Part IB — Analysis II

Theorems with proof

Based on lectures by N. Wickramasekera

Notes taken by Dexter Chua

Michaelmas 2015

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 $\mathbb{R}^n$ 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

0 Introduction

1 Uniform convergence

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 ε, 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

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

So done.

Now suppose (f_n) is uniformly Cauchy. Then $(f_n(x))$ is Cauchy for all x. So it converges. Let

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

We want to show that $f_n \to f$ uniformly. 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 $|f_n(x) - f(x)| \leq \frac{\varepsilon}{2} < \varepsilon$. So done. □

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.

In particular, if f_n are continuous everywhere, then f is continuous everywhere.

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 δ 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. If we are given a specific sequence of functions and are asked to prove that they converge uniformly, we usually take the pointwise limit and show that the convergence is uniform. However, given a general function, this is usually not helpful. Instead, we can use the Cauchy criterion by showing that the sequence is uniformly Cauchy.

We want to find an N such that $n, m > N$ implies $\sup |f_n - f_m| < \varepsilon$. We want to relate this to the derivatives. We might want to use the fundamental theorem of algebra for this. However, we don't know that the derivative is integrable! So instead, we go for the mean value theorem.

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 ε, 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. However, we know that $g_n \to g$ pointwise on $[a, b]$, and we know that g_n are all continuous.

To conclude that g is continuous, we have to show that the convergence is uniform. To show that g_n converges uniformly, we rely on the Cauchy criterion and the mean value theorem.

For $x \neq y$, we know that

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

for some $t \in [x, y]$. This also holds for $x = y$, since $g_n(y) - g_m(y) = f_n'(y) - f_m'(y)$ by definition.

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

$$|g_n(x) - g_m(x)| \leq \sup |f_n' - f_m'| < \varepsilon.$$

So

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

i.e. 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.

(i) Let $f_n, g_n : E \to \mathbb{R}$, be sequences, and $f_n \to f$, $g_n \to g$ uniformly on E. Then for any $a, b \in \mathbb{R}$, $af_n + bg_n \to af + bg$ uniformly.

(ii) Let $f_n \to f$ uniformly, and let $g : E \to \mathbb{R}$ is bounded. Then $gf_n : E \to \mathbb{R}$ converges uniformly to gf.

Proof.

(i) Easy exercise.

(ii) Say $|g(x)| < M$ for all $x \in E$. Then

$$|(gf_n)(x) - (gf)(x)| \leq M|f_n(x) - f(x)|.$$

So

$$\sup_E |gf_n - gf| \leq M \sup_E |f_n - f| \to 0.$$ □

2 Series of functions

2.1 Convergence of series

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

Proof. Again, we don't have a candidate for the limit. So we use the Cauchy criterion.

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. □

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

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

Taking supremum, we have

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

So done by the Cauchy criterion. □

2.2 Power series

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.

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

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

We say that the sum converges locally absolutely uniformly inside circle of convergence, i.e. for every point $y \in (a - R, a + R)$, there is some open interval around y on which the sum converges absolutely uniformly.

These results hold for complex power series as well, but for concreteness we will just do it for real series.

Proof. See IA Analysis I for (i) and (ii).

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

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

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

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

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

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 $R_1 \leq R$.

Suppose that the inequality is strict, i.e. $R_1 < R$, 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 $n|c_n|r_1^{n-1} \leq |c_n|r^n$ for sufficiently large n. So we must have $R_1 = R$.

(ii) Let $f_n(x) = \sum\limits_{j=0}^{n} c_j(x-a)^j$. Then $f_n'(x) = \sum\limits_{j=1}^{n} jc_j(x-a)^{j-1}$. 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. $\square$

3 Uniform continuity and integration

3.1 Uniform continuity

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

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. □

3.2 Applications to Riemann integrability

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

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

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.

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

Since f is integrable, for arbitrary ε', 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 δ, then $\sup_{I_j} g \circ f - \inf_{I_j} g \circ f$ will be less that ε. 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

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

So

$$\sup_{I_j} g \circ f - \inf_{I_j} g \circ f \leq \varepsilon.$$

Hence we know that

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

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

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

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

Corollary. A continuous function $g : [a, b] \to \mathbb{R}$ is integrable.

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

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

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

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

Hence for any partition P,

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

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

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

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

Using the Cauchy-Schwarz inequality, we get

$$\begin{aligned}&\leq \int_a^b \|\mathbf{v}\|\|\mathbf{f}\|(x)\;\mathrm{d}x\\ &= \|\mathbf{v}\|\int_a^b \|\mathbf{f}\|\;\mathrm{d}x.\end{aligned}$$

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

3.3 Non-examinable fun*

Theorem (Weierstrass Approximation Theorem*)**.** If $f : [0,1] \to \mathbb{R}$ is continuous, then there exists a sequence of polynomials (p_n) such that $p_n \to f$ uniformly. In fact, the sequence can be given by

$$p_n(x) = \sum_{k=0}^{n} f\left(\frac{i}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}.$$

These are known as *Bernstein polynomials.*

Proof. For convenience, let

$$p_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}.$$

First we need a few facts about these functions. Clearly, $p_{n,k}(x) \geq 0$ for all $x \in [0,1]$. Also, by the binomial theorem,

$$\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = (x+y)^n.$$

So we get

$$\sum_{k=0}^{n} p_{n,k}(x) = 1.$$

Differentiating the binomial theorem with respect to x and putting $y = 1-x$ gives

$$\sum_{k=0}^{n} \binom{n}{k} k x^{k-1} (1-x)^{n-k} = n.$$

We multiply by x to obtain

$$\sum_{k=0}^{n} \binom{n}{k} k x^{k} (1-x)^{n-k} = nx.$$

In other words,

$$\sum_{k=0}^{n} k p_{n,k}(x) = nx.$$

Differentiating once more gives

$$\sum_{k=0}^{n} k(k-1) p_{n,k}(x) = n(n-1)x^2.$$

Adding these two results gives

$$\sum_{k=0}^{n} k^2 p_{n,k}(x) = n^2x^2 + nx(1-x).$$

We will write our results in a rather weird way:

$$\sum_{k=0}^{n} (nx-k)^2 p_{n,k}(x) = n^2x^2 - 2nx \cdot nx + n^2x^2 + nx(1-x) = nx(1-x). \quad (*)$$

This is what we really need.

Now given ε, since f is continuous, f is uniformly continuous. So pick δ such that $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$.

Since $\sum p_{n,k}(x) = 1$, $f(x) = \sum p_{n,k}(x)f(x)$. Now for each *fixed* x, we can write

$$\begin{aligned}
|p_n(x) - f(x)| &= \left|\sum_{k=0}^{n} \left(f\left(\frac{k}{n}\right) - f(x)\right) p_{n,k}(x)\right| \\
&\leq \sum_{k=0}^{n} \left|f\left(\frac{k}{n}\right) - f(x)\right| p_{n,k}(x) \\
&= \sum_{k:|x-k/n|<\delta} \left(\left|f\left(\frac{k}{n}\right) - f(x)\right| p_{n,k}(x)\right) \\
&\qquad + \sum_{k:|x-k/n|\geq\delta} \left(\left|f\left(\frac{k}{n}\right) - f(x)\right| p_{n,k}(x)\right) \\
&\leq \varepsilon \sum_{k=0}^{n} p_{n,k}(x) + 2\sup_{[0,1]}|f| \sum_{k:|x-k/n|>\delta} p_{n,k}(x) \\
&\leq \varepsilon + 2\sup_{[0,1]}|f| \cdot \frac{1}{\delta^2} \sum_{k:|x-k/n|>\delta} \left(x - \frac{k}{n}\right)^2 p_{n,k}(x) \\
&\leq \varepsilon + 2\sup_{[0,1]}|f| \cdot \frac{1}{\delta^2} \sum_{k=0}^{n} \left(x - \frac{k}{n}\right)^2 p_{n,k}(x) \\
&= \varepsilon + \frac{2\sup|f|}{\delta^2 n^2} nx(1-x) \\
&\leq \varepsilon + \frac{2\sup|f|}{\delta^2 n}
\end{aligned}$$

Hence given any ε and δ, we can pick n sufficiently large that that $|p_n(x) - f(x)| < 2\varepsilon$. This is picked independently of x. So done. □

Theorem (Lebesgue's theorem on the Riemann integral*)**.** Let $f : [a,b] \to \mathbb{R}$ be a bounded function, and let $\mathcal{D}_f$ be the set of points of discontinuities of f. Then f is Riemann integrable *if and only if* $\mathcal{D}_f$ has measure zero.

4 $\mathbb{R}^n$ as a normed space

4.1 Normed spaces

Lemma (Cauchy-Schwarz inequality (for integrals))**.** 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 $f = 0$ (since f is continuous). So 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

$$\phi(t) = \int_a^b (g - tf)^2\ \mathrm{d}t \geq 0.$$

for every t. We can expand this as

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

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

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

So done. □

Proposition. If $\|\cdot\|$ and $\|\cdot\|'$ are Lipschitz equivalent norms on a vector space V, then

(i) A subset $E \subseteq V$ is bounded with respect to $\|\cdot\|$ if and only if it is bounded with respect to $\|\cdot\|'$.

(ii) A sequence x_k converges to x with respect to $\|\cdot\|$ if and only if it converges to x with respect to $\|\cdot\|'$.

Proof.

(i) This is direct from definition of equivalence.

(ii) Say we have a, b such that $a\|\mathbf{y}\| \leq \|\mathbf{y}\|' \leq b\|\mathbf{y}\|$ for all $\mathbf{y}$. So

$$a\|\mathbf{x}_k - \mathbf{x}\| \leq \|\mathbf{x}_k - \mathbf{x}\|' \leq b\|\mathbf{x}_k - \mathbf{x}\|.$$

So $\|\mathbf{x}_k - \mathbf{x}\| \to 0$ if and only if $\|\mathbf{x}_k - \mathbf{x}\|' \to 0$. So done. □

Proposition. Let $(V, \|\cdot\|)$ be a normed space. Then

(i) If $\mathbf{x}_k \to \mathbf{x}$ and $\mathbf{x}_k \to \mathbf{y}$, then $\mathbf{x} = \mathbf{y}$.

(ii) If $\mathbf{x}_k \to \mathbf{x}$, then $a\mathbf{x}_k \to a\mathbf{x}$.

(iii) If $\mathbf{x}_k \to \mathbf{x}$, $\mathbf{y}_k \to \mathbf{y}$, then $\mathbf{x}_k + \mathbf{y}_k \to \mathbf{x} + \mathbf{y}$.

Proof.

(i) $\|\mathbf{x}-\mathbf{y}\| \leq \|\mathbf{x}-\mathbf{x}_k\| + \|\mathbf{x}_k - \mathbf{y}\| \to 0$. So $\|\mathbf{x}-\mathbf{y}\| = 0$. So $\mathbf{x} = \mathbf{y}$.

(ii) $\|a\mathbf{x}_k - a\mathbf{x}\| = |a|\|\mathbf{x}_k - \mathbf{x}\| \to 0$.

(iii) $\|(\mathbf{x}_k + \mathbf{y}_k) - (\mathbf{x}+\mathbf{y})\| \leq \|\mathbf{x}_k - \mathbf{x}\| + \|\mathbf{y}_k - \mathbf{y}\| \to 0$. □

Proposition. Convergence in $\mathbb{R}^n$ (with respect to, say, the Euclidean norm) is equivalent to coordinate-wise convergence, i.e. $\mathbf{x}^{(k)} \to \mathbf{x}$ 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

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

So done □

Theorem (Bolzano-Weierstrass theorem in $\mathbb{R}^n$)**.** Any bounded sequence in $\mathbb{R}^n$ (with, say, the Euclidean norm) 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. □

4.2 Cauchy sequences and completeness

Proposition. Any convergent sequence is Cauchy.

Proof. If $\mathbf{x}_k \to \mathbf{x}$, then

$$\|\mathbf{x}_k - \mathbf{x}_\ell\| \leq \|\mathbf{x}_k - \mathbf{x}\| + \|\mathbf{x}_\ell - \mathbf{x}\| \to 0 \text{ as } k, \ell \to \infty.$$ □

Proposition. A Cauchy sequence is bounded.

Proof. By definition, there is some N such that for all $n \geq N$, we have $\|\mathbf{x}_N - \mathbf{x}_n\| < 1$. So $\|\mathbf{x}_n\| < 1 + \|\mathbf{x}_N\|$ for $n \geq N$. So, for all n,

$$\|\mathbf{x}_n\| \leq \max\{\|\mathbf{x}_1\|, \cdots, \|\mathbf{x}_{N-1}\|, 1 + \|\mathbf{x}_N\|\}. \qquad \square$$

Proposition. If a Cauchy sequence has a subsequence converging to an element $\mathbf{x}$, then the whole sequence converges to $\mathbf{x}$.

Proof. Suppose $\mathbf{x}_{k_j} \to \mathbf{x}$. Since $(\mathbf{x}_k)$ is Cauchy, given $\varepsilon > 0$, we can choose an N such that $\|\mathbf{x}_n - \mathbf{x}_m\| < \frac{\varepsilon}{2}$ for all $n, m \geq N$. We can also choose j_0 such that $k_{j_0} \geq n$ and $\|\mathbf{x}_{k_{j_0}} - \mathbf{x}\| < \frac{\varepsilon}{2}$. Then for any $n \geq N$, we have

$$\|\mathbf{x}_n - \mathbf{x}\| \leq \|\mathbf{x}_n - \mathbf{x}_{k_{j_0}}\| + \|\mathbf{x} - \mathbf{x}_{k_{j_0}}\| < \varepsilon. \qquad \square$$

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.

Proof. This follows directly from definition. $\square$

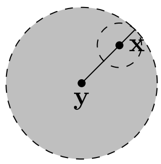
Theorem. $\mathbb{R}^n$ (with the Euclidean norm, say) is complete.

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. $\square$

Proposition. $B_r(\mathbf{y}) \subseteq V$ is an open subset for all $r > 0$, $\mathbf{y} \in V$.

Proof. Let $\mathbf{x} \in B_r(\mathbf{y})$. Let $\rho = r - \|\mathbf{x} - \mathbf{y}\| > 0$. Then $B_\rho(\mathbf{x}) \subseteq B_r(\mathbf{y})$.



$\square$

Proposition. Let $E \subseteq V$. Then E contains all of its limit points if and only if $V \setminus E$ is open in V.

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. □

Proposition. Let $E \subseteq V$. Then E contains all of its limit points if and only if $V \setminus E$ is open in V.

Proof. ($\Rightarrow$) Suppose E contains all its limit points. To show $V \setminus E$ is open, we 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. □

4.3 Sequential compactness

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 $\mathbb{R}^n$ (with, say, the Euclidean norm), then if K is closed and bounded, then K is compact.

Proof.

(i) Let K be compact. Boundedness is easy: 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 any 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

Theorem. Let $(V, \|\cdot\|)$, $(V', \|\cdot\|')$ be normed spaces, $E \subseteq V$, $f : E \to \mathbf{V}'$. Then f is continuous at $\mathbf{y} \in E$ if and only if 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. □

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

(i) $f(K)$ is compact in V'

(ii) $f(K)$ is closed and bounded

(iii) If $V' = \mathbb{R}$, then the function attains its supremum and infimum, i.e. there is some $\mathbf{y}_1, \mathbf{y}_2 \in K$ such that

$$f(\mathbf{y}_1) = \sup\{f(\mathbf{y}) : \mathbf{y} \in K\}, \quad f(\mathbf{y}_2) = \inf\{f(\mathbf{y}) : \mathbf{y} \in K\}.$$

Proof.

(i) 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.

(ii) This follows directly from (i), since every compact space is closed and bounded.

(iii) If F is any bounded subset of $\mathbb{R}$, then either $\sup F \in F$ or $\sup F$ is a limit point of F (or both), by definition of the supremum. If F is closed and bounded, then any limit point must be in F. So $\sup F \in F$. Applying this fact to $F = f(K)$ gives the desired result, and similarly for infimum. □

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 $(V, \|\cdot\|_2)$.

Proof. $\|\cdot\|_2$ is well-defined since $x_1, \cdots, x_n$ are uniquely determined by $\mathbf{x}$ (by (a certain) 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.

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
$$\begin{aligned}\|\mathbf{x}\| &= \left\|\sum_{j=1}^{n} x_j \mathbf{v}_j\right\|\\ &\leq \sum |x_j| \|\mathbf{v}_j\|\\ &\leq \|\mathbf{x}\|_2 \left(\sum_{j=1}^{n} \|\mathbf{v}_j\|^2\right)^{\frac{1}{2}}\end{aligned}$$
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, i.e. 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

5.1 Preliminary definitions

Proposition. The limit of a convergent sequence is unique.

Proof. Same as that of normed spaces. □

5.2 Topology of metric spaces

Proposition. Let (X, d) be a metric space. Then $x_k \to x$ if and only if for every neighbourhood V of x, there exists some K such that $x_k \in V$ for all $k \geq K$. Hence convergence is a topological notion.

Proof. ($\Rightarrow$) Suppose $x_k \to X$, and let V be any neighbourhood of x. Since V is open, by definition, there exists some ε such that $B_\varepsilon(x) \subseteq V$. By definition of convergence, there is some K such that $x_k \in B_\varepsilon(x)$ for $k \geq K$. So $x_k \in V$ whenever $k \geq K$.

($\Rightarrow$) Since every open ball is a neighbourhood, this direction follows directly from definition. □

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

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

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

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

Proof.

(i) Let $U = \bigcup_\alpha V_\alpha$, where each V_α is open. If $x \in U$, then $x \in V_\alpha$ for some α. Since V_α 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_α 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, $B_1(x) \subseteq X$. □

Proposition. A subset is closed if and only if its complement is open.

Proof. Exactly the same as that of normed spaces. It is useful to observe that $y \in X$ is a limit point of E if and only if $(B_r(y) \setminus \{y\}) \cap E \neq \emptyset$ for all $r > 0$. □

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

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

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

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

Proof. By taking complements of the result for open subsets. □

Proposition. Let (X, d) be a metric space and $x \in X$. Then the singleton $\{x\}$ is a closed subset, and hence any finite subset is closed.

Proof. Let $y \in X \setminus \{x\}$. So $d(x, y) > 0$. Then $B_{d(y,x)}(x) \subseteq X \setminus \{x\}$. So $X \setminus \{x\}$ is open. So $\{x\}$ is closed.

Alternatively, since $\{x\}$ has no limit points, it contains all its limit points. So it is closed. □

5.3 Cauchy sequences and completeness

Proposition. Let (X, d) be a metric space. Then

(i) Any convergent sequence is Cauchy.

(ii) If a Cauchy sequence has a convergent subsequence, then the original sequence converges to the same limit.

Proof.

(i) If $x_k \to x$, then

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

as $m, n \to \infty$.

(ii) Suppose $x_{k_j} \to x$. Since (x_k) is Cauchy, given $\varepsilon > 0$, we can choose an N such that $d(x_n, x_m) < \frac{\varepsilon}{2}$ for all $n, m \geq N$. We can also choose j_0 such that $k_{j_0} \geq n$ and $d(x_{k_{j_0}}, x) < \frac{\varepsilon}{2}$. Then for any $n \geq N$, we have

$$d(x_n, x) \leq d(x_n, x_{k_{j_0}}) + d(x, x_{k_{j_0}}) < \varepsilon.$$ □

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. □

5.4 Compactness

Theorem. All compact spaces are complete and bounded.

Proof. Let (X, d) be a compact metric space. Let (x_k) be Cauchy in X. By compactness, it has some convergent subsequence, say $x_{k_j} \to x$. So $x_k \to x$. So it is complete.

If (X, d) is not bounded, by definition, for any x_0, there is a sequence (x_k) such that $d(x_k, x_0) > k$ for every k. But then (x_k) cannot have a convergent subsequence. Otherwise, if $x_{k_j} \to x$, then

$$d(x_{k_j}, x_0) \leq d(x_{k_j}, x) + d(x, x_0)$$

and is bounded, which is a contradiction. □

Theorem. (non-examinable) Let (X, d) be a metric space. Then X is compact if and only if X is complete and totally bounded.

Proof. ($\Leftarrow$) Let X be complete and totally bounded, $(y_i) \in X$. For every $j \in \mathbb{N}$, there exists a finite set of points E_j such that every point is within $\frac{1}{j}$ of one of these points.

Now since E_1 is finite, there is some $x_1 \in E_1$ such that there are infinitely many y_i's in $B(x_1, 1)$. Pick the first y_i in $B(x_1, 1)$ and call it y_{i_1}.

Now there is some $x_2 \in E_2$ such that there are infinitely many y_i's in $B(x_1, 1) \cap B(x_2, \frac{1}{2})$. Pick the one with smallest value of $i > i_1$, and call this y_{i_2}. Continue till infinity.

This procedure gives a sequence $x_i \in E_i$ and subsequence (y_{i_k}), and also

$$y_{i_n} \in \bigcap_{j=1}^{n} B\left(x_j, \frac{1}{j}\right).$$

It is easy to see that (y_{i_n}) is Cauchy since if $m > n$, then $d(y_{i_m}, y_{i_n}) < \frac{2}{n}$. By completeness of X, this subsequence converges.

($\Rightarrow$) Compactness implying completeness is proved above. Suppose X is not totally bounded. We show it is not compact by constructing a sequence with no Cauchy subsequence.

Suppose ε is such that there is no finite set of points $x_1, \cdots, x_N$ with

$$X = \bigcup_{i=1}^{N} B_\varepsilon(x_i).$$

We will construct our sequence iteratively.

Start by picking an arbitrary y_1. Pick y_2 such that $d(y_1, y_2) \geq \varepsilon$. This exists or else $B_\varepsilon(y_1)$ covers all of X.

Now given $y_1, \cdots, y_n$ such that $d(y_i, y_j) \geq \varepsilon$ for all $i, j = 1, \cdots, n$, $i \neq j$, we pick y_{n+1} such that $d(y_{n+1}, y_j) \geq \varepsilon$ for all $j = 1, \cdots, n$. Again, this exists, or else $\bigcup_{i=1}^{n} B_\varepsilon(y_i)$ covers X. Then clearly the sequence (y_n) is not Cauchy. So done. □

5.5 Continuous functions

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.

Proof. We are going to prove by contradiction. 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. □

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): The argument for this is the same as for 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 δ 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 ε, 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 δ such that $B_\delta(y) \subseteq U$. So we get
$$B_\delta(y) \subseteq f^{-1}(B_\varepsilon(f(y))).$$
So we get continuity. □

Corollary. A function $f : (X, d) \to (X', d')$ is continuous if $f^{-1}(V)$ is open in X whenever V is open in X'.

Proof. Follows directly from the equivalence of (i) and (iii) in the theorem above. □

5.6 The contraction mapping theorem

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, i.e. there is a unique x such that $f(x) = x$.

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

Proof. We first focus on the case where f itself is a contraction.

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 $d(x, y) = 0$, i.e. $x = y$.

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. $\square$

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

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

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.

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 ε 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 $\mathbf{f}$ is differentiable, since if a continuous $\mathbf{f}$ 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 $X \to X$, i.e. $T\mathbf{g} \in X$ 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

$$\kappa(b - a) < 1, \tag{$\dagger$}$$

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

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

i.e.

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

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

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

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

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

i.e. $T^{(m)}$ 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. □

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

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

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 $B = A$. □

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 $\mathbf{f} = (f_1, f_2, \cdots, f_m) : U \to \mathbb{R}^m$, 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 $\mathbf{f}$ is differentiable at $\mathbf{a}$, then all partial derivatives $D_j f_i(\mathbf{a})$ exist for $j = 1, \cdots, n$; $i = 1, \cdots, m$, and are given by
$$D_j f_i(\mathbf{a}) = Df_i(\mathbf{a})\mathbf{e}_j.$$

(vii) If $A = (A_{ij})$ be the matrix representing $D\mathbf{f}(\mathbf{a})$ with respect to the standard basis for $\mathbb{R}^n$ and $\mathbb{R}^m$, i.e. 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) Exercise on example sheet 4.

(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) We've proved this in the previous discussion.

(vi) We've proved this in the previous discussion.

(vii) 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. □

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 $m = 1$, by the long proposition. 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

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

6.2 The operator norm

Proposition.

(i) $\|A\| < \infty$ for all $A \in \mathcal{L}$.

(ii) $\|\cdot\|$ is indeed a norm on $\mathcal{L}$.

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

(iv) $\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^n$.

(v) 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 is since A is continuous and $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}$ is compact.

(ii) The only non-trivial part is the triangle inequality. We have

$$\begin{aligned}\|A + B\| &= \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x} + B\mathbf{x}\| \\ &\leq \sup_{\|\mathbf{x}\|=1} (\|A\mathbf{x}\| + \|B\mathbf{x}\|) \\ &\leq \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| + \sup_{\|\mathbf{x}\|=1} \|B\mathbf{x}\| \\ &= \|A\| + \|B\|\end{aligned}$$

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

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

(iv) Immediate from above.

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

Proposition.

(i) If $A \in L(\mathbb{R}, \mathbb{R}^m)$, then A can be written as $Ax = x\mathbf{a}$ for some $\mathbf{a} \in \mathbb{R}^m$. Moreover, $\|A\| = \|\mathbf{a}\|$, where the second norm is the Euclidean norm in $\mathbb{R}^n$

(ii) If $A \in L(\mathbb{R}^n, \mathbb{R})$, then $A\mathbf{x} = \mathbf{x} \cdot \mathbf{a}$ for some fixed $\mathbf{a} \in \mathbb{R}^n$. Again, $\|A\| = \|\mathbf{a}\|$.

Proof.

(i) Set $A(1) = \mathbf{a}$. Then by linearity, we get $Ax = xA(1) = x\mathbf{a}$. Then we have

$$\|A\mathbf{x}\| = |x|\|\mathbf{a}\|.$$

So we have
$$\frac{\|A\mathbf{x}\|}{|x|} = \|\mathbf{a}\|.$$

(ii) Exercise on example sheet 4. $\square$

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. The proof is very easy if we use the little o notation. 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

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

6.3 Mean value inequalities

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. $\square$

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 we can apply the previous theorem, and get

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

Corollary. Let $\mathbf{f} : B_r(\mathbf{a}) \subseteq \mathbb{R}^n \to \mathbb{R}^m$ have $D\mathbf{f}(\mathbf{x}) = 0$ for all $\mathbf{x} \in B_r(\mathbf{a})$. Then $\mathbf{f}$ is constant.

Proof. Apply the mean value inequality with $M = 0$. $\square$

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 ε such that $B_\varepsilon(\gamma(s)) \subseteq U$ since U is open. By continuity of γ, there is a δ 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 previous corollary, 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 $g(0) = g(1)$, i.e. $f(\mathbf{a}) = f(\mathbf{b})$. $\square$

6.4 Inverse function theorem

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. ($\Rightarrow$) 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

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

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

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

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

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.

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 $\mathbf{x} \in \overline{B_r(\mathbf{a})}$, i.e. $T(\mathbf{x}) = \mathbf{x}$. 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 $\mathbf{y} \in W$,

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

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 $\|\mathbf{v}\| = 0$, i.e. $\mathbf{v} = \mathbf{0}$. 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. ☐

6.5 2nd order derivatives

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 $D_iD_j\mathbf{f}(\mathbf{x})$ and $D_jD_i\mathbf{f}(\mathbf{x})$ exist for all $\mathbf{x} \in B_\rho(\mathbf{a})$ and are continuous at $\mathbf{a}$. Then in fact
$$D_iD_j\mathbf{f}(\mathbf{a}) = D_jD_i\mathbf{f}(\mathbf{a}).$$

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

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

Proposition. If $f : U \to \mathbb{R}^m$ is differentiable in U such that $D_i D_j \mathbf{f}(\mathbf{x})$ exists in a neighbourhood of $\mathbf{a} \in U$ and are continuous at $\mathbf{a}$, then $D\mathbf{f}$ is differentiable at $\mathbf{a}$ and

$$D^2 \mathbf{f}(\mathbf{a})(\mathbf{u}, \mathbf{v}) = \sum_j \sum_i D_i D_j \mathbf{f}(\mathbf{a}) u_i v_j.$$

is a symmetric bilinear form.

Proof. This follows from the fact that continuity of second partials implies differentiability, and the symmetry of mixed partials. □

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

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2} D^2 f(\mathbf{h}, \mathbf{h}) + E(\mathbf{h}),$$

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

$$g(1) = g(0) + g'(0) + \frac{1}{2} g''(s)$$

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

In other words,

$$\begin{aligned} f(\mathbf{a} + \mathbf{h}) &= f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2} D^2 f(\mathbf{a} + s\mathbf{h})(\mathbf{h}, \mathbf{h}) \\ &= f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2} D^2 f(\mathbf{a})(\mathbf{h}, \mathbf{h}) + E(\mathbf{h}), \end{aligned}$$

where

$$E(\mathbf{h}) = \frac{1}{2} \left(D^2 f(\mathbf{a} + s\mathbf{h})(\mathbf{h}, \mathbf{h}) - D^2 f(\mathbf{a})(\mathbf{h}, \mathbf{h}) \right).$$

By definition of the operator norm, we get

$$|E(\mathbf{h})| \leq \frac{1}{2}\|D^2 f(\mathbf{a} + s\mathbf{h}) - D^2 f(\mathbf{a})\|\|\mathbf{h}\|^2.$$

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. □