathbb{R}$ can move into a metric space unchanged:

**Theorem.** Let $(X, d)$ be a compact metric space, and $(X', d')$ is any metric space. If $f : X \to X'$ be continuous, then $f$ is uniformly continuous.

This is exactly the same proof as what we had for the $[0, 1]$ case.

*Proof.* Suppose $f : X \to X'$ is not uniformly continuous. Since $f$ is not uniformly continuous, there is some $\varepsilon > 0$ such that for all $\delta = \frac{1}{n}$, there is some $x_n, y_n$ such that $d(x_n, y_n) < \frac{1}{n}$ but $d'(f(x_n), f(y_n)) > \varepsilon$.

By compactness of $X$, $(x_n)$ has a convergent subsequence $(x_{n_i}) \to x$. Then we also have $y_{n_i} \to x$. So by continuity, we must have $f(x_{n_i}) \to f(x)$ and $f(y_{n_i}) \to f(x)$. But $d'(f(x_{n_i}), f(y_{n_i})) > \varepsilon$ for all $n_i$. This is a contradiction. □

Remember what this theorem was in $\mathbb{R}$? If $f$ is in a closed and bounded interval, then continuity implies uniform continuity. Now we prove the continuity equivalence:

**Theorem.** Let $(X, d)$ and $(X', d')$ be metric spaces, and $f : X \to X'$. Then the following are equivalent:

(i) $f$ is continuous at $y$.

(ii) $f(x_k) \to f(y)$ for every sequence $(x_k)$ in $X$ with $x_k \to y$.

(iii) For every neighbourhood $V$ of $f(y)$, there is a neighbourhood $U$ of $y$ such that $U \subseteq f^{-1}(V)$.

*Proof.*

- (i) ⇔ (ii): This is the same as normed spaces.
- (i) ⇒ (iii): Let $V$ be a neighbourhood of $f(y)$. Then by definition there is $\varepsilon > 0$ such that $B_\varepsilon(f(y)) \subseteq V$. By continuity of $f$, there is some $\delta$ such that

  $$B_\delta(y) \subseteq f^{-1}(B_\varepsilon(f(y))) \subseteq f^{-1}(V).$$

  Set $U = B_\varepsilon(y)$ and done.
- (iii) ⇒ (i): for any $\varepsilon$, use the hypothesis with $V = B_\varepsilon(f(y))$ to get a neighbourhood $U$ of $y$ such that

  $$U \subseteq f^{-1}(V) = f^{-1}(B_\varepsilon(f(y))).$$

  Since $U$ is open, there is some $\delta$ such that $B_\delta(y) \subseteq U$. So we get

  $$B_\delta(y) \subseteq f^{-1}(B_\varepsilon(f(y))).$$

  So we get continuity.

□

# 6 The Contraction Mapping Theorem, aka Not Met and Top

## 6.1 Contraction Mapping Theorem

Now we finally finish the introduction to Met and Top and come back to Analysis. This section is EXTREMELY important and has been tested multiple times. make sure you understand what this is really talking about.

**Definition** (Contraction mapping)**.** Let $(X, d)$ be metric space. A mapping $f : X \to X$ is a *contraction* if there exists some $\lambda$ with $0 \leq \lambda < 1$ such that

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

**Theorem** (Contraction mapping theorem)**.** Let $X$ be a (non-empty) complete metric space, and if $f : X \to X$ is a contraction, then $f$ has a *unique* fixed point, ie. there is a unique $x$ such that $f(x) = x$.

Moreover, if $f : X \to X$ is a function such that $f^{(m)} : X \to X$ (ie. $f$ composed with itself $m$ times) is a contraction for some $m$, then $f$ has a unique fixed point.

**Note.** This theorem is maximal in the sense that no condition can be relaxed. It is not enough to assume $d(f(x)), f(y)) < d(x, y)$ for all $x, y$. Also, completeness is necessary (otherwise functions like $\frac{x}{2}$ just work on $(0, 1)$).

Also, the proof of this is not hard to understand. Existence is established by reapplying $f$ over and over, as you will see $\lambda < 1$ is very important.

*Proof.* Uniqueness is straightforward. By assumption, there is some $0 \leq \lambda < 1$ such that

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

for all $x, y \in X$. If $x$ and $y$ are both fixed points, then this says

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

This is possible only if $d(x, y) = 0$, ie. $x = y$.

To prove existence, the idea is to pick a point $x_0$ and keep applying $f$. Let $x_0 \in X$. We define the sequence $(x_n)$ inductively by

$$x_{n+1} = f(x_n).$$

We first show that this is Cauchy. For any $n \geq 1$, we can compute

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

Since this is true for any $n$, for $m > n$, we have

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \cdots + d(x_{n+1}, x_n) \\ &= \sum_{j=n}^{m-1} d(x_{j+1}, x_j) \\ &= \sum_{j=n}^{m-1} \lambda^j d(x_1, x_0) \leq d(x_1, x_0) \sum_{j=n}^{\infty} \lambda^j = \frac{\lambda^n}{1-\lambda} d(x_1, x_0). \end{aligned}$$

Note that we have again used the property that $\lambda < 1$.

This implies $d(x_m, x_n) \to 0$ as $m, n \to \infty$. So this sequence is Cauchy. By the completeness of $X$, there exists some $x \in X$ such that $x_n \to x$. Since $f$ is a contraction, it is continuous. So $f(x_n) \to f(x)$. However, by definition $f(x_n) = x_{n+1}$. So taking the limit on both sides, we get $f(x) = x$. So $x$ is a fixed point.

Now suppose that $f^{(m)}$ is a contraction for some $m$. Hence by the first part, there is a unique $x \in X$ such that $f^{(m)}(x) = x$. But then

$$f^{(m)}(f(x)) = f^{(m+1)}(x) = f(f^{(m)}(x)) = f(x).$$

So $f(x)$ is also a fixed point of $f^{(n)}(x)$. By uniqueness of fixed points, we must have $f(x) = x$. Since any fixed point of $f$ is clearly a fixed point of $f^{(n)}$ as well, it follows that $x$ is the unique fixed point of $f$. □

## 6.2 Picard-Lindelöf Existence Theorem

Now we introduce a theorem that *seems* to be unrelated:

**Notation.** For $\mathbf{x}_0 \in \mathbb{R}^n$, $R > 0$, we let

$$\overline{B_R(\mathbf{x}_0)} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\|_2 \leq R\}.$$

Then the theorem says

**Theorem** (Picard-Lindelöf existence theorem)**.** Let $\mathbf{x}_0 \in \mathbb{R}^n$, $R > 0$, $a < b$, $t_0 \in [a, b]$. Let $\mathbf{F} : [a, b] \times \overline{B_R(\mathbf{x}_0)} \to \mathbb{R}^n$ be a continuous function satisfying

$$\|\mathbf{F}(t, \mathbf{x}) - \mathbf{F}(t, \mathbf{y})\|_2 \leq \kappa \|\mathbf{x} - \mathbf{y}\|_2$$

for some fixed $\kappa > 0$ and all $t \in [a, b], \mathbf{x} \in \overline{B_R(\mathbf{x}_0)}$. In other words, $F(t, \cdot) : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz on $\overline{B_R(\mathbf{x}_0)}$ with the same Lipschitz constant for every $t$. Then

(i) There exists an $\varepsilon > 0$ and a unique differentiable function $\mathbf{f} : [t_0 - \varepsilon, t_0 + \varepsilon] \cap [a, b] \to \mathbb{R}^n$ such that

$$\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}t} = \mathbf{F}(t, \mathbf{f}(t)) \qquad (*)$$

and $\mathbf{f}(t_0) = \mathbf{x}_0$.

(ii) If

$$\sup_{[a,b] \times \overline{B_R(\mathbf{x}_0)}} \|\mathbf{F}\|_2 \leq \frac{R}{b - a},$$

then there exists a unique differential function $\mathbf{f} : [a, b] \to \mathbb{R}^n$ that satisfies the differential equation and boundary conditions above.

Wait, what? What is this? Here is an (attempted) breakdown of what it is actually saying:

**Note.**

- Now consider a first order differential equation $\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}t} = \mathbf{F}(t, \mathbf{f}(t))$. We want to know if there are solutions.
- Now we want a local solution knowing that the initial condition is $\mathbf{f}(t_0) = x_0$. Then, if this differential function satisfies $\|\mathbf{F}(t, \mathbf{x}) - \mathbf{F}(t, \mathbf{y})\|_2 \leq \kappa \|\mathbf{x} - \mathbf{y}\|_2$, which is just a Lipschitz relation, then we have some cool results. NOTE this has to be satisfied by the *same* $\kappa$ by *all* $t$.

- What are the cool results? Well (i) tells us a local solution exists for some closed ball around $t_0$. $(ii)$ tells us if some further condition is satisfied, we have a *unique* solution on the whole space we want.

Now we prove it.

*Proof.* First, note that (ii) implies (i). We know that

$$\sup_{[a,b]\times\overline{B_R(\mathbf{x})}} \|\mathbf{F}\|$$

is bounded since it is a continuous function on a compact domain. So we can pick an $\varepsilon$ such that

$$2\varepsilon \leq \frac{R}{\sup_{[a,b]\times\overline{B_R(\mathbf{x})}} \|\mathbf{F}\|}.$$

Then writing $[t_0 - \varepsilon, t_0 + \varepsilon] \cap [a, b] = [a_1, b_1]$, we have

$$\sup_{[a_1,b_1]\times\overline{B_R(\mathbf{x})}} \|\mathbf{F}\| \leq \sup_{[a,b]\times\overline{B_R(\mathbf{x})}} \|\mathbf{F}\| \leq \frac{R}{2\varepsilon} \leq \frac{R}{b_1 - a_1}.$$

So (ii) implies there is a solution on $[t_0 - \varepsilon, t_0 + \varepsilon] \cap [a, b]$. Hence it suffices to prove (ii).

To apply the contraction mapping theorem, we need to convert this into a fixed point problem. The key is to reformulate the problem as an integral equation. We know that a differentiable $\mathbf{f} : [a, b] \to \mathbb{R}^n$ satisfies the differential equation $(*)$ if and only if $\mathbf{f} : [a, b] \to \overline{B_R(\mathbf{x}_0)}$ is continuous and satisfies

$$\mathbf{f}(t) = \mathbf{x}_0 + \int_{t_0}^{t} \mathbf{F}(s, \mathbf{f}(s))\ \mathrm{d}s$$

by the fundamental theorem of calculus. Note that we don't require **f** is differentiable, since if a continuous **f** satisfies this equation, it is automatically differentiable by the fundamental theorem of calculus. This is very helpful, since we can work over the much larger vector space of continuous functions, and it would be easier to find a solution.

We let $X = C([a, b], \overline{B_R(\mathbf{x}_0)})$. We equip $X$ with the supremum metric

$$\|\mathbf{g} - \mathbf{h}\| = \sup_{t\in[a,b]} \|\mathbf{g}(t) - \mathbf{h}(t)\|_2.$$

We see that $X$ is a closed subset of the complete metric space $C([a, b], \mathbb{R}^n)$ (again taken with the supremum metric). So $X$ is complete. For every $\mathbf{g} \in X$, we define a function $T\mathbf{g} : [a, b] \to \mathbb{R}^n$ by

$$(T\mathbf{g})(t) = \mathbf{x}_0 + \int_{t_0}^{t} \mathbf{F}(s, \mathbf{g}(s))\ \mathrm{d}s.$$

Our differential equation is thus

$$\mathbf{f} = T\mathbf{f}.$$

So we first want to show that $T$ is actually mapping $X \to X$, ie. $T\mathbf{g} \in X$ whenever $\mathbf{g} \in X$, and then prove it is a contraction map.

We have

$$\begin{aligned}\|T\mathbf{g}(t) - \mathbf{x}_0\|_2 &= \left\|\int_{t_0}^t \mathbf{F}(s, \mathbf{g}(s))\;\mathrm{d}s\right\|\\ &\leq \left|\int_{t_0}^t \|\mathbf{F}(s, \mathbf{g}(s))\|_2\;\mathrm{d}s\right|\\ &\leq \sup_{[a,b]\times \overline{B_R(\mathbf{x}_0)}} \|\mathbf{F}\| \cdot |b - a|\\ &\leq R\end{aligned}$$

Hence we know that $T\mathbf{g}(t) \in \overline{B}_R(\mathbf{x}_0)$. So $T\mathbf{g} \in X$.

Next, we need to show this is a contraction. However, it turns out $T$ need not be a contraction. Instead, what we have is that for $\mathbf{g}_1, \mathbf{g}_2 \in X$, we have

$$\begin{aligned}\|T\mathbf{g}_1(t) - T\mathbf{g}_2(t)\|_2 &= \left\|\int_{t_0}^t \mathbf{F}(s, \mathbf{g}_1(s)) - \mathbf{F}(s, \mathbf{g}_2(s))\;\mathrm{d}s\right\|_2\\ &\leq \left|\int_{t_0}^t \|\mathbf{F}(s, \mathbf{g}_1(s)) - \mathbf{F}(s, \mathbf{g}_2(s))\|_2\;\mathrm{d}s\right|\\ &\leq \kappa(b - a)\|\mathbf{g}_1 - \mathbf{g}_2\|_\infty\end{aligned}$$

by the Lipschitz condition on $F$. If we indeed have

$$\kappa(b - a) < 1, \tag{†}$$

then the contraction mapping theorem gives an $f \in X$ such that

$$T\mathbf{f} = \mathbf{f},$$

ie.

$$\mathbf{f} = \mathbf{x}_0 + \int_{t_0}^t \mathbf{F}(s, \mathbf{f}(s))\;\mathrm{d}s.$$

However, we do not necessarily have (†). There are many ways we can solve this problem. Here, we can solve it by finding an $m$ such that $T^{(m)} = T \circ T \circ \cdots \circ T : X \to X$ is a contraction map. We will in fact show that this map satisfies the bound

$$\sup_{t\in[a,b]} \|T^{(m)}\mathbf{g}_1(t) - T^{(m)}\mathbf{g}_2(t)\| \leq \frac{(b - a)^m\kappa^m}{m!} \sup_{t\in[a,b]} \|\mathbf{g}_1(t) - \mathbf{g}_2(t)\|. \tag{‡}$$

The key is the $m!$, since this grows much faster than any exponential. Given this bound, we know that for sufficiently large $m$, we have

$$\frac{(b - a)^m\kappa^m}{m!} < 1,$$

ie. $T^{(m)}$ is a contraction. So by the contraction mapping theorem, the result holds.

So it only remains to prove the bound. To prove this, we prove instead the pointwise bound: for any $t \in [a, b]$, we have

$$\|T^{(m)}\mathbf{g}_1(t) - T^{(m)}\mathbf{g}_2(t)\|_2 \leq \frac{(|t - t_0|)^m\kappa^m}{m!} \sup_{s\in[t_0,t]} \|\mathbf{g}_1(s) - \mathbf{g}_2(s)\|.$$

From this, taking the supremum on the left, we obtain the bound (‡).

To prove this pointwise bound, we induct on $m$. We wlog assume $t > t_0$. We know that for every $m$, the difference is given by

$$\begin{aligned}\|T^{(m)}g_1(t) - T^{(m)}g_2(t)\|_2 &= \left\| \int_{t_0}^{t} F(s, T^{(m-1)}g_1(s)) - F(s, T^{(m-1)}g_2(s))\ \mathrm{d}s \right\|_2. \\ &\leq \kappa \int_{t_0}^{t} \|T^{(m-1)}g_1(s) - T^{(m-1)}g_2(s)\|_2\ \mathrm{d}s.\end{aligned}$$

This is true for all $m$. If $m = 1$, then this gives

$$\|Tg_1(t) - Tg_2(t)\| \leq \kappa(t - t_0) \sup_{[t_0, t]} \|g_1 - g_2\|_2.$$

So the base case is done.

For $m \geq 2$, assume by induction the bound holds with $m - 1$ in place of $m$. Then the bounds give

$$\begin{aligned}\|T^{(m)}g_1(t) - T^{(m)}g_2(t)\| &\leq \kappa \int_{t_0}^{t} \frac{k^{m-1}(s - t_0)^{m-1}}{(m-1)!} \sup_{[t_0, s]} \|g_1 - g_2\|_2\ \mathrm{d}s \\ &\leq \frac{\kappa^m}{(m-1)!} \sup_{[t_0, t]} \|g_1 - g_2\|_2 \int_{t_0}^{t} (s - t_0)^{m-1}\ \mathrm{d}s \\ &= \frac{\kappa^m (t - t_0)^m}{m!} \sup_{[t_0, t]} \|g_1 - g_2\|_2.\end{aligned}$$

So done. □

Phew. A summary of the tactic used and some notes:

**Note.**

- The clever thing is to rearrange this into a integral function of **f**. Then we prove it is a valid mapping, then a contraction. Then we apply the fixed-point theorem to get a solution.
- Note that to get the factor of $m!$ to ptove that it is a contraction, we had to actually perform the integral. In general, this is a good strategy if we want tight bounds. Instead of bounding

  $$\left| \int_a^b f(x)\ \mathrm{d}x \right| \leq (b - a) \sup |f(x)|,$$

  we write $f(x) = g(x)h(x)$, where $h(x)$ is something easily integrable. Then we can have a bound

  $$\left| \int_a^b f(x)\ \mathrm{d}x \right| \leq \sup |g(x)| \int_a^b |h(x)|\ \mathrm{d}x.$$

# 7 Differentiation from $\mathbb{R}^m$ to $\mathbb{R}^n$

This section basically discusses how to extend differentiation into vector-valued functions. Here are some definitions:

## 7.1 Definitions

**Definition** (Limit of function)**.** Let $E \subseteq \mathbb{R}^n$ and $f : E \to \mathbb{R}^m$. Let $\mathbf{a} \in \mathbb{R}^n$ be a limit point of $E$, and let $\mathbf{b} \in \mathbb{R}^m$. We say

$$\lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = \mathbf{b}$$

if for every $\varepsilon > 0$, there is some $\delta > 0$ such that

$$(\forall \mathbf{x} \in E)\ 0 < \|\mathbf{x} - \mathbf{a}\| < \delta \Rightarrow \|f(\mathbf{x}) - \mathbf{b}\| < \varepsilon.$$

**Definition** (Differentiation in $\mathbb{R}^n$)**.** Let $U \subseteq \mathbb{R}^n$ be open, $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$. We say $\mathbf{f}$ is differentiable at a point $\mathbf{a} \in U$ if there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\mathbf{f}(\mathbf{a} + \mathbf{h}) - \mathbf{f}(\mathbf{a}) - A\mathbf{h}}{\|\mathbf{h}\|} = 0.$$

We call $A$ the *derivative of* $\mathbf{f}$ *at* $a$. We write the derivative as $D\mathbf{f}(\mathbf{a})$.

**Note.** Note that we have moved the usual "derivative" $A$ to the left hand side. This is to prevent a modulus issue when dividing by the modulus of $h$. If we didn't move it, the equation does not make sense in $n = m = 1$.

**Proposition** (Uniqueness of derivative)**.** Derivatives are unique.

*Proof.* Suppose $A, B : \mathbb{R}^n \to \mathbb{R}^m$ both satisfy the condition

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\mathbf{f}(\mathbf{a} + \mathbf{h}) - \mathbf{f}(\mathbf{a}) - A\mathbf{h}}{\|\mathbf{h}\|} = \mathbf{0}$$
$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\mathbf{f}(\mathbf{a} + \mathbf{h}) - \mathbf{f}(\mathbf{a}) - B\mathbf{h}}{\|\mathbf{h}\|} = \mathbf{0}.$$

By the triangle inequality, we get

$$\|(B - A)\mathbf{h}\| \leq \|\mathbf{f}(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}\| + \|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - B\mathbf{h}\|.$$

So

$$\frac{\|(B - A)\mathbf{h}\|}{\|\mathbf{h}\|} \to 0$$

as $h \to 0$. We set $\mathbf{h} = t\mathbf{u}$ in this proof to get

$$\frac{\|(B - A)t\mathbf{u}\|}{\|t\mathbf{u}\|} \to 0$$

as $t \to 0$. Since $(B - A)$ is linear, we know

$$\frac{\|(B - A)t\mathbf{u}\|}{\|t\mathbf{u}\|} = \frac{\|(B - A)\mathbf{u}\|}{\|\mathbf{u}\|}.$$

So $(B - A)\mathbf{u} = \mathbf{0}$ for all $\mathbf{u} \in \mathbb{R}^n$. So $B = A$. ☐

Ok, so we know that derivatives exist and they seem to be well-defined. But how do I even calculate this? The thing is that we have to look at directional derivatives, the derivative going in a specific direction (because otherwise the value is not well-defined):

**Definition** (Directional derivative)**.** We write

$$D_{\mathbf{u}}\mathbf{f}(\mathbf{a}) = \lim_{t\to 0} \frac{\mathbf{f}(\mathbf{a}+t\mathbf{u}) - \mathbf{f}(\mathbf{a})}{t} = \frac{\mathrm{d}}{\mathrm{d}t}||_{t=0}\mathbf{f}(\mathbf{a}+t\mathbf{u}).$$

whenever this limit exists. We call $D_{\mathbf{u}}\mathbf{f}(\mathbf{a})$ the *directional derivative* of $\mathbf{f}$ at $\mathbf{a} \in U$ in the direction of $\mathbf{u} \in \mathbb{R}^n$.

Now partial derivatives are just special cases of the directional derivative:

**Definition** (Partial derivative)**.** The *$j$th partial derivative* of $f : U \to \mathbb{R}$ at $\mathbf{a} \in U$ is

$$D_{\mathbf{e}_j} f(\mathbf{a}) = \lim_{t\to\infty} \frac{f(\mathbf{a}+t\mathbf{e}_j) - f(\mathbf{a})}{t},$$

when the limit exists. We often write this as

$$D_{\mathbf{e}_j} f(\mathbf{a}) = D_j f(\mathbf{a}) = \frac{\partial f}{\partial x_j}.$$

## 7.2 Properties for Derivatives

Everything here is just really proving that "common sense" still applies. But note that $(ii)$ is particularly useful, as we can consider functions sent to real numbers only.

**Proposition.** Let $U \subseteq \mathbb{R}^n$ be open, $\mathbf{a} \in U$.

(i) If $\mathbf{f} : U \to \mathbb{R}^m$ is differentiable at $\mathbf{a}$, then $\mathbf{f}$ is continuous at $\mathbf{a}$.

(ii) If we write $\mathbf{f} = (f_1, f_2, \cdots, f_n) : U \to \mathbb{R}^m$, where each $f_i : U \to \mathbb{R}$, then $\mathbf{f}$ is differentiable at $\mathbf{a}$ if and only if each $f_j$ is differentiable at $\mathbf{a}$ for each $j$.

(iii) If $f, g : U \to \mathbb{R}^m$ are both differentiable at $\mathbf{a}$, then $\lambda\mathbf{f} + \mu\mathbf{g}$ is differentiable at $\mathbf{a}$ with

$$D(\lambda\mathbf{f} + \mu\mathbf{g})(\mathbf{a}) = \lambda D\mathbf{f}(\mathbf{a}) + \mu D\mathbf{g}(\mathbf{a}).$$

(iv) If $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, then $A$ is differentiable for any $\mathbf{a} \in \mathbb{R}^n$ with

$$DA(\mathbf{a}) = A.$$

(v) If $\mathbf{f}$ is differentiable at $\mathbf{a}$, then the directional derivative $D_{\mathbf{u}}\mathbf{f}(\mathbf{a})$ exists for all $\mathbf{u} \in \mathbb{R}^n$, and in fact

$$D_{\mathbf{u}}\mathbf{f}(\mathbf{a}) = D\mathbf{f}(\mathbf{a})\mathbf{u}.$$

(vi) If $A = (A_{ij})$ be the matrix representing $D\mathbf{f}(\mathbf{a})$ with respect to the standard basis for $\mathbb{R}^n$ and $\mathbb{R}^m$, ie. for any $\mathbf{h} \in \mathbb{R}^n$,

$$D\mathbf{f}(\mathbf{a})\mathbf{h} = A\mathbf{h}.$$

Then $A$ is given by

$$A_{ij} = \langle D\mathbf{f}(\mathbf{a})\mathbf{e}_j, \mathbf{b}_i\rangle = D_j\mathbf{f}_i(a).$$

where $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$, and $\{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$ is the standard basis for $\mathbb{R}^m$.

*Proof.*

(i) By definition, if $\mathbf{f}$ is differentiable, then as $\mathbf{h} \to \mathbf{0}$, we know

$$\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})\mathbf{h} \to \mathbf{0}.$$

Since $D\mathbf{f}(\mathbf{a})\mathbf{h} \to \mathbf{0}$ as well, we must have $\mathbf{f}(\mathbf{a}+\mathbf{h}) \to \mathbf{f}(\mathbf{h})$.

(ii) Expand the definition.

(iii) We just have to check this directly. We have

$$\begin{aligned}&\frac{(\lambda\mathbf{f}+\mu\mathbf{g})(\mathbf{a}+\mathbf{h}) - (\lambda\mathbf{f}+\mu\mathbf{g})(\mathbf{a}) - (\lambda D\mathbf{f}(\mathbf{a}) + \mu D\mathbf{g}(\mathbf{a}))}{\|\mathbf{h}\|}\\ &= \lambda\frac{\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})\mathbf{h}}{\|\mathbf{h}\|} + \mu\frac{\mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a}) - D\mathbf{g}(\mathbf{a})\mathbf{h}}{\|\mathbf{h}\|}.\end{aligned}$$

which tends to $0$ as $\mathbf{h} \to \mathbf{0}$. So done.

(iv) Since $A$ is linear, we always have $A(\mathbf{a}+\mathbf{h}) - A(\mathbf{a}) - A\mathbf{h} = \mathbf{0}$ for all $\mathbf{h}$.

(v) Compare the two forms.

(vi) This follows from the general result for linear maps: for any linear map represented by $(A_{ij})_{m\times n}$, we have

$$A_{ij} = \langle A\mathbf{e}_j, \mathbf{e}_i\rangle.$$

Applying this with $A = D\mathbf{f}(\mathbf{a})$ and note that for any $\mathbf{h} \in \mathbb{R}^n$,

$$D\mathbf{f}(\mathbf{a})\mathbf{h} = (D\mathbf{f}_1(\mathbf{a})\mathbf{h}, \cdots, D\mathbf{f}_m(\mathbf{a})\mathbf{h}).$$

So done.

□

**Note.** Important! The existence of partial/directional derivatives at a point, even in all directions, does NOT guarantee the existence of a derivative at the point. Here is an example:

**Example.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x,y) = \begin{cases} \frac{x^3}{y} & y \neq 0 \\ 0 & y = 0 \end{cases}$$

Then for $\mathbf{u} = (u_1, u_2) \neq \mathbf{0}$ and $t \neq 0$, we can compute

$$\frac{f(\mathbf{0}+t\mathbf{u}) - f(\mathbf{0})}{t} = \begin{cases} \frac{tu_1^3}{u_2} & u_2 \neq 0 \\ 0 & u_2 = 0 \end{cases}$$

So

$$D_{\mathbf{u}}f(\mathbf{0}) = \lim_{t\to 0}\frac{f(\mathbf{0}+t\mathbf{u}) - f(\mathbf{0})}{t} = 0,$$

and the directional derivative exists. However, the function is not differentiable at $0$, since it is not even continuous at $0$, as

$$f(\delta, \delta^4) = \frac{1}{\delta}$$

diverges as $\delta \to 0$.

This is because you need to consider ALL directions for a derivative. The concept is similar to $|x|$, where derivatives exist as you approach 0 from left and right, but at 0, there is no derivative.

**Theorem.** Let $U \subseteq \mathbb{R}^n$ be open, $\mathbf{f} : U \to \mathbb{R}^m$. Let $\mathbf{a} \in U$. Suppose there exists some open ball $B_r(\mathbf{a}) \subseteq U$ such that

(i) $D_j \mathbf{f}_i(\mathbf{x})$ exists for every $\mathbf{x} \in B_r(\mathbf{a})$ and $1 \leq i \leq m, j \leq 1 \leq n$

(ii) $D_j \mathbf{f}_i$ are continuous at $\mathbf{a}$ for all $1 \leq i \leq m, j \leq 1 \leq n$.

Then $\mathbf{f}$ is differentiable at $\mathbf{a}$.

*Proof.* It suffices to prove for $m = 1$, as by $(ii)$ above, we know that we only need to consider functions that send to real numbers. For each $\mathbf{h} = (h_1, \cdots, h_n) \in \mathbb{R}^n$, we have

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \sum_{j=1}^{n} f(\mathbf{a} + h_1\mathbf{e}_1 + \cdots + h_j\mathbf{e}_j) - f(\mathbf{a} + h_1\mathbf{e}_1 + \cdots + h_{j-1}\mathbf{e}_{j-1}).$$

Now for convenience, we can write

$$\mathbf{h}^{(j)} = h_1\mathbf{e}_1 + \cdots + h_j\mathbf{e}_j = (h_1, \cdots, h_j, 0, \cdots, 0).$$

Then we have

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) &= \sum_{j=1}^{n} f(\mathbf{a} + \mathbf{h}^{(j)}) - f(a + \mathbf{h}^{(j-1)}) \\ &= \sum_{j=1}^{n} f(\mathbf{a} + \mathbf{h}^{(j-1)} + h_j\mathbf{e}_j) - f(\mathbf{a} + \mathbf{h}^{(j-1)}). \end{aligned}$$

Note that in each term, we are just moving along the coordinate axes. Since the partial derivatives exist, the mean value theorem of single-variable calculus applied to

$$g(t) = f(\mathbf{a} + \mathbf{h}^{(j-1)} + t\mathbf{e}_j)$$

on the interval $t \in [0, h_j]$ allows us to write this as

$$\begin{aligned} &f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) \\ &= \sum_{j=1}^{n} h_j D_j f(\mathbf{a} + \mathbf{h}^{(j-1)} + \theta_j h_j \mathbf{e}_j) \\ &= \sum_{j=1}^{n} h_j D_j f(\mathbf{a}) + \sum_{j=1}^{n} h_j \Big(D_j f(\mathbf{a} + \mathbf{h}^{(j-1)} + \theta_j h_j \mathbf{e}_j) - D_j f(\mathbf{a})\Big) \end{aligned}$$

for some $\theta_j \in (0, 1)$.

Note that $D_j f(\mathbf{a} + \mathbf{h}^{(j-1)} + \theta_j h_j \mathbf{e}_j) - D_j f(\mathbf{a}) \to 0$ as $\mathbf{h} \to 0$ since the partial derivatives are continuous at $\mathbf{a}$. So the second term is $o(\mathbf{h})$. So $f$ is differentiable at $\mathbf{a}$ with

$$Df(\mathbf{a})\mathbf{h} = \sum_{j=1}^{n} D_j f(\mathbf{a}) h_j.$$

□

Therefore, if we can prove it has *continuous* partial derivatives, then it is differentiable.

## 7.3 Operator Norm

What? I thought we left Met and Top. Turns out we are back.

**Definition** (Operator norm)**.** The *operator norm* on $\mathcal{L} = L(\mathbb{R}^n; \mathbb{R}^m)$ is defined by

$$\|A\| = \sup_{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1} \|A\mathbf{x}\|.$$

Now we can check it is actually a (finite) norm. And it has some interesting properties, listed below:

**Proposition.**

(i)
$$\|A\| = \sup_{\mathbb{R}^n \setminus \{0\}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}.$$

Equivalently, $\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$.

(ii) Let $A \in L(\mathbb{R}^n; \mathbb{R}^m)$ and $B \in L(\mathbb{R}^m; \mathbb{R}^p)$. Then $BA = B \circ A \in L(\mathbb{R}^n; \mathbb{R}^p)$ and

$$\|BA\| \leq \|B\|\|A\|.$$

*Proof.*

(i) This follows from linearity of $A$, and for any $\mathbf{x} \in \mathbb{R}^n$, we have

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

(ii)
$$\|BA\| = \sup_{\mathbb{R}^n \setminus \{0\}} \frac{\|BA\mathbf{x}\|}{\|\mathbf{x}\|} \leq \sup_{\mathbb{R}^n \setminus \{0\}} \frac{\|B\|\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \|B\|\|A\|.$$

□

Now why do we pick this norm? Since every norm is Lipschitz equivalent, we just pick the one that is helpful to prove some welcomed results, like the chain rule and mean value theorem :

**Theorem** (Chain rule)**.** Let $U \subseteq \mathbb{R}^n$ be open, $\mathbf{a} \in U$, $\mathbf{f} : U \to \mathbb{R}^m$ differentiable at $\mathbf{a}$. Moreover, $V \subseteq \mathbb{R}^m$ is open with $\mathbf{f}(U) \subseteq V$ and $\mathbf{g} : V \to \mathbb{R}^p$ is differentiable at $\mathbf{f}(\mathbf{a})$. Then $\mathbf{g} \circ \mathbf{f} : U \to \mathbb{R}^p$ is differentiable at $\mathbf{a}$, with derivative

$$D(\mathbf{g} \circ \mathbf{f})(\mathbf{a}) = D\mathbf{g}(\mathbf{f}(\mathbf{a}))\ D\mathbf{f}(\mathbf{a}).$$

*Proof.* Let $A = D\mathbf{f}(\mathbf{a})$ and $B = D\mathbf{g}(\mathbf{f}(\mathbf{a}))$. By differentiability of $\mathbf{f}$, we know

$$\begin{aligned} \mathbf{f}(\mathbf{a} + \mathbf{h}) &= \mathbf{f}(\mathbf{a}) + A\mathbf{h} + o(\mathbf{h}) \\ \mathbf{g}(\mathbf{f}(\mathbf{a}) + \mathbf{k}) &= \mathbf{g}(\mathbf{f}(\mathbf{a})) + B\mathbf{k} + o(\mathbf{k}) \end{aligned}$$

Now we have

$$\begin{aligned}\mathbf{g}\circ\mathbf{f}(\mathbf{a}+\mathbf{h}) &= \mathbf{g}(\mathbf{f}(\mathbf{a}) + \underbrace{A\mathbf{h}+o(\mathbf{h})}_{\mathbf{k}})\\ &= \mathbf{g}(\mathbf{f}(\mathbf{a})) + B(A\mathbf{h}+o(\mathbf{h})) + o(A\mathbf{h}+o(\mathbf{h}))\\ &= \mathbf{g}\circ\mathbf{f}(\mathbf{a}) + BA\mathbf{h} + B(o(\mathbf{h})) + o(A\mathbf{h}+o(\mathbf{h})).\end{aligned}$$

We just have to show the last term is $o(\mathbf{h})$, but this is true since $B$ and $A$ are bounded. By boundedness,

$$\|B(o(\mathbf{h}))\| \leq \|B\|\|o(\mathbf{h})\|.$$

So $B(o(\mathbf{h})) = o(\mathbf{h})$. Similarly,

$$\|A\mathbf{h}+o(\mathbf{h})\| \leq \|A\|\|\mathbf{h}\| + \|o(\mathbf{h})\| \leq (\|A\|+1)\|\mathbf{h}\|$$

for sufficiently small $\|\mathbf{h}\|$. So $o(A\mathbf{h}+o(\mathbf{h}))$ is in fact $o(\mathbf{h})$ as well. Hence

$$\mathbf{g}\circ\mathbf{f}(\mathbf{a}+\mathbf{h}) = \mathbf{g}\circ\mathbf{f}(\mathbf{a}) + BA\mathbf{h} + o(\mathbf{h}).$$

□

### 7.3.1 Mean Value Theorem

To prove the MVT in multiple dimensions, we first consider the case where the domain is just a subset of $\mathbb{R}$. Then we generalize our proof. This way it is easier to see the argument.

**Theorem.** Let $\mathbf{f}: [a,b] \to \mathbb{R}^m$ be continuous on $[a,b]$ and differentiable on $(a,b)$. Suppose we can find some $M$ such that for all $t \in (a,b)$, we have $\|D\mathbf{f}(t)\| \leq M$. Then

$$\|\mathbf{f}(b) - \mathbf{f}(a)\| \leq M(b-a).$$

*Proof.* Let $\mathbf{v} = \mathbf{f}(b) - \mathbf{f}(a)$. We define

$$g(t) = \mathbf{v}\cdot\mathbf{f}(t) = \sum_{i=1}^m v_i f_i(t).$$

Since each $f_i$ is differentiable, $g$ is continuous on $[a,b]$ and differentiable on $(a,b)$ with

$$g'(t) = \sum v_i f_i'(t).$$

Hence, we know

$$|g'(t)| \leq \left|\sum_{i=1}^m v_i f_i'(t)\right| \leq \|\mathbf{v}\| \left(\sum_{i=1}^n f_i'^2(t)\right)^{1/2} = \|\mathbf{v}\|\|D\mathbf{f}(t)\| \leq M\|\mathbf{v}\|.$$

We now apply the mean value theorem to $g$ to get

$$g(b) - g(a) = g'(t)(b-a)$$

for some $t \in (a,b)$. By definition of $g$, we get

$$\mathbf{v}\cdot(\mathbf{f}(b) - \mathbf{f}(a)) = g'(t)(b-a).$$

By definition of $\mathbf{v}$, we have

$$\|\mathbf{f}(b) - \mathbf{f}(a)\|^2 = |g'(t)(b-a)| \leq (b-a)M\|\mathbf{f}(b) - \mathbf{f}(a)\|.$$

If $\mathbf{f}(b) = \mathbf{f}(a)$, then there is nothing to prove. Otherwise, divide by $\|\mathbf{f}(b) - \mathbf{f}(a)\|$ and done. □

We now apply this to prove the general version.

**Theorem** (Mean value inequality). Let $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{f} : B_r(\mathbf{a}) \to \mathbb{R}^m$ be differentiable on $B_r(\mathbf{a})$ with $\|D\mathbf{f}(\mathbf{x})\| \leq M$ for all $\mathbf{x} \in B_r(\mathbf{a})$. Then

$$\|\mathbf{f}(\mathbf{b}_1) - f(\mathbf{b}_2)\| \leq M\|\mathbf{b}_1 - \mathbf{b}_2\|$$

for any $\mathbf{b}_1, \mathbf{b}_2 \in B_r(\mathbf{a})$.

*Proof.* We will reduce this to the previous theorem.

Fix $\mathbf{b}_1, \mathbf{b}_2 \in B_r(\mathbf{a})$. Note that

$$t\mathbf{b}_1 + (1-t)\mathbf{b}_2 \in B_r(\mathbf{a})$$

for all $t \in [0,1]$. Now consider $\mathbf{g} : [0,1] \to \mathbb{R}^m$.

$$\mathbf{g}(t) = \mathbf{f}(t\mathbf{b}_1 + (1-t)\mathbf{b}_2).$$

By the chain rule, $\mathbf{g}$ is differentiable and

$$\mathbf{g}'(t) = D\mathbf{g}(t) = (D\mathbf{f}(t\mathbf{b}_1 + (1-t)\mathbf{b}_2))(\mathbf{b}_1 - \mathbf{b}_2)$$

Therefore

$$\|D\mathbf{g}(t)\| \leq \|D\mathbf{f}(t\mathbf{b}_1 + (1-t)\mathbf{b}_2)\|\|\mathbf{b}_1 - \mathbf{b}_2\| \leq M\|\mathbf{b}_1 - \mathbf{b}_2\|.$$

Now using the previous theorem with $\mathbf{g}(t)$, we have:

$$\|\mathbf{g}(1) - \mathbf{g}(0)\| \leq M\|\mathbf{b}_1 - \mathbf{b}_2\|.$$

But $\|\mathbf{g}(1) - \mathbf{g}(0)\| = \|\mathbf{f}(\mathbf{b}_1) - \mathbf{f}(\mathbf{b}_2)\|$, so we are done. □

**Definition** (Path-connected subset). A subset $E \subseteq \mathbb{R}^n$ is *path-connected* if for any $\mathbf{a}, \mathbf{b} \in E$, there is a continuous map $\gamma : [0,1] \to E$ such that

$$\gamma(0) = \mathbf{a}, \quad \gamma(1) = \mathbf{b}.$$

The following theorem extends the Mean value theorem in the case that the derivative is always 0.

**Theorem.** Let $U \subseteq \mathbb{R}^m$ be open and path-connected. Then for any differentiable $\mathbf{f} : U \to \mathbb{R}^m$, if $D\mathbf{f}(\mathbf{x}) = 0$ for all $\mathbf{x} \in U$, then $\mathbf{f}$ is constant on $U$.

*Proof.* We are going to use the fact that $\mathbf{f}$ is locally constant. wlog, assume $m = 1$. Given any $\mathbf{a}, \mathbf{b} \in U$, we show that $f(\mathbf{a}) = f(\mathbf{b})$. Let $\gamma : [0,1] \to U$ be a (continuous) path from $\mathbf{a}$ to $\mathbf{b}$. For any $s \in (0,1)$, there exists some $\varepsilon$ such that $B_\varepsilon(\gamma(s)) \subseteq U$ since $U$ is open. By continuity of $\gamma$, there is a $\delta$ such that $(s-\delta, s+\delta) \subseteq [0,1]$ with $\gamma((\sigma - \delta, s+\delta)) \subseteq B_\varepsilon(\gamma(s)) \subseteq U$.

Since $f$ is constant on $B_\varepsilon(\gamma(s))$ by the mean value theorem applied at $M = 0$, we know that $g(t) = f \circ \gamma(t)$ is constant on $(s-\delta, s+\delta)$. In particular, $g$ is differentiable at $s$ with derivative 0. This is true for all $s$. So the map $g : [0,1] \to \mathbb{R}$ has zero derivative on $(0,1)$ and is continuous on $(0,1)$. So $g$ is constant. So $g(0) = g(1)$, ie. $f(\mathbf{a}) = f(\mathbf{b})$. □

## 7.4 Inverse Function Theorem

Now this is also one of the most important theorems of this course. We first define a few terms:

**Definition** ($C^1$ function)**.** Let $U \subseteq \mathbb{R}^n$ be open. We say $\mathbf{f} : U \to \mathbb{R}^m$ is $C^1$ on $U$ if $\mathbf{f}$ is differentiable at each $\mathbf{x} \in U$ and

$$D\mathbf{f} : U \to L(\mathbb{R}^n, \mathbb{R}^m)$$

is continuous.

**Proposition.** Let $U \subseteq \mathbb{R}^n$ be open. Then $\mathbf{f} = (f_1, \cdots, f_n) : U \to \mathbb{R}^n$ is $C^1$ on $U$ if and only if the partial derivatives $D_j f_i(\mathbf{x})$ exists for all $\mathbf{x} \in U$, $1 \leq i \leq n$, $1 \leq j \leq n$, and $D_j f_i : U \to \mathbb{R}$ are continuous.

*Proof.* $(\mathbb{R}ightarrow)$ Differentiability of $\mathbf{f}$ at $\mathbf{x}$ implies $D_j f_i(\mathbf{x})$ exists and is given by

$$D_j f_i(\mathbf{x}) = \langle D\mathbf{f}(\mathbf{x})\mathbf{e}_j, \mathbf{b}_i \rangle,$$

where $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ and $\{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$ are the standard basis for $\mathbb{R}^n$ and $\mathbb{R}^m$.

So we know

$$|D_j f_i(\mathbf{x}) - D_j f_i(\mathbf{y})| = |\langle (D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{y}))\mathbf{e}_j, \mathbf{b}_i \rangle| \leq \|D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{y})\|$$

since $\mathbf{e}_j$ and $\mathbf{b}_i$ are unit vectors. Hence if $D\mathbf{f}$ is continuous, so is $D_j f_i$.

($\Leftarrow$) Since the partials exist and are continuous, by our previous theorem, we know that the derivative $D\mathbf{f}$ exists. To show $D\mathbf{f} : U \to L(\mathbb{R}^m; \mathbb{R}^n)$ is continuous, note the following general fact:

For any linear map $A \in L(\mathbb{R}^n; \mathbb{R}^m)$ represented by $(a_{ij})$ so that $A\mathbf{h} = a_{ij}h_j$, then for $\mathbf{x} = (x_1, \cdots, x_n)$, we have

$$\|A\mathbf{x}\|^2 = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} A_{ij} x_j \right)^2$$

By Cauchy-Schwarz, we have

$$\begin{aligned} &\leq \sum_{i=1}^{m} \left( \sum_{j=1}^{n} a_{ij}^2 \right)^2 \left( \sum_{j=1}^{n} x_j^2 \right)^2 \\ &= \|\mathbf{x}\|^2 \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2. \end{aligned}$$

Dividing by $\|\mathbf{x}\|^2$, we know

$$\|A\| \leq \sqrt{\sum \sum a_{ij}^2}.$$

Applying this to $A = D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{y})$, we get

$$\|D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{y})\| \leq \sqrt{\sum \sum (D_j f_i(\mathbf{x}) - D_j f_i(\mathbf{x}))^2}.$$

So if all $D_j f_i$ are continuous, then so is $D\mathbf{f}$. $\square$

Finally, we can get to the inverse function theorem.

**Theorem** (Inverse function theorem)**.** Let $U \subseteq \mathbb{R}^n$ be open, and $\mathbf{f} : U \to \mathbb{R}^m$ be a $C^1$ map. Let $\mathbf{a} \in U$, and suppose that $D\mathbf{f}(\mathbf{a})$ is invertible as a linear map $\mathbb{R}^n \to \mathbb{R}^n$. Then there exists open sets $V, W \subseteq \mathbb{R}^n$ with $\mathbf{a} \in V$, $\mathbf{f}(\mathbf{a}) \in W$, $V \subseteq U$ such that

$$\mathbf{f}|_V : V \to W$$

is a bijection. Moreover, the inverse map $\mathbf{f}|_V^{-1} : W \to V$ is also $C^1$.

These functions have names:

**Definition** (Diffeomorphism)**.** Let $U, U' \subseteq \mathbb{R}^n$ are open, then a map $\mathbf{g} : U \to U'$ is a *diffeomorphism* if it is $C^1$ with a $C^1$ inverse.

**Note.** The inverse function theorem basically says: if $\mathbf{f}$ is $C^1$ and $D\mathbf{f}(\mathbf{a})$ is invertible, then $\mathbf{f}$ is a local diffeomorphism at $\mathbf{a}$.

*Proof.* By replacing $\mathbf{f}$ with $(D\mathbf{f}(\mathbf{a}))^{-1}\mathbf{f}$ (or by rotating our heads and stretching it a bit), we can assume $D\mathbf{f}(\mathbf{a}) = I$, the identity map. By continuity of $D\mathbf{f}$, there exists some $r > 0$ such that

$$\|D\mathbf{f}(\mathbf{x}) - I\| < \frac{1}{2}$$

for all $\mathbf{x} \in \overline{B_r(\mathbf{a})}$. By shrinking $r$ sufficiently, we can assume $\overline{B_r(\mathbf{a})} \subseteq U$. Let $W = B_{r/2}(\mathbf{f}(\mathbf{a}))$, and let $V = \mathbf{f}^{-1}(W) \cap B_r(\mathbf{a})$.

That was just our setup. There are three steps to actually proving the theorem.

**Claim.** $V$ is open, and $\mathbf{f}|_V : V \to W$ is a bijection.

Since $\mathbf{f}$ is continuous, $\mathbf{f}^{-1}(W)$ is open. So $V$ is open. To show $\mathbf{f}|_V : V \to W$ is bijection, we have to show that for each $\mathbf{y} \subseteq W$, then there is a *unique* $\mathbf{x} \in V$ such that $\mathbf{f}(\mathbf{x}) = \mathbf{y}$. We are going to use the contraction mapping theorem to prove this. This statement is equivalent to proving that for each $\mathbf{y} \in W$, the map $T(\mathbf{x}) = \mathbf{x} - \mathbf{f}(\mathbf{x}) + \mathbf{y}$ has a unique fixed point $\mathbf{x} \in V$.

Let $\mathbf{h}(\mathbf{x}) = \mathbf{x} - \mathbf{f}(\mathbf{x})$. Then note that

$$D\mathbf{h}(\mathbf{x}) = I - D\mathbf{f}(\mathbf{x}).$$

So by our choice of $r$, for every $\mathbf{x} \in \overline{B_r(\mathbf{a})}$, we must have

$$\|D\mathbf{h}(\mathbf{x})\| \leq \frac{1}{2}.$$

Then for any $\mathbf{x}_1, \mathbf{x}_2 \in \overline{B_r(\mathbf{a})}$, we can use the mean value inequality to estimate

$$\|\mathbf{h}(\mathbf{x}_1) - \mathbf{h}(\mathbf{x}_2)\| \leq \frac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|.$$

Hence we know

$$\|T(\mathbf{x}_1) - T(\mathbf{x}_2)\| = \|\mathbf{h}(\mathbf{x}_1) - \mathbf{h}(\mathbf{x}_2)\| \leq \frac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|.$$

Finally, to apply the contraction mapping theorem, we need to pick the right domain for $T$, namely $\overline{B_r(\mathbf{a})}$.

For any $\mathbf{x} \in \overline{B_r(\mathbf{a})}$, we have

$$\begin{aligned}\|T(\mathbf{x}) - \mathbf{a}\| &= \|\mathbf{x} - \mathbf{f}(\mathbf{x}) + \mathbf{y} - \mathbf{a}\| \\ &= \|\mathbf{x} - \mathbf{f}(\mathbf{x}) - (\mathbf{a} - \mathbf{f}(\mathbf{a})) + \mathbf{y} - \mathbf{f}(\mathbf{a})\| \\ &\leq \|\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{a})\| + \|\mathbf{y} - \mathbf{f}(\mathbf{a})\| \\ &\leq \frac{1}{2}\|\mathbf{x} - \mathbf{a}\| + \|\mathbf{y} - \mathbf{f}(\mathbf{a})\| \\ &< \frac{r}{2} + \frac{r}{2} \\ &= r.\end{aligned}$$

So $T : \overline{B_r(\mathbf{a})} \to B_r(\mathbf{a}) \subseteq \overline{B_r(\mathbf{a})}$. Since $\overline{B_r(\mathbf{a})}$ is complete, $T$ has a unique fixed point $\mathbf{x} \in \overline{B_r(\mathbf{a})}$, ie. $T(\mathbf{x}) = \mathbf{x}$. Finally, we need to show $\mathbf{x} \in B_r(\mathbf{a})$, since this is where we want to find our fixed point. But this is true, since $T(\mathbf{x}) \in B_r(\mathbf{a})$ by above. So we must have $\mathbf{x} \in B_r(\mathbf{a})$. Also, since $f(\mathbf{x}) = \mathbf{y}$, we know $\mathbf{x} \in f^{-1}(W)$. So $\mathbf{x} \in V$.

So we have shown that for each $\mathbf{y} \in V$, there is a unique $\mathbf{x} \in V$ such that $\mathbf{f}(\mathbf{x}) = \mathbf{y}$. So $\mathbf{f}|_V : V \to W$ is a bijection.

We have done the hard work now. It remains to show that $\mathbf{f}|_V$ is invertible with $C^1$ inverse.

**Claim.** The inverse map $\mathbf{g} = \mathbf{f}|_V^{-1} : W \to V$ is Lipschitz (and hence continuous). In fact, we have

$$\|\mathbf{g}(\mathbf{y}_1) - \mathbf{g}(\mathbf{y}_2)\| \leq 2\|\mathbf{y}_1 - \mathbf{y}_2\|.$$

For any $\mathbf{x}_1, \mathbf{x}_2 \in V$, by the triangle inequality, know

$$\begin{aligned}\|\mathbf{x}_1 - \mathbf{x}_2\| - \|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\| &\leq \|(\mathbf{x}_1 - \mathbf{f}(\mathbf{x}_1)) - (\mathbf{x}_2 - \mathbf{f}(\mathbf{x}_2))\| \\ &= \|\mathbf{h}(\mathbf{x}_1) - \mathbf{h}(\mathbf{x}_0)\| \\ &\leq \frac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|.\end{aligned}$$

Hence, we get

$$\|\mathbf{x}_1 - \mathbf{x}_2\| \leq 2\|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\|.$$

Apply this to $\mathbf{x}_1 = \mathbf{g}(\mathbf{y}_1)$ and $\mathbf{x}_2 = \mathbf{g}(\mathbf{y}_2)$, and note that $\mathbf{f}(\mathbf{g}(\mathbf{y}_j)) = \mathbf{y}_j$ to get the desired result.

**Claim.** $\mathbf{g}$ is in fact $C^1$, and moreover, for all $\mathbf{y} \in W$,

$$D\mathbf{g}(\mathbf{y}) = D\mathbf{f}(\mathbf{g}(\mathbf{y}))^{-1}. \qquad (*)$$

Note that if $\mathbf{g}$ where differentiable, then its derivative must be given by $(*)$, since by definition, we know

$$\mathbf{f}(\mathbf{g}(\mathbf{y})) = \mathbf{y},$$

and hence the chain rule gives

$$D\mathbf{f}(\mathbf{g}(\mathbf{y})) \cdot D\mathbf{g}(\mathbf{y}) = I.$$

Also, we immediately know $D\mathbf{g}$ is continuous, since it is the composition of continuous functions (the inverse of a matrix is given by polynomial expressions of the components). So we only need to check that $D\mathbf{f}(\mathbf{g}(\mathbf{y}))^{-1}$ satisfies the definition of the derivative.

First we check that $D\mathbf{f}(\mathbf{x})$ is indeed invertible for every $\mathbf{x} \in \overline{B_r(\mathbf{a})}$. We use the fact that

$$\|D\mathbf{f}(\mathbf{x}) - I\| \leq \frac{1}{2}.$$

If $D\mathbf{f}(\mathbf{x})\mathbf{v} = \mathbf{0}$, then we have

$$\|\mathbf{v}\| = \|D\mathbf{f}(\mathbf{x})\mathbf{v} - \mathbf{v}\| \leq \|D\mathbf{f}(\mathbf{x}) - I\|\|\mathbf{v}\| \leq \frac{1}{2}\|\mathbf{v}\|.$$

So we must have $\|\mathbf{v}\| = 0$, ie. $\mathbf{v} = \mathbf{0}$. So $\ker D\mathbf{f}(\mathbf{x}) = \{\mathbf{0}\}$. So $D\mathbf{f}(\mathbf{g}(\mathbf{y}))^{-1}$ exists.

Let $\mathbf{x} \in V$ be fixed, and $\mathbf{y} = \mathbf{f}(\mathbf{x})$. Let $\mathbf{k}$ be small and

$$\mathbf{h} = \mathbf{g}(\mathbf{y} + \mathbf{k}) - \mathbf{g}(\mathbf{y}).$$

In other words,

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) = \mathbf{k}.$$

Since $\mathbf{g}$ is invertible, whenever $\mathbf{k} \neq \mathbf{0}$, $\mathbf{h} \neq \mathbf{0}$. Since $\mathbf{g}$ is continuous, as $\mathbf{k} \to \mathbf{0}$, $\mathbf{h} \to \mathbf{0}$ as well.

We have

$$\begin{aligned}
&\frac{\mathbf{g}(\mathbf{y} + \mathbf{k}) - \mathbf{g}(\mathbf{y}) - D\mathbf{f}(\mathbf{g}(\mathbf{y}))^{-1}\mathbf{k}}{\|\mathbf{k}\|} \\
&= \frac{\mathbf{h} - D\mathbf{f}(\mathbf{g}(\mathbf{y}))^{-1}\mathbf{k}}{\|\mathbf{k}\|} \\
&= \frac{D\mathbf{f}(\mathbf{x})^{-1}(D\mathbf{f}(\mathbf{x})\mathbf{h} - \mathbf{k})}{\|\mathbf{k}\|} \\
&= \frac{-D\mathbf{f}(\mathbf{x})^{-1}(\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{x})\mathbf{h})}{\|\mathbf{k}\|} \\
&= -D\mathbf{f}(\mathbf{x})^{-1}\left(\frac{\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{x})\mathbf{h}}{\|\mathbf{h}\|} \cdot \frac{\|\mathbf{h}\|}{\|\mathbf{k}\|}\right) \\
&= -D\mathbf{f}(\mathbf{x})^{-1}\left(\frac{\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{x})\mathbf{h}}{\|\mathbf{h}\|} \cdot \frac{\|\mathbf{g}(\mathbf{y} + \mathbf{k}) - \mathbf{g}(\mathbf{y})\|}{\|(\mathbf{y} + \mathbf{k}) - \mathbf{y}\|}\right).
\end{aligned}$$

As $\mathbf{k} \to \mathbf{0}$, $\mathbf{h} \to \mathbf{0}$. The first factor $-D\mathbf{f}(\mathbf{x})^{-1}$ is fixed; the second factor tends to $\mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$; the third factor is bounded by 2. So the whole thing tends to $\mathbf{0}$. So done. □

## 7.5 2nd Derivatives

Now, naturally, we extend our discussion to second derivatives.

**Definition** (2nd derivative)**.** Let $U \subseteq \mathbb{R}^n$ be open, $\mathbf{f} : U \to \mathbb{R}^m$ be differentiable. Then $D\mathbf{f} : U \to L(\mathbb{R}^n; \mathbb{R}^m)$. We say $D\mathbf{f}$ is *differentiable* at $\mathbf{a} \in U$ if there exists $A \in L(\mathbb{R}^n; L(\mathbb{R}^n; \mathbb{R}^m))$ such that

$$\lim_{\mathbf{h} \to 0} = \frac{1}{\|\mathbf{h}\|}(D\mathbf{f}(\mathbf{a} + \mathbf{h}) - D\mathbf{f}(\mathbf{a}) - A\mathbf{h}) = 0.$$

Using this, we can prove that orders don't matter in mixed partials:

**Theorem** (Symmetry of mixed partials). Let $U \subseteq \mathbb{R}^n$ be open, $\mathbf{f} : U \to \mathbb{R}^m$, $\mathbf{a} \in U$, and $\rho > 0$ such that $B_\rho(\mathbf{a}) \subseteq U$.

Let $i, j \in \{1, \cdots, n\}$ be fixed and suppose that $D_i D_j \mathbf{f}(\mathbf{x})$ and $D_j D_i \mathbf{f}(\mathbf{x})$ exist for all $\mathbf{x} \in B_\rho(\mathbf{a})$ and are continuous at $\mathbf{a}$. Then in fact

$$D_i D_j \mathbf{f}(\mathbf{a}) = D_j D_i \mathbf{f}(\mathbf{a}).$$

The proof is quite short, when we know what to do.

*Proof.* wlog, assume $m = 1$. If $i = j$, then there is nothing to prove. So assume $i \neq j$.

Let

$$g_{ij}(t) = f(\mathbf{a} + t\mathbf{e}_i + t\mathbf{e}_j) - f(\mathbf{a} + t\mathbf{e}_i) - f(\mathbf{a} + t\mathbf{e}_j) + f(\mathbf{a}).$$

Then for each fixed $t$, define $\phi : [0, 1] \to \mathbb{R}$ by

$$\phi(s) = f(\mathbf{a} + st\mathbf{e}_i + t\mathbf{e}_j) - f(\mathbf{a} + st\mathbf{e}_i).$$

Then we get

$$g_{ij}(t) = \phi(1) - \phi(0).$$

By the mean value theorem and the chain rule, there is some $\theta \in (0, 1)$ such that

$$g_{ij}(t) = \phi'(\theta) = t\Big(D_i(\mathbf{a} + \theta t\mathbf{e}_i + t\mathbf{e}_j) - D_i(\mathbf{a} + \theta t\mathbf{e}_i)\Big).$$

Now apply mean value theorem to the function

$$s \mapsto D_i(\mathbf{a} + \theta t\mathbf{e}_i + st\mathbf{e}_j),$$

there is some $\eta \in (0, 1)$ such that

$$g_{ij}(t) = t^2 D_j D_i f(\mathbf{a} + \theta t\mathbf{e}_i + \eta t\mathbf{e}_j).$$

We can do the same for $g_{ji}$, and find some $\tilde{\theta}, \tilde{\eta}$ such that

$$g_{ij}(t) = t^2 D_i D_j f(a + \tilde{\theta} t\mathbf{e}_i + \tilde{\eta} t\mathbf{e}_j).$$

Since $g_{ij} = g_{ji}$, we get

$$t^2 D_j D_i f(\mathbf{a} + \theta t\mathbf{e}_i + \eta t\mathbf{e}_j) = t^2 D_i D_j f(a + \tilde{\theta} t\mathbf{e}_i + \tilde{\eta} t\mathbf{e}_j).$$

Divide by $t^2$, and take the limit as $t \to 0$. By continuity of the partial derivatives, we get

$$D_j D_i f(\mathbf{a}) = D_i D_j f(\mathbf{a}).$$

□

**Theorem** (Second-order Taylor's theorem). Let $f : U \to \mathbb{R}$ be $C^2$, ie. $D_i D_j f(\mathbf{x})$ are continuous for all $\mathbf{x} \in U$. Let $\mathbf{a} \in U$ and $B_r(\mathbf{a}) \subseteq U$. Then

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + D^2 f(\mathbf{h}, \mathbf{h}) + E(\mathbf{h}),$$

where $E(\mathbf{h}) = o(\|\mathbf{h}\|^2)$.

*Proof.* Consider the function

$$g(t) = f(\mathbf{a} + t\mathbf{h}).$$

Then the assumptions tell us $g$ is twice differentiable. By the 1D Taylor's theorem, we know

$$g(1) = g(0) + g'(0) + g''(s)$$

for some $s \in [0, 1]$.

In other words,

$$\begin{aligned} f(\mathbf{a} + \mathbf{h}) &= f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + D^2 f(\mathbf{a} + s\mathbf{h})(\mathbf{h}, \mathbf{h}) \\ &= f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + D^2 f(\mathbf{a})(\mathbf{h}, \mathbf{h}) + E(\mathbf{h}), \end{aligned}$$

where

$$E(\mathbf{h}) = D^2 f(\mathbf{a} + s\mathbf{h})(\mathbf{h}, \mathbf{h}) - D^2 f(\mathbf{a})(\mathbf{h}, \mathbf{h}).$$

By definition of the operator norm, we get

$$|E(\mathbf{h})| \leq \|D^2 f(\mathbf{a} + s\mathbf{h}) - D^2 f(\mathbf{a})\| \|\mathbf{h}\|^2.$$

By continuity of the second derivative, as $\mathbf{h} \to \mathbf{0}$, we get

$$\|D^2 f(\mathbf{a} + s\mathbf{h}) - D^2 f(\mathbf{a})\| \to 0.$$

So $E(\mathbf{h}) = o(\|\mathbf{h}\|^2)$. So done. ☐

# 8 Met and Top Extension

Now we enter the world of Met and Top. Again. The Whole Met and Top course can be summarized by Section 5 and this Section, so any terms/definitions/theorems already in Section 5 would not be repeated here again.

## 8.1 Topological Spaces

When we defined continuity, we showed that this condition is equivalent to the fact that $f^{-1}(U)$ is open whenever $U$ is open. So really what we need is the set of open subsets, not the metric:

**Definition** (Topological space)**.** A *topological space* is a set $X$ (the space) together with a set $\mathcal{U} \subseteq \mathbb{P}(X)$ (the topology) such that:

(i) $\emptyset, X \in \mathcal{U}$

(ii) If $V_\alpha \in \mathcal{U}$ for all $\alpha \in A$, then $\displaystyle\bigcup_{\alpha \in A} V_\alpha \in \mathcal{U}$.

(iii) If $V_1, \cdots, V_n \in \mathcal{U}$, then $\displaystyle\bigcap_{i=1}^{n} V_i \in \mathcal{U}$.

The elements of $X$ are the *points*, and the elements of $\mathcal{U}$ are the open subsets of $X$.

**Definition** (Induced topology)**.** Let $(X, d)$ be a metric space. Then the topology *induced by* $d$ is the set of all open sets of $X$ under $d$.

But not all topological spaces are induced by metrics. Here are some important examples:

**Example.**

(i) Let $X$ be any set.

  (a) $\mathcal{U} = \{\emptyset, X\}$ is the *coarse topology* on $X$.

  (b) $\mathcal{U} = \{A \subseteq X : X \setminus A$ is finite or $A = \emptyset\}$ is the *cofinite topology* on $X$.

(ii) Let $X = \mathbb{R}$, and $\mathcal{U} = \{(a, \infty) : a \in \mathbb{R}\}$ is the *right order topology* on $\mathbb{R}$.

Now we redefine continuous functions in the appropriate topological definition:

**Definition** (Continuous function)**.** Let $f : X \to Y$ be a map of topological spaces. Then $f$ is *continuous* if $f^{-1}(U)$ is open in $X$ whenever $U$ is open in $Y$.

As an example, every function is continuous if $Y$ has the coarse topology. Now we define a "bijection" in topological spaces:

**Definition** (Homeomorphism)**.** $f : X \to Y$ is a *homeomorphism* if

(i) $f$ is a bijection

(ii) Both $f$ and $f^{-1}$ are continuous

Two spaces are *homeomorphic* if there exists a homeomorphism between them, and we write $X \simeq Y$.

This definition is similar to the group isomorphism, but we require explicitly that the inverse function is continuous, because it isn't guaranteed. Similarly, this can be easily shown to be an equivalence relation.

Note that we can show that two spaces are homeomorphic easily: Just write down the function. But it is *much* harder to show that they are not.

### 8.1.1 Convergence

Continuing, we will define convergence in a topologically-friendly way:

**Definition** (Convergent sequence)**.** A sequence $x_n \to x$ if for every open neighbourhood $U$ of $x$, $\exists N$ such that $x_n \in U$ for all $n > N$.

Note that this does not guarantee uniqueness of limits! In fact it is nowhere near (Let $U$ be the coarse topology, then every sequence converges to every number)

But there is a class of topological spaces that behave nicely:

**Definition** (Hausdorff space)**.** A topological space $X$ is *Hausdorff* if for all $x_1, x_2 \in X$ with $x_1 \neq x_2$, there exists open neighbourhoods $U_1$ of $x_1$, $U_2$ of $x_2$ such that $U_1 \cap U_2 = \emptyset$.

**Lemma.** If $X$ is Hausdorff, $x_n$ is a sequence in $X$ with $x_n \to x$ and $x_n \to x'$, then $x = x'$, ie. limits are unique.

*Proof.* Suppose the contrary that $x \neq x'$. Then by definition of Hausdorff, there exists open neighbourhoods $U, U'$ of $x, x'$ respectively with $U \cap U' = \emptyset$.

Since $x_n \to x$ and $U$ is a neighbourhood of $x$, by definition, there is some $N$ such that whenever $n > N$, we have $x_n \in U$. Similarly, since $x_n \to x'$, there is some $N'$ such that whenever $n > N'$, we have $x_n \in U'$.

This means that whenever $n > \max(N, N')$, we have $x_n \in U$ and $x_n \in U'$. So $x_n \in U \cap U'$. □

### 8.1.2 Closure and Interior

**Definition.** Let $X$ be a topological space and $A \subseteq X$. Define

$$\mathcal{C}_A = \{C \subseteq X : A \subseteq C \text{ and } C \text{ is closed in } X\}$$

Then the *closure* of $A$ in $X$ is

$$\bar{A} = \bigcap_{C \in \mathcal{C}_A} C.$$

We are literally taking the intersection of all closed sets that are bigger than $A$, so it should not be surprising to see that:

**Proposition.** $\bar{A}$ is the smallest closed subset of $X$ which contains $A$.

*Proof.* Let $K \subseteq X$ be a closed set containing $A$. Then $K \in \mathcal{C}_a$. So $\bar{A} = \bigcap_{C \in \mathcal{C}_A} C \subseteq K$. So $\bar{A} \subseteq K$. □

Now, let $L(A)$ denote the set of limit points, then we can use this to find the closure of subsets:

**Corollary.** Suppose $C \subseteq X$ is closed and $A \subseteq C$ and $C \subseteq L(A)$, Then $C = \bar{A}$.

*Proof.* $C \subseteq L(A) \subseteq \bar{A} \subseteq C$, where the last step is since $\bar{A}$ is the smallest closed set containing $A$. So $C = L(A) = \bar{A}$. □

**Definition** (Dense subset)**.** $A \subseteq X$ is *dense* in $X$ if $\bar{A} = X$.

This literally begs the over-used example: $\mathbb{Q}$ is dense in $\mathbb{R}$.
Now we define the interior of $A$:

**Definition** (Interior)**.** Let $A \subseteq X$, and let

$$\mathcal{O}_A = \{U \subseteq X : U \subseteq A, U \text{ is open in } X\}.$$

The *interior* of $A$ is

$$\mathbb{Z}(A) = \bigcup_{U \in \mathcal{O}} U.$$

It is the largest open subset of $A$ contained in $A$ (with proof similar to the closure one). Now the following relates the interior and closure:

**Proposition.** $X \setminus \mathbb{Z}(A) = \overline{X \setminus A}$

*Proof.* $U \subseteq A \Leftrightarrow (X \setminus U) \supseteq (X \setminus A)$. Also, $U$ open in $X \Leftrightarrow X \setminus U$ is closed in $X$.
So the complement of the largest open subset of $X$ contained in $A$ will be the smallest closed subset containing $X \setminus A$. □

### 8.1.3 New Topologies from Old

This section explores the possibilities of creating new topologies from the old one.

**Definition** (Subspace topology)**.** Let $X$ is a topological space and $Y \subseteq X$. The *subspace topology* on $Y$ is given by: $V$ is an open subset of $Y$ if there is some $U$ open in $X$ such that $V = Y \cap U$.

**Proposition.** The subspace topology is a topology.

*Proof.*

(i) Since $\emptyset$ is open in $X$, $\emptyset = Y \cap \emptyset$ is open in $Y$.
Since $X$ is open in $X$, $Y = Y \cap X$ is open in $Y$.

(ii) If $V_\alpha$ is open in $Y$, then $V_\alpha = Y \cap U_\alpha$ for some $U_\alpha$ open in $X$. Then

$$\bigcup_{\alpha \in A} V_\alpha = \bigcup_{\alpha \in A} (Y \cap U_\alpha) = Y \cap \left( \bigcup_{\alpha \in U} U_\alpha \right).$$

Since $\bigcup U_\alpha$ is open in $X$, so $\bigcup V_\alpha$ is open in $Y$.

(iii) If $V_i$ is open in $Y$, then $V_i = Y \cap U_i$ for some open $U_i \subseteq X$. Then

$$\bigcap_{i=1}^{n} V_i = \bigcap_{i=a}^{n} (Y \cap U_i) = Y \cap \left( \bigcap_{i=a}^{n} U_i \right).$$

Since $\bigcap U_i$ is open, $\bigcap V_i$ is open.

□

**Proposition.** If $Y$ has the subspace topology, $f : Z \to Y$ is continuous iff $\iota \circ f : Z \to X$ is continuous.

*Proof.* ($\Rightarrow$) If $U \subseteq X$ is open, then $\iota^{-1}(U) = Y \cap U$ is open in $Y$. So $\iota$ is continuous. So if $f$ is continuous, so is $\iota \circ f$.

($\Leftarrow$) Suppose we know that $\iota \circ f$ is continuous. Given $V \subseteq Y$ is open, we know that $V = Y \cap U = \iota^{-1}(U)$. So $f^{-1}(V) = f^{-1}(\iota^{-1}(U))) = (\iota \circ f)^{-1}(U)$ is open since $\iota \circ f$ is continuous. So $f$ is continuous. □

**Definition** (Product topology)**.** Let $X$ and $Y$ be topological spaces. The *product topology* on $X \times Y$ is given by:

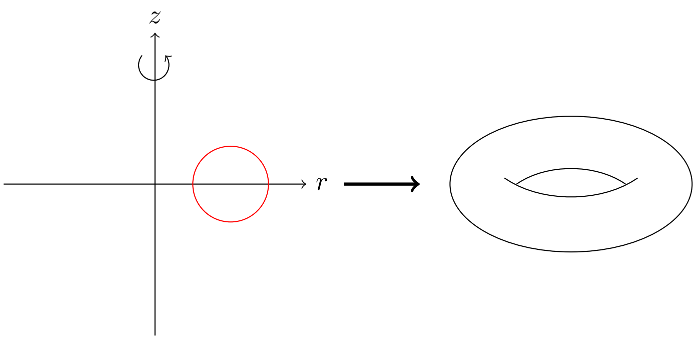
$U \subseteq X \times Y$ is open if: for every $(x, y) \in U$, there exists $V_x \subseteq X, W_y \subseteq Y$ open neighbourhoods of $x$ and $y$ such that $V_x \times W_y \subseteq U$.

This deserves a good example from our old friend: The Torus. Graphic thanks to Dexter.

**Example.** Let $A \subseteq \{(r, z) : r > 0\} \subseteq \mathbb{R}^2$, $R(A)$ be the set obtained by rotating $A$ around the $z$ axis. Then $R(A) \simeq S \times A$ by

$$(x, y, z) = (\mathbf{v}, z) \mapsto (\hat{\mathbf{v}}, (|\mathbf{v}|, z)).$$

In particular, if $A$ is a circle, then $R(A) \simeq S^1 \times S^1 = T^2$ is the two-dimensional torus.



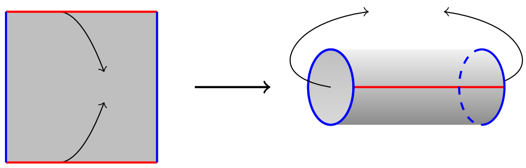
As we have a product, we shall have a quotient!

**Definition** (Quotient topology)**.** If $X$ is a topological space, the quotient topology on $X/\sim$ is given by: $U$ is open in $X/\sim$ if $\pi^{-1}(U)$ is open in $X$.

We can think of the quotient as "gluing" the points identified by $\sim$ together. The defining property is $f : X/\sim \to Y$ is continuous iff $f \circ \pi : X \to Y$ is continuous.

**Example.** Let $X = [0,1] \times [0,1]$ with $\sim$ given by $(0,y) \sim (1,y)$ and $(x,0) \sim (x,1)$, then $X/\sim \simeq S^1 \times S^1 = T^2$, by, say

$$(x,y) \mapsto \big((\cos 2\pi x, \sin 2\pi x), (\cos 2\pi y, \sin 2\pi y)\big)$$



Similar to a vector space, we can define a "basis" for the topology:

**Definition** (Basis)**.** Let $\mathcal{U}$ be a topology on $X$. A subset $\mathcal{B} \subseteq \mathcal{U}$ is a *basis* if "$U \in \mathcal{U}$ iff $U$ is a union of sets in $\mathcal{B}$".

**Example.** $\{V \times W : V \subseteq X, W \subseteq Y \text{ are open}\}$ is a basis for the product topology for $X \times Y$.

## 8.2 Connectivity

**Definition** (Connected space)**.** A topological space $X$ is *disconnected* if $X$ can be written as $A \cup B$, where $A$ and $B$ are disjoint, non-empty open subsets of $X$. We say $A$ and $B$ *disconnect* $X$. A space is *connected* if it is not disconnected.

Now we would like to prove "common sense" with this definition:

**Theorem.** $[0,1]$ is connected.

*Proof.* Suppose $A$ and $B$ disconnect $[0,1]$. wlog, assume $1 \in B$. Since $A$ is non-empty, $\alpha = \sup A$ exists. Then either

- $\alpha \in A$. Then $\alpha < 1$, since $1 \in B$. Since $A$ is open, $\exists \varepsilon > 0$ with $B_\varepsilon(\alpha) \subseteq A$. So $\alpha + \frac{\varepsilon}{2} \in A$, contradicting supremality of $\alpha$
- $\alpha \notin A$. Then $\alpha \in B$. Since $B$ is open, $\exists \varepsilon > 0$ such that $B_\varepsilon(\alpha) \subseteq B$. Then $a \leq \alpha - \varepsilon$ for all $a \in A$. This contradicts $\alpha$ being the *least* upper bound of $A$.

Either option gives a contradiction. So $A$ and $B$ cannot exist and $[0,1]$ is connected. □

Now we would prove the Intermediate Value Theorem. But before that, we need this proposition:

**Proposition.** If $f : X \to Y$ is continuous and $X$ is connected, then $\operatorname{Im} f$ is also connected.

*Proof.* Suppose $A$ and $B$ disconnect $\operatorname{Im} f$. We will show that $f^{-1}(A)$ and $f^{-1}(B)$ disconnect $X$.

Since $A, B \subseteq \operatorname{Im} f$ are open, we know that $A = \operatorname{Im} f \cap A'$ and $B = \operatorname{Im} f \cap B'$ for some $A', B'$ open in $Y$. Then $f^{-1}(A) = f^{-1}(A')$ and $f^{-1}(B) = f^{-1}(B)$ are open in $X$.

Since $A, B$ are non-empty, $f^{-1}(A)$ and $f^{-1}(B)$ are non-empty. Also, $f^{-1}(A) \cap f^{-1}(B) = f^{-1}(A \cap B) = f^{-1}(\emptyset) = \emptyset$. Finally, $A \cup B = \operatorname{Im} f$. So $f^{-1}(A) \cup f^{-1}(B) = f^{-1}(A \cup B) = X$.

So $f^{-1}(A)$ and $f^{-1}(B)$ disconnect $X$, contradicting our hypothesis. So $\operatorname{Im} f$ is connected. □

Now we would prove the IVT:

**Theorem** (Intermediate value theorem)**.** Suppose $f : X \to \mathbb{R}$ is continuous and $X$ is connected. If $\exists x_0, x_1$ such that $f(x_0) < 0 < f(x_1)$, then $\exists x \in X$ with $f(x) = 0$.

*Proof.* Suppose no such $x$ exists. Then $0 \notin \operatorname{Im} f$ while $0 > f(x_0) \in \operatorname{Im} f$, $0 < f(x_1) \in \operatorname{Im} f$. Then $\operatorname{Im} f$ is disconnected (from our previous example), contradicting $X$ being connected. □

### 8.2.1 Path Connectivity

Now we can also use the notion of connectivity (which is also intuitive) that says if some space is connected, we can draw a path around it:

**Definition** (Path)**.** Let $X$ be a topological space, and $x_0, x_1 \in X$. Then a *path* from $x_0$ to $x_1$ is a continuous function $\gamma : [0, 1] \mapsto X$ such that $\gamma(0) = x_0, \gamma_1 = x_1$.

**Definition** (Path connectivity)**.** A topological space $X$ is *path connected* if for all points $x_0, x_1 \in X$, there is a path from $x_0$ to $x_1$.

This form of connectivity is strictly stronger than connectivity:

**Proposition.** If $X$ is path connected, then $X$ is connected.

*Proof.* Suppose $X$ is path connected but not connected. Then there is a continuous surjective $f : X \to \{0, 1\}$. Choose $x_0, x_1$ with $f(x_0) = 0$, $f(x_1) = 1$. Since $X$ is path connected, there is a path $\gamma : [0, 1] \to X$ with $\gamma(0) = x_0, \gamma(1) = x_1$. Then $f \circ \gamma : [0, 1] \to \{0, 1\}$ is continuous and surjective. So $[0, 1]$ is not connected, which is a contradiction. □

Now we can use connectivity to show that some spaces are not homeomorphic to each other:

**Example.** $[0, 1] \not\simeq (0, 1)$. Suppose it were. Let $f : [0, 1]$ be a homeomorphism. Let $A = (0, 1]$. Then $f|_A : (0, 1] \to (0, 1) \setminus \{f(0)\}$ is a homeomorphism. But $(0, 1]$ is connected while $(0, 1) \setminus \{f(0)\}$ is disconnected. Contradiction.

Similarly, $[0, 1) \not\simeq [0, 1]$ and $[0, 1) \not\simeq (0, 1)$.

### 8.2.2 Components

If the space is disconnected, then we can separate the space into different components that are connected. To do that, we need an equivalence relation:

**Lemma.** Define $x \sim y$ if there is a path from $x$ to $y$ in $X$. Then $\sim$ is an equivalence relation. So equivalence classes of the relation "$x \sim y$ if there is a path from $x$ to $y$" are *path components* of $X$.

This defines the *path* connected components. But what if we only want normal connectedness? Then we first need to introduce a theorem:

**Proposition.** Suppose $Y_\alpha \subseteq X$ is connected for all $\alpha \in T$ and that $\bigcap_{\alpha \in T} Y_\alpha \neq \emptyset$. Then $Y = \bigcup_{\alpha \in T} Y_\alpha$ is connected.

*Proof.* Suppose the contrary that $A$ and $B$ disconnect $Y$. Then $A$ and $B$ are open in $Y$. So $A = Y \cap A'$ and $B = Y \cap B'$, where $A', B'$ are open in $X$. For any fixed $\alpha$, let

$$A_\alpha = Y_\alpha \cap A = Y_\alpha \cap A', \quad B_\alpha = Y_\alpha \cap B = Y_\alpha \cap B'.$$

Then they are open in $Y_\alpha$. Since $Y = A \cup B$, we have

$$Y_\alpha = Y \cap Y_\alpha = (A \cup B) \cap Y_\alpha = A_\alpha \cup B_\alpha.$$

Since $A \cap B = \emptyset$, we have

$$A_\alpha \cap B_\alpha = Y_\alpha \cap (A \cap B) = \emptyset.$$

So $A_\alpha, B_\alpha$ are disjoint. So $Y_\alpha$ is connected but is the disjoint union of open subsets $A_\alpha, B_\alpha$.

By definition of connectivity, this can only happen if $A_\alpha = \emptyset$ or $B_\alpha = \emptyset$.

However, by assumption, $\displaystyle\bigcap_{\alpha \in T} Y_\alpha \neq \emptyset$. So pick $y \in \displaystyle\bigcap_{\alpha \in T} Y_\alpha$. Since $y \in Y$, either $y \in A$ or $y \in B$. wlog, assume $y \in A$. Then $y \in Y_\alpha$ for all $\alpha$ implies that $y \in A_\alpha$ for all $\alpha$. So $A_\alpha$ is non-empty for all $\alpha$. So $B_\alpha$ is empty for all $\alpha$. So $B = \emptyset$. So $A$ and $B$ did not disconnect $Y$ after all. Contradiction. □

We can now define the *connected component*:

**Definition** (Connected component)**.** For $x \in X$, define

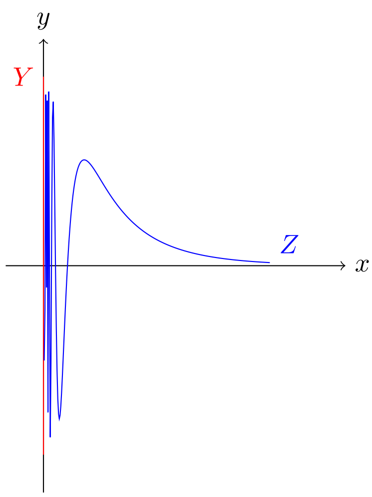
$$\mathcal{C}(x) = \{A \subseteq X : x \in A \text{ and } A \text{ is connected}\}.$$

Then $C(x) = \displaystyle\bigcup_{A \in \mathcal{C}(x)} A$ is the *connected component* of $x$.

Now the following example illustrates the difference between connected and path-connected:

**Example.** Let $Y = \{(0, y) : y \in R\} \subseteq \mathbb{R}^2$ be the $y$ axis.

Let $Z = \{(x, \frac{1}{x} \sin \frac{1}{x}) : x \in (0, \infty)\}$.




Let $X = Y \cup Z \subseteq \mathbb{R}^2$. We claim that $Y$ and $Z$ are the path components of $X$. Since $Y$ and $Z$ are individually path connected, it suffices to show that there is no continuous $\gamma : [0, 1] \to X$ with $\gamma(0) = (0, 0)$, $\gamma(1) = (1, \sin 1)$.

Suppose $\gamma$ existed. Then the function $\pi_2 \circ \gamma : [0, 1] \to \mathbb{R}$ projecting the path to the $y$ direction is continuous. So it is bounded. Let $M$ be such that $\pi_2 \circ \gamma(t) \leq M$ for all $t \in [0, 1]$. Let $W = X \cap (\mathbb{R} \times (-\infty, M])$ be the part of $X$ that lies below $y = M$. Then $\operatorname{Im} \gamma \subseteq W$.

However, $W$ is disconnected: pick $t_0$ with $\frac{1}{t_0} \sin \frac{1}{t_0} > M$. Then $W \cap ((-\infty, t_0) \times \mathbb{R})$ and $W \cap ((t_0, \infty) \times \mathbb{R})$ disconnect $W$. This is a contradiction, since $\gamma$ is continuous and $[0, 1]$ is connected.

We also claim that $X$ is connected: suppose $A$ and $B$ disconnect $X$. Then since $Y$ and $Z$ are connected, either $Y \subseteq A$ or $Y \subseteq B$; $Z \subseteq A$ or $Z \subseteq B$. If both $Y \subseteq A, Z \subseteq A$, then $B = \emptyset$, which is not possible.

So wlog assume $A = Y, B = Z$. This is also impossible, since $Y$ is not open in $X$ as it is not a union of balls (any open ball containing a point in $Y$ will also contain a point in $X$). Hence $X$ must be connected.

We end this section by providing a special case when being connected means being path-connected:

**Proposition.** If $U \subseteq \mathbb{R}^n$ is open and connected, then it is path-connected.

*Proof.* Let $A$ be a path component of $U$. We first show that $A$ is open.

Let $a \in A$. Since $U$ is open, $\exists \varepsilon > 0$ such that $B_\varepsilon(a) \subseteq U$. We know that $B_\varepsilon(a) \simeq \mathbb{Z}(D^n)$ is path-connected (eg. use line segments connecting the points). Since $A$ is a path component and $a \in A$, we must have $B_\varepsilon(a) \subseteq A$. So $A$ is an open subset of $U$.

Now suppose $b \in U \setminus A$. Then since $U$ is open, $\exists \varepsilon > 0$ such that $B_\varepsilon(b) \subseteq U$. Since $B_\varepsilon(b)$ is path-connected, so if $B_\varepsilon(b) \cap A \neq \emptyset$, then $B_\varepsilon(b) \subseteq A$. But this implies $b \in A$, which is a contradiction. So $B_\varepsilon(b) \cap A = \emptyset$. So $B_\varepsilon(b) \subseteq U \setminus A$. Then $U \setminus A$ is open.

So $A, U \setminus A$ are disjoint open subsets of $U$. Since $U$ is connected, we must have $U \setminus A$ empty (since $A$ is not). So $U = A$ is path-connected. $\square$

## 8.3 Compactness

We arrive at the last topic in this course. Compactness. What is like generalizing the fact about being "closed and bounded" in $\mathbb{R}$. But the actual definition is weird and one takes time to actually understand it:

**Definition** (Open cover)**.** Let $\mathcal{U} \subseteq \mathbb{P}(X)$ be a topology on $X$. An *open cover* of $X$ is a subset $\mathcal{V} \subseteq \mathcal{U}$ such that

$$\bigcup_{V \in \mathcal{V}} V = X.$$

We say $\mathcal{V}$ *covers* $X$.

If $\mathcal{V}' \subseteq \mathcal{V}$, and $\mathcal{V}'$ covers $X$, then we say $\mathcal{V}'$ is a *subcover* of $V$.

**Definition** (Compact space)**.** A topological space $X$ is *compact* if every open cover $\mathcal{V}$ of $X$ has a finite subcover $\mathcal{V}' = \{V_1, \cdots, V_n\} \subseteq \mathcal{V}$.

Let us look at some examples to get out head around this definition:

**Example.**

(i) Let $X = \mathbb{R}$ and $\mathcal{V} = \{(-R, R) : R \in \mathbb{R}, R > 0\}$. Then this is an open cover with no finite subcover. So $\mathbb{R}$ is not compact. Hence all open intervals are not compact since they are homeomorphic to $\mathbb{R}$.

(ii) Let $X = [0, 1] \cap \mathbb{Q}$. Let

$$U_n = X \setminus (\alpha - 1/n, \alpha + 1/n).$$

for some irrational $\alpha$ in $(0, 1)$ (eg. $\alpha = \sqrt{2}^{-1}$).

Then $\bigcup_{n>0} U_n = X$ since $\alpha$ is irrational. Then $\mathcal{V} = \{U_n : n \in \mathbb{Z} > 0\}$ is an open cover of $X$. Since this has no finite subcover, $X$ is not compact.

**Theorem.** $[0, 1]$ is compact.

*Proof.* Suppose $\mathcal{V}$ is an open cover of $[0, 1]$. Let

$$A = \{a \in [0, 1] : [0, a] \text{ has a finite subcover of } \mathcal{V}\}.$$

First show that $A$ is non-empty. Since $V$ covers $[0, 1]$, in particular, there is some $V_0$ that contains 0. So $\{0\}$ has a finite subcover $V_0$. So $0 \in A$.

Next we note that by definition, if $0 \leq b \leq a$ and $a \in A$, then $b \in A$.

Now let $\alpha = \sup A$. Suppose $\alpha < 1$. Then $\alpha \in [0, 1]$.

Since $\mathcal{V}$ covers $X$, let $\alpha \in V_\alpha$. Since $V_\alpha$ is open, there is some $\varepsilon$ such that $B_\varepsilon(\alpha) \subseteq V_\alpha$. By definition of $\alpha$, we must have $\alpha - \varepsilon/2 \in A$. So $[0, \alpha - \varepsilon/2]$ has a finite subcover. Add $V_\alpha$ to that subcover to get a finite subcover of $[0, \alpha + \varepsilon/2]$. Contradiction.

So we must have $\alpha = \sup A = 1$.

Now we argue as before: $\exists V_1 \in \mathcal{V}$ such that $1 \in V_1$ and $\exists \varepsilon > 0$ with $(1 - \varepsilon, 1] \subseteq V_q$. Since $1 - \varepsilon \in A$, there exists a finite $\mathcal{V}' \subseteq V$ which covers $[0, 1 - \varepsilon/2]$. Then $\mathcal{W} = \mathcal{V}' \cup \{V_1\}$ is a finite subcover of $V$. □

**Proposition.** If $X$ is compact and $C$ is closed subset of $X$, then $C$ is also compact.

*Proof.* To prove this, given an open cover of $C$, we need to find a finite subcover. To do so, we need to first convert it into an open cover of $X$. We can do so by adding $X \setminus C$, which is open since $C$ is closed. Then since $X$ is compact, we can find finite subcover of this, which we can convert back to a finite subcover of $C$.

Formally, suppose $\mathcal{V}$ is an open cover of $C$. Say $\mathcal{V} = \{V_\alpha : \alpha \in T\}$. For each $\alpha$, since $V_\alpha$ is open in $C$, $V_\alpha = C \cap V'_\alpha$ for some $V'_\alpha$ open in $X$. Also, since $\bigcup_{\alpha \in T} V_a = C$, we have $\bigcup_{\alpha \in T} V'_\alpha \supseteq C$.

Since $C$ is closed, $U = X \setminus C$ is open in $X$. So $\mathcal{W} = \{V'_\alpha : \alpha \in T\} \cup \{U\}$ is an open cover of $X$. Since $X$ is compact, $\mathcal{W}$ has a finite subcover $\mathcal{W}' = \{V'_{\alpha_1}, \cdots, V'_{\alpha_n}, U\}$ ($U$ may or may not be in there, but it doesn't matter). Now $U \cap C = \emptyset$. So $\{V_{\alpha_1}, \cdots, V_{\alpha_n}\}$ is a finite subcover of $C$. □

**Proposition.** Let $X$ be a Hausdorff space. If $C \subseteq X$ is compact, then $C$ is closed in $X$.

*Proof.* Let $U = X \setminus C$. We will show that $U$ is open.

For any $x$, we will find a $U_x$ such that $U_x \subseteq U$ and $x \in U_x$. Then $U = \bigcup_{x \in U} U_x$ will be open since it is as union of open sets.

To construct $U_x$, fix $x \in U$. Since $X$ is Hausdorff, so for each $y \in C$, $\exists U_{xy}, W_{xy}$ open neighbourhoods of $x$ and $y$ respectively with $U_{xy} \cap W_{xy} = \emptyset$.

Then $\mathcal{W} = \{W_{xy} \cap C : y \in C\}$ is an open cover of $C$. Since $C$ is compact, there exists a finite subcover $\mathcal{W}' = \{W_{xy_1} \cap C, \cdots, W_{xy_n} \cap C\}$.

Let $U_x = \bigcap_{i=1}^n U_{xy_i}$. Then $U_x$ is open since it is a finite intersection of open sets. To show $U_x \subseteq U$, note that $W_x = \bigcup_{i=1}^n W_{xy_i} \supseteq C$ since $\{W_{xy_i} \cap C\}$ is an open cover. We also have $W_x \cap U_x = \emptyset$. So $U_x \subseteq U$. So done. □

Now we relate compactness to closedness:

**Definition** (Bounded metric space)**.** A metric space $(X, d)$ is *bounded* if there exists $M \in \mathbb{R}$ such that $d(x, y) \leq M$ for all $x, y \in X$.

**Proposition.** A compact metric space $(X, d)$ is bounded.

*Proof.* Pick $x \in X$. Then $V = \{B_r(x) : r \in \mathbb{R}^+\}$ is an open cover of $X$. Since $X$ is compact, there is a finite subcover $\{B_{r_1}(x), \cdots, B_{r_n}(x)\}$.

Let $R = \max\{r_1, \cdots, r_n\}$. Then $d(x, y) < R$ for all $y \in X$. So for all $y, z \in X$,

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

So $X$ is bounded. □

**Theorem** (Heine-Borel)**.** $C \subseteq \mathbb{R}$ is compact iff $C$ is closed and bounded.

*Proof.* Since $\mathbb{R}$ is a metric space (hence Hausdorff), $C$ is also a metric space.

So if $C$ is compact, $C$ is closed in $\mathbb{R}$, and $C$ is bounded, by our previous two propositions.

Conversely, if $C$ is closed and bounded, then $C \subseteq [-N, N]$ for some $N \in \mathbb{R}$. Since $[-N, N] \simeq [0, 1]$ is compact, and $C = C \cap [-N, N]$ is closed in $[-N, N]$, $C$ is compact. □

Now compactness is preserved under continuous functions:

**Proposition.** If $f : X \to Y$ is continuous and $X$ is compact, then $\operatorname{Im} f \subseteq Y$ is also compact.

*Proof.* Suppose $\mathcal{V} = \{V_\alpha : \alpha \in T\}$ is an open cover of $\operatorname{Im} f$. Since $V_\alpha$ is open in $\operatorname{Im} f$, we have $V_\alpha = \operatorname{Im} f \cap V'_\alpha$, where $V'_\alpha$ is open in $Y$. Then

$$W_\alpha = f^{-1}(V_\alpha) = f^{-1}(V'_\alpha)$$

is open in $X$. If $x \in X$ then $f(x)$ is in $V_\alpha$ for some $\alpha$, so $x \in W_\alpha$. Thus $\mathcal{W} = \{W_\alpha : \alpha \in T\}$ is an open cover of $X$.

Since $X$ is compact, so there's a finite subcover $\{W_{\alpha_1}, \cdots, W_{\alpha_n}\}$ of $\mathcal{W}$.

Since $V_\alpha \subseteq \operatorname{Im} f$, $f(W_\alpha) = f(f^{-1}(V_\alpha)) = V_\alpha$. So $\{V_{\alpha_1}, \cdots, V_{\alpha_n}\}$ is a finite subcover of $V$. □

**Theorem** (Maximum value theorem)**.** If $f : X \to \mathbb{R}$ is continuous and $X$ is compact, then $\exists x \in X$ such that $f(x) \geq f(y)$ for all $y \in X$.

*Proof.* Since $X$ is compact, $\operatorname{Im} f$ is compact. Let $\alpha = \max\{\operatorname{Im} f\}$. Then $\alpha \in \operatorname{Im} f$. So $\exists x \in X$ with $f(x) = \alpha$. Then by definition $f(x) \geq f(y)$ for all $y \in X$. □

### 8.3.1 Products and Quotients

**Theorem.** If $X$ and $Y$ are compact, then so is $X \times Y$.

*Proof.* First consider the special type of open cover $\mathcal{V}$ of $X \times Y$ such that every $U \in \mathcal{V}$ has the form $U = V \times W$, where $V \subseteq X$ and $W \subseteq Y$ are open.

For every $(x, y) \in X \times Y$, there is $U_{xy} \in \mathcal{V}$ with $(x, y) \in U_{xy}$. Write

$$U_{xy} = V_{xy} \times W_{xy},$$

where $V_{xy} \subseteq X$, $W_{xy} \subseteq Y$ are open, $x \in V_{xy}, y \in W_{xy}$.

Fix $x \in X$. Then $\mathcal{W}_x = \{W_{xy} : y \in Y\}$ is an open cover of $Y$. Since $Y$ is compact, there is a finite subcover $\{W_{xy_1}, \cdots, W_{xy_n}\}$.

Then $V_x = \bigcap_{i=1}^n V_{xy_i}$ is a finite intersection of open sets. So $V_x$ is open in $X$. Moreover, $\mathcal{V}_x = \{U_{xy_1}, \cdots, U_{xy_n}\}$ covers $V_x \times Y$.

Now $\mathcal{O} = \{V_x : x \in X\}$ is an open cover of $X$. Since $X$ is compact, there is a finite subcover $\{V_{x_1}, \cdots, V_{x_m}\}$. Then $\mathcal{V}' = \bigcup_{i=1}^m \mathcal{V}_{x_i}$ is a finite subset of $\mathcal{V}$, which covers all of $X \times Y$.

In the general, case, suppose $\mathcal{V}$ is an open cover of $X \times Y$. For each $(x, y) \in X \times Y$, $\exists U_{xy} \in \mathcal{V}$ with $(x, y) \in U_{xy}$. Since $U_{xy}$ is open, $\exists V_{xy} \in X, W_{xy} \subseteq Y$ open with $V_{xy} \times W_{xy} \subseteq U_{xy}$ and $x \in V_{xy}, y \in W_{xy}$.

Then $\mathcal{Q} = \{V_{xy} \times W_{xy} : (x, y) \in (X, Y)\}$ is an open cover of $X \times Y$ of the type we already considered above. So it has a finite subcover $\{V_{x_1y_1} \times W_{x_1y_1}, \cdots, V_{x_ny_n} \times W_{x_ny_n}\}$. Now $V_{x_iy_i} \times W_{x_iy_i} \subseteq U_{x_iy_i}$. So $\{U_{x_1y_1}, \cdots, U_{x_ny_n}\}$ is a finite subcover of $X \times Y$. □

For quotients, it is easy to prove that a quotient of a compact space is compact, since every open subset in the quotient space can be projected back to an open subset of the original space. So we project the cover of the quotient space back to the normal one, and take the finite subcover there. Done!

Therefore, we would prove some interesting properties of compact quotients. But first, some lemmas:

**Proposition.** Suppose $f : X \to Y$ is a continuous bijection. If $X$ is compact and $Y$ is Hausdorff, then $f$ is a homeomorphism.

*Proof.* We show that $f^{-1}$ is continuous. To do this, it suffices to show $(f^{-1})^{-1}(C)$ is closed in $Y$ whenever $C$ is closed in $X$. By hypothesis, $f$ is a bijection . So $(f^{-1})^{-1}(C) = f(C)$.

Supposed $C$ is closed in $X$. Since $X$ is compact, $C$ is compact. Since $f$ is continuous, so $f(C) = (\operatorname{Im} f|_C)$ is compact. Since $Y$ is Hausdorff and $f(C) \subseteq Y$ is compact, $f(C)$ is closed. □

**Corollary.** Suppose $f : X/{\sim} \to Y$ is a bijection, $X$ is compact, $Y$ is Hausdorff, and $f \circ \pi$ is continuous, then $f$ is a homeomorphism.

*Proof.* Since $X$ is compact and $\pi : X \mapsto X/{\sim}$ is continuous, $\operatorname{Im} \pi \subseteq X/{\sim}$ is compact. Since $f \circ \pi$ is continuous, $f$ is continuous. So we can apply the proposition. □

**Example.** Let $X = D^2$ and $A = S^1 \subseteq X$. Then $f : X/A \mapsto S^2$ by $(r, \theta) \mapsto (1, \pi r, \theta)$ in spherical coordinates is a homeomorphism.

We can check that $f$ is a continuous bijection and $D^2$ is compact. So $X/A \simeq S^2$.

## 8.4 Sequential compactness

Now there is another definition of compactness, in terms of sequences. We would just show here that it is equivalent to compactness in metric spaces.

**Definition** (Sequential compactness)**.** A topological space $X$ is *sequentially compact* if every sequence $(x_n)$ in $X$ has a convergent subsequence (that converges to a point in $X$!).

But first, a lemma:

**Lemma.** Let $(x_n)$ be a sequence in a metric space $(X, d)$ and $x \in X$. Then $(x_n)$ has a subsequence converging to $x$ iff for every $\varepsilon > 0$, $x_n \in B_\varepsilon(x)$ for infinitely many $n$ $(*)$.

*Proof.* If $(x_{n_i}) \to x$, then for every $\varepsilon$, we can find $I$ such that $i > I$ implies $x_{n_j} \in B_\varepsilon(x)$ by definition of convergence. So $(*)$ holds.

Now suppose $(*)$ holds. We will construct a sequence $x_{n_i} \to x$ inductively. Take $n_0 = 0$. Suppose we have defined $x_{n_0}, x_{n_{i-1}}$.

By hypothesis, $x_n \in B_{1/i}(x)$ for infinitely many $n$. Take $n_i$ to be smallest such $n$ with $n_i > n_{i-1}$.

Then $d(x_{n_i} x) < \frac{1}{i}$ implies that $x_{n_i} \to x$. □

**Theorem.** $(X, d)$ is a compact *metric space* iff $X$ is sequentially compact.

*Proof.* ($\Rightarrow$) Suppose $x_n$ is a sequence in $X$ with no convergent subsequence. Then for any $y \in X$, there is no subsequence converging to $y$. By lemma, there exists $\varepsilon > 0$ such that $x_n \in B_\varepsilon(y)$ for only finitely many $n$.

Let $U_y = B_\varepsilon(y)$. Now $\mathcal{V} = \{U_y : y \in X\}$ is an open cover of $X$. Since $X$ is compact, there is a finite subcover $\{U_{y_1}, \cdots, U_{y_m}\}$. Then $x_n \in \bigcup_{i=1}^m U_{y_i} = X$ for only finitely many $n$. This is nonsense, since $x_n \in X$ for all $n$!

So $x_n$ must have a convergent subsequence.

($\Leftarrow$) This is much trickier. To prove this, we first prove the following claim:

**Claim.** A sequentially compact space contains a countable dense set.

*Proof.* Let $K$ be sequentially compact. Assume $K$ is infinite and note that $K$ must be bounded. We now inductively construct a dense sequence: Choose arbitrary $p_1$. Then having chosen $p_1, \cdots, p_n$, let $\delta_n = \sup_{p \in K} \min_{i \leq n} d(p, p_n)$. Then let $p_{n+1}$ be such a point such that $d(p_{n+1}, p_i) \geq \frac{\delta_n}{2}$. for all $i$. Now $\{p_n\}$ has a convergent subsequence, so we may find, for each $\epsilon > 0$ integers $m, n$ with $m < n$ and $d(p_m, p_n) < \epsilon$. So $\delta_{n-1} < 2\epsilon$. So each $p \in K$ is at most $2\epsilon$ away of some $p_i$. As $\epsilon$ is arbitrary, we have created a dense set. □

Now let $V_\alpha$ be an open cover of $K$. Let $\{p_n\}$ be a dense sequence, and consider the family $F$ of neighbourhoods $B_r(p_n)$ with $r \in \mathbb{Q}$ that are contained in some $G_\alpha$. Now $F$ is countable. We claim it is an open cover. Let $p \in K$ and $G_\alpha$ containing $p$. Then $G_\alpha \supset B_s(p)$ for some $s > 0$. As $\{p_n\}$ is dense, we have $d(p, p_n) < \frac{s}{2}$ for some $n$. Then for every rational $r$ with $d(p, p_n) < r < s - d(p, p_n)$. we have $p \in B_r(p_n) \subset B_s(p) \subset G_\alpha$. So $F$ is a countable cover of $K$.

Then we prove that every countable open cover has a finite subcover. Let $G_i$ be such a cover. Then we claim $K = \cup_{i=1}^n G_i$ for some $n$. Assume not. Let $p_n$ be a point in the complement of $\cup_{i=1}^n G_i$, and $p$ be a limit in a subsequence of $\{p_n\}$. Then $p \in K$ and $p \in G_N$. But for $n > N$, $p_n \notin G_N$ by construction, so contradiction. Thus, $F$ has a finite subcover, so $V_\alpha$ has a finite subcover. □