Analysis II

Notes made by Finley Cooper

20th January 2026

Contents

1 Uniform Convergence

For a subset $E \subseteq \mathbb{R}$, have a sequence $f_n : E \to \mathbb{R}$. What does it mean for the sequence (f_n) to converge? The most basic notion for any $x \in E$ require that the sequence of real numbers $f_n(x)$ to converge in $\mathbb{R}$. If this holds we can defined a new function $f : E \to \mathbb{R}$ by setting each value to the limit of the function.

Definition. (Pointwise limit) We say that (f_n) converges *pointwise* if for all x in its domain we have that

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

converges. We write that $f_n \to f$ pointwise.

Are properties such as continuity, differentiability integrability, preserved in the limit? We'll use an example to show that continuity is not preserved.

We can see this by taking a sequence of functions which converge to a step function by taking tighter and tighter curvers which get steeper and steeper. For example take,

$$f_n : [-1, 1] \to \mathbb{R}, \quad f_n(x) = x^{\frac{1}{2n+1}}.$$

So in the limit we get that

$$f_n(x) \to f(x) = \begin{cases} 1 & 0 < x \leq 1 \\ 0 & x = 0 \\ -1 & -1 \leq x < 0 \end{cases}$$

which is not continious.

For an example where integability is not preserved, let $q_1, q_2, q_3, \ldots$ be an enumeration of $\mathbb{Q} \cap [0, 1]$ and define

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

so we get $f_n(x)$ continious everywhere on $[0, 1]$ apart from a finite number of points, then f_n is integrable on $[0, 1]$ (IA Analysis I). But,

$$\lim_{n\to\infty} f_n(x) = \mathbf{1}_{\mathbb{Q}}(x)$$

which we know is not integrable.

If $f_n \to f$ pointwise, f_n integrable, f integrable, does it follow that $\int f_n \to \int f$? (Spoiler: No) For example take f_n to be a 'spike' with height n and width $\frac{2}{n}$, concretely,

$$f_n(x) = \begin{cases} n^2 x & 0 \leq x \leq \frac{1}{n} \\ n^2(\frac{2}{n} - x) & \frac{1}{n} \leq x \leq \frac{2}{n} \\ 0 & \text{otherwise} \end{cases}$$

So the integral of f_n over $[0, 1]$ is 1, but we can see that f_n converges pointwise to zero. So $\int_0^1 f_n \to 1$ but $\int_0^1 f \to 0$.

So we need a better (stronger) notion for the convergence of a sequence of functions. We can't use something too strong, such as $f_n \to f$ if f_n is eventually f for large enough n. We've got to find something inbetween. This is uniform convergence.

Definition. (Uniform convergence) Let $f_n, f : E \to \mathbb{R}$, for $n \in \mathbb{N}$. We say that (f_n) converges *uniformly* on E if the following holds. For all $\varepsilon > 0$, $\exists N = N(\varepsilon)$ such that for every $n \geq N$ and for every $x \in E$ we have that $|f_n(x) - f(x)| < \varepsilon$.

Remark. This statement is equivalent to the following,

$$\forall \varepsilon > 0, \exists N = N(\varepsilon), \text{ s.t. } \forall n \geq N, \sup_{x \in E} |f_n(x) - f(x)| < \varepsilon.$$

Comparing this to pointwise convergence, $\forall x \in E$ and $\forall \varepsilon > 0$, $\exists N = N(\varepsilon, x)$ such that $n \geq N \implies |f_n(x) - f(x)| < \varepsilon$. So we can change our N value for each individual x. However we can't in uniform convergence, which makes this is stronger statement.

Hence we see Uniform convergence $\implies$ Pointwise convergence. This gives a nice way to compute uniform limits. If a function doesn't converge pointwise then we know it doesn't converge uniformly. If we know a sequence of functions converges pointwise to some limit function, then this function must be the limit of the uniform limit, if it exists.

Definition. (Uniformly Cauchy) Let $f_n : E \to \mathbb{R}$ be a sequence of functions. We say that (f_n) is *uniformly Cauchy* on E if

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

Theorem. (Cauchy criterion for uniform convergence) Let (f_n) be a sequence of functions with $f_n : E \to \mathbb{R}$. The (f_n) converges uniformly on E if and only if (f_n) is uniformly Cauchy on E.

Proof. Suppose that (f_n) is a sequence converging uniformly in E to some function f. Given some $\varepsilon > 0$, there is a N such that $\sup_{x \in E} |f_n(x) - f(x)| < \varepsilon$ for all $n \geq N$. By the triangle inequality $\forall x \in E$, picking $n, m \geq N$,

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

hence (f_n) is uniformly Cauchy.
For the converse, suppose that (f_n) is a sequence uniformly Cauchy in E. Then the sequence of real numbers $(f_n(x))$ is Cauchy so by IA Analysis I, this sequence has a limit, call it $f(x)$. So (f_n) converges pointwise to f. Now we check that $f_n \to f$ uniformly on E. Pick any $\varepsilon > 0$ and note that by the hypothesis that (f_n) is uniformly Cauchy, there exists a number N such that for all $n, m \geq N$ we have $|f_n(x) - f_m(x)| < \varepsilon$. Fix $n \geq N$ and let $m \to \infty$ in this. So since $f_m(x)$ converges to $f(x)$ pointwise, we get that

$$|f_n(x) - f(x)| \leq \varepsilon$$

hence (f_n) converges uniformly in E. □

For an example consider $f_n : \mathbb{R} \to \mathbb{R}$ defined by $f_n(x) = \frac{x}{n}$. So $f_n \to 0$ pointwise on $\mathbb{R}$. But $|f_n - 0|$ is unbounded so the suprenum doesn't exist so f_n does not converge uniformly on $\mathbb{R}$. However if we restrict the domain of f_n to $[-a, a]$ then we get uniform convergence.

Theorem. (Continuity is preserved under uniform limits) Let $f_n, f : [a, b] \to \mathbb{R}$. Suppose that (f_n) converges to f uniformly on $[a, b]$. If $x \in [a, b]$ is such that f_n is continuous at x for all $n \in \mathbb{N}$, then f is continuous at x.

Proof. Let $\varepsilon > 0$ by uniform convergence of $f_n \to f$ we have some $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\sup_{y \in [a,b]} |f_n(y) - f(y)| < \varepsilon$$

. By continuity of f_N at x we have $\delta = \delta(N, x, \varepsilon) > 0$ s.t. $y \in [a, b], |x - y| < \delta \implies |f_N(y) - f_N(x)| < \varepsilon$.
Then $y \in [a, b], |x - y| < \delta$ we] have

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

Hence f is continuous at x. $\square$

It is instructive to see where this proof goes wrong if we only assume that (f_n) converges to f pointwise.

Corollary. (Uniform limits of continuous functions are continuous) If $f_n, f : [a, b] \to \mathbb{R}$, and $f_n \to f$ uniformly on $[a, b]$ and if f_n is continuous on $[a, b]$ for every n then f is continuous on $[a, b]$.

Proof. Immediate from the previous theorem. $\square$

From now on we will denote $C([a, b]) = \{f : [a, b] \to \mathbb{R} : f \text{ is continuous on } [a, b]\}$.

Theorem. Let (f_n) be a uniformly Cauchy sequence of functions in $C([a, b])$ the it converges to a function in $C([a, b])$.

Proof. Trivial from our theorems earlier proved. $\square$

Theorem. (Uniform convergence implies convergence of integrals) For $f_n, f : [a, b] \to \mathbb{R}$ be such that f_n, f are bounded and integrable on $[a, b]$. If $f_n \to f$ uniformly on $[a, b]$ then

$$\int_a^b f_n(x)\mathrm{d}x \to \int_a^b f(x)\mathrm{d}x$$

Remark. The assumption that f is integrable is redundant. We will see later that integrability of f_n implies that f is integrable if $f_n \to f$ uniformly

Proof.

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

by assumption.

1.1 Differentation and uniform convergence

This is more subtle if $f_n \to f$ uniformly on some interval and if f_n are differentiable it does not follow that

(i) That f is differentiable.

(ii) Even if f is differentiable that $f_n'(x) \to f(x)$.

We can view this in the example of $f_n : [-1,1] \to \mathbb{R}$ with $f_n(x) = |x|^{1+\frac{1}{n}}$. Hence we have that

$$\lim_{x\to 0} \frac{f_n(x) - f_n(0)}{x} = \lim_{x\to 0} \operatorname{sgn}(x^{\frac{1}{n}}) = 0$$

So f_n is differentialbe at 0 with $f_n(0) = 0$ and clearly f_n is differentiable everywhere where $x = 0$ too. We can check that $f_n \to |x|$ uniformly. But $|x|$ is not differentiable at $x = 0$.

Now consider the example $f_n : \mathbb{R} \to \mathbb{R}$ with

$$f_n(x) = \frac{\sin(nx)}{\sqrt{n}}.$$

So $f_n \to 0$ uniformly on $\mathbb{R}$. So we have a differentiable limit but $f_n'(x) = \sqrt{n}\cos(nx)$ which is not convergent as $n \to \infty$. So we don't have $f_n'(x) \to f'(x)$ pointwise on $\mathbb{R}$.

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

(i) $f_n' \to g$ uniformly for some function $g : [a,b] \to \mathbb{R}$.

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

Then (f_n) converges uniformly to some function $f : [a,b] \to \mathbb{R}$ where f is differentiable everywhere on $[a,b]$ and $f'(x) = g(x)$ for all $x \in [a,b]$.

This proves that

$$\left(\lim_{n\to\infty} f_n\right)' = \lim_{n\to\infty} f_n'$$

i.e. we can exchange the derivative and limit in this case.

Remark. If we assume that f_n' are continuous, then the proof is more straightforward and can be based on the fundamental theorem of calculus.

Proof. By the mean value theorem applied to the difference $(f_n - f_m)$ we have that for any $x \in [a, b]$

$$\begin{aligned} f_n(x) - f_m(x) &= f_n(c) - f_m(c) + (x - c)(f_n - f_m)'(x_{n,m}) \\ \implies |f_n(x) - f_m(x)| &\le |f_n(c) - f_m(c)| + (b - a)|f_n'(x_{n,m}) - f_m'(x_{n,m})| \\ \implies \sup |f_n - f_m| &< |f_n(c) - f_m(c)| + (b - a) \sup |f_n' - f_m'| \to 0 \end{aligned}$$

as $n \to \infty$. So (f_n) is uniformly Cauchy and hence there is an $f : [a, b] \to \mathbb{R}$ s.t. $f_n \to f$ uniformly.

For the next part fix some $y \in [a, b]$. Define

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

Now we only have to estabilish that h is continuous at y to show that f is differentiable at y with $f'(y) = g(y)$. Let

$$h_n(x) = \begin{cases} \frac{f_n(x)-f_n(y)}{x-y} & x \ne y \\ f_n'(y) & x = y \end{cases}$$

then since f_n is differentiable at y we see that h_n is continuous on $[a, b]$. The pointwise limit of (h_n) is h almost by definition since $f_n' \to g$ at $x = y$. Since the uniform limit of sequence of continuous functions is continuous, we just need to show that (h_n) is uniformly Cauchy on $[a, b]$ since the limit must be h since it converges pointwise to h.

$$h_n(x) - h_m(x) = \begin{cases} \frac{(f_n-f_m)(x)-(f_n-f_m)(y)}{x-y} & x \ne y \\ (f_n' - f_m')(y) & x = y \end{cases}.$$

By the mean value theorem,

$$\begin{aligned} &h_n(x) - h_m(x) = \begin{cases} (f_n - f_m)'(x_{n,m}) \text{ for some } x_{n,m} \text{ between } x \text{ and } y & x \ne y \\ (f_n - f_m)'(y) & x = y \end{cases} \\ &\sup_{[a,b]} |h_n - h_m| \le \sup_{[a,b]} |f_n' - f_m'| \to 0 \end{aligned}$$

as $n, m \to \infty$. So (h_n) is uniformly Cauchy so we're done. □

Remark. f_n' need not be continuous consider

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & x \ne 0 \\ 0 & x = 0 \end{cases}$$

the f is differentiable on $[-1, 1]$ with f'(x) not continuous at $x = 0$ and we can take $f_n(x) = f(x)$ for all n (or $f_n(x) = f(x) + \frac{x}{n}$.

We have a shorter proof of the above theorem, assuming that (f_n') are continuous in addition to the hypothesis. For any $x \in [a, b]$ we can write

$$f_n(x) = f_n(c) + \int_c^x f_n'(t)\mathrm{d}t$$

by FTC. Then

$$\begin{aligned}|f_n(x) - f_m(x)| &= \left| f_n(c) - f_m(c) + \int_c^x (f_n'(c) - f_m'(c))\mathrm{d}t \right| \\ &\leq |f_n(c) - f_m(c)| + \sup_{t \in [a,b]} |f_n'(t) - f_m'(t)|(b-a) \to 0\end{aligned}$$

as $n, m \to \infty$. So (f_n) is uniformly Cauchy, hence converges uniformly.

Note that

$$\int_c^x f_n'(t)\mathrm{d}t \to \int_c^x g(t)\mathrm{d}t$$

by uniform convergence of $f_n' \to g$ which implies g is continuous and hence also integrable. We can let $n \to \infty$ the first equation for $f_n(x)$ which gives that

$$f(x) = f(c) + \int_c^x g(x)\mathrm{d}t$$

So we can take the derivative of both sides giving that $f'(x) = g(x) = \lim f_n'(x)$. □

Proposition. If $f_n, g_n : E \to \mathbb{R}$ with $f_n \to f$ uniformly on E and $g_n \to g$ uniformly on E then $f_n + g_n$ converges uniformly to $f + g$ on E, and if $h : E \to \mathbb{R}$ is a bounded function then $hf_n \to hf$ uniformly on E also.

Proof. On the example sheet.

2 Series of functions

Definition. (Convergence of a series of functions) Let $g_n : E \to \mathbb{R}$ for $n \in \mathbb{N}$ then write

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

defined pointwise. Then we say that that,

(i) The series of functions $\sum_{n=1}^{\infty} g_n$ is convergent at a point $x \in E$ if the sequence of partial sums $(f_n(x))$ converges.

(ii) The series of functions $\sum_{n=1}^{\infty} g_n$ uniformly on E if the sequence (f_n) converges uniformly on E.

(iii) $\sum_{n=1}^{\infty} g_n$ converges absolutely at $x \in E$ if the series $\sum_{n=1}^{\infty} |g_n(x)|$ converges.

(iv) $\sum_{n=1}^{\infty} g_n$ converges absolutely uniformly on E if $\sum_{n=1}^{\infty} |g_n|$ converges uniformly on E.

We know from IA Analysis I that absolutely convergence $\implies$ convergence for a sequences in $\mathbb{R}$. From this we have that if $\sum_{n=1}^{\infty} g_n$ converges absolutely at a point $x \in E$ then $\sum_{n=1}^{\infty} g_n$ converges at x. Similiar to this we have the following proposition relating absolute uniform convergence and uniform convergence.

Proposition. (Absolute uniform convergence implies uniform convergence) If $g_n : E \to \mathbb{R}$ and if $\sum_{n=1}^{\infty} g_n$ converges absolutely uniformly on E then $\sum_{n=1}^{\infty} g_n$ converges uniformly on E.

Proof. Let $f_n = \sum_{i=1}^{n} g_i$ Then

$$\begin{aligned}|f_n(x) - f_m(x)| &= \left|\sum_{i=m+1}^{n} g(i)\right| \\ &= \sum_{i=m+1}^{n} |g_i(x)| = h_n(x) - h_m(x), \quad \text{where } h_n(x) = \sum_{i=1}^{n} |g_i(x)| \\ \sup_{x\in E} |f_n(x) - f_m(x)| &\le \sup_{x\in E} |h_n(x) - h_m(x)| \to 0\end{aligned}$$

as $n, m \to \infty$ so (f_n) converges uniformly on E. □

Remark. Uniform convergence and absolute pointwise convergence aren't enough to conclude that the series convergence absolutely uniformly.

Theorem. (Weierstrass M-test) Let $g_n : E \to \mathbb{R}$ be a sequence of functions and suppose that $\exists M_n$ such that

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

and that

$$\sum_{n=1}^{\infty} M_n$$

converges. Then

$$\sum_{n=1}^{\infty} g_n$$

converges absolutely uniformly on E.

Proof. Let

$$h_n(x) = \sum_{j=1}^{n} |g_n(x)|$$

for $n > m$,

$$\begin{aligned}h_n(x) - h_m(x) = \sum_{j=m+1}^{n} |g_j(x)| &\le \sum_{j=k+1}^{n} M_j = \sum_{j=1}^{n} M_j - \sum_{j=1}^{m} M_j \\ \implies \sup_{x\in E} |h_n(x) - h_m(x)| &\le \left|\sum_{j=1}^{n} M_j - \sum_{j=1}^{m} M_j\right| \quad \forall n, m\end{aligned}$$

by assumption the right hand side $\to 0$ since $\sum_{j=1}^{\infty} M_j$ is convergent, hence (h_n) is uniformly Cauchy hence converges uniformly.

2.1 Power series

We'll now specialise to the case where $g_n(x) = c_n(x-a)^n$ for $a, c_n \in \mathbb{R}$. This gives a real power series.

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

(i) If $|x-a| < R$ then the power series converges absolutely.

(ii) If $|x-a| > R$ then the power series diverges.

(iii) R is given by

$$R = \frac{1}{\limsup_{n\to\infty} |c_n|^{\frac{1}{n}}}$$

where if the limit is zero, then $R = \infty$.

(iv) For any $r \in (0, R)$ we have the power series converges uniformly on $[a-r, a+r]$, in particular the function that the power series converges to is continuous on $(a-R, a+R)$.

Proof. The proof for (i), (ii), and (iii) are in IA Analysis I. We'll just prove (iv). Note first that the power series converges absolutely at $x = a + r$ i.e. we have that

$$\sum_{n=0}^{\infty} |c_n| r^n$$

is convergent. Since $|c_n(x-a)^n| \leq |c_n| r^n$ for any $x \in [a-r, a+r]$ we can apply the Weierstrass M-test with $M_n = |c_n| r^n$ to conclude that the series

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

converges absolutely uniformly on $[a-r, a+r]$. It follows that f is continuous. at any point in $(a-R, a+R)$ by picking r small enough.

Remark. (Boundary behaviour. Let

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

with power series boundary R with $0 < R < \infty$. If the power series converges at one of the boundary points of the interval of convergence, say at $x = a + R$ i.e. $\sum_{n=0}^{\infty} c_n R^n$ is convergent then

$$\lim_{x\to a+R} f(x) = \sum_{n=0}^{\infty} c_n R^n$$

so f extends to $(a-R, a+R]$ as a continuous function.

Moreover, under the same conditions that $\sum_{n=0}^{\infty} c_n R^n$ converges we have that the series converges uniformly on $[a-r, a+r]$ for any $r \in (0, R)$. Same discussion applies at the endpoint $a - R$.

Theorem. (Differentation of power series) Let $\sum_{n=0}^{\infty} c_n(x-a)^n$ be a power series with radius of convergent $R > 0$. Let

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

defined on $(a-R, a+R)$. We have the following

(i) The derived series

$$\sum_{n=1}^{\infty} nc_n(x-a)^{n-1}$$

has radius of convergent R.

(ii) f is differentiable on $(a-R, a+R)$ with

$$f'(x) = \sum_{n=1}^{\infty} nc_n(x-a)^{n-1} \quad \forall x \in (a-R, a+R)$$

Proof. Before we prove the theorem let's give a definition we've seen slightly before.

Definition. If (a_n) is a sequence of reals let

$$p_n = \sup\{a_m : m \geq n\}$$
$$q_n = \inf\{a_m : m \geq n\}.$$

Then we define

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} p_n$$
$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} q_n.$$

which exists in $\mathbb{R} \cup \{\infty\}$ since (q_n) and (p_n) are monotone.

$$\limsup_{n\to\infty}(n|c_n|)^{\frac{1}{n}} = \limsup_{n\to\infty} |c_n|^{\frac{1}{n}}$$

since we have that $\lim_{n\to\infty} n^{\frac{1}{n}} = 1$. So we have (i).

Define $f_n(x) = \sum_{j=0}^{j} c_j(x-a)^j$ is clearly differentiable on $\mathbb{R}$ with $f_n'(x) = \sum_{j=1}^{n} jc_j(x-a)^{j-1}$. By (i) we have that $f_n'(x)$ converges uniformly on $[a-r, a+r]$ for all $r < R$ and $f_n(a) = c_0 \forall n$ so $(f_n(a))$ converges. So the limit is differentiable in $[a-r, a+r]$, with

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

. $\square$

If we have a power series $\sum_{n=1}^{\infty} c_n(x-a)^n$ we say the power series converges *locally uniformly* on the interval of convergence $(a-R, a+R)$ i.e. for all $0 < r < R$ the power series converges uniformly on $[a-r, a+r]$.

Remark. By repeatedly applying the above theorem we get that if $f(x) = \sum_{n=1}^{\infty} c_n(x-a)^n$ has radius of convergence $R > 0$ then f is differentiable to any order $k \in \mathbb{N}$ in $(a-R, a+R)$ and the kth derivative is given by

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)c_n(x-a)^{(n-k)}.$$

Plugging in $x = a$ we get that

$$c_n = \frac{f^{(k)}(a)}{k!}.$$

This says that f is uniquely determined by its values in an arbitrarily small interval around the point $x = a$ since that's all we need to capture it's derivatives and form its power series.

3 Uniform continuity and Riemann integrability

3.1 Uniform continuity

Definition. (Uniform continuity) Let $E \subseteq \mathbb{R}$ and let $f : E \to \mathbb{R}$. We say that f is *Uniformly continuous* on E if $\forall \varepsilon > 0$ there exists a $\delta > 0$ such that $\forall x, y \in E$ we have that

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

This differs from our usual definition of continuity. We require some δ to work for *any* $x, y \in E$ given some ε, rather than picking a δ for each ε and x value. Clearly uniform continuity implies continuity but the converse is not true. For an example consider $f(x) = \frac{1}{x}$ on $(0,1)$. Clearly continuous at each x, but not uniformly continuous since it gets too steep around 0.

Not even boundedness and continuity is enough for uniform continuity, consider $\sin\left(\frac{1}{x}\right)$, take $x_n = \frac{1}{2n\pi}$ and $y_n = \frac{1}{2n+\frac{1}{2})\pi}$ then $|f(x) - f(y)| = 1$, so no δ works, we can always choose an n large enough.

Theorem. Let $[a,b]$ be a closed, bounded interval and $f : [a,b] \to \mathbb{R}$ a continuous function. Then f is uniformly continuous.

Proof. Argue by contradiction. Suppose that f is not uniformly continuous, so there exists an $\varepsilon > 0$ such that for all $\delta > 0$ there is a pair of points $x, y \in [a,b]$ such that $|y-x| < \delta$ but $|f(x) - f(y)| \geq \varepsilon$. Now let $\delta_n = \frac{1}{n}$, so we get a sequence of functions x_n and y_n satisfying the above for each δ_n. By Bolzano-Weiestrass, there exists a subsequence (x_{n_k}) that converges to a point $x \in [a,b]$.

$$|x - y_{n_k}| \leq |x - x_{n_k}| + |x_{n_k} - y_{n_k} \leq |x - y_{n_k}| + \frac{1}{n_k} \to 0 \text{ as } n \to \infty$$

By the continuity of f at x we get $f(x_{n_k}) \to f(x)$ and $f(y_{n_k}) \to f(x)$. But this is contradiction since $f(x)$ and $f(y_{n_k})$ are always seperated by some distance ε. $\square$

We can actually strengthen this theorem.

Theorem. Let $f : [a,b] \to \mathbb{R}$ where $-\infty < a < b < \infty$ be any function. Suppose that there is a collection $\mathcal{C}$ of open intevals $I \subseteq \mathbb{R}$ such that if

$$F = [a,b] \setminus \bigcup_{I \in \mathcal{C}} I$$

then f is continuous at every point in F (i.e. the set of discontinuities is contained in the union). Then $\forall \varepsilon > 0 \exists \delta > 0$ s.t. $x \in F, y \in [a,b]$, with $|x-y| < \delta \implies |f(x) - f(y)| < \varepsilon$.

Proof. Same as above, using the fact that F is *closed* so it contains all of its limit points.

Let's show some applications of uniform continuity.

3.2 Riemann Integration

We'll do a quick recap of Riemann integration. For full proofs, look at IA Analysis I. Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Say that $m \leq f(x) \leq M$ for $m, M \in \mathbb{R}$. Let $P = \{a_0 = a, a_1, a_2, \ldots, a_n = b\}$ be a partition of the interval $[a,b]$ with $a_0 < a_1 < \cdots < a_n$. We will write $P = \{a_0 = a < a_1 < \cdots < a_n = b\}$ as shorthand.

We write that $I_j = [a_j, a_{j+1}]$ for $0 \leq j < n$. Define the upper sum of f with P as

$$U(P,f) = \sum_{j=0}^{n-1} (a_{j+1} - a_j) \sup_{I_j} f$$

and the lower sum of f with P as

$$L(P,f) = \sum_{j=0}^{n-1} (a_{j+1} - a_j) \inf_{I_j} f.$$

We can see immediately that $m(b-a) \leq L(P,f) \leq U(P,f) \leq M(b-a)$. When we refine the partition by adding finitely many new points the uppers sum decreases or stays the same, and the lower sum increases or stays the same. So now we can define the upper and lower Riemann integral as

$$I^*(f) = \inf_P U(P,f)$$
$$I_*(f) = \sup_P L(P,f).$$

We say that f is Riemann integrable if $I^*(f) = I_*(f)$. We denote

$$\int_a^b f(x) \mathrm{d}x$$

as this common value.

Theorem. (Riemann criterion for integrability) For $f : [a,b] \to \mathbb{R}$ bounded, f is integrable if and only if for all $\varepsilon > 0$ there exists a partition P of $[a,b]$ such that

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

Proof. In IA Analysis I.

Theorem. Let $f : [a, b] \to [A, B]$ be integrable and $g : [A, B] \to \mathbb{R}$ continuous. Then the composite function $g \circ f : [a, b] \to \mathbb{R}$ is integrable.

We may ask does this hold is we switch the order? i.e. given the both conditions is $f \circ g$ always be integrable?

Proof. Since g is continuous in a bounded interval, it is uniformly continuous. Given any $\varepsilon > 0$ there is a δ such that $x, y \in [A, B]$ with $|y - x| < \delta \implies |g(x) - g(y)| < \varepsilon$. We also have by integrability that there exists a partition P such that $U(P, f) - L(P, f) < \varepsilon'$ for all $\varepsilon' > 0$.

$$U(P, g \circ f) - L(P, g \circ f) = \sum (a_{j+1} - a_j) \left(\sup_{I_j} g \circ f - \inf_{I_j} g \circ f \right)$$

Take $J = \left\{ j : \sup_{I_j} f - \inf_{I_j} f \leq \delta \right\}$. For any $j \in J$ for all $x, y \in I_J$ we must have that

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

Hence we get that

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

so

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

which gives that

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

hence it suffices to make the sum over the js not in J small enough. We know that

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

so if we pick $\varepsilon' = \varepsilon\delta$ we get that

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

Corollary. If f is continuous then it is integrable

Proof. Apply the theorem with $g = \text{id}$ which is clearly integrable. $\square$

Theorem. (Uniform limits of integrable functions are integrable) Suppose we have $f_n : [a,b] \to \mathbb{R}$ be a sequence of Riemann integrable functions and $f_n \to f$ uniformly. Then f is bounded, Riemann integrable and

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

Proof.

$$\sup_{[a,b]} |f| \leq \sup_{[a,b]} |f - f_n| + \sup_{[a,b]} |f_n| \leq 1 + \sup_{[a,b]} |f_n|$$

for n sufficently large (setting $\varepsilon = 1$). Hence f is bounded.

Let $P = \{a_0, \dots, a_m\}$ be a partition of $[a,b]$. Given some $\varepsilon > 0$ and consider

$$\begin{aligned}
U(P,f) - L(P,f) &= \sum_{j=0}^{m-1} (a_{j+1} - a_j) \left(\sup_{I_j} f - \inf_{I_j} f \right) \\
&= \sum_{j=0}^{m-1} (a_{j+1} - a_j) \left(\sup_{I_j} (f - f_n + f_n) - \inf_{I_j} (f - f_n + f_n) \right) \\
&\leq \sum_{j=0}^{m-1} (a_{j+1} - a_j) \left(\sup_{I_j} (f - f_n) + \sup_{I_j} (f_n) - \inf_{I_j} (f - f_n) - \inf_{I_j} (f_n) \right) \\
&\leq U(P, f_n) - L(P, f_n) + 2(a-b) \sup_{[a,b]} |f - f_n|
\end{aligned}$$

So for our $\varepsilon > 0$ choose some N such that $2(b-a)\sup_{[a,b]} |f - f_N| \leq \frac{\varepsilon}{2}$ by uniform convergence. Now also choose a partition P such that $U(P, f_N) - L(P, f_N) < \frac{\varepsilon}{2}$ since f_N is Riemann integrable. Hence $U(P,f) - L(P,f) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$ for any $\varepsilon > 0$ so f is integrable by the Riemann criterion. The last part have been proved previously in the course. $\square$

Non-examinable We'll now prove an equivalent condition for a function to be Riemann integrable. First we'll set up some frameworks. For a function $f : [a,b] \to \mathbb{R}$ bounded, we use $\mathcal{D}_f$ to denote its set of discontinuities. We know that there are functions with $\mathcal{D}_f$ non-empty which are still Riemann integrable, such as Thomae's function which has $\mathcal{D}_f = \mathbb{Q}$. We also know that all monotone functions are integrable. What condition on $\mathcal{D}_f$ do we need for integrability?

Definition. (Null set) A subset $\mathcal{R} \subseteq \mathbb{R}$ is said to be a *null set* (or a set of *Lebesgue measure zero*) if $\forall \varepsilon > 0$ there exists an at most countable collection of open intevals $I_j = (a_i, b_i)$ such that

$$\mathcal{D} \subseteq \bigcup_{i=1}^{n} I_i$$

and

$$\sum_{j=1}^{\infty} |I_j| \leq \varepsilon$$

where $|I_j| = b_j - a_j$.

We have a few examples of null sets.

(i) The empty set and singleton sets are null.

(ii) Any subset of small enough sets are null.

(iii) Any countable union of null sets is null (namely $\mathbb{Q}$ is a null set and any other countable set like the algebraic numbers).

(iv) The (standard) Cantor set is a null set even though it's uncountable.

(v) However not every set is a null set, every (open or closed) interval is not a null set.

Now for the big theorem completely characterising Riemann integrable functions.

Theorem. (Lebesgue's theorem on the Riemann integral) Let $f : [a, b] \to \mathbb{R}$ bounded. Then f is Riemann integrable if and only if $\mathcal{D}_f$ is a null set.

Proof. See Part II Probability and Measure.

Remark. Many results on Riemann integration are direct corollaries from Lebesgue's thereom. For example from IA Analysis I Example Sheet 3 we know that the set of discontinuities for a monotone function is countable. Hence for a monotone function $\mathcal{D}$ is a null set and thus f is Riemann integrable.
Also if $f : [a, b] \to [A, B]$ is integrable, and $g : [A, B] \to \mathbb{R}$ is continuious, then $g \circ f$ is integrable can be proved too. Clearly, $g \circ f$ is bounded, since g is bounded. Since f is integrable, we know that $\mathcal{D}_f$ is null. But $\mathcal{D}_{g\circ f} \subseteq \mathcal{D}_f$ (since if f is continuous at x then $g \circ x$ is continuous at x since g is continuous. Hence $\mathcal{D}_{g\circ f}$ is null, so $g \circ f$ is integrable by Lebesgue's theorem.
Finially if we have a sequence $f_n : [a, b] \to \mathbb{R}$ integrable converging uniformly to f, then f is integrable. We can also prove this using Lebesgue's theorem since $\mathcal{D}_{f_n}$ is null by Lebesgue's theorem for all $n \in \mathbb{N}$, so $\bigcup_{n\in\mathbb{N}} \mathcal{D}_{f_n}$ is null too. But $\mathcal{D}_f \subseteq \bigcup_{n\in\mathbb{N}} \mathcal{D}_{f_n}$ (Example Sheet 1), so it's also null, hence f is integrable.

Here's a new result that can be deduced from Lebesgue's theorem.

Corollary. If $f : [a, b] \to \mathbb{R}$ is integrable, then $|f|$ is integrable on $[a, b]$. Moreover

$$\int_a^b |f| = 0 \iff f = 0 \text{ except on a null set}$$

So there exists a null set $N \subseteq [a, b]$ such that $f(x) = 0$ for all $x \in [a, b] \setminus N$. We say that $f = 0$ *almost everywhere* on $[a, b]$.

Proof. Exercise. □

This concludes our non-examinable interlude.

4 Metric and Normed Spaces

Definition. (Metric Space) Let X be any set. A *metric* (or distance function) on X is a function, $d : X \times X \to \mathbb{R}$ satisfying the following for any $x, y, z \in X$,

(i) $d(x, y) \geq 0$ and $d(x, y) = 0 \iff x = y$;

(ii) $d(x, y) = d(y, x)$;

(iii) $d(x, y) \leq d(x, z) + d(z, y)$.

We call such a pair (X, d) a *metric space.*

Definition. (Normed Space) Let V be a real vector space. A *norm* on V is a function $|| \cdot || : V \to \mathbb{R}$ satisfying for any $x, y \in V$, and any $\lambda \in \mathbb{R}$,

(i) $||x|| \geq 0$ and $||x|| = 0 \iff x = 0$;

(ii) $||\lambda x|| = |\lambda| \cdot ||x||$;

(iii) $||x + y|| \leq ||x|| + ||y||$.

We say that $(V, || \cdot ||)$ is a *normed space.*

Proposition. If $(V, || \cdot ||)$ is a normed space, and if $d : V \times V \to \mathbb{R}$ is defined by $d(x, y) = ||x - y||$, then (V, d) is a metric space.

Proof. Exercise. □

Let's go over a few examples.

(i) (Finite dimensional normed spaces) The prototypical example of a normed space is the Euclidean space $\mathbb{R}^n = \{(x_1, x_2, \dots\} : x_i \in \mathbb{R}\}$, with $n \in \mathbb{N}$ fixed. We know that this defines a vector space and we can define various different norms on the vector space. Taking $V = \mathbb{R}^n$ with its usual vector space structure, we can define several useful norms on V:

(a) The Euclidean norm (or the ℓ_2-norm, defined by

$$||x||_{\ell_2} = ||x||_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}.$$

All the requirements are simple to prove apart from the triangle inequality.

$$\begin{aligned} ||x - y||_2^2 &= \sum_{i=1}^{n} (x_i + y_i)^2 \\ &= ||x||_2^2 + ||y||_2^2 + 2 \sum_{i=1}^{n} x_i y_i \\ &\leq ||x||_2^2 + ||y||_2^2 + 2||x||_2^2 ||y||_2^2 \quad \text{by Cauchy-Schwarz} \\ &= ||x||_2^2 + ||y||_2^2 \end{aligned}$$

(b) The ℓ_1-norm on $\mathbb{R}^n$ defined as $||x||_1 = \sum_{i=1}^{n} |x_i|$.

(c) We can also define the ℓ_∞-norm as $||x||_\infty = \sup\{|x_i| : 1 \leq i \leq n\}$.

Remark. More generally, for $x \in \mathbb{R}^n$ we can define the ℓ_p-norm as

$$||x||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)$$

and it turns out that for $p \leq 1$ this is indeed a norm (however the triangle inequality is non-trivial to proof). Moreover if we let $p \to \infty$ we recover $||x||_\infty$.

(ii) ($\mathbb{R}^\mathbb{N}$) Let's look at the infinite sequences of real numbers. We write $\mathbb{R}^\mathbb{N}$ for the set of real sequences $(x_k)_{k\in\mathbb{N}}$. This is vector space under addition defined by termwise addition and scalar multiplication.

(a) We look at the space $\ell_1 = \{(x_k) \in \mathbb{R}^\mathbb{N} : \sum_{k=1}^\infty |x_i| < \infty\}$. This is a linear subspace of $\mathbb{R}^\mathbb{N}$. We can turn this into a normed space by defining,

$$||x||_{\ell_1} = ||x||_1 = \sum_{k=1}^{\infty} |x_k|.$$

(b) Likewise, $\ell_2 = \{(x_k) \in \mathbb{R}^\mathbb{N} : \sum_{i=1}^\infty x_i^2 < \infty\}$ is a linear subspace of $\mathbb{R}^\mathbb{N}$ and define the ℓ_2 norm as

$$||x||_{\ell_2} = ||x||_2 = \left(\sum_{i=1}^{\infty} |x_i|^2 \right)^{\frac{1}{2}}.$$

(c) We can also define $\ell_\infty = \{(x_k) \in \mathbb{R}^\mathbb{N} : \sup_{k\geq 1} |x_k| < \infty\}$ i.e. the space of bounded sequences. We can define define the ℓ_∞ norm as

$$||x||_{\ell_\infty} = ||x||_\infty = \sup_{k\geq 1} |x_k|.$$

Remark. More generally, let $\ell_p = \{(x_k) \in \mathbb{R}^\mathbb{N} : \sum_{i=1}^\infty |x_i|^p \leq \infty$. Then ℓ_p is a subspace of $\mathbb{R}^\mathbb{N}$, and the ℓ_p-norm is defined as

$$||x||_p = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{\frac{1}{p}}.$$

And again we can see that the ℓ_∞ norm is the limit of the ℓ_p norm as $p \to \infty$.

(iii) (The space of continuous functions in a bounded, closed interval) Define a vector space as

$$V = C([a, b]) = \{f : [a, b] \to \mathbb{R} : f \text{ is continuous}\}.$$

This forms a vector space under pointwise addition and scalar multiplication of functions.

(a) The L^1-norm

$$||f||_{L^1([a,b])} = ||f||_{L^1} = ||f||_1 = \int_a^b |f(x)|\mathrm{d}x.$$

(b) The L^2-norm

$$||f||_{L^2} = ||f||_2 = \left(\int_a^b |f(x)|^2 \mathrm{d}x \right)^{\frac{1}{2}}.$$

(c) The L^∞-norm or the *uniform norm*

$$||f||_{L^\infty} = ||f||_\infty = \sup_{x\in[a,b]} |f(x)|.$$

It is easy to check that these are norms, we just need to know the Cauchy-Schwarz theorem for integrals to prove the triangle inequality for L^2.

$$\int_a^b |f\cdot g| \leq \left(\int_a^b |f|^2\right)^{\frac{1}{2}} \left(\int_a^b |g|^2\right)^{\frac{1}{2}}.$$

Which we can prove by considering,

$$\varphi(t) = \int_a^b (|f| - t|g|)^2 \geq 0,$$

which is a quadratic in t, so using the fact that the discriminant is non-positive (with the $g = 0$ case being trivial).

Remark. By the previous proposition, all of these examples are naturally metric spaces. For example the *Euclidean metric* is $d_E(x,y) = ||x-y||_2$ on $\mathbb{R}^n$ for $x, y \in \mathbb{R}^n$ and the *uniform metric* on $C([a,b])$, $d(f,g) = ||f-g||_\infty = \sup_{x\in[a,b]} |f(x)-g(x)|$.

Remark. Integral norms such as L^1, are more naturally defined on the larger space of integrable functions, $\mathcal{R}([a,b])$ which is a vector space under pointwise addition and scalar multiplication, where we have that $C([a,b]) \subseteq \mathcal{R}([a,b])$. However we have a problem since there are functions in $\mathcal{R}([a,b])$ which have norm of zero but aren't the zero function. But by a previous corollary of Lebesgue's theorem, we have that

$$\int_a^b |f| = 0 \implies f = 0 \text{ almost everywhere on } [a,b].$$

So we can turn $\mathcal{R}([a,b])$ into a normed space by defining an equivalence relation on the space by setting $f \sim g$ if $f = g$ almost everywhere. Then the space $\mathcal{R}([a,b])/\sim$ is now a normed vector space. Everything is well-defined independent of equivalence class representatives by the corollary of Lebesgue's theorem.

(iv) (Discrete metric) For any set X let

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

(v) (New metrices from given ones) Let (X,d) be a mretric space

(a) Define $g : X \times X \to \mathbb{R}$ by $g(x,y) = \min\{1, d(x,y)\}$.

(b) $h : X \times X \to \mathbb{R}$ by $h(x,y) = \frac{d(x,y)}{1+d(x,y)}$.

(c) Take $X = \mathbb{R}^2$, then define

$$d(x,y) = \begin{cases} ||x-y||_2 & \text{if } x = ty \text{ for some } t \in \mathbb{R} \\ ||x||_2 + ||y||_2 & \text{otherwise} \end{cases}.$$

This is the French railways metric or the SNCF metric.

Definition. (Metric subspaces) If (X, d) is a metric space, let $Y \subseteq X$ be any subset. Then the restriction

$$d\,|_{Y \times Y} : Y \times Y \to \mathbb{R}$$

is a metric on Y called the *induced metric* or the *subset metric*.

4.1 Open and closed subsets

We will now look at the very important definitions of open and closed subsets in a metric space.

Definition. (Open ball) Let (X, d) be a metric space. Then for any $a \in X$ and any $r > 0$. The *open ball* with radius r and centre a is the set

$$B_r(a) = \{x \in X : d(x, a) < r\}.$$

This is our abstraction of ε-neighbourhoods on $\mathbb{R}$ for general metric spaces.

Definition. (Open set) Let (X, d) be a metric space. Then a subset $U \subseteq X$ is *open* if for all $a \in U$ there exists a radius $r > 0$ such that $B_r(a) \subseteq U$.

Definition. (Closed set) Let (X, d) be a metric space. Then a subset $E \subseteq X$ is *closed* if $X \setminus E$ is open.

The property of being open or closed for a subset is relative to the containing ambient space. For example consider $X = \mathbb{R}$ with the Euclidean metric. Then consider $Y = [0, 1) \cup \{2\}$ with the induced metric. We can see that $[0, 1)$ is neither open or closed in X. Looking at Y however $[0, 1)$ is both open and closed in Y.

Proposition. Let (X, d) be a metric space. Then
- (i) Any open ball $B_r(a)$ is an open set;
- (ii) Any singleton $\{x\}$, $x \in X$ is closed.

Proof. Let $y \in B_r(a)$, let $r_1 = r - d(y, a)$. Then $r_1 > 0$ since $y \in B_r(a)$, so $d(y, a) < r$. Take some $z \in B_{r_1}(y)$. So

$$d(z, a) \leq d(z, y) + d(y, a) < r_1 + d(y, a) = r,$$

hence $z \in B_r(a)$, so $B_{r_1}(y) \leq B_r(a)$.

For the second part, take a point $z \in X \setminus \{x\}$, then $r = d(x, z) > 0$. So $x \notin B_r(z)$, therefore $B_r(z) \subseteq X \setminus \{x\}$. Hence $X \setminus \{x\}$ is open, so $\{x\}$ is closed. □

Theorem. Let (X, d) be a metric space. We have the following,
- (i) The union of any (possibly uncountable) collection of open set is open.
- (ii) The intersection of any finite collection of open sets is open.
- (iii) The empty set, $\emptyset$, and the whole set, X, are both open.

Proof. Exericse. □

By taking complements of sets we get th By taking complements of sets we get the corresponding theorem.

Theorem. Let (X, d) be a metric space. We have the following,

(i) The intersection of any (possibly uncountable) collection of closed sets is closed.

(ii) The union of any finite collection of closed sets is closed.

(iii) The empty set, $\emptyset$, and the whole set, X, are both closed.

Note that the "finite" is important for part (ii) of both theorems. For example consider the metric space $\mathbb{R}$ over the Euclidean metric, then

$$\bigcap_{n=1}^{\infty}\left(-\frac{1}{n}, \frac{1}{n}\right) = \{0\}$$

which is not open. Similarly we have that

$$\bigcup_{n=1}^{\infty}\left(\mathbb{R} \setminus \left(-\frac{1}{n}, \frac{1}{n}\right)\right) = \mathbb{R} \setminus \{0\}$$

which is not closed.

Definition. (Convergence of sequences in metric spaces) Let (X, d) be a metric space. A sequence (x_k) in X is said to converge to a point $x \in X$ if $d(x_k, x) \to 0$ as $k \to \infty$. So $\forall \varepsilon > 0$, $\exists N$ such that $k > N \implies d(x_k, x) < \varepsilon \iff x_k \in B_\varepsilon(x)$.

It's clear from the ε definition if $x_k \to x$ we must have that $x_{n_k} \to x$ for any subsequence x_{n_k}.

Proposition. (Uniquness of the limits) Let (X, d) be a metric space, and (x_k) be a sequence in X with $x_k \to x$ and $x_k \to y$, then $x = y$.

Proof. We have that

$$d(x, y) \le d(x_k, y) + d(x_k, x) \to 0$$

as $k \to \infty$. Hence $d(x, y) = 0$, so $x = y$. □

Remark. It is possible that the same sequence in X has different limits with respect to different metrices on X. For example take $X = \mathbb{R}$ and let $f : \mathbb{R} \to \mathbb{R}$ with

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

Then if we take $d_1(x, y) = |f(x) - f(y)|$. Then the sequence $x_k = \frac{1}{k} \to 0$ in the Euclidean metric, but $x_k \to 1$ in the d_1 metric. This example seems a bit contrived, but this can happen with respect to two different norms (only if the normed space is infinite dimensional). This is because any two norms on a finite dimensional vector space are Lipschitz equivalent (more to come later).

Proposition. (Convergence in $\mathbb{R}^n$ with the ℓ_2 norm) Convergence in $\mathbb{R}^n$ with respect to the Euclidean norm is equivalent to the convergence of the coordinates (as real numbers). Formally, if $x^{(k)} = \left(x_1^{(k)}, \dots, x_n^{(k)}\right)$ is a sequence in $\mathbb{R}^n$ with $k \in \mathbb{N}$, and $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, then

$$x^{(k)} \to x \text{ in } ||\cdot||_2 \iff x_j^{(k)} \to x_j \ \in \{1, 2, \dots, n\}.$$

Proof. Fix some $\varepsilon > 0$ The there exists some N such that $k \geq N \implies ||x^{(k)} - x||_2 < \varepsilon$. So

$$\sum_{j=1}^{n} (x_j^{(k)} - x_j)^2 < \varepsilon^2$$

so $\left|x_j^{(k)} - x_j\right| < \varepsilon$ for all j with $k \geq N$.

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

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

So we're done. $\square$

Remark. The convergence in $(C([a,b]), ||\cdot||_\infty$ in just uniform convergence as we've seen earlier.

Definition. (Bounded subset) Let (X, d) be a metric space. A subset $E \subseteq X$ is *bounded* if $E \subseteq B_R(a)$ for some $a \in X$ and some $R > 0$.

Theorem. (Bolzano-Weierstrass in $\mathbb{R}^n$) Every bounded sequence in $\mathbb{R}^n$ with respect to the Euclidean metric has a convergent subsequence.

Proof. Proceed by induction on n. We know the base case on $\mathbb{R}$ holds by standard Bolzano-Weierstrass in IA Analaysis I. Let $n \geq 2$, and assume by induction that the theorem holds in $\mathbb{R}^{n-1}$. Let $(x^{(k)}$ be a bounded sequence in $\mathbb{R}^n$, say $||x^{(k)}||_2^2 \leq R^2$ for some R and all k. Write $x^{(k)} = (x_1^{(k)}, \dots, x_{n-1}^{(k)}, x_n^{(k)})$ and let $y^{(k)} = (x_1^{(k)}, \dots, x_{n-1}^{(k)}) \in \mathbb{R}^{n-1}$. So $||y^{(k)}||_2^2 + |x_n^{(k)}|^2 \leq R^2$. So $y^{(k)}$ is a bounded sequence in $\mathbb{R}^{n-1}$, hence by the induction hypothesis, there exists a subsequence (k_j) of (k) and a point $y \in \mathbb{R}^{n-1}$ such that $y^{(k_j)} \to y$. Also by Bolzano-Weierstrass in $\mathbb{R}$, there is a further subsequence $\left(x_n^{(k_{j_\ell})}\right)$ of $(x_n^{k_j})$ that converges to say $y_n \in \mathbb{R}$. Then we know that

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

Hence we're finished. $\square$

Let's show an example where Bolzano-Weierstrass doesn't hold in the infinite dimensional case. Let's look at the the metric space $(\ell^\infty, ||\cdot||_\infty)$. If we let $e_j^{(k)} = \delta_{jk}$ be the sequence with a 1 in the kth component and 0 all other components which is clearly bounded. We know that $e_k^{(k)} \to 0$ for all fixed j, and hence $e^{(k)}$ converges componentwise to the zero sequence. However $e^{(k)}$ does not converge to the zero element since $||e^{(k)} - 0||_\infty = 1$ for all k. Hence this also doesn't have a convergent subsequence for the same reason.

Remark. In fact for normed spaces $(V, ||\cdot||)$, the Bolzano-Weierstrass property (every bounded sequence has a convergent subsequence) is equivalent to the space being finite dimensional.

Definition. (Limit point) If (X, d) is a metric space and we have a subset $E \subseteq X$ and a point $x \in X$, then say that x is a *limit point* of E if there is a sequence $(x_k) \in E$ with $x_k \neq x$ for all k and $x_k \to x$.

Definition. (Isolated point) If (X, d) is a metric space and we have a subset $E \subseteq X$ then $x \in E$ is a *isolated point* of E if $x \in E$ and x is not a limit point of E. Equivalently there exists a $r > 0$ such that $E \cap B_r(x) = \{x\}$.

Definition. (Closure) If (X, d) is a metric space and we have a subset $E \subseteq X$ then the *closure* of E denoted as $\bar{E}$ is the union of E and all of its limit points. (i.e. it's all the points which are the limit of some sequence in E.)

Definition. (Interior point) If (X, d) is a metric space and we have a subset $E \subseteq X$ then $x \in X$ is a *interior point* if there exists a $r > 0$ such that

$$B_r(x) \subseteq E.$$

Definition. (Interior) If (X, d) is a metric space and we have a subset $E \subseteq X$ then the *interior* of E denoted as $\mathring{E}$ is the set of interior points of E.

Let's look at an example. Take $\mathbb{R}$ with the Euclidean metric and take the set $E = [0, 1) \cup \{2\}$. We can see that the limit points are $[0, 1]$, the closure is $\{1\}$, the interior points are $(0, 1)$, the isolated points are $\{2\}$.

What about $E = \mathbb{Q}$? Then there are no isolated points, the closure is $\mathbb{R}$, there are no interior points, and the limit points are $\mathbb{R}$.

Proposition. Let (X, d) be a metric space. For any $E \subseteq X$, the $\bar{E}$ is a closed set and in fact

$$\bar{E} = \bigcap_{\substack{E \subseteq F \\ F \text{closed}}} F$$

Proof. Example Sheet 2 □

Remark. What this is really saying is that the closure $\mathring{E}$ is the smallest set (by inclusion) which is closed and contains E.

Proposition. Let (X, d) be a metric space and let $E \subseteq X$. Then the following statements are equivalent,

(i) If (x_k) is a sequence in E with $(x_k) \to x \in X$, then $x \in E$;

(ii) $E = \bar{E}$;

(iii) E is closed in X.

Proof. Follows directly from the definitions. $\square$

4.2 Cauchy sequences and completeness

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

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

Proposition. Let (X, d) be a metric space. Then we have the following,

(i) Any convergent sequence is Cauchy;

(ii) Any Cauchy sequence is bounded;

(iii) If (x_k) is a Cauchy sequence that has a convergent subsequence, converging to $x \in X$, then the whole sequence convergence to x.

Proof.

(i) If $x_k \to x$, then we have that

$$d(x_m, x_n) \leq d(x_m, x) + d(x_n, x) \to 0$$

if $m, n \to \infty$. Hence (x_k) is Cauchy.

(ii) If (x_k) is Cauchy, then there is some N such that for all $n, m \geq N$ we have that

$$d(x_n, x_m) < 1$$

So if we take $r = \max\{1, d(x_1, x_n), d(x_2, x_n), \ldots, d(x_{n-1}, x_n)\}$, then the sequence is contained in the ball centred at x_n with radius $r + 1$.

(iii) Suppose we have a sequence $x_{k_n} \to x$. We have that (x_k) is Cauchy, so given some $\varepsilon > 0$ we can choose N with $d(x_n, x_m) < \varepsilon$ for all $n, m \geq N$. We can also choose n_0 such that $k_{n_0} \geq n$ and $d(x_{k_{n_0}}, x) < \varepsilon$. Hence for any $n \geq N$ we have that

$$d(x_n, x) \leq d(x_n, x_{k_{n_0}}) + d(x, x_{k_{n_0}}) < 2\varepsilon. \quad \square$$

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

Definition. (Complete normed space) A normed space $(V, ||\cdot||)$ is *complete* if V with the metric defined by $||\cdot||$ is complete.

Theorem. $(\mathbb{R}^n, ||\cdot||_2)$ is complete.

Proof. If $(x^{(k)})$ is a Cauchy sequence in $\mathbb{R}^n$ then by directly applying the definition, each coordinate is itself a Cauchy sequence of real numbers. Hence by the completeness of $\mathbb{R}$ (IA Analysis I), we know that $(x_{(k)})$ converges componentwise some $L \in \mathbb{R}^n$, hence since we have componentwise convergence, we also have convegence in the Euclidean norm. Hence $(x^{(k)})$ converges to $\mathbb{R}^n$, so the metric space $\mathbb{R}^n$ with the Euclidean metric is complete. $\square$

Theorem. Any finite dimensional normed space is complete.

Proof. Follows from the previous theorem and the equivalence of norms.

Theorem. The metric space $(C[a,b], ||\cdot||_\infty)$ is complete.

We've proved that uniformly Cauchy implies uniform convergence. Since the uniform limit of continuous functions are continuous we have that all Cauchy sequences converge in $C[a,b]$. $\square$

Theorem. The metric spaces $(\ell_1, ||\cdot||_1)$, $(\ell_2, ||\cdot||_2)$, and $(\ell_\infty ||\cdot||_\infty)$ are complete.

Proof. We'll just prove the $(\ell_\infty, ||\cdot||_\infty)$ case. The rest is in Example Sheet 2.

When does a subset of a metric space remain complete as a subspace with the induced metric.

Theorem. Let (X,d) be a complete metric space, and $Y \subseteq X$ any subset. Then $(Y, d|_Y)$ is complete if and only if Y is closed.

Proof. Suppose $(Y, d|_Y)$ closed, then let (X_k) be a sequence in Y, with (x_k) Cauchy. Then $x_k \to x \in X$ by completeness in X. By the closure of Y, we have that $x \in \bar{Y} = Y$. Conversely suppose that $(Y, d|_Y)$ is complete then let (x_k) be a sequence in Y with $x_k \to x \in X$. Now (x_k) is Cauchy in X hence in Y as well. By completeness $x_k \to z \in Y$. By uniqueness of the limit, $x \in Y$, so Y is closed. $\square$

We can see an example of this, we'll show the that L^1 is complete with respect to Riemann integration.

(i) $(C([a,b]), L^1)$ is complete (proof on Example Sheet 2).

(ii) Define $\tilde{\mathcal{R}}([a,b]) = \mathcal{R}([a,b])/\sim$, where $f \sim g$ if $f = g$ except for a null set. So we have all addition and scalar multiplication defined still, and the L^1 limit is still well-defined by Lebesgue's theorem.

Theorem. If V is a finite dimensional real vector space and if $||\cdot||, ||\cdot||'$ are two norms on V then $||\cdot||, ||\cdot||'$ are *Lipchitz-equivalent* (i.e there are constants $C_1, C_2 > 0$ such that

$$C||x||' \leq ||x|| \leq C||x||'$$

for all $x \in V$.

Definition. (Sequential Compactness) A metric space (X, d) is *sequentially compact* if every sequence (x_n) is the space has a convergent subsequence, $x_{k_j} \to x$. A subset $K \subseteq X$ is sequentially compact if (K, d) is sequentially compact.

Remark. Another notion of compactness exists which applies more generally to topological spaces. Metric spaces are examples of topological spaces, and a metric space is sequentially compact if and only if the induced topology is compact. In this cases we just write compact to mean sequentially compact.

Theorem. A subset $S \subseteq \mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof. If the subset is closed and bounded then by Bolzano-Weiestrass and the definition of closed we can prove the space is compact. Conversely let S be comapct. Then for each $x_k \to x \in \mathbb{R}^n$, $x_j \to x \in S$ by uniqueness of the limit and the subsequence goes to the some limit $z, x \in S$. Suppose S unbounded. Then pick x_n such that $d(x_n, a) > n$. Then (x_n) has no covergent subsequence.

Theorem. If (X, d) is compact, then (X, d) is bounded and complete.

Proof. Any Cauchy sequence has a convergent subsequence, hence converges, so the metric space is complete. Suppose that the metric space is not bounded. Then take the sequence x_n such that $d(x_n, x_0) \geq n$. This clearly has no convergent subsequence, contradiction, hence (X, d) is bounded.

Theorem. If $K \subseteq X$ is a compact subset of a metric space X, then K is closed and bounded.

The converse of this theorem is actually false. Let $D_r = \overline{B_r}$. Consider D_r in ℓ_∞ by looking at $e^{(k)} = (0, 0, \dots, 1, \dots)$. We can see that $e^{(k)}$ has no convergent subsequence. But D_r is complete, closed, bounded, but not compact.

Theorem. $K \subseteq \mathbb{R}^n$ with the Euclidean metric is compact if and only if K is closed and bounded.

Proof. Suppose that K is closed and bounded. Let (x_k) be a sequence in K. Then since K is bounded, we can apply Bolzano-Weiestrass to get some subsequence (x_{k_j}) and $x \in \mathbb{R}^n$ with $x_{k_j} \to x$. Since K is closed, we know that $x \in K$, so K is compact. We're already proven the converse for general metric space, so it applies here. $\square$

Definition. (Totally bounded) Let (X, d) be a metric space. We say that (X, d) is *totally bounded* if for every $\varepsilon > 0$ there is a finite set $\{x_1, \dots, x_n\} \in X$ such that

$$X \subseteq \bigcup_{j=1}^{B} {}_{\varepsilon}(x_j).$$

Theorem. (X,d) is compact if and only if (X,d) is complete and totally bounded.

4.3 Continuous mappings between metric spaces

Definition. (Continuous) Let (X,d), (X',d') be metric spaces, and suppose we have a function $f : X \to X'$. We say that f is *conmtinuous* at $x \in X$ if $\forall \varepsilon > 0$ there exists a $\delta > 0$ such that
$$d(y,x) < \delta \implies d'(f(y), f(x)) < \varepsilon.$$
Equivalently we have that $\forall \varepsilon > 0,\ \exists \delta > 0$ such that $f(B_\delta(x)) \subseteq B_\varepsilon(f(x))$.

Theorem. (Sequential definition of continuity) Let $f : X \to X'$. Then f is continuous at x if and only if for every sequence $x_n \to x$ we have that $f(x_n) \to f(x)$.

Proof. Suppose that f is continuous at x. Let $x_n \to x$ be a sequence in X. Then for a given $\varepsilon > 0$ we have some $\delta > 0$ and $N > 0$ such that $n \geq N \implies d(x_n, x) < \delta \implies d'(f(x_n), f(x)) < \varepsilon$. Conversely if f is not continuous at x. Then there exists $\varepsilon > 0$ such that $\forall n \geq 1$, there $\exists x_n$ such that $d(x_n, x) < \frac{1}{n}$, but $d'(f(x_n), f(x)) \geq \varepsilon$. And so $x_n \to x$ but $f(x_n) \to f(x)$ which is a contradiction, hence f continuous at x. $\square$

Theorem. Let (X,d) and (X',d') be metric spaces. Then the following are equivalent.
(i) f is continuous;
(ii) For every sequence $x_n \to x$, $f(x_n) \to f(x)$;
(iii) $f^{-1}(v) = \{x \in X : f(x) \in V\}$ is open in X for every open $V \subseteq X'$.

Proof. Exercise.

Definition. (Uniformly continuous) Let $f : X \to X'$. We say that f is *uniformly continuous* if there exists some $\varepsilon > 0$ such that there exists $\delta > 0$ with $\forall x, y \in X$
$$d(x,y) < \delta \implies d'(f(x), f(y)) < \varepsilon.$$
Equivalently $B_\delta(x) \subseteq f^{-1}(B_\varepsilon(f(x)))$.

Definition. We say that $f : X \to X'$ is Lipschitz if there exists an L such that $\forall x, y \in X$,
$$d'(f(x), f(y)) \leq Ld(x,y).$$

Theorem. If $f : X \to X'$ is continuous, and (X,d) is compact and (X',d') is any metric space then,
(i) f is uniformly continuous;
(ii) $f(X)$ is a compact subspace of X';
(iii) $f(X)$ is closed and bounded;

(iv) If $X' = \mathbb{R}$, and d' is the Euclidean metric, then f attains its supremum and infimum.

Proof. We'll prove the statement part by part.

(i) Argue exactly as we did in the proof for $\mathbb{R}$, replacing the closed interval $[a, b]$ for a general compact metric space.

(ii) Let (y_k) be a sequence in $f(X)$. So $y_k = f(x_k)$ for some sequence x_k in X. By compactness of X there exists a subsequence x_{k_j} in X such that $x_{k_j} \to x \in X$ as $j \to \infty$. By the continuity of f we know that $y_{k_j} = f(x_{k_j}) \to f(x) \in f(X)$ as $j \to \infty$.

(iii) Done by $f(X)$ compact.

(iv) By the extreme value of theorem, f continuous implies that $f(X)$ is bounded and closed, hence f attains its supremum and infimu. $\square$

4.4 Equivalence of metrics and norms

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

Definition. (Topologically equivalent) Two metrics d, d' on X are *topologically equivalent* if they induce the same topology. So $U \subseteq X$ is open with resepct to d if and only if U is open with resepct to d'.

Definition. (Lipschitz equivalent) Two metrices d, d' on X are *lipschitz equivalent* if there exists fixed $a, b > 0$ such that

$$ad(x, y) \leq d'(x, y) \leq bd(x, y)$$

for all $x, y \in X$.

Remark. We can make some remarks about these definitions.

(i) Topological equivalence and Lipschitz equivalence parition the metrics, hence they form a equivalence relation.

(ii) d, d' being Lipschitz equivalent $\implies$ d, d' are topologically equivalent. But the converse is not true. Lipschitz equivalence is a stronger property than topological equivalence.

Let's look at an example to show this. Take $X = \mathbb{R}$ and $d(x, y) = |x-y|$, $d'(x, y) = \min\{1, |x-y|\}$. These are topologically equivalent metrices but $d(n, 0) = n$ and $d'(n, 0) \leq 1$ hence they're not Lipschitz equivalent.

A notion or property concerning a metric space X is a topological property if it depends on the topology on X and not the specific metric inducing the topology. For example the convergence of sequences is a topological property but the boundedness of a set is not a topological property. However compactness if a topological property. Completeness is not a topological property.

Definition. The norms $||\cdot||'$ on a vector space V are Lipschitz equivalent if there are $a, b > 0$ such that

$$a||x|| \leq ||x||' \leq b||x|| \quad \forall x \in V$$

We write $B_R^{||\cdot||}(0) = \{x \in V : ||x|| < R\}$.

Proposition. $||\cdot||$ and $||\cdot||'$ are Lipschitz equiavelnt if and only if there exists $r, R > 0$ such that

$$B_r^{||\cdot||}(0) \subseteq B_1^{||\cdot||'}(0) \subseteq B_R^{||\cdot||}(0)$$

Proof. Exercise.

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

Proof. Let $\dim V = n$. For $\{e_1, \dots, e_n\}$ a basis for V, define the Euclidean norm $||\cdot||_2$ on V by $||x||_2 = \left(\sum_{j=1}^n a_j^2\right)^{\frac{1}{2}}$. Fix $||\cdot||$ some other norm on V. For $x = \sum_{j=1}^n x_j e_j$,

$$\begin{aligned} ||x|| \leq \sum_{j=1}^n ||x_j \cdot e_j|| &= \sum_{j=1}^n |x_j|||e_j|| \\ &= \left(\sum_{j=1}^n |x_j|^2\right)^{\frac{1}{2}} \left(\sum_{j=1}^n ||e_j||^2\right)^{\frac{1}{2}} \end{aligned}$$

So $||x|| < b||x||_2$ for all $x \in V$ where $b = \left(\sum_{j=1}^n ||e_j||^2\right)^{\frac{1}{2}}$. To prove the other side, let

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

So need to show that $a \leq ||\hat{x}||$ for all $||\hat{x}||_2 = 1$. By $S = \{x \in V : ||x||_2 = 1\}$ compact we need to show that $x \to ||x||$ is continuous which can be shown to complete the proof. $\square$

Remark. Lipschitz equivalence of two norms, $||\cdot||$ and $||\cdot||'$ on V is the statement that the identity map

$$\mathrm{id}_V : (V, ||\cdot||) \to (V, ||\cdot||')$$

is Lipschitz (recall that a function is Lipschitz if $||f(x) - f(y)|| \leq L||x - y||$ for some $L \in \mathbb{R}$).

Remark. In infinite dimensional vector spaces, inequivalent norms exist. For example see that $\ell_1 \subseteq \ell_\infty$ since absolutely convergent sequences are bounded. So the restriction of ℓ_∞ on the ℓ_1 space is not Lipschitz equivalent to ℓ_1 (take the sequence $x^(k) = (1, 1, \dots, 1, 0, 0, \dots)$). In fact equivalence of all norms on a $\mathbb{R}$ vector space characterises finite dimensionality.

5 Differentiation in $\mathbb{R}^n$

In this secxtion we'll always take the norm $||\cdot||$ on $\mathbb{R}^n$ to be the Euclidean norm.

First we need the following (slightly different definition of the limit).

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

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

means that $\forall \varepsilon > 0$ there exists $\delta > 0$ such that $x \in E$ with $0 < ||x - a|| < \delta \implies ||f(x) - b|| < \varepsilon$.

We do not ask that f be defined at $x = a$.

Given some function $f : U \to \mathbb{R}^m$, with $U \subseteq \mathbb{R}^n$ and a point $a \in U$, how can we define the derivative of f at a. If $n = m = 1$ then say $f : (b, c) \to \mathbb{R}$ and $a \in (b, c)$ we know that

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

If $n \geq 2$ we might try to define the derivative as

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{||h||}. \tag{$\star$}$$

However, if we take $n = 1$ in we would have $\lim_{h\to 0} \frac{f(a+h)-f(a)}{|h|}$ which is very different to the definition of the derivative we've been using for years. It also sucks since this definition says that $y = x$ isn't differentiable at 0 and $y = |x|$ is sifferentiable at 0. This are not properties we want and so ($\star$) is not the correct notion we're looking for.

Let's look at the 1 dimensional situation a bit closer.

$$\begin{aligned} f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} &\iff \lim_{h \to 0} \frac{f(a+h) - f(a) - f'(a)h}{h} = 0 \\ &\iff \lim_{h \to 0} \frac{f(a+h) - f(a) - f'(a)h}{|h|} = 0 \\ &\iff \lim_{h \to 0} \frac{f(a+h) - f(a) - Ah}{|h|} = 0 \quad \text{for some } A \in \mathbb{R}. \end{aligned}$$

This is the correct notion we're looking for.

Definition. (Differentiable) Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}^m$ and let $a \in U$. Then we say that f is *differentiable* at a if there is a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

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

We say that A is the derivative of f at a and write $A = Df(a)$.

Remark. Since we assume that U is open, there is aball $B_r(a) \subseteq U$, hence $a + h \in U$ for any h with $||h||$ sufficiently small.

Proposition. Let $U \subseteq \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^m$.

(i) If f is differentiable at $a \in U$, then $Df(a)$ is unique.

(ii) If f is differentiable at $a \in U$, then f is continuous at a.

(iii) Write $f = (f_1, f_2, \ldots, f_m)$, where $f_j : U \to \mathbb{R}$, then f is differentiable at $a \in U$ if and only if f_j is differentiable at a and

$$Df(a) = \begin{pmatrix} Df_1(a) \\ \vdots \\ Df_m(a) \end{pmatrix}$$

.

Proof. Exercise.

Remark. Differentiability of f of a point a means that the affine function $x \to f(x) + Df(a)(x-a)$ is a good approximation to $f(x)$ near $f(a)$. In fact if we look at the definition of $Df(a)$ this is exactly what it's saying (replacing h with $x - a$). The deviation of $f(x)$ from $f(a) + Df(a)(x-a)$ tends to zero faster than x tends to a.

Let's introduce the little o notation formally now.

We write $o(x)$ for $x \in \mathbb{R}^n$ to denote any function with the property that

$$\frac{o(x)}{||x||} \to 0$$

as $x \to 0$. So in this notation differentiability of f at a is the statement that $f(a+h) = f(a) + Df(a)h + o(h)$.

5.1 Directional Derivatives

Definition. (Directional derivative) Let $f : U \to \mathbb{R}^m$ with $U \subseteq \mathbb{R}^n$ open. Let $a \in U$. The *directional derivative* of f at a in the direction of u (where $u \in \mathbb{R}^n$ is fixed), is

$$D_u f(a) = \lim_{t \to 0} \frac{f(a+tu) - f(a)}{t}$$

if the limit exists.

Proposition. If f is differentiable at $a \in U$, then for any $u \in \mathbb{R}^n$, the directional derivative $D_u f(a)$ exists and is equal to $Df(a)u$. In particular the map $u \to D_u f(a)$ is linear.

Proof. We know that

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

So that $h = tu$ with $t \to 0$ hence we have that

$$\lim_{t \to 0} \frac{f(a+tu) - f(a) - t(Df(a)u)}{t||u||} = 0$$

so we recover that $Df(a)u = D_u f(a)$. $\square$

A special case of this is partial derivatives (usually in the case with $m = 1$. Let $\{e_i\}$ be the standard basis for $\mathbb{R}^n$. Take $u = e_j$. Then

$$D_{e_j} f(a) = \lim_{t \to 0} \frac{f(a + te_j) - f(a)}{t}$$

is the jth partial derivative of f (given the limit exists). We notate this by $\frac{\partial f}{\partial x_j}$ or $D_j f(a)$.

So if $f \to U \to \mathbb{R}$ is differentiable at $a \in U$, then the partial derivatives $D_j f(a)$ all exist and $Df(a) = (D_1 f(a), \cdots, D_n f(a))$. More generally,

Proposition. If $U \subseteq \mathbb{R}^n$ is open and $f : U \to \mathbb{R}^m$, is differentiable at $a \in U$ then the partial derivatives $D_j f_i(a)$ all exist and the matrix for $Df(a)$ is given by $A = (D_j f_i(a))$.

Proof. By the previous discussion and proposition.

Remark. In the case where $m = 1$, $Df(a)$ is often called the gradient of f at a.

We saw that differentiability of a implies existance of all directoinal derivatives and hence existance of all partials. Let's see an example to show that existance of all partials doesn't imply differentiability. Take the function

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

So both partials are zero, but the derivative in the $(1, 1)$ direction does not exist. Hence the function is not differentiable.

Even if all directional derivatives exist, we *still* may not have differentiability. Take $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

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

Calculating the directional derivatives take $u = (u_1, u_2)$ and $t \neq 0$. Then

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

So $D_u f(0)$ exists and is zero for all directions u. But the function is not continuous at the origin hence not differentiable.

Let's see a third example. Let $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} \frac{x^3}{x^2 + y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}.$$

We have that $|f(x, y)| \leq |x|$ so f is continuous everywhere. In fact through direct calculation we can see that all the partials exist at the origin. In fact for any $u = (u_1, u_2)$ we have that

$$\frac{f(0 + tu) - f(0)}{t} = \frac{u_1^3}{u_1^2 + u_2^2}.$$

So all directional derivatives exist, but the map from $u \to D_u f(0,0)$ is not linear in u hence f can't be differentiable at the origin (we don't have that $D_u f(0,0) = Df(0,0)u$. Alternatively if f were differentiable at the origin, then the derivative $Df(0,0)h = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = h_1$. Hence

$$\frac{f(0+h) - f(0) - Df(0)h}{||h||} = -\frac{h_1 h_2^2}{(h_1^2 + h_2^2)^{3/2}}$$

which does not tend to zero as $||h|| \to 0$. So f is not differentiable at the origin.

Theorem. (Continuity of partial derivatives guarantees differentiability) Let $f : U \to \mathbb{R}^m$ with $U \subseteq \mathbb{R}^n$ open. Let $a \in U$ and suppose that for some $B_r(a) \in U$ the partial derivatives $D_j f_i(x)$ exist for every $x \in B_r(a)$. Suppose also that $D_j f_i$ is continuous at a for every $1 \leq i \leq n$ and $1 \leq j \leq n$. Then f is differentiable at a.

Proof. Assume *wlog* that $m = 1$ by taking coordinatewise functions. Let $h = (h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$. Then we have that

$$\begin{aligned} &f(a+h) - f(a) = \\ &\qquad \sum_{j=2}^{n} (f(a + h_1 e_1 + h_2 e_2 + \cdots + h_{j-1} e_{j-1} + h_j e_j) - f(a + h_1 e_1 + \cdots + h_{j-1} e_{j-1})) \\ &\qquad + f(a + h_1 e_1) - f(a). \end{aligned}$$

We can then apply the mean value theorem to the function $g(t) = f(a+h_1e_1+\cdots+h_{j-1}e_{j-1}+te_j)$ for $t \in [0, h_j]$. Then we get $f(a+h) - f(a) = \sum_{j=1}^{m} h_j D_j f(a + h_1 e_1 + \cdots h_{j-1} e_{j-1} + \theta_j h_j e_j)$ for some $\theta_j \in (0,1)$.

$$\begin{aligned} f(a+h) - f(a) - \sum_{j=1}^{n} h_j D_j f(a) &= \sum_{j=1}^{n} h_j (D_j f(a + h_1 e_1 + \cdots \theta_j h_j e_j) - D_j f(a)) \\ &= o(h) \end{aligned}$$

since the bracketed expression goes to zero. So f is differentiable at a with $Df(a)$ as we've show. ☐

For some of the next results we need a convenient norm on the vector space $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. First we note that any map $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is Lipschitz hence continuous. Let $S = \{x \in \mathbb{R}^n : ||x|| = 1\}$, S is compact (closed and bounded), so a map A is bounded and attains its supremum on the sphere.

Definition. (Operator norm on $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$) For $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and define

$$||A||_{\text{op}} = \sup_{|x|=1} ||A(x)|| = \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{||A(x)||}{||x||}.$$

Remark. We can make some remarks about this norm.

(i) $||A||_{\text{op}}$ is a norm on $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

(ii) $\forall x \in \mathbb{R}^n$, $||A(x)|| \leq ||A||_{\text{op}} ||x||$.

(iii) If $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $B \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p)$ then $B \circ A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ and $||BA||_{\text{op}} = ||B||_{\text{op}}||A||_{\text{op}}$.

(iv) Since $\dim \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) = nm < \infty$ any norm on $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is equivalent to the operator norm. We will use the operator norm instead because of properties (ii) and (iii)