Analysis and Geometry: Supplementary Notes

November 4, 2016

CONTENTS

1 WEEK 1

1.1 VECTOR SPACES

The most familiar example of a vector space will be $\mathbb{R}^n$. Elements of $\mathbb{R}^n$ can be added together and they can be scaled, and we have rules about how these two operations interact. For example, scaling the sum of two elements should be the same as summing the individually scaled elements. As we will see, there are many other useful sets (eg. sets of sequences, sets of functions) that share this same basic structure and so we need an abstract and precise definition.

Definition 1.1. A ***vector space*** over $\mathbb{R}$ is a non-empty set V with two operations $V \times V \to V$ and $\mathbb{R} \times V \to V$, denoted by $(x, y) \mapsto x+y$ and $(\alpha, x) \mapsto \alpha x$ respectively, which satisfy the following, for all $x, y, z \in V$ and all $\alpha \in \mathbb{R}$:

- $x + y = y + x$
- $(x + y) + z = x + (y + z)$
- $\exists \mathbf{0} \in V$ s.t. $x + \mathbf{0} = x$
- $\exists (-x) \in V$ s.t. $x + (-x) = \mathbf{0}$
- $\alpha(x + y) = \alpha x + \alpha y$
- $\alpha(\beta x) = (\alpha\beta)x$

- $1x = x$
- $0x = \mathbf{0}$ for all $x, y, z \in V$ and all $\alpha, \beta \in \mathbb{R}$.

Remarks.

- Students who have studied group theory will notice that the first four properties make $(V, +)$ an abelian group.
- It is possible to have vector spaces over the complex numbers, or indeed any field. If instead of a field we use a ring with identity we have what is called a *module*.

Example 1.2. The space of quadratic polynomials with real coefficients. Let $\alpha \in \mathbb{R}$,$u = u_1 + u_2 t + u_3 t^2$ and $v = v_1 + v_2 t + v_3 t^2$, then define

$$\alpha u := \alpha u_1 + \alpha u_2 t + \alpha u_3 t^2, \quad u + v := u_1 + v_1 + (u_2 + v_2)t + (u_3 + v_3)t^2$$

Example 1.3. The space ℓ of sequences $(x_1, x_2, \ldots, x_i, \ldots)$ with $x_i \in \mathbb{R}$. Addition and scalar multiplication are the same as for $\mathbb{R}^n$.

Example 1.4. The space $\mathrm{Map}([a, b], \mathbb{R})$ of maps from the interval $[a, b] \subset \mathbb{R}$ to $\mathbb{R}$. Addition and scalar multiplication are defined pointwise, i.e. given $f, g \in \mathrm{Map}([a, b], \mathbb{R})$, we define αf by $(\alpha f)(t) := \alpha f(t)$ and $f + g$ by $(f + g)(t) := f(t) + g(t)$.

✍ 1.5. For at least one of the above verify that the properties of a vector space are satisfied.

✍ 1.6. Recall that a subset of a vector space V is called a *subspace* if it is closed under addition and scalar multiplication. Show that the set of continuous maps $C([a, b], \mathbb{R})$ is a subspace of $\mathrm{Map}([a, b], \mathbb{R})$.

Definition 1.7. A map $L : V \to W$ between vector spaces is called a *linear map* (or transformation or operator) if it preserves the vector space structure, i.e:

$$L(\alpha u + \beta v) = \alpha L(u) + \beta L(v)$$

for all $\alpha, \beta \in \mathbb{R}$ and $u, v \in V$. If L is also a bijection it is called an *isomorphism* and we say that V and W are *isomorphic*.

✍ 1.8. Show that the vector space from Example 1.2 is isomorphic to $\mathbb{R}^3$.

Definition 1.9.

- A finite set of vectors $\{v_1, v_2 \dots v_k\}$ in V is *linearly dependent* if there exist $\alpha_1, \alpha_2, \dots, \alpha_k$, not all zero, such that $\sum_{i=1}^k \alpha_i v_i = \mathbf{0}$.
- If a finite set of vectors is not linearly dependent then it is *linearly independent*.
- An *infinite* set of vectors is linearly independent if every finite subset is linearly independent.
- A vector space is n-*dimensional* (n finite) if it contains a set of n linearly independent vectors but no set of $n+1$ linearly independent vectors. Otherwise, it is infinite dimensional.

Example 1.10. The space $\mathrm{Map}([a,b],\mathbb{R})$ is infinite dimensional. For any n we can define $f_i(t) = t^i$ for $i = 0, 1, \dots, n$ and then $\{f_i : i = 0 \dots n\}$ is a set of $n+1$ linearly independent vectors. To see that they are linearly independent note that if $\sum_{i=0}^n \alpha_i f_i = \mathbf{0}$ (the zero function) then $\alpha_0 + \alpha_1 t + \ldots + \alpha_n t^n \equiv 0$ for every value of t and therefore every $\alpha_i = 0$. This also shows that $C([a,b],\mathbb{R})$ is infinite dimensional, because the f_i are continuous.

Definition 1.11.

- The *span* of a set S of vectors in V is the set of all linear combinations $\alpha_1 v_1 + \ldots + \alpha_k v_k$ where $\alpha_1, \dots \alpha_k \in \mathbb{R}$ and $v_1, \dots, v_k \in S$.
- A linearly independent subset $B \subset V$ with span equal to V is called a *basis* for V.

If B is a basis for V, then every non-zero $v \in V$ has a *unique* representation as a linear combination of finitely many elements of B.

✍ 1.12. Prove that every finite dimensional vector space has a basis, and that every finite n-dimensional vector space is isomorphic to $\mathbb{R}^n$.

1.2 NORMED VECTOR SPACES

Definition 1.13. Let V be a real vector space. A function $V \to \mathbb{R}$ denoted $v \mapsto \|v\|$ is called a *norm* on V if

- $\|v\| = 0$ iff $v = \mathbf{0}$
- $\|\alpha v\| = |\alpha| \|v\|$

- $\|u+v\| \leq \|u\| + \|v\|$

and then $(V, \|.\|)$ is called a *normed space*.

✍ 1.14. Show that a norm is always non-negative.

In most examples verifying the first two properties of a norm is straightforward, but checking the triangle inequality requires some classical inequalities:

- *Young's inequality:* for $p > 1, \frac{1}{p} + \frac{1}{q} = 1$, and $a, b \geq 0$

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

- *Hölder's inequality for sums:* for $x, y \in \mathbb{R}^n$ and $p > 1, \frac{1}{p} + \frac{1}{q} = 1$

$$\textstyle\sum_i |x_i y_i| \leq (\sum_i |x_i|^p)^{\frac{1}{p}} (\sum_i |y_i|^q)^{\frac{1}{q}}$$

  The special case $p = q = 2$ is called the Cauchy-Schwarz inequality.

- *Minkowski's inequality for sums*: for $x, y \in \mathbb{R}^n$ and $p \geq 1$

$$(\textstyle\sum_i |x_i + y_i|^p)^{\frac{1}{p}} \leq (\sum_i |x_i|^p)^{\frac{1}{p}} + (\sum_i |y_i|^p)^{\frac{1}{p}}$$

- *Hölder's inequality for integrals*:

$$\textstyle\int_a^b |f(t)g(t)|dt \leq \left(\int_a^b |f(t)|^p dt\right)^{\frac{1}{p}} \left(\int_a^b |g(t)|^q\right)^{\frac{1}{q}}$$

  where $f, g : [a, b] \to \mathbb{R}$ are such that the integrals exist.

- *Minkowski's inequality for integrals*:

$$\left(\textstyle\int_a^b |f(t) + g(t)|^p dt\right)^{\frac{1}{p}} \leq \left(\int_a^b |f(t)|^p dt\right)^{\frac{1}{p}} + \left(\int_a^b |g(t)|^p dt\right)^{\frac{1}{p}}$$

✍ 1.15. (Difficult) Prove the above inequalities. (*Hints.* For Young's inequality, susbstitute $b = ta^{p/q}$ and use some calculus. For Hölder's, set $\bar{x} = (|x_1|, \ldots, |x_n|)/\|x\|_p$, $\bar{y} = (|y_1|, \ldots, |y_n|)/\|y\|_p$ and use Young's inequality. For Minkowski's, you will need Hölder's. For the integral versions try to imitate what was done for sums.)

Example 1.16. Euclidean space: $\mathbb{R}^n$ with the standard norm

$$\|x\| = \sqrt{x_1^2 + \ldots + x_n^2}$$

This is a special case of the norm

$$\|x\|_p := (|x_1|^p + \ldots + |x_n|^p)^{\frac{1}{p}}$$

for which the triangle inequality corresponds to Minkowski's inequality.

Example 1.17. On the space of sequences ℓ (Example 1.3) we define, for $1 \le p < \infty$

$$\|x\|_p := (\textstyle\sum_{i=1}^{\infty} |x_i|^p)^{\frac{1}{p}}$$

and $\ell^p := \{x \in \ell : \|x\|_p < \infty\}$ is a normed space (Minkowski's inequality still holds with $n = \infty$). Notice that ℓ^1 is the space of absolutely convergent series. We can also define

$$\|x\|_\infty := \sup_{1 \le i < \infty} |x_i|$$

and $\ell^\infty := \{x \in \ell : \|x\| < \infty\}$ is a normed space.

Example 1.18. $C([a,b],\mathbb{R})$ with the uniform norm (a.k.a. sup norm)

$$|f|_0 := \sup_{t \in [a,b]} |f(t)|$$

The sup (in fact, max) exists by the extreme value theorem.

✍ 1.19. Check that $|.|_0$ is a norm.

Example 1.20. For $1 \le p < \infty$ and $f \in C([a,b],\mathbb{R})$ define

$$\|f\|_p := \left(\int_a^b |f(t)|^p dt\right)^{\frac{1}{p}}$$

This norm is known as the L^p norm. The triangle inequality is just the Minkowski inequality for integrals.

Every normed space has an ***induced metric*** given by $d(u,v) := \|u - v\|$. It is non-negative by exercise 1.14 and symmetric by the the commutativity of vector space addition. For the triangle inequality: $d(u,v) = \|u - v\| = \|u - w + w - v\| \le \|u - w\| + \|w - v\| = d(u,w) + d(w,v)$. It therefore also follows that every normed space has a topology.

✍ 1.21. Using the metric space definition of continuity prove that in a normed space scalar multiplication by a fixed $\alpha \in \mathbb{R}$ is continuous at $\mathbf{0}$.

2 WEEK 2

2.1 MORE NORMED SPACES

Definition 2.1. A linear map $A : V \to W$ between normed spaces is called *bounded* if there exists a constant $k \geq 0$ such that for all $v \in V$:

$$\|Av\| \leq k\|v\|$$

(It is common practice to omit the brackets when denoting a linear map, i.e. Av instead of $A(v)$. Also, we have adopted a common abuse of notation here whereby $\|\cdot\|$ denotes both the norm on V and the norm on W. It is clear from the arguments $Av \in W$,$v \in V$ which norm is intended.)

Proposition 2.2. Let $A : V \to W$ be a linear map between normed spaces. Then the following conditions are equivalent:

(a) A is continuous at $\mathbf{0}$,

(b) A is continuous,

(c) A is bounded.

Proof. (a) iff (c): Assuming A is continuous at $\mathbf{0}$, then using the open ball definition of continuity, for all $\varepsilon > 0$ there exists $\delta > 0$ such that $v \in B_\delta(\mathbf{0}) \implies Av \in B_\varepsilon(\mathbf{0})$. In terms of the respective norms this translates to $\|v\| < \delta \implies \|Av\| < \varepsilon$. Choose $\varepsilon = 1$, so there exists δ_1 such that $\|v\| < \delta_1 \implies \|Av\| < 1$. In particular, for any v the vector $\frac{\delta_1}{2\|v\|}v$ has norm $\frac{\delta_1}{2} < \delta_1$, and so it satisfies

$$\|A\tfrac{\delta_1}{2\|v\|}v\| < 1$$

which rearranges to $\|Av\| < \frac{2}{\delta_1}\|v\|$, i.e. A is bounded. Conversely, if A is bounded with constant $k > 0$ (if $k = 0$ then continuity is trivial), given $\varepsilon > 0$ we choose $\delta = \frac{\varepsilon}{k}$. Then if $\|v\| < \delta$, since A is bounded we have

$$\|Av\| \leq k\|v\| < k\frac{\varepsilon}{k} = \varepsilon.$$

✍ 2.3. Prove (a) implies (b).

□

Example 1.4 actually holds more generally: if V, W are vector spaces then $\text{Map}(V, W)$ is an infinite dimensional vector space with operations defined by $(f+g)(v) := f(v)+g(v)$ and $\alpha f)(v) := \alpha f(v)$. The linear maps $L(V, W)$ form a subspace of $\text{Map}(V, W)$.

✍ 2.4. Prove that if V, W are finite dimensional then so is $L(V, W)$.

Example 2.5. If V, W are normed spaces then we define another subspace $B(V, W)$ consisting of the bounded (and therefore continuous) linear maps. This space can be given a norm, for $A \in B(V, W)$ define:

$$|A| := \sup_{v \in V - \{\mathbf{0}\}} \frac{\|Av\|}{\|v\|} = \sup_{\|v\|=1} \|Av\|$$

The case $W = \mathbb{R}$ is important enough to have its own name: $B(V, \mathbb{R})$ is known as the ***dual space*** of V and is usually denoted V^*.

Proposition 2.6. Write E^n for $\mathbb{R}^n$ with the Euclidean norm. Every linear map $E^n \to E^m$ is continuous, i.e. $L(E^n, E^m) = B(E^n, E^m)$.

Proof. Let $f : E^n \to E^m$ be a linear map. Using the standard bases we can always find a matrix representation of f, i.e.

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix} = \begin{pmatrix} a_{11} & & \\ & \ddots & \\ & & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Now matrix multiplication can be written $f_i(x) = \sum_{j=1}^n a_{ij}x_j$. If we let $k = \max_{i,j} |a_{ij}|$ then by the triangle inequality, and then Cauchy-Schwarz,

$$|f_i(x)| \leq k(\textstyle\sum_j |x_j|) = k(\sum_j 1|x_j|) \leq k(\sum_j |x_j|^2)^{\frac{1}{2}}(\sum_j 1^2)^{\frac{1}{2}} = \sqrt{n}k\|x\|$$

hence $\|f(x)\|^2 = \sum_i |f_i(x)|^2 \leq \sum_i nk^2\|x\|^2 = n^2k^2\|x\|^2$. It follows that f is bounded, and therefore continuous (Prop. 2.2). □

The above result $L(V, W) = B(V, W)$ also holds for any *finite dimensional* vector spaces V, W, but this will be easier to prove once we introduce compactness. It is not true for infinite dimensional vector spaces.

✍ 2.7. Find an example of an *unbounded* linear operator.

2.2 INNER PRODUCT SPACES

Another familiar structure on $\mathbb{R}^n$ which generalises to other vector spaces is the dot product: $x \cdot y = x_1 y_1 + \ldots + x_n y_n$. It is an example of what is called an inner product (or sometimes scalar product).

Definition 2.8. An ***inner product*** on a real vector space V is a map $V \times V \to \mathbb{R}$ denoted $(u, v) \mapsto \langle u, v \rangle$ which satisfies:

(i) $\langle u, u \rangle \geq 0$ and $\langle u, u \rangle = 0 \iff u = \mathbf{0}$

(ii) $\langle u, v \rangle = \langle v, u \rangle$

(iii) $\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle$

for all $u, v, w \in V$, and $\alpha, \beta \in \mathbb{R}$.

There are also technical terms for each of these properties. Maps which satisfy (i) and (ii) are called ***positive semi-definite*** and ***symmetric*** respectively. The third property can be called linearity in the first argument, i.e. for fixed $w \in V$, the map $\langle \cdot, w \rangle : V \to \mathbb{R}$ is linear. By the symmetry property it is also linear in the second argument: $\langle u, \alpha v + \beta w \rangle = \langle \alpha v + \beta w, u \rangle = \alpha \langle v, u \rangle + \beta \langle w, u \rangle = \alpha \langle u, v \rangle + \beta \langle u, w \rangle$, and so we say it is ***bilinear***. In summary, an inner product is a positive semi-definite, symmetric bilinear map $V \times V \to \mathbb{R}$.

Lemma 2.9. Every inner product on V satisfies

$$\langle u, v \rangle^2 \leq \langle u, u \rangle \langle v, v \rangle \tag{1}$$

for all $u, v \in V$.

Proof. If $v = \mathbf{0}$ then the inequality holds trivially, so assume $v \neq \mathbf{0}$. By (i) and bilinearity, for any $\alpha \in \mathbb{R}$:

$$0 \leq \langle u - \alpha v, u - \alpha v \rangle = \langle u, u \rangle - 2\alpha \langle u, v \rangle + \alpha^2 \langle v, v \rangle$$

Let $\alpha = \frac{\langle u, v \rangle}{\langle v, v \rangle}$ and substitute into the above to get

$$0 \leq \langle u, u \rangle - 2\frac{\langle u, v \rangle^2}{\langle v, v \rangle} + \frac{\langle u, v \rangle^2}{\langle v, v \rangle} = \langle u, u \rangle - \frac{\langle u, v \rangle^2}{\langle v, v \rangle}$$

and now rearrange. □

A norm can be derived from any inner product according to $\|v\| := \sqrt{\langle v, v\rangle}$, and using this notation (1) is written:

$$|\langle u, v\rangle| \leq \|u\|\|v\|$$

which is the (general) *Cauchy-Schwarz inequality.*

✍ 2.10. Verify that the norm derived from an inner product is indeed a norm.

Any norm which is induced by an inner product according to the rule above must satisfy the *parallelogram equality*:

$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2)$$

(to confirm this calculate $\langle u+v, u+v\rangle + \langle u-v, u-v\rangle$). There are norms which *do not* satisfy this equality, and so not all normed spaces are inner product spaces (see examples below). If we are given a norm which does satisfy the parallelogram equality, the corresponding inner product can be recovered by the *polarization identity*:

$$\langle u, v\rangle = \tfrac{1}{4}(\|u+v\|^2 - \|u-v\|^2)$$

(this identity can also be verified by straightforward calculation).

Example 2.11. The sequence space ℓ^2 (cf. Example 1.17) has the inner product $\langle x, y\rangle := \sum_{i=1}^{\infty} x_i y_i$. This is well defined (i.e. the sum converges) by the Hölder inequality for sums (which is actually Cauchy-Schwarz for this inner product) and the assumption that x, y are in ℓ^2.

Example 2.12. ℓ^p with $p \neq 2$ is *not* an inner product space. The norm fails to satisfy the parallelogram inequality: suppose $x = (1, 1, 0, 0, \ldots), y = (1, -1, 0, 0, \ldots)$, then $\|x\|_p = 2^{\frac{1}{p}} = \|y\|_p$ and $\|x+y\|_p = 2 = \|x-y\|_p$.

Example 2.13. For $f, g \in C([a, b], \mathbb{R})$ define

$$\langle f, g\rangle := \int_a^b f(t)g(t)dt$$

The corresponding norm is the L^2 norm from example 1.20, and so this is called the L^2 inner product. Cauchy-Schwarz for this inner product is the Hölder inequality for integrals.

We remark that $C([a,b],\mathbb{R})$ is not the biggest function space on which the L^2 norm and inner product is well defined. For example, any piecewise continuous function has an L^2 norm. We will have more to say about this later.

✍ 2.14. Show that the uniform norm $|\cdot|_0$ (example 1.18) is not derived from an inner product.

3 WEEK 3

3.1 CONVERGENCE, COMPACTNESS AND COMPLETENESS

Definition 3.1. Let $(x_n) = (x_1, x_2, \ldots, x_n, \ldots)$ be a sequence in a metric space (X, d). We say the sequence is ***convergent*** and has ***limit*** $x \in X$ (notation $x_n \to x$ or $\lim_{n\to\infty} x_n = x$) if $d(x, x_n) \to 0$, i.e. for all $\varepsilon > 0$ there exists N such that $d(x, x_n) < \varepsilon$ for all $n > N$.

Lemma 3.2. Every convergent sequence is bounded.

Proof. We recall that a subset $U \subset X$ is bounded if its diameter:

$$\sup_{x,y\in U} d(x,y)$$

is finite. A sequence is bounded if the corresponding subset of X is bounded. Suppose $x_n \to x$, then for $\varepsilon = 1$ there exists N such that $d(x_n, x) < 1$ for all $n > N$. Let $a = \max_{n\le N} d(x_n, x)$, then for all n, m we have

$$d(x_n, x_m) \le d(x_n, x) + d(x, x_m) < 2\max\{a, 1\},$$

and so the sequence is bounded. □

Lemma 3.3. The limit of a convergent sequence is unique.

Proof. Suppose $x_n \to x$ and also $x_n \to y$, then

$$d(x,y) \le d(x, x_n) + d(x_n, y)$$

for all n. But for all $\varepsilon > 0$, there exist N_1, N_2 such that the right hand side of the above inequality is less than 2ε when $n > \max\{N_1, N_2\}$. Hence for all $\varepsilon > 0$ we have $d(x,y) < \varepsilon$, i.e. $d(x,y) = 0$ and $x = y$. □

✍ 3.4. Prove that if $x_n \to x$ and $y_n \to y$ then $d(x_n, y_n) \to d(x, y)$.

Proposition 3.5. A subset $U \subset X$ is closed iff for every convergent sequence $x_n \to x$ with $x_n \in U$ for all n we also have $x \in U$.

Proof. First suppose U is closed and $x_n \to x$ with $x_n \in U$. If $x \in X\backslash U$, then there exists $B_\varepsilon(x) \subset X\backslash U$ because $X\backslash U$ is open. But by convergence there exists an N such that $x_n \in B_\varepsilon(x)$ for all $n > N$, i.e. $x_n \notin U$, which contradicts an assumption, so $x \in U$. Conversely suppose $x_n \to x$ with

$x_n \in U$ implies $x \in U$. We need to prove $X \backslash U$ is open. Suppose it is not, then there exists $a \in X \backslash U$ such that for all $\varepsilon > 0$ the intersection $B_\varepsilon(a) \cap U$ is non-empty. So we can construct a sequence by choosing $x_n \in B_{1/n}(a) \cap U$. By construction $x_n \in U$ and $x_n \to a$, however $a \notin U$ which contradicts an assumption, so $X \backslash U$ must be open and U is closed. □

Definition 3.6. A subset U of a metric space X is called (sequentially[1]) ***compact*** if every sequence $(x_n) \subset U$ has a convergent subsequence with limit in U

Note: a subsequence is obtained by deleting elements of the original sequence, the order is not changed.

✍ 3.7. All compact sets are closed.

Lemma 3.8. Every sequence (x_n) in $\mathbb{R}$ has a monotone subsequence.

Proof. We call x_n a *trough* if $x_m > x_n$ for all $m > n$. Consider the three cases:

- If there are no troughs then there exists $x_{n_i} \leq x_1$, and since x_{n_i} is not a trough there exists $x_{n_{i+1}} \leq x_{n_i}$. Thus we construct a non-increasing subsequence.
- If there are finitely many troughs, delete everything before the last trough and proceed as if there are no troughs.
- If there are infinitely many troughs $x_{n_1} < x_{n_2} < \dots$ then they form an increasing subsequence.

□

Theorem 3.9. (*Bolzano-Weierstrass Theorem*) Every closed and bounded interval $[a, b] \subset \mathbb{R}$ is compact.

Proof. By Lemma 3.8 any sequence in $[a, b]$ has a monotone subsequence (which is bounded), so by the monotone sequences theorem this subsequence converges. Since the interval is closed, it contains the limit. □

[1] Later in the main notes you will see a more general definition of compactness which applies to topological spaces, and agrees with this one in metric spaces.

Definition 3.10. A sequence (x_n) in a metric space X is called a ***Cauchy sequence*** if for all $\varepsilon > 0$ there exists $N > 1$ such that $d(x_n, x_m) < \varepsilon$ for all $m, n \geq N$.

✍ 3.11. Every Cauchy sequence is bounded.

✍ 3.12. Every convergent sequence is a Cauchy sequence (the converse is not always true).

Example 3.13. Let $X = (0, 1]$ with the standard Euclidean metric from $\mathbb{R}$ (derived from absolute value). Then the sequence $x_n = 1/n$ is Cauchy but not convergent in X (although it is convergent in $\mathbb{R}$). This demonstrates that convergence is not a property of the sequence alone, but also depends on the metric space containing the sequence.

Example 3.14. Consider the rational numbers $\mathbb{Q}$ with the standard metric (derived from absolute value). The sequence of rationals defined recursively by $x_{n+1} := \frac{x_n}{2} + \frac{1}{x_n}$, $x_1 = 1$, is Cauchy but not convergent in $\mathbb{Q}$. (This sequence computes approximations of $\sqrt{2}$ and is sometimes called the Babylonian method).

Definition 3.15. A metric space is called ***complete*** if every Cauchy sequence converges.

Lemma 3.16. If X is compact then it is complete.

Proof. Let (x_n) be a Cauchy sequence. Since X is compact there is a convergent subsequence $x_{n_k} \to x$, and we will show that $x_n \to x$ also. Given $\varepsilon > 0$ there exists N such that $d(x_n, x_m) < \varepsilon/2$ for all $m, n > N$, and there exists K such that $d(x_{n_k}, x) < \varepsilon/2$ for all $k > K$. Hence for $n \geq N$, choosing $k > K$ sufficiently large so that $n_k > N$:

$$d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_k}, x) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

i.e. $x_n \to x$. □

Example 3.14 showed that $\mathbb{Q}$ is not complete.

Proposition 3.17. $\mathbb{R}$ is complete.

Proof. Consider a Cauchy sequence (x_n) in $\mathbb{R}$. By exercise 3.11 (x_n) is bounded, and therefore it is contained in a closed interval $[a, b]$. By the Bolzano-Weierstrass theorem this interval is compact, and then it is complete by 3.16 and the sequence converges. □

✍ 3.18. (Difficult) ℓ^p is complete.

Proposition 3.19. $C([a,b],\mathbb{R})$ is complete with respect to the metric induced by the uniform norm $|\cdot|_0$

Proof. Let (x_n) be a Cauchy sequence, i.e. given $\varepsilon > 0$ there exists N such that for all $m, n > N$

$$d(x_n, x_m) = \sup_{t\in[a,b]} |x_n(t) - x_m(t)| < \varepsilon$$

Hence for any fixed $t_0 \in [a,b]$ we also have $|x_n(t_0) - x_m(t_0)| < \varepsilon$. So $(x_n(t_0))$ is a Cauchy sequence in $\mathbb{R}$, which is complete, so $(x_n(t_0))$ converges (this is called *pointwise* convergence). For each $t \in [a,b]$ define $x(t) := \lim_{n\to\infty} x_{n(t)}$. We will first show that x is continuous, and then that $x_n \to x$ in the uniform norm.

Suppose x has a point of discontinuity at t_0, i.e. there exists $\varepsilon > 0$ such that for all $\delta > 0$ there exists $t_1 \in B_\delta(t_0)$ with $|x(t_1) - x(t_0)| > \varepsilon$. By pointwise convergence there exists n such that $|x_n(t_1) - x(t_1)| < \varepsilon/3$ and $|x_n(t_0) - x(t_0)| < \varepsilon/3$. Then using the triangle inequality

$$\begin{aligned}\varepsilon < |x(t_1) - x(t_0)| &= |x(t_1) - x_n(t_1) + x_n(t_1) - x_n(t_0) + x_n(t_0) - x(t_0)| \\ &< |x(t_1) - x_n(t_1)| + |x_n(t_1) - x_n(t_0)| + |x_n(t_0) - x(t_0)| \\ &< \varepsilon/3 + |x_n(t_1) - x_n(t_0)| + \varepsilon/3\end{aligned}$$

From this we conclude $\varepsilon/3 < |x_n(t_1) - x_n(t_0)|$, and therefore t_0 is a point of discontinuity of x_n. But this contradicts the assumption that $x_n \in C([a,b],\mathbb{R})$, so x must be continuous.

Now suppose that x_n does not converge to x. Then there is an $\varepsilon > 0$ such that for all N there exists $n > N$ with $\sup_{t\in[a,b]} |x_n(t) - x(t)| > \varepsilon$. But then there also exists $t_0 \in [a,b]$ such that $|x_n(t_0) - x(t_0)| > \varepsilon/2$, which contradicts pointwise convergence. □

Proposition 3.20. $C([a,b],\mathbb{R})$ is *not* complete in the (metric derived from the) L^2 norm.

Proof. To make the notation lighter we let $[a,b] = [0,1]$. We will construct a Cauchy sequence in $C([0,1],\mathbb{R})$ which converges in the space of *piecewise-continuous* functions (also with the L^2 inner product). Define

$$f_n(t) := \begin{cases} 0 & 0 \le t < \frac{1}{2} - \frac{1}{2n} \\ nt + \frac{1}{2}(1-n) & \frac{1}{2} - \frac{1}{2n} \le t < \frac{1}{2} + \frac{1}{2n} \\ 1 & \frac{1}{2} + \frac{1}{2n} \le t < 1 \end{cases}$$

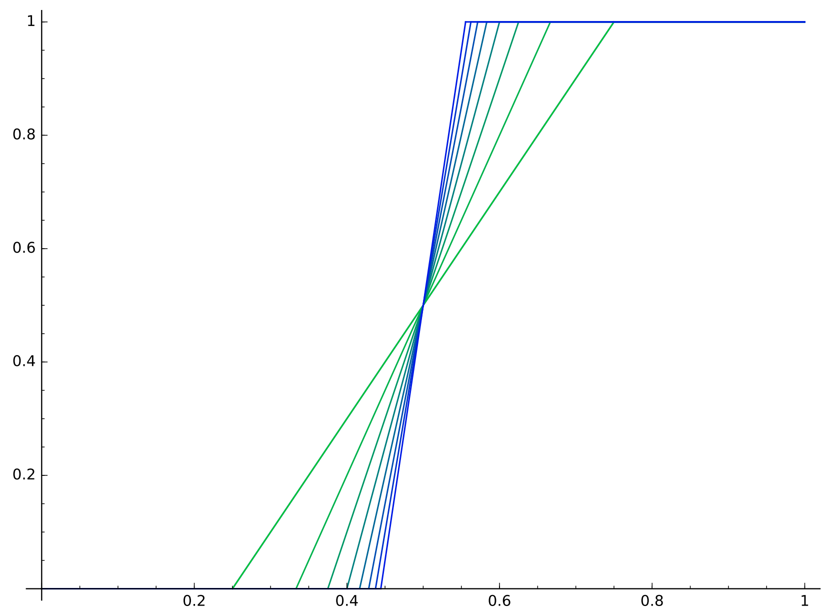


Figure 1: A plot of f_n for $n = 2 \dots 10$. Blue represents higher values of n.

so (f_n) is a sequence of functions which are all in $C([0,1],\mathbb{R})$. Since any convergent sequence is Cauchy, if we can prove that (f_n) converges in the L^2-norm to a function h which is *not* continuous, then we have a Cauchy sequence in $C([0,1],\mathbb{R}), \|\cdot\|_2$. Moreover, by uniqueness of limits we will have shown that this sequence does not converge (in $C([0,1],\mathbb{R}), \|\cdot\|_2$). We will therefore have a counterexample proving $C([0,1],\mathbb{R}), \|\cdot\|_2$ is not complete. Define

$$h(t) := \begin{cases} 0 & 0 \le t \le \frac{1}{2} \\ 1 & \frac{1}{2} < t \le 1 \end{cases}$$

and observe that $f_n(t) - h(t) = 0$ except when $\frac{1}{2} - \frac{1}{2n} < t < \frac{1}{2} + \frac{1}{2n}$, and notice also that $(f_n(t) - h(t))^2 < 1$. Thus

$$\begin{aligned} \|f_n - h\|_2^2 = \int_0^1 (f_n(t) - h(t))^2 dt &= \int_{\frac{1}{2}-\frac{1}{2n}}^{\frac{1}{2}+\frac{1}{2n}} (f_n(t) - h(t))^2 dt \\ &< \int_{\frac{1}{2}-\frac{1}{2n}}^{\frac{1}{2}+\frac{1}{2n}} 1 dt = \frac{1}{n} \end{aligned}$$

and therefore $\lim_{n\to\infty} \|f_n - h\|_2 = 0$. $\square$

Note that this proof is easily adapted to general intervals $[a, b]$, and also the L^p norm.

4 WEEK 4

4.1 COMPLETION OF METRIC SPACES

Definition 4.1. A map $f : X \to Y$ between metric spaces is called an ***isometry*** if for all $x, y \in X$,

$$d_Y(f(x), f(y)) = d_X(x, y)$$

The metric spaces X and Y are called ***isometric*** if there exists an isometry $X \to Y$ which is bijective.

✍ 4.2. Describe all the isometries from $\mathbb{R}^2$ to itself (with the Euclidean metric).

✍ 4.3. Show that a linear map $A : V \to W$ between normed spaces is an isometry iff $\|Av\| = \|v\|$

Definition 4.4. A subset M of a metric space X is called ***dense*** in X if the closure $\bar{M}$ (i.e. the smallest closed subset of X containing M) is equal to X.

Lemma 4.5. *M is dense in X iff for all $x \in X$ and $\varepsilon > 0$, $B_\varepsilon(x) \cap M$ is non-empty.*

Proof. If for some $x \in X$ and $\varepsilon > 0$ we have $B_\varepsilon(x) \cap M = \emptyset$, then $X \backslash B_\varepsilon(x)$ is a closed proper subset of X which contains M, i.e. $\bar{M} \neq X$. Conversely if there exists $m \in B_\varepsilon(x) \cap M$ for all $x \in X$ and $\varepsilon > 0$, then for each $x \in X$ we can construct a sequence (m_n) in M with $m_n \in B_{1/n}(x) \cap M$. Since $m_n \to x$ we must have $x \in \bar{M}$ by Proposition 3.5. □

✍ 4.6. $\mathbb{Q}$ is dense in $\mathbb{R}$.

The following inequality will be useful in the proof of Theorem 4.8.

Lemma 4.7. *Let $u, v, w, x \in X$, then*

$$|d(u, v) - d(w, x)| \leq d(u, w) + d(v, x). \tag{2}$$

Proof. By the triangle inequality $d(u, v) \leq d(u, w) + d(w, x) + d(x, v)$ and therefore $d(u, v) - d(w, x) \leq d(u, w) + d(x, v)$. Similarly $d(w, x) - d(u, v) \leq d(w, u) + d(v, x)$, and the result follows by combining these two inequalities. □

Theorem 4.8. *Given any metric space X, d there exists a complete metric space $\hat{X}, \hat{d}$ such that X is isometric to a dense subspace of $\hat{X}$. This subspace $\hat{X}$ is called the completion of X and is unique up to isometries (i.e. if $\tilde{X}$ is another completion then $\tilde{X}$ and $\hat{X}$ are isometric).*

Proof. First we will construct $\hat{X}, \hat{d}$ out of equivalence classes of Cauchy sequences. Suppose (x_n) and (x'_n) are Cauchy sequences in X and define an equivalence relation

$$(x_n) \sim (x'_n) \text{ iff } \lim_{n\to\infty} d(x_n, x'_n) = 0.$$

We write $[(x_n)]$ for the equivalence class of (x_n). Define $\hat{X}$ to be the set of all equivalence classes of $\sim$, and for $\hat{x} = [(x_n)], \hat{y} = [(y_n)]$ in $\hat{X}$ we define

$$\hat{d}(\hat{x}, \hat{y}) := \lim_{n\to\infty} d(x_n, y_n)$$

There are several things to check to confirm that $\hat{d}$ is well-defined: the limit must always exist and be independent of the particular choices $(x_n), (y_n)$ representing the equivalence classes, and $\hat{d}$ must satisfy the properties of a metric. For existence of the limit, we use (2)

$$|d(x_m, y_m) - d(x_n, y_n)| \leq d(x_n, x_m) + d(y_m, y_n)$$

Since $(x_n), (y_n)$ are Cauchy sequences, for all $\varepsilon > 0$ there exists N such that the left hand side of the above is less than ε. This means the sequence $(d(x_n, y_n))$ is Cauchy and therefore converges by completeness of $\mathbb{R}$, and so $\lim_{n\to\infty} d(x_n, y_n)$ exists. To see that the limit does not depend on the choices of $(x_n), (y_n)$ suppose we also have $(x_n)' \in \hat{x}$ and $(y'_n) \in \hat{y}$. Using (2) again we have

$$|d(x_n, y_n) - d(x'_n, y'_n)| \leq d(x_n, x'_n) + d(y_n, y'_n)$$

Since $(x_n) \sim (x'_n)$ and $(y_n) \sim (y'_n)$, in the limit $n \to \infty$ the above inequality gives $\lim_{n\to\infty} |d(x_n, y_n) - d(x'_n, y'_n)| = 0$ and therefore $\lim_{n\to\infty} d(x_n, y_n) = \lim_{n\to\infty} d(x'_n, y'_n)$. To see that $\hat{d}$ is a metric, observe that non-negativity and symmetry follow directly from the properties of d. Moreover $\hat{d}(\hat{x}, \hat{y}) = 0 = \lim_{n\to\infty} d(x_n, y_n)$ iff $(x_n) \sim (y_n)$ and $\hat{x} = \hat{y}$. The triangle inequality for $\hat{d}$ follows by taking the limit of the triangle inequality for d.

Next we will construct an isometry $T : X \to T(X) \subset \hat{X}$ by $b \mapsto \hat{b} := [(b, b, \ldots)]$, i.e. $\hat{b}$ is the equivalence class of a constant (Cauchy) sequence. This map is injective and if $\hat{c} = [(c, c, \ldots)]$ then

$$\hat{d}(\hat{b}, \hat{c}) = \lim_{n\to\infty} d(b, c) = d(b, c)$$

so T is indeed an isometry. To show that $T(X)$ is dense we will use the criterion provided by Lemma 4.5 and prove that for all $\varepsilon > 0$ and any $\hat{x} =$

$[(x_n)] \in X$ the open ball $B_\varepsilon(\hat{x})$ contains an element of $T(X)$. Indeed since (x_n) is Cauchy there exists N such that $d(x_n, x_N) < \varepsilon/2$ for all $n > N$, and then $\hat{x}_N := [(x_N, x_N, \ldots)] \in T(X)$ satisfies

$$\hat{d}(\hat{x}, \hat{x}_N) = \lim_{n\to\infty} d(x_n, x_N) \leq \varepsilon/2 < \varepsilon$$

i.e. $\hat{x}_N \in T(X) \cap B_\varepsilon(\hat{x})$.

We now show that $\hat{X}$ is complete. Let $(\hat{x}_n)$ be a Cauchy sequence in $\hat{X}$. Since $T(X)$ is dense in $\hat{X}$, for each $\hat{x}_n$ there exists $z_n \in X$ with $\hat{z}_n := T(z_n) \in B_{1/n}(\hat{x}_n)$. In this way we construct a sequence $(\hat{z}_n)$ in $T(X)$ which satisfies

$$\hat{d}(\hat{z}_m, \hat{z}_n) \leq \hat{d}(\hat{z}_m, \hat{x}_m) + \hat{d}(\hat{x}_m, \hat{x}_n) + \hat{d}(\hat{x}_n, \hat{z}_n) \leq \tfrac{1}{m} + d(\hat{x}_m, \hat{x}_n) + \tfrac{1}{n}$$

which shows that $(\hat{z}_n)$ is Cauchy (because $(\hat{x}_n)$ is). Now since T is an isometry the corresponding sequence (z_n) in X is also Cauchy, and so it is a member of an equivalence class in $\hat{X}$. We will write $\hat{x} := [(z_n)]$, this is our candidate for the limit of $(\hat{x}_n)$. Since $\hat{z}_n \in B_{1/n}(\hat{x}_n)$ we have

$$\hat{d}(\hat{x}_n, \hat{x}) \leq \hat{d}(\hat{x}_n, \hat{z}_n) + \hat{d}(\hat{z}_n, \hat{x}) \leq \tfrac{1}{n} + \hat{d}(\hat{z}_n, \hat{x})$$

Now recalling that $\hat{x} = [(z_m)]$ and $\hat{z}_n = [(z_n, z_n, \ldots)]$, and the definition of $\hat{d}$,

$$\hat{d}(\hat{x}_n, \hat{x}) \leq \tfrac{1}{n} + \lim_{m\to\infty} d(z_n, z_m)$$

and now taking $\lim_{n\to\infty}$ shows $\hat{x}_n \to \hat{x}$.

Finally, let us prove that the completion is unique up to isometry. Suppose $\tilde{T} : X \to \tilde{X}$ is another completion. Then $T\tilde{T}^{-1} : \tilde{T}(X) \to T(X)$ is an isometry. We define an isometry $\phi : \tilde{X} \to \hat{X}$ as follows. For any $\tilde{x} \in \tilde{X}$, since $\tilde{T}(X)$ is dense there exists $(\tilde{x}_n)$ in $\tilde{T}(X)$ such that $\tilde{x}_n \to \tilde{x}$. Then because $T\tilde{T}^{-1}$ is an isometry, $(T\tilde{T}^{-1}\tilde{x}_n)$ is a Cauchy sequence and therfore has a limit $\hat{x} =: \phi(\tilde{x})$. Before proving that this is a well defined function (i.e. it does not depend on the choice of $(\tilde{x}_n)$) we note that given $\tilde{x}, \tilde{y}$:

$$\hat{d}(\phi(\tilde{x}), \phi(\tilde{y})) = \lim_{n\to\infty} d(\tilde{T}^{-1}(\tilde{x}_n), \tilde{T}^{-1}\tilde{y}_n) = \lim_{n\to\infty} \tilde{d}(\tilde{x}_n, \tilde{y}_n) = \tilde{d}(\tilde{x}, \tilde{y})$$

by definition of $\hat{d}$, the fact that $\tilde{T}$ is an isometry, and exercise 3.4. This shows that ϕ is an isometry and also that it is well defined. Indeed if $\tilde{x}_n \to \tilde{x}$, $\tilde{x}'_n \to \tilde{x}$ and $T\tilde{T}^{-1}\tilde{x}_n \to \hat{x}$, $T\tilde{T}^{-1}\tilde{x}'_n \to \hat{x}'$, then the above equality implies that $\hat{d}(\hat{x}, \hat{x}') = \tilde{d}(\tilde{x}, \tilde{x}) = 0$. So ϕ is a well defined isometry. In fact an inverse can be constructed by the same method so it is also bijective and $\tilde{X}$ is isometric to $\hat{X}$. □

Example 4.9. We have already noted that $\mathbb{Q}$ is dense in $\mathbb{R}$, but we did not specify exactly how $\mathbb{R}$ should be constructed. In fact one way of defining $\mathbb{R}$ is as the completion of $\mathbb{Q}$.

✍ 4.10. (difficult) The metric space completion of a normed space can be given the structure of a normed space (note that this requires a vector space structure as well as a norm on the space of equivalence classes of Cauchy sequences).

5 WEEK 5

5.1 BANACH AND HILBERT SPACES

Definition 5.1. A normed space which is complete with respect to the metric derived from the norm is called a *Banach space*. An inner product space which is complete with respect to the metric derived from the (norm derived from the) inner product is called a *Hilbert space*.

Example 5.2. We will prove below that every finite dimensional normed space is a Banach space.

Example 5.3. Proposition 3.20 showed that $C([a,b],\mathbb{R}), \|\cdot\|_2$ is not complete. The completion is denoted $L^2([a,b],\mathbb{R})$, and called the space of square integrable functions. More generally $L^p([a,b],\mathbb{R})$ can be defined as the completion of $C([a,b],\mathbb{R}), \|\cdot\|_p$. This is a Banach space for each $p \geq 1$, and a Hilbert space when $p = 2$.

Note that according to the proof of Prop. 3.20, $L^1([a,b],\mathbb{R})$ must contain all the piecewise continuous functions. Moreover it must then also contain the limits of Cauchy sequences of piecewise continuous functions, some of which are not Riemann integrable, and yet have a well-defined L^1 norm. We might use this as motivation for developing a method of integration which does apply to such functions. Such a method does exist but it is beyond the scope of this course. It is called *Lebesgue integration* and $L^1([a,b],\mathbb{R})$ can also be defined as the set of Lebesgue integrable functions.

Lemma 5.4. Let $\{v_1,\ldots,v_n\}$ be a linearly independent set of vectors in a normed space $V, \|\cdot\|$. Then there exists a constant $c > 0$ such that for any scalars $\alpha^1,\ldots,\alpha^n$ (the superscript is denoting an index here, not an exponent):

$$\|\alpha^1 v_1 + \ldots + \alpha^n v_n\| \geq c(|\alpha^1| + \ldots + |\alpha^n|)$$

Proof. Let $s = (|\alpha^1| + \ldots + |\alpha^n|)$. If $s = 0$ then the inequality holds for any s, so suppose $s \neq 0$ and define $\beta^i := \alpha^i/s$. Dividing each side of the inequality by s we see that an equivalent statement is that for any $\beta^i,\ldots,\beta^n$ with $\sum_i |\beta^i| = 1$ there exists $c > 0$ such that

$$\|\sum_i \beta^i v_i\| \geq c$$

Suppose this statement is false. Then there is a sequence

$$(\beta_m) = (\beta_m^1 v_1 + \ldots + \beta_m^n v_n)$$

in V with $\sum_i |\beta^i_m| = 1$ such that $\|\beta_m\| \to 0$. Since each coefficient sequence (β^i_m) is bounded $|\beta^i_m| \leq 1$, by the Bolzano-Weierstrass theorem 3.9 there is a convergent subsequence $\beta^i_{m_k} \to \beta^i$ for each i. To avoid nested subscripts we introduce some new notation: a subsequence of (β_m) will be denoted $(\beta_{m\in I})$ where $I \subset \mathbb{N}$, and then a subsequence of a subsequence can be denoted $(\beta_{m\in J})$ where $J \subset I$.

Take a *diagonal subsequence* of (β_m) as follows: start with a subsequence $(\beta_{m\in I_1})$ such that $\beta^1_{m\in I_1} \to \beta^1$, then by Bolzano-Weierstrass this subsequence has a subsequence $(\beta_{m\in I_2 \subset I_1})$ for which the second coefficient also converges $\beta^2_{m\in I_2} \to \beta^2$. Continuing in this manner we construct a subsequence $(\beta_{m\in I_n})$ such that $\beta^i_{m\in I_n} \to \beta^i$ for all i, and therefore $\beta_{m\in I_n} \to \beta = \sum_i \beta^i v_i$ because

$$\|\beta_m - \beta\| = \|\sum_i (\beta^i_m - \beta^i) v_i\| \leq \sum_i |\beta^i_m - \beta^i| \|v_i\|.$$

Now since $|\beta^i_{m\in I_n}| \to |\beta^i|$ and $\sum_i |\beta^i_m| = 1$ it follows that $\sum_i |\beta^i| = 1$. But then there exists i such that $\beta^i \neq 0$, which by the linear independence of $\{v_1, \ldots, v_n\}$ implies that $\beta \neq 0$, and then $\|\beta_{m\in I_n}\| \to \|\beta\| \neq 0$ contradicts the assumption that $\|\beta_m\| \to 0$ (a subsequence of a convergent sequence must have the same limit).

□

Proposition 5.5. Every finite dimensional normed space is complete (and therefore a Banach space).

Proof. Let (x_m) be a Cauchy sequence in an n-dimensional normed space $V, \|\cdot\|$. Let $\{e_1, \ldots e_n\}$ be a basis for V, so $x_m = x^1_m e_1 + \ldots + x^n_m e_n$. Since (x_m) is Cauchy, for all $\varepsilon > 0$ there exists N such that if $m, \ell > N$ then by Lemma 5.4

$$\varepsilon > \|x_m - x_\ell\| = \|\sum_i (x^i_m - x^i_\ell) e_i\| \geq c(\sum_i |x^i_m - x^i_\ell|)$$

for some $c > 0$. Therefore $|x^i_m - x^i_\ell| < \varepsilon/c$, i.e. (x^i_m) is Cauchy. Since $\mathbb{R}$ is complete (Prop. 3.17) (x^i_m) converges $x^i_m \to x^i$ for each i. Let $x := x^1 e_1 + \ldots + x^n e_n$ be the corresponding vector. Then from

$$\|x_m - x\| = \|\sum_i (x^i_m - x^i) e_i\| \leq \sum_i |x^i_m - x^i| \|e_i\|$$

and $|x^i_m - x^i| \to 0$ we have $\|x_m - x\| \to 0$, i.e. $x_m \to x$. □

We have already seen that there exist infinite dimensional normed spaces which are not Banach spaces, however according to exercise 4.10 every normed space is isometric to a subspace of a Banach space.

5.2 EQUIVALENT SPACES

Definition 5.6. Two metrics d and e on a metric space X are said to be ***equivalent metrics*** if there exist positive constants c_1, c_2 such that for all $x, y \in X$

$$c_1 d(x,y) \leq e(x,y) \leq c_2 d(x,y)$$

Two norms $|\cdot|$ and $\|\cdot\|$ on a vector space V are called ***equivalent norms*** if there exist positive constants c_1, c_2 such that for all $v \in V$

$$c_1\|v\| \leq |v| \leq c_2\|v\|$$

We say two normed spaces $V, |\cdot|$ and $W, \|\cdot\|$ are equivalent if there is a linear isomorphism $A : V \to W$ and positive constants c_1, c_2 such that

$$c_1|v| \leq \|Av\| \leq c_2|v|.$$

✍ 5.7. Equivalent norms induce equivalent metrics, and equivalent metrics induce *the same* topology.

✍ 5.8. If d, e are equivalent metrics on X, then X, d and X, e have the same Cauchy sequences.

✍ 5.9. Equivalent normed spaces are homeomorphic (Prop. 2.2 will be useful).

Proposition 5.10. All finite n-dimensional normed (Banach) spaces are equivalent.

Proof. Let $V, |\cdot|$ and $W, \|\cdot\|$ be n-dimensional Banach spaces, with bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_n\}$ respectively. Define (as you might have in exercise 1.12) an isomorphism $\phi : V \to W$ by

$$\phi x = \phi(x^1 v_1 + \ldots + x^n v_n) := x^1 w_1 + \ldots + x^n w_n$$

Then we want to prove there exist $c_1, c_2 > 0$ such that $c_1|x| \leq \|\phi x\| \leq c_2|x|$ for all $x \in V$. If we let $k = \max_j \|w_j\|$ then

$$\|\phi x\| = \|x^1 w_2 + \ldots + x^n w_n\| \leq \sum_i |x^i| \|w_i\| \leq k \sum_i |x_i|$$

and then by Lemma 5.4 there exists $c > 0$ such that $\|\phi x\| \leq \frac{k}{c}|x|$. To find c_1 we can apply exactly the same argument to ϕ^{-1} and obtain k', c' such that $|\phi^{-1}y| \leq \frac{k'}{c'}\|y\|$ for all $y \in W$. Setting $y = \phi(x)$ and rearranging gives $\frac{c'}{k'}|x| \leq \|\phi x\|$. □

In particular it follows that any two norms on a finite dimensional vector space are equivalent, and then by exercise 5.7 it follows that there is only *one* topology on a finite dimensional vector space which comes from a norm. These statements do not hold for infinite dimensional spaces.

✍ 5.11. Show that on $C([a, b], \mathbb{R})$ the uniform norm $|\cdot|_0$ and the L^2 norm $\|\cdot\|$ are not equivalent (Hint: suppose there exists c such that $|f|_0 \leq c\|f\|_2$ and construct a continuous function which contradicts this inequality).

✍ 5.12. Let $\mathcal{T}^0, \mathcal{T}^2$ be the respective topologies on $C([a, b], \mathbb{R})$ induced by the uniform norm and the L^2 norm. Show that $\mathcal{T}^0$ is ***finer*** than $\mathcal{T}^2$, i.e. $\mathcal{T}^2 \subset \mathcal{T}^0$.

5.3 SUMMARY

The following diagram summarises the inclusions between the various kinds of spaces we have covered.

		topological spaces		
		$\cup$		
complete metric spaces	$\subset$	metric spaces		
$\cup$		$\cup$		
Banach spaces	$\subset$	normed vector spaces	$\subset$	vector spaces
$\cup$		$\cup$		
Hilbert spaces	$\subset$	inner product spaces		

6 WEEK 6

6.1 ORTHONORMAL SETS

We will work with Hilbert spaces in this section, even though most of the definitions and results apply in any inner product space.

Definition 6.1. Let $\mathcal{H}, \langle \cdot, \cdot \rangle$ be a Hilbert space.

- We say $x, y \in \mathcal{H}$ are *orthogonal* ($x \perp y$) if $\langle x, y \rangle = 0$.
- A set F of vectors in $\mathcal{H}$ is called an *orthogonal set* if all distinct pairs $e, f \in F, e \neq f$ are orthogonal.
- An orthogonal set F with $\|e\| = 1$ for all $e \in F$ is called *orthonormal*.

Lemma 6.2. An orthogonal set is linearly independent.

Proof. Suppose F is an orthogonal set and $\sum_{i=1}^n \alpha_i e_i = 0$ for some $e_1, \ldots, e_n \in F$ and $\alpha_i \in \mathbb{R}$. Then for any $j = 1, \ldots, n$

$$0 = \langle \sum \alpha_i e_i, e_j \rangle = \sum \alpha_i \langle e_i, e_j \rangle = \alpha_j \|e_j\|^2$$

because $\langle e_i, e_j \rangle = 0$ when $i \neq j$, and therefore $\alpha_j = 0$ for all j. □

Example 6.3. The standard basis for $\mathbb{R}^n$:

$$e_1 = (1, 0, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$$

is orthonormal with respect to the dot product. Similarly the sequence space ℓ^2 (cf. example 1.17) contains the set:

$$\{(1, 0, 0, \ldots), (0, 1, 0, 0, \ldots), (0, 0, 1, 0, 0, \ldots), \ldots\}$$

which is orthonormal in the inner product: $v \cdot w := \sum_{i=1}^\infty v_i w_i$ for $v, w \in \ell^2$.

Example 6.4. Consider $f_n(x) := \cos nx \in L^2([0, \pi], \mathbb{R})$ for $n = 0, 1, \ldots$, then we will show that $\{f_0, f_1, \ldots\}$ is an orthogonal set in L^2. For $n \geq 1$:

$$\langle f_0, f_n \rangle = \int_0^\pi 1 \cos nx dx = \left. \frac{-\sin nx}{n} \right|_0^\pi = 0$$

For $n \neq m$, and $n, m \geq 1$, using the trigonometric identity $2\cos\theta\cos\phi = \cos(\theta - \phi) + \cos(\theta + \phi)$:

$$\begin{aligned}
\langle f_n, f_m \rangle &= \int_0^\pi \cos(nx)\cos(mx)dx \\
&= \tfrac{1}{2}\int_0^\pi (\cos(n+m)x + \cos(n-m)x)dx \\
&= \frac{-1}{2(n+m)}\sin(n+m)x\bigg|_0^\pi - \frac{1}{2(n-m)}\sin(n-m)x\bigg|_0^\pi = 0
\end{aligned}$$

To construct an orthonormal set we need the norms:

$$\|f_n\|^2 = \int_0^\pi \cos^2 nx dx = \tfrac{1}{2}\int_0^\pi (1 + \cos 2nx)dx = \frac{\pi}{2}, \|f_0\|^2 = \int_0^\pi 1 dx = \pi$$

so $\{e_0 = \frac{1}{\sqrt{\pi}}, e_n = \sqrt{\frac{2}{\pi}}\cos nx, n \geq 1\}$ is an orthonormal set in $L^2([0,\pi],\mathbb{R})$.

Proposition 6.5. Let $\{v_1, \ldots, v_k\}$ be a set of linearly independent vectors in $\mathcal{H}$. Then there exists an orthonormal set $\{e_1, \ldots, e_n\}$ with $\operatorname{Span}\{v_1, \ldots, v_k\} = \operatorname{Span}\{e_1, \ldots, e_n\}$

Proof. (Gram-Schmidt algorithm). Normalise the first vector $e_1 := v_1/\|v_1\|$ and define

$$\tilde{e}_2 := v_2 - \langle v_2, e_1 \rangle e_1$$

(i.e. v_2 minus its component in the e_1 direction). Note that $\tilde{e}_2$ is non-zero: if not then $v_2 = \langle v_2, e_1 \rangle e_1 = \langle v_2, e_1 \rangle v_1/\|v_1\|$ and v_2 is a scalar multiple of v_1, but they are supposed to be linearly independent. We therefore define $e_2 := \tilde{e}_1/\|\tilde{e}_2\|$, and e_1, e_2 are orthonormal:

$$\langle e_1, e_2 \rangle = \tfrac{1}{\|\tilde{e}_2\|}\langle e_1, \tilde{e}_2 \rangle = \tfrac{1}{\|\tilde{e}_2\|}\langle e_1, v_2 - \langle v_2, e_1 \rangle e_1 \rangle = \tfrac{1}{\|\tilde{e}_2\|}(\langle e_1, v_2 \rangle - \langle v_2, e_1 \rangle) = 0$$

Suppose inductively that $\{e_1, \ldots, e_k\}$ $(k < n)$ is an orthonormal set which spans the same subspace as $\{v_1, \ldots, v_k\}$. Then we define

$$\tilde{e}_{k+1} := v_{k+1} - \sum_{i=1}^{k} \langle v_{k+1}, e_i \rangle e_i$$

Once again $\tilde{e}_{k+1} \neq 0$ or else $\{v_1 \ldots v_n\}$ will be linearly dependent, and so we define $e_{k+1} := \tilde{e}_{k+1}/\|\tilde{e}_{k+1}\|$. It remains to check that e_{k+1} thus defined is orthogonal to each e_j for $j \leq k$:

$$\begin{aligned}\langle e_{k+1}, e_j\rangle &= \tfrac{1}{\|\tilde{e}_{k+1}\|}(\langle v_{k+1}, e_j\rangle - \textstyle\sum_{i=1}^{k}\langle v_{k+1}, e_i\rangle\langle e_i, e_j\rangle)\\ &= \tfrac{1}{\|\tilde{e}_{k+1}\|}(\langle v_{k+1}, e_j\rangle - \langle v_{k+1}, e_j\rangle) = 0\end{aligned}$$

□

✍ 6.6. Recall from exercise 1.10 that $\{f_0(t) = 1, f_1(t) = t, \ldots, f_j(t) = t^j, \ldots\}$ is a linearly independent set in $C([-1,1], \mathbb{R})$, and therefore also in $L^2([-1,1], \mathbb{R})$. Use the Gram-Schmidt algorithm to orthonormalise the first four functions. If you would like to check your answers look up the Legendre polynomials online.

Definition 6.7. Let $A \subset \mathcal{H}$ be a non-empty subset. The ***orthogonal complement*** of A in $\mathcal{H}$ is $A^{\perp} := \{v \in \mathcal{H} : \langle v, a\rangle = 0 \text{ for all } a \in A\}$.

✍ 6.8. If A contains an open ball then $A^{\perp} = \{0\}$.

✍ 6.9. Let $\sum_{i=1}^{\infty} x_i$ be a series in a Banach space which is absolutely convergent, i.e. $\sum_{i=1}^{\infty}\|x_i\|$ converges. Show that $\sum_{i=1}^{\infty} x_i$ also converges.

Theorem 6.10. (Bessel's inequality) Let $\{e_1, e_2, \ldots\}$ be an orthonormal sequence in $\mathcal{H}$, then for all $v \in \mathcal{H}$:

$$\sum_{n=1}^{\infty}\langle v, e_n\rangle^2 \leq \|v\|^2$$

Proof. Consider the partial sum $v_k = \sum_{n=1}^{k}\langle v, e_n\rangle e_n \in \mathcal{H}$. Then

$$\begin{aligned}\|v_k\|^2 = \langle v_k, v_k\rangle &= \left\langle \textstyle\sum_{n=1}^{k}\langle v, e_n\rangle e_n, \sum_{m=1}^{k}\langle v, e_m\rangle e_m \right\rangle\\ &= \textstyle\sum_{n=1}^{k}\sum_{m=1}^{k}\langle v, e_n\rangle\langle v, e_m\rangle\langle e_n, e_m\rangle\\ &= \textstyle\sum_{n=1}^{k}\langle v, e_n\rangle^2\end{aligned}$$

Furthermore

$$\begin{aligned}\|v - v_k\|^2 = \langle v - v_k, v - v_k\rangle &= \|v\|^2 - 2\langle v, v_k\rangle + \|v_k\|^2\\ &= \|v\|^2 - 2\textstyle\sum_{n=1}^{k}\langle v, e_n\rangle\langle v, e_n\rangle + \sum_{n=1}^{k}\langle v, e_n\rangle^2\\ &= \|v\|^2 - \sum_{n=1}^{k}\langle v, e_n\rangle^2\end{aligned}$$

It then follows that $\sum_{n=1}^{k}\langle v, e_n\rangle^2 = \|v\|^2 - \|v - v_k\|^2 \leq \|v\|^2$, and taking the limit $k \to \infty$ gives Bessel's inequality (the limit exists by the monotone sequences theorem).

□

Definition 6.11. An orthonomal set F in $\mathcal{H}$ is called ***total*** if Span F is dense in $\mathcal{H}$.

In some texts a total orthonormal set is called a *complete orthonormal set* or an *orthonormal basis*. Note however that the latter is not consistent with our definition 1.11.

Theorem 6.12. Let $\{e_1, e_2, \ldots\}$ be an orthonormal set in $\mathcal{H}$. Then the following are equivalent:

(a) $\{e_1, e_2, \ldots\}$ is total in $\mathcal{H}$,

(b) $\{e_1, e_2, \ldots\}^\perp = \{0\}$,

(c) For any $x \in \mathcal{H}$ we have $x = \sum_{n=1}^{\infty}\langle x, e_n\rangle e_n$.

Proof. (a) implies (b): let $Y = \text{Span}\{e_1, e_2, \ldots\}$ and assume $\overline{Y} = \mathcal{H}$. If $x \in \{e_1, e_2, \ldots\}^\perp$ then for any linear combination $\sum_{j=1}^{n}\alpha_j e_j$ we have $\langle x, \sum \alpha_j e_j\rangle = \sum \alpha_j\langle x, e_j\rangle = 0$, i.e. $x \in Y^\perp$. Since Y is dense there exists $x_k \to x$ with each $x_k \in Y$, and then $\|x\|^2 = \langle x, x\rangle = \lim_{k\to\infty}\langle x, x_k\rangle = 0$ and $x = 0$.

(b) implies (c): Let $x \in \mathcal{H}$, and label the partial sums $s_k := \sum_{n=1}^{k}\langle x, e_n\rangle e_n$ and $\sigma_k := \sum_{n=1}^{k}\langle x, e_n\rangle^2$. By Bessell's inequality σ_k converges, and it is therefore Cauchy. Now supposing $j < k$, and using orthonormality,

$$\|s_k - s_j\|^2 = \|\sum_{n=j}^{k}\langle x, e_n\rangle e_n\|^2 = \sum_{n=j}^{k}\langle x, e_n\rangle^2 = |\sigma_k - \sigma_j|$$

From this it follows that (s_k) is also Cauchy, and therefore converges because $\mathcal{H}$ is complete. Now consider $y = x - \sum_{n=1}^{\infty}\langle x, e_n\rangle e_n$. Then $\langle y, e_j\rangle = \langle x, e_j\rangle - \langle x, e_j\rangle\|e_j\|^2 = 0$ for all $j \geq 1$, so $y \in \{e_1, e_2, \ldots\}^\perp = \{0\}$ and $x = \sum_{n=1}^{\infty}\langle x, e_n\rangle e_n$.

✍ 6.13. (c) implies (a).

□

Note: it is important to remember here that $x = \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n$ means precisely $\lim_{k\to\infty} \|x - \sum_{n=1}^{k} \langle x, e_n \rangle e_n\| = 0$. This is not the same as saying x is in the span of $\{e_1, e_2, \dots\}$.

Example 6.14. The orthonormal set $\{e_n = (0, \dots, 0, 1, 0 \dots), n = 1, 2, \dots\}$ (i.e. the nth element of e_n is 1) from example 6.3 is total in ℓ^2. Indeed if $x = (x_1, \dots, x_n, \dots) \in \ell^2$ is perpendicular to each e_n then $0 = \langle x, e_n \rangle = x_n$ for all n, and $x = 0$. So $\{e_n = (0, \dots, 0, 1, 0 \dots), n = 1, 2, \dots\}$ is a total orthonormal set by Theorem 6.12.

7 WEEK 7

7.1 ORTHONORMAL SETS CONTINUED

Example 7.1. The orthonormal set $\{e_0 = \frac{1}{\sqrt{\pi}}, e_n = \sqrt{\frac{2}{\pi}} \cos nx, n \geq 1\}$ from example 6.4 is total in $L^2([0,\pi],\mathbb{R})$ (we don't have the tools for a proof of this yet, so it will come later as an exercise). Note that when we write $x = \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n$ in this case the convergence is in the L^2 norm, which *does not* imply convergence in the uniform norm. Indeed, you might remember from first year that if a function f has points of jump discontinuity its Fourier series does not converge uniformly to f.

Proposition 7.2. (Parseval's relation) Let $\{e_1, e_2, \dots\}$ be a total orthonormal set in a Hilbert space $\mathcal{H}$. Then

$$\langle x, y \rangle = \sum_{n=1}^{\infty} \langle x, e_n \rangle \langle e_n, y \rangle$$

for all $x, y \in \mathcal{H}$, and in particular $\|x\|^2 = \sum_{n=1}^{\infty} \langle x, e_n \rangle^2$.

Proof. By theorem 6.12 part (c), and continuity of the inner product, we have

$$\begin{aligned} \langle x, y \rangle &= \lim_{k\to\infty} \left\langle \sum_{n=1}^{k} \langle x, e_n \rangle e_n, \sum_{m=1}^{k} \langle y, e_m \rangle e_m \right\rangle \\ &= \lim_{k\to\infty} \sum_{n=1}^{k} \sum_{m=1}^{k} \langle x, e_n \rangle \langle y, e_m \rangle \langle e_n, e_m \rangle \\ &= \lim_{k\to\infty} \sum_{n=1}^{k} \langle x, e_n \rangle \langle e_n, y \rangle \\ &= \sum_{n=1}^{\infty} \langle x, e_n \rangle \langle e_n, y \rangle \end{aligned}$$

✍ 7.3. Prove that this sum converges.

□

Definition 7.4. A metric space X is called ***separable*** if it has a countable dense subset.

Example 7.5. $\mathbb{R}$ is separable: $\mathbb{Q}$ is countable and dense.

✍ 7.6. Prove that ℓ^2 is separable. (Hint: consider the set of rational linear combinations of elements in the total orthonormal set from example 6.14).

Proposition 7.7. Let $\mathcal{H}$ be an infinite dimensional Hilbert space. Then $\mathcal{H}$ is separable iff it contains a total orthonormal set.

Proof. If $\mathcal{H}$ is separable then there exists a dense countable set

$$W = \{w_1, w_2, \ldots\} \subset \mathcal{H}$$

from which we construct a linearly independent subset $W' \subset W$ by induction: assuming $\{v_1, v_2, \ldots, v_k\} \subset W$ is linearly independent, let $v_{k+1} = w_j$ where w_j is the first element of W which is not in $\mathrm{Span}\{v_1, \ldots v_k\}$. Now $W \subset \mathrm{Span}\, W = \mathrm{Span}\, W'$ so $\mathrm{Span}\, W'$ is dense. Applying the Gram-Schmidt algorithm to W' we obtain an orthonormal sequence $\{e_1, e_2, \ldots\}$ with $\mathrm{Span}\{e_1, e_2, \ldots\} = \mathrm{Span}\, W'$, which is dense so $\{e_1, e_2, \ldots\}$ is total.

Conversely suppose $\{e_1, e_2, \ldots\}$ is a total orthonormal sequence. Consider $U = \{\sum_{j=1}^{n} r_j e_j : n \geq 1, r_1, \ldots r_n \in \mathbb{Q}\}$, the set of linear combinations with rational coefficients. This set is countable, we show it is also dense. Let $x \in \mathcal{H}, \varepsilon > 0$, then because $\mathrm{Span}\{e_1, e_2, \ldots\}$ is dense there exists a real ($\alpha_j \in \mathbb{R}$) linear combination $y = \sum_{j=1}^{n} \alpha_j e_j$, such that $\|x - y\| < \varepsilon/2$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$ there exist $r_j \in \mathbb{Q}$ such that $|r_j - \alpha_j| < \varepsilon/2n$. Then $\|y - \sum_{j=1}^{n} r_j e_j\| < \varepsilon/2$ and it follows by the triangle inequality that $\|x - \sum_{j=1}^{n} r_j e_j\| < \varepsilon$, which shows that U is dense. □

Theorem 7.8. If $\mathcal{H}$ and $\mathcal{F}$ are infinite dimensional separable Hilbert spaces then they are isometrically isomorphic.

Proof. Let $\{e_1, e_2, \ldots\}$ and $\{f_1, f_2, \ldots\}$ be total orthonormal sets for $\mathcal{H}$ and $\mathcal{F}$. If $x \in \mathcal{H}$ then $x = \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n$ by Theorem 6.12. We define $T : \mathcal{H} \to \mathcal{F}$ by $Tx := \sum_{n=1}^{\infty} \langle x, e_n \rangle f_n$ (this sum converges by the same reasoning used in Theorem 6.12). Then T is linear and using Parseval's relation $\|Tx\|^2 = \sum_{n=1}^{\infty} \langle x, e_n \rangle^2 = \|x\|^2$, i.e. T is an isometry. It then also follows that T is injective, because $Tx = Ty$ implies $0 = \|Tx - Ty\| = \|T(x-y)\| = \|x - y\|$. Moreover, T is onto: given $y \in \mathcal{F}$ we can set $x = \sum_{n=1}^{\infty} \langle y, f_n \rangle e_n$, and then $Tx = y$. □

7.2 WEIERSTRASS APPROXIMATION THEOREM

Definition 7.9. A ***Dirac sequence*** is a sequence (K_n) of functions $\mathbb{R} \to \mathbb{R}$ such that

(1) $K_n(x) \geq 0$ for all n, x.

(2) K_n is piecewise continuous on every finite interval and $\int_{-\infty}^{\infty} K_n(x)dx = 1$.

(3) Given ε, δ, there exists N such that if $n \geq N$ then $\int_{-\infty}^{-\delta} K_n + \int_{\delta}^{\infty} K_n < \varepsilon$

Example 7.10. Define

$$K_n(t) = \begin{cases} \frac{(1-t^2)^n}{c_n} & |t| \leq 1 \\ 0 & |t| > 1 \end{cases}$$

where $c_n := \int_{-1}^{1}(1-t^2)^n dt$ ensures (2) is satisfied, and also (1). For (3) we need to estimate c_n. Since K_n is even:

$$\frac{c_n}{2} = \int_0^1 (1-t^2)^n dt = \int_0^1 (1+t)^n (1-t)^n dt \geq \int_0^1 (1-t)^n dt = \frac{1}{n+1}$$

so $c_n \geq 2/(n+1)$. Then given $\delta > 0$:

$$\int_\delta^1 K_n(t)dt = \int_\delta^1 \frac{(1-t^2)^n}{c_n} \leq \int_\delta^1 \frac{n+1}{2}(1-\delta^2)^n dt = \frac{n+1}{2}(1-\delta^2)^n(1-\delta)$$

but $(n+1)(1-\delta^2)^n \to 0$ as $n \to \infty$ (proof: let $1+t = 1/(1-\delta^2)$, then $t > 0$ and $(1+t)^n = 1 + nt + \frac{n(n-1)}{2}t^2 + \ldots \geq \frac{n(n-1)}{2}t^2$, thus $(n+1)(1-\delta^2)^n = \frac{n+1}{(1+t)^n} \leq \frac{2(n+1)}{n(n-1)t^2} \to 0$). Since K_n is even, $\int_{-1}^{-\delta} K_n(t)dt \to 0$ also, and (3) holds.

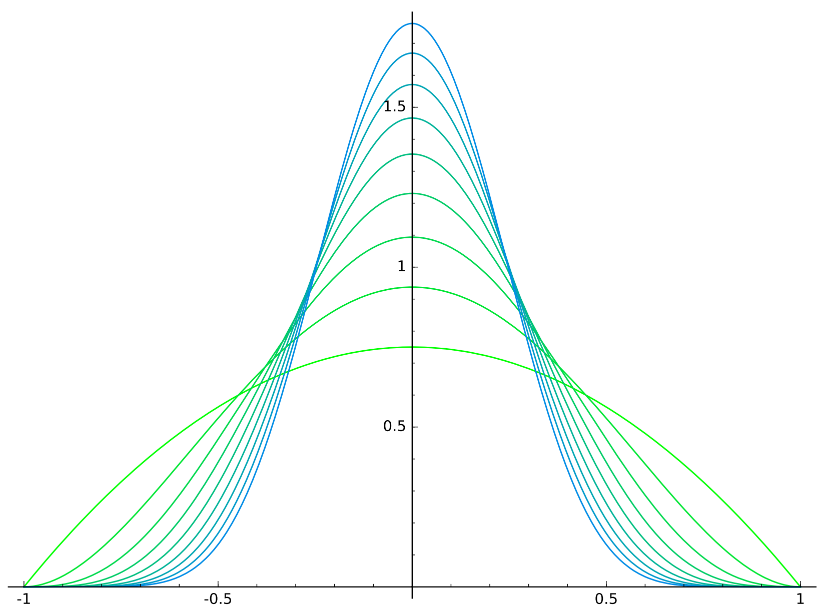


Figure 2: A plot of K_n for $n = 2 \ldots 10$. Blue represents higher values of n.

Definition 7.11. A function $f : X \to Y$ between metric spaces is called *uniformly continuous* if for all $\varepsilon > 0$, there exists $\delta > 0$ such that $f(x) \in B_\varepsilon(f(x_0))$ for all x, x_0 with $x \in B_\delta(x_0)$.

The difference between uniform continuity and continuity is that for the latter δ is allowed to vary depending on x_0.

✍ 7.12. Show that a continuous function $f : [a, b] \to \mathbb{R}$ is uniformly continuous.

Proposition 7.13. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function which is zero except on (a, b), and (K_n) a Dirac sequence. Define $f_n(x) := K_n * f := \int_{-\infty}^{\infty} f(t)K_n(x-t)dt$, then f_n converges in the uniform norm to f on $[a, b]$.

Proof. Substituting $t \to x - t$ gives $f_n(x) = \int_{-\infty}^{\infty} f(x-t)K_n(t)dt$, and from property (2) of Dirac sequences $f(x) = f(x)\int_{-\infty}^{\infty} K_n(t)dt = \int_{-\infty}^{\infty} f(x)K_n(t)dt$, and therefore

$$f_n(x) - f(x) = \int_{-\infty}^{\infty} (f(x-t) - f(x))K_n(t)dt \tag{3}$$

Since f is uniformly continuous on $[a, b]$ by 7.12, given $\varepsilon > 0$ there exists δ such that

$$|f(x-t) - f(x)| < \varepsilon \tag{4}$$

for any $x \in [a, b]$, when $|t| < \delta$. Moreover setting $M = \sup_{t\in[a,b]} |f(t)|$, by property (3) there exists N such that if $n \geq N$ then

$$\int_{-\infty}^{-\delta} K_n + \int_{\delta}^{\infty} K_n < \varepsilon/2M \tag{5}$$

Now from (3)

$$|f_n(x) - f(x)| \leq (\int_{-\infty}^{-\delta} + \int_{-\delta}^{\delta} + \int_{\delta}^{\infty})|f(x-t) - f(x)|K_n(t)dt$$

To estimate the first and third integrals, note that (by the triangle ineq.) $|f(x-t) - f(x)| \leq 2M$ and thus

$$(\int_{-\infty}^{-\delta} + \int_{\delta}^{\infty})|f(x-t) - f(x)|K_n(t)dt \leq 2M(\int_{-\infty}^{-\delta} + \int_{\delta}^{\infty})K_n(t)dt < \varepsilon$$

using (5). For the middle integral we use (4)

$$\int_{-\delta}^{\delta} |f(x-t) - f(x)| K_n(t) dt \le \varepsilon \int_{-\delta}^{\delta} K_n \le \varepsilon$$

Hence $|f_n(x) - f(x)| < 2\varepsilon$ for all $x \in [a, b]$, and $f_n \to f$ uniformly.

□

Theorem 7.14. (Weierstrass approximation) Let $f \in C([a, b], \mathbb{R})$ and $\varepsilon > 0$. Then there exists a polynomial function $p(t) = \alpha_0 + \alpha_1 t + \ldots + \alpha_n t^n$ such that $|f - p|_0 < \varepsilon$, i.e. the polynomial functions are dense in $C([a, b], \mathbb{R})$.

Proof. First note that we can reparametrize any $f \in C([a, b], \mathbb{R})$ by $x = (b-a)u + a, 0 \le u \le 1$ and then $g(u) := f((b-a)u + a)$ is an element of $C([0, 1], \mathbb{R})$. If we can find a polynomial $p(u)$ such that $|g(u) - p(u)| < \varepsilon$ for $u \in [0, 1]$ then since $u = (x-a)/(b-a)$ we have $|f(x) - p(\frac{x-a}{b-a})| < \varepsilon$. But $p(\frac{x-a}{b-a})$ is a polynomial in x so the result follows for f if it is proved for g. It is therefore sufficient to prove the theorem for $[a, b] = [0, 1]$. Moreover, suppose $[a, b] = [0, 1]$ and let $h(x) = f(x) - f(0) - x(f(1) - f(0))$, then we have $h(0) = 0 = h(1)$, and if we can approximate h by a polynomial p we can approximate f by the polynomial $p + f(0) + x(f(1) - f(0))$. Therefore it is sufficient to prove the theorem for f satisfying $f(0) = 0 = f(1)$.

So consider $f \in C([0, 1], \mathbb{R})$ with $f(0) = 0 = f(1)$ and extend it to all of $\mathbb{R}$ by $f(x) = 0$ outside $[0, 1]$. We use the Dirac sequence from example 7.10, and by Prop. 7.13

$$f_n(x) := \int_{-\infty}^{\infty} f(t) K_n(x-t) dt = \int_0^1 f(t) K_n(x-t) dt$$

converges uniformly to f on $[0, 1]$. It remains to prove that f_n is a polynomial. By definition (example 7.10) $K_n(x-t)$ is a polynomial in t and x of order $2n$, so we write

$$K_n(x-t) = g_0(t) + g_1(t)x + \ldots + g_{2n}(t)x^{2n}$$

and then

$$f_n(x) = a_0 + a_1 x + \ldots + a_{2n} x^{2n}$$

where $a_i = \int_0^1 f(t) g_i(t) dt$.

□

✍ 7.15. Show that $C([a,b],\mathbb{R})$ is separable.

✍ 7.16. (difficult) Use the Weierstrass approximation theorem to prove that the orthonormal set $\{e_0 = \frac{1}{\sqrt{\pi}}, e_n = \sqrt{\frac{2}{\pi}}\cos nx, n \geq 1\}$ from example 6.4 is total.

8 WEEK 8

8.1 CONTRACTION MAPPINGS

Definition 8.1. A map $f : X \to Y$ between metric spaces X, d and Y, e is called ***Lipschitz*** if there exists $c > 0$ such that for all $x_1, x_2 \in X$:

$$e(f(x_1), f(x_2)) \leq cd(x_1, x_2)$$

The smallest such c is called the ***Lipschitz constant*** of f.

✍ 8.2. Show that every Lipschitz function is uniformly continuous.

If $f : \mathbb{R} \to \mathbb{R}$ then the Lipschitz condition is $|f(x_1) - f(x_2)| \leq c|x_1 - x_2|$ or:

$$\frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} \leq c$$

which tells us that the absolute slope is bounded by c. So we can picture this condition as follows:

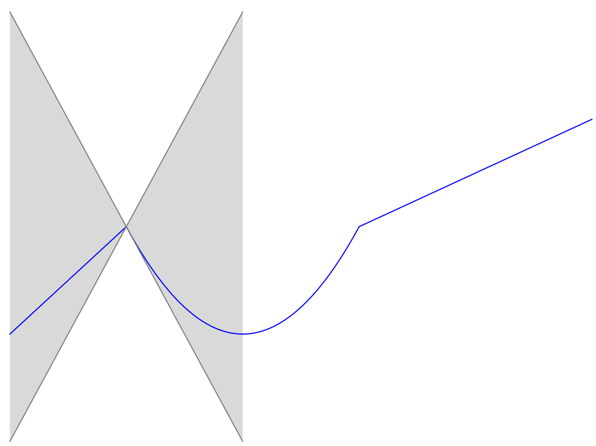


Figure 3: Lipschitz condition for $f : \mathbb{R} \to \mathbb{R}$: there is a double cone with slope c such that the graph of f remains outside the cone as the vertex is moved along the graph.

✍ 8.3. Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable on all of $\mathbb{R}$. Then f is Lipschitz iff $|f'(x)|$ is bounded.

Definition 8.4. If X, d is a metric space then $f : X \to X$ is called a ***contraction mapping*** if it is Lipschitz with constant $c < 1$, i.e.

$$d(f(x), f(y)) \leq cd(x, y), \quad c < 1$$

for all $x, y \in X$. A ***fixed point*** of f is a point $x \in X$ such that $f(x) = x$.

Theorem 8.5. (Banach fixed point theorem) Let X be a complete metric space and $f : X \to X$ a contraction mapping. Then f has a unique fixed point.

Proof. The basic idea: for a fixed point $d(f(x), x) = 0$, so guess any x_0 and find the distance $d(f(x_0), x_0)$. Unless we are very lucky this distance will be non-zero, so we update our guess to $x_1 := f(x_0)$ and find the distance $d(f(x_1), x_1) = d(f(f(x_0)), f(x_0))$. Since f is a contraction mapping this distance is smaller than $d(f(x_0), x_0)$, so x_1 is closer to being a fixed point than x_0. We prove that by iterating this process we converge to a fixed point.

Choose $x_0 \in X$ and define the sequence (x_n) by successive iterations of f, i.e.

$$\begin{aligned} x_1 &= f(x_0), \\ x_2 &= f(x_1) = f(f(x_0)) =: f^2(x_0) \\ &\vdots \\ x_n &= f(x_{n-1}) = f^n(x_0) \\ &\vdots \end{aligned}$$

We are going to prove that this sequence is Cauchy. By repeated instances of the Lipschitz property

$$\begin{aligned} d(x_{m+1}, x_m) &= d(f^m(x_1), f^m(x_0)) \\ &\leq cd(f^{m-1}(x_1), f^{m-1}(x_0)) \leq \ldots \leq c^m d(x_1, x_0) \end{aligned}$$

Now for $n > m$, using the above inequality, the triangle inequality, and the formula for the geometric series:

$$\begin{aligned} d(x_n, x_m) &\leq d(x_m, x_{m+1}) + \ldots + d(x_{n-1}, x_n) \\ &\leq (c^m + c^{m+1} + \ldots c^{n-1}) d(x_0, x_1) \\ &\leq c^m (\sum_{i=1}^{\infty} c^i) d(x_0, x_1) = c^m \frac{1}{1-c} d(x_0, x_1) \end{aligned}$$

From which we see that (x_n) is Cauchy: since $c < 1$ we have $c^m \to 0$, so given $\varepsilon > 0$ there exists N such that $c^m < \varepsilon(1-c)/d(x_0, x_1)$ for all $m > N$, and then $d(x_n, x_m) < \varepsilon$ for all $n > m > N$. Now because X is assumed to be complete we have $x_n \to x \in X$. To see that x is actually a fixed point:

$$d(f(x), x) \leq d(f(x), x_n) + d(x_n, x) \leq cd(x, x_{n-1}) + d(x_n, x) \underset{n\to\infty}{\longrightarrow} 0$$

so $d(f(x), x) = 0$ and $f(x) = x$. Finally, we show that it is unique. Suppose $f(x) = x$ and $f(y) = y$. Then

$$d(x, y) = d(f(x), f(y)) \leq cd(x, y)$$

and because $c < 1$ this inequality implies $d(x, y) = 0$ and $x = y$. □

8.2 EXISTENCE AND UNIQUENESS OF SOLUTIONS TO ODE

As a first application of the Banach fixed point theorem we will prove Picard's theorem on existence and uniqueness of solutions to ordinary differential equations (ODE). This theorem gives sufficient conditions on a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that the *initial value problem*

$$\frac{dx}{dt} = f(t, x), \quad x(t_0) = x_0 \in \mathbb{R} \tag{6}$$

has a unique solution.

We are going to need the following property of compact sets.

✍ 8.6. Show that if $f : X \to Y$ is a continuous map from a compact topological space, then the image $f(X)$ is a compact subset of Y.

Theorem 8.7. Let f be continuous on the domain

$$D := \{(t, x) \in \mathbb{R}^2 : |t - t_0| \leq \alpha, |x - x_0| \leq \beta\}$$

and suppose f is Lipschitz in its second variable with constant k (independent of t), i.e.

$$|f(t, x_1) - f(t, x_2)| \leq k|x_1 - x_2|$$

for all $(t, x_1), (t, x_2) \in D$. Then there exists $\varepsilon > 0$ such that (6) has a unique solution $x : [t_0 - \varepsilon, t_0 + \varepsilon] \to \mathbb{R}$.

Proof. Instead of 6 we consider the equivalent integral equation

$$x(t) = x_0 + \int_0^t f(\tau, x(\tau))d\tau \tag{7}$$

Assume $\varepsilon < \alpha$ so that we remain in D, and let $U \subset C([t_0 - \varepsilon, t_0 + \varepsilon], \mathbb{R})$ be such that the map $T : U \to C([t_0 - \varepsilon, t_0 + \varepsilon], \mathbb{R}), x \mapsto Tx$ defined by

$$Tx(t) := x_0 + \int_0^t f(\tau, x(\tau))d\tau$$

is well defined. At minimum this means that $x \in U$ must satisfy $|x(t) - x_0| < \beta$ so that $(t, x(t)) \in D$ and Tx is continuous, but we will define U more precisely after some more observations. Notice that any fixed point of T will be a solution to (7) and therefore (6). However in order to apply the Banach fixed point theorem we need to restrict T to an appropriate domain U such that U is complete, $T(U) \subset U$, and T is a contraction mapping on U. To this end, note that D is compact by the Heine-Borel theorem (4.13 in Springham's notes), and therefore by exercise 8.6 $f(D)$ is closed and bounded and $c := \max_{(t,x)\in D} |f(t, x)|$ exists. Assuming $\varepsilon < \min\{\alpha, \frac{\beta}{c}, \frac{1}{k}\}$, we will show that the domain

$$U := \{x \in C([t_0 - \varepsilon, t_0 + \varepsilon], \mathbb{R}) : |x(t) - x_0| < c\varepsilon\}$$

has the desired properties.

First observe that U is closed: suppose $x_n \to x$ is a convergent sequence with $x_n \in U$, then

$$|x(t) - x_0| \leq |x(t) - x_n(t)| + |x_n(t) - x_0| \leq |x - x_n|_0 + c\varepsilon$$

therefore $x \in U$ and U is closed by Proposition 3.5. Since a closed subset of a complete space is itself complete (prove as an exercise), U is complete.

To check that $T(U) \subset U$:

$$|Tx(t) - x_0| = |\int_{t_0}^t f(\tau, x(\tau))d\tau| \leq \int_{t_0}^t |f(\tau, x(\tau))|d\tau \leq c|t - t_0| \leq c\varepsilon$$

and therefore $Tx \in U$.

Finally, T is a contraction on U because for $x, y \in U$

$$\begin{aligned}
|Tx(t) - Ty(t)| &= |\int_{t_0}^{t} f(\tau, x(\tau)) - f(\tau, y(\tau)) d\tau| \\
&\leq k \int_{t_0}^{t} |x(\tau) - y(\tau)| d\tau \\
&\leq k|t - t_0||x - y|_0 \\
&\leq \varepsilon k |x - y|_0
\end{aligned}$$

Taking the supremum it follows that $|Tx - Ty|_0 \leq \varepsilon k |x - y|_0$ and T is a contraction by the assumption $\varepsilon < \frac{1}{k}$. Since T is a contraction on U it has a unique fixed point $x \in U \subset C([t_0 - \varepsilon, t_0 + \varepsilon], \mathbb{R})$ which satisfies (7) and therefore (6).

□

9 WEEK 9

9.1 DIFFERENTIATION

We recall that the derivative of a function $f : \mathbb{R} \to \mathbb{R}$ at $a \in \mathbb{R}$ is defined as

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

if the limit exists. An equivalent statement is

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - hf'(a)}{h} = 0 \tag{8}$$

Loosely speaking, this means that as h gets small we have $f(a+h) - f(a) - hf'(a) \approx 0$. Substituting $h = x - a$ and rearranging

$$f(x) \approx f(a) + f'(a)(x - a)$$

i.e. when $|x - a|$ is small $f(x)$ can be approximated by the tangent line at $f(a)$, which is an affine function . Similarly, recall that a function $f : \mathbb{R}^2 \to \mathbb{R}$ can be approximated by the tangent plane at $f(a_1, a_2)$ (if it exists):

$$f(x_1, x_2) \approx f(a_1, a_2) + (\partial_1 f(a_1, a_2), \partial_2 f(a_1, a_2)) \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \end{pmatrix}$$

More succinctly: $x, a \in \mathbb{R}^2$

$$f(x) \approx f(a) + Df(a) \cdot (x - a)$$

This property is often described by saying that the derivative is the linear map which gives the best affine approximation to the original function. But in what sense is it the "best"? Consider a general affine approximation $f(x) \approx f(a) + A \cdot (x - a)$, substituting $h = x - a$ (which means that $h \in \mathbb{R}^2$ now)

$$f(a + h) \approx f(a) + A \cdot h$$

The error in this approximation will be a function of h

$$E(h) := \|f(a + h) - f(a) - A \cdot h\|$$

The derivative as we know it (i.e. (8)) can be characterised by requiring that $E(h)$ should be ***of order*** h: meaning

$$\lim_{h \to 0} \frac{E(h)}{\|h\|} \to 0$$

Intuitively, as h gets small the error gets smaller quicker. It turns out that if a linear map A which gives this behaviour for $E(h)$ exists then it is unique (see below), and this is the precise sense in which the derivative is "best".

Definition 9.1. Let E, F be Banach spaces and $U \subset E$. A function $f : U \to F$ is ***differentiable at*** $a \in U$ if there is a continuous linear transformation $Df(a) : E \to F$, called a ***derivative***, such that

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

Note: $h \in E$, the norm in the denominator is the E norm, and the norm in the numerator is from F.

Theorem 9.2. If $f : U \to F$ is differentiable at $a \in U$ then the derivative is unique.

Proof. Suppose a derivative $Df(a)$ exists and $\lambda : E \to F$ also satisfies

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

Let $d(h) = f(a+h) - f(a)$, then

$$\begin{aligned}\lim_{h\to 0} \frac{\|\lambda h - Df(a)h\|}{\|h\|} &= \lim_{h\to 0} \frac{\|\lambda h - d(h) + d(h) - Df(a)h\|}{\|h\|} \\ &\le \lim_{h\to 0} \frac{\|d(h) - \lambda h\|}{\|h\|} + \lim_{h\to 0} \frac{\|d(h) - Df(a)h\|}{\|h\|} \\ &= 0\end{aligned}$$

Then let $h = tx$, where $x \in E$ and $t \in \mathbb{R}$,

$$0 = \lim_{t\to 0} \frac{\|\lambda(tx) - Df(a)(tx)\|}{\|tx\|} = \lim_{t\to 0} \frac{\|t\lambda x - tDf(a)x\|}{|t|\|x\|} = \frac{\|\lambda x - Df(a)x\|}{\|x\|}$$

hence $\lambda x = Df(a)x$ for all $x \in E$. □

Definition 9.3. $f : U \to F$ is ***differentiable*** if it is differentiable at every $u \in U$.

✍ 9.4. Prove that if $f : U \to F$ is differentiable at $a \in U$ then it is continuous at a, and conclude that if $f : U \to F$ is differentiable then U is open. (Hint: Proposition 2.2).

Theorem 9.5. (Chain rule). If $f : U \subset E \to F$ is differentiable at a and $g : F \to G$ is differentiable at $f(a)$ then $g \circ f : U \to G$ is differentiable at a and

$$D(g \circ f)(a) = Dg(f(a)) \circ D(f(a))$$

Proof. Let $b = f(a), \lambda = Df(a), \mu = Dg(b)$, and define

$$E_f(x) := f(x) - f(a) - \lambda(x - a) \tag{9}$$
$$E_g(y) := g(y) - g(b) - \mu(y - b) \tag{10}$$
$$E_{g\circ f}(x) := g \circ f(x) - g \circ f(a) - \mu \circ \lambda(x - a) \tag{11}$$

so that by definition of the derivative

$$\lim_{x\to a} \frac{\|E_f(x)\|}{\|x - a\|} = 0 \tag{12}$$
$$\lim_{y\to b} \frac{\|E_g(y)\|}{\|y - b\|} = 0 \tag{13}$$

and to prove the chain rule we need to show that $\lim_{x\to a} \frac{\|E_{g\circ f}(x)\|}{\|x-a\|} = 0$. Using (9):

$$\begin{aligned} E_{g\circ f}(x) &= g(f(x)) - g(b) - \mu(f(x) - f(a) - E_f(x)) \\ &= g(f(x)) - g(b) - \mu(f(x) - b) + \mu E_f(x) \\ &= E_g(f(x)) + \mu E_f(x) \end{aligned}$$

So by the triangle inequality and squeeze theorem, it is enough to prove

$$\lim_{x\to a} \frac{\|E_g(f(x))\|}{\|x - a\|} = 0 \tag{14}$$
$$\lim_{x\to a} \frac{\|\mu E_f(x)\|}{\|x - a\|} = 0 \tag{15}$$

For the latter: by Proposition 2.2 there is a constant k such that $\|\mu E_f(x)\| \leq k\|E_f(x)\|$ and then the result follows by (12) and the squeeze theorem. As for (14), from (13) we know that for all $\varepsilon > 0$ there exists δ such that if $\|f(x) - b\| < \delta$ then

$$\frac{\|E_g(f(x))\|}{\|f(x) - b\|} < \varepsilon \tag{16}$$

Moreover, since f is continuous there exists δ_1 such that if $\|x-a\| < \delta_1$ then $\|f(x)-b\| < \delta$ and then rearranging (16)

$$\begin{aligned} \|E_g(f(x))\| < \varepsilon\|f(x)-b\| &= \varepsilon\|E_f(x)+\lambda(x-a)\| \\ &\leq \varepsilon\|E_f(x)\| + \varepsilon k\|x-a\| \end{aligned}$$

where we have used Proposition 2.2 again. Now overall we have: for all $\varepsilon > 0$ there exists δ_1 such that if $\|x-a\| < \delta_1$ then

$$\frac{\|E_g(f(x))\|}{\|x-a\|} < \varepsilon\frac{\|E_f(x)\|}{\|x-a\|} + \varepsilon k$$

which together with (9) implies (14). □

Proposition 9.6. Some other properties of the derivative:

(1) If $f : E \to F$ is a constant function (i.e. for some $y \in F$ we have $f(x) = y$ for all $x \in E$) then $Df(a) = \mathbf{0}$.

(2) If $f : E \to F$ is a continuous linear transformation then $Df(a) = f$.

(3) If $f : \mathbb{R}^n \to \mathbb{R}^m$ then f is differentiable at $a \in \mathbb{R}^n$ iff each component function $f^i : \mathbb{R}^n \to \mathbb{R}$ is differentiable at a, and

$$Df(a) = (Df^1(a), \ldots, Df^m(a)) : \mathbb{R}^n \to \mathbb{R}^m$$

(4) If $f, g : \mathbb{R}^n \to \mathbb{R}$ are differentiable at a then

$$\begin{aligned} D(f+g)(a) &= Df(a) + Dg(a) \\ D(fg)(a) &= g(a)Df(a) + f(a)Dg(a) \end{aligned}$$

✍ 9.7. Prove Proposition 9.6 (they are not all easy).

Recall the definition of a partial derivative of $f : \mathbb{R}^n \to \mathbb{R}$ at $a \in \mathbb{R}^n$:

$$\partial_i f(a) := \lim_{h\to 0} \frac{f(a^1, \ldots, a^i + h, \ldots a^n) - f(a)}{h} \tag{17}$$

The next theorem shows how the partial derivatives are related to Df.

Theorem 9.8. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a then the partial derivatives of component functions $\partial_j f^j(a)$ exist for all $1 \le i \le m, 1 \le j \le n$ and $Df(a)$ is the linear map represented by the $m \times n$ matrix with ijth entry $\partial_j f^i(a)$ (the *Jacobian matrix*).

Proof. First consider the case $m = 1$, so $f : \mathbb{R}^n \to \mathbb{R}$. Define $\phi_j : \mathbb{R} \to \mathbb{R}^n$ by $\phi_j(x) := (a^1, \dots, x, \dots, a^n)$, i.e. with x in the jth coordinate. Then by (17) and the chain rule:

$$\begin{aligned}
\partial_j f(a) &= \lim_{h\to 0} \frac{(f\circ\phi_j)(a^j + h) - (f\circ\phi_j)(a^j)}{h} \\
&= D(f\circ\phi_j)(a^j) \\
&= Df(\phi_j(a^j)) \circ D\phi_j(a^j) \\
&= Df(a) \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \end{pmatrix}
\end{aligned}$$

where $D\phi_j(a_j)$ has been calculated using the properties in Proposition 9.6. So $\partial_j f(a)$ exists and is the jth entry of the $1 \times n$ matrix $[Df(a)]$. The theorem now follows for $m \ge 1$ by Proposition 9.6 part (3).

□

✍ 9.9. Find an example proving the converse to Theorem 9.8 does not hold.

10 WEEK 10/11

10.1 HIGHER ORDER DERIVATIVES

In Example 2.5 we saw that the space $B(V, W)$ of continuous linear maps between normed vector spaces $V \to W$ is itself a normed space.

✍ 10.1. (difficult) If W is a Banach space then $B(V, W)$ is a Banach space.

Definition 10.2. Let V_1, V_2, W be normed vector spaces and $b : V_1 \times V_2 \to W$. We say b is ***bilinear*** if for all $v_1 \in V_1, v_2 \in V_2$

- $b(\cdot, v_2) : V_1 \to W$, and
- $b(v_1, \cdot) : V_2 \to W$

are both linear. Similarly ***multilinear*** maps $V_1 \times \ldots \times V_k \to W$ are linear in each argument, and when the vector spaces are finite dimensional they are often called ***tensors***.

Definition 10.3. The ***direct sum*** $V_1 \oplus V_2$ of two vector spaces is the cartesian product $V_1 \times V_2$ with the vector space structure defined by $(x_1, x_2) + (y_1, y_2) := (x_1 + y_1, x_2 + y_2)$ and $\alpha(x_1, x_2) := (\alpha x_1, \alpha x_2)$.

✍ 10.4. Show that $\|(v_1, v_2)\| := \|v_1\| + \|v_2\|$ defines a norm on $V_1 \oplus V_2$, and if V_1, V_2 are Banach spaces then so is $V_1 \oplus V_2$. Prove that the topology on $V_1 \times V_2$ induced by this norm is the product topology.

✍ 10.5. Show that a bilinear map is *not* linear on the direct sum, and vice versa.

✍ 10.6. Prove that a bilinear map $b : V_1 \times V_2 \to W$ is continuous iff there exists $c > 0$ such that $\|b(v_1, v_2)\| \leq c\|v_1\|\|v_2\|$ (in which case we say it is bounded).

We will use the notation $B(V_1, V_2; W)$ for the space of continuous (bounded) bilinear maps. This space has a norm given by

$$|b| := \sup_{(v_1,v_2)\in V_1\times V_2} \frac{\|b(v_1, v_2)\|}{\|v_1\|\|v_2\|}$$

✍ 10.7. Show that the map $\phi : B(V_1, V_2; W) \to B(V_1, B(V_2, W))$ defined by $\phi(b)(v_1) := b(v_1, \cdot) : V_1 \to W$ is a linear homeomorphism.

✍ 10.8. Extend the previous result to show that

$$B(V_1, \dots, V_k; W) \cong B(V_1, \dots B(V_{k-1}, B(V_k, W)) \dots)$$

where $B(V_1, \dots, V_k; W)$ is the space of continuous *multilinear* maps.

Definition 10.9. Let $f : U \subset E \to F$ be a differentiable function and define the map $Df : U \to B(E, F)$ by $x \mapsto Df(x)$. If Df is a continuous map then we say that f is ***continuously differentiable*** or ***of class*** C^1. If Df is also continuously differentiable, i.e. $D^2 f := D(Df) : U \to B(E, B(E, F)) = B(E, E; F)$ is continuous, then we say f is of class C^2. If $D^k f : U \to B(E, \dots E; F)$ (defined recursively) exists and is continuous then f is of class C^k, and if $D^k f$ exists and is continuous for all $k \geq 1$ then f is called C^∞ or ***smooth***.

When $E = \mathbb{R}^n$ and $F = \mathbb{R}^m$ the above definition of C^k is equivalent to requiring (as in lectures) that all partial derivatives up to order k exist and are continuous. We will not give a proof here, but the basic idea is the same as in the proof of Theorem 9.8:

✍ 10.10. Let $e_i, i = 1, \dots n$ be the standard basis vectors in $\mathbb{R}^n$. Show that if $f : \mathbb{R}^n \to \mathbb{R}^m$ is C^1 then $(D^2 f(x)(e_i, e_j))_k = \partial_i \partial_j f^k(x)$ for all $k = 1 \dots m$. (Here f^k is the kth component function of f and $(D^2 f(x)(e_i, e_j))_k$ is the kth component of $D^2 f(x)(e_i, e_j) \in \mathbb{R}^m$.)

10.2 INVERSE FUNCTION THEOREM

Recall the Newton-Raphson (NR) method for solving $f(x) = 0$ for x where $f : \mathbb{R} \to \mathbb{R}$ is a differentiable function: guess x_0, hopefully $f'(x_0) \neq 0$, calculate $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$, then iterate $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$, and hopefully we find that $|x_{n+1} - x_n| \to 0$ (there is a nice animation on wikipedia). Suppose instead we want to find an inverse for f, i.e. for a given y_0 we want to solve $f(x) = y_0$. Then we can simply apply the NR-method to $f(x) - y_0$, i.e. calculate $x_1 = x_0 - \frac{f(x_0) - y_0}{f'(x_0)} \dots$. We write

$$F_{y_0}(x) = x - \frac{f(x) - y_0}{f'(x)}$$

so that $x_n = F_{y_0}(x_{n-1})$, and convergence of the NR method is equivalent to convergence of successive iterations of F_{y_0}, which calls to mind the Banach fixed point theorem. For $f : \mathbb{R} \to \mathbb{R}$ the inverse function theorem says that if

$f'(a) \neq 0$, x is sufficiently close to a, and y is sufficiently close to $f(a)$, then this process converges and gives an inverse for each y sufficiently close to a. Briefly, we say that f is *locally invertible* in a neighbourhood of a.

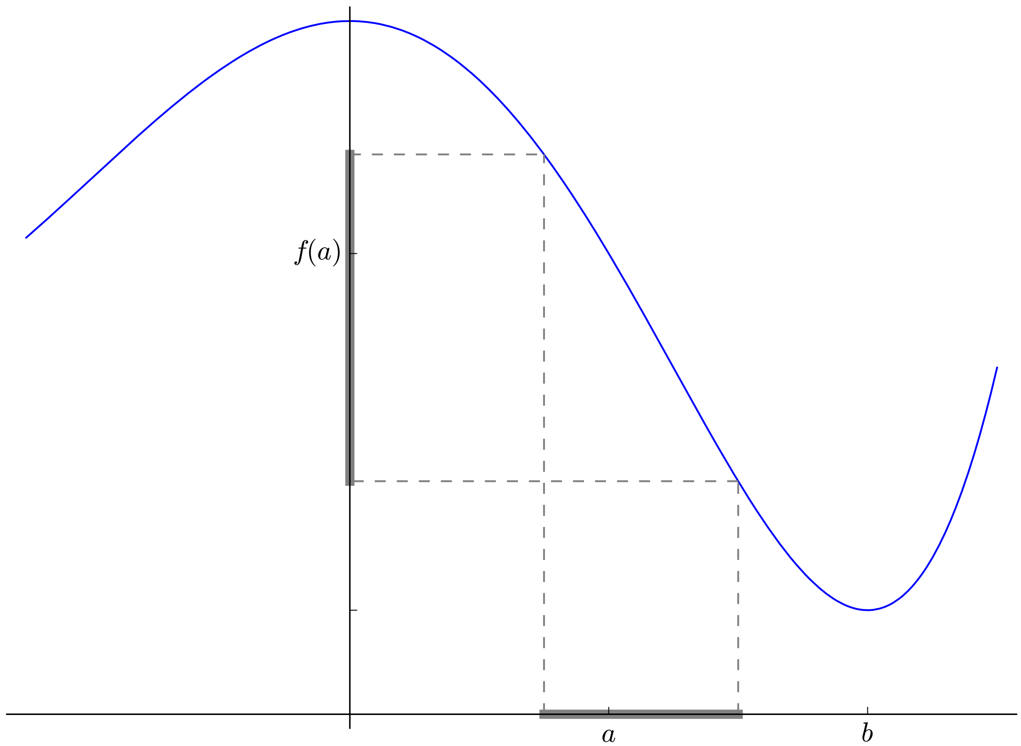


Figure 4: $f'(a) \neq 0$ so there is a neighbourhood of a which is mapped bijectively to a neighbourhood of $f(a)$. Notice that f will not be invertible in any neighbourhood of b.

We are going to prove the inverse function theorem for differentiable maps $f : \mathbb{R}^n \to \mathbb{R}^n$. The basic idea is the same as above but dividing by $f'(x)$ no longer makes sense. We will need to use $Df(x)^{-1}$ instead, and replace the condition $f'(a) \neq 0$ with "$Df(a)$ is invertible". First some preliminary results.

Lemma 10.11. (Mean value inequality) Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ be a C^1 function, and $V \subset U$ a convex subset such that $\sup_{x \in V} |Df(x)|$ exists. Then for all $x_1, x_2 \in V$

$$\|f(x_1) - f(x_2)\| \leq \sup_{x \in V} |Df(x)| \|x_1 - x_2\|$$

Proof. Since V is convex the path $\sigma(t) := tx_1 + (1-t)x_2$ is in V. By the chain rule

$$\frac{d}{dt} f(\sigma(t)) = Df(\sigma(t)) \frac{d\sigma}{dt} = Df(\sigma(t)(x_1 - x_2)$$

Integrating with respect to t:

$$f(x_1) - f(x_2) = \int_0^1 Df(\sigma(t))(x_1 - x_2)$$

therefore

$$\begin{aligned} \|f(x_1) - f(x_2)\| &\leq \| \int_0^1 Df(\sigma(t))(x_1 - x_2) dt \| \\ &\leq \int_0^1 \|Df(\sigma(t))(x_1 - x_2)\| dt \\ &\leq \int_0^1 |Df(\sigma(t))| \|x_1 - x_2\| dt \\ &\leq \sup_{x \in V} |Df(x)| \|x_1 - x_2\| \end{aligned}$$

□

Corollary 10.12. For any $a \in V$

$$\|f(x_1) - f(x_2) - Df(a)(x_1 - x_2)\| \leq \sup_{x \in V} |Df(x) - Df(a)| \|x_1 - x_2\|$$

Proof. Apply Lemma 10.11 to the function $g(x) := f(x) - Df(a)x$. □

Theorem 10.13. (Inverse function theorem) Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ be a C^k function. If for some $a \in U$ we have that $Df(a)$ is invertible, then there is an open set $V \subset U$ containing a such that the restriction $f|V \to f(V) \subset \mathbb{R}^n$ has a C^k inverse $g : f(V) \to V$, and $Dg(f(x)) = (Df(x))^{-1}$.

Proof. Write $y_0 = f(a)$, $\lambda = Df(a)$ and define $F_{y_0} : U \to \mathbb{R}^n$ by

$$F_{y_0}(x) := x - \lambda^{-1}(f(x) - y_0)$$

We will prove that F_{y_0} is a contraction map (on a subset of U). For $x_1, x_2 \in U$

$$\begin{aligned} F_{y_0}(x_1) - F_{y_0}(x_2) &= x_1 - x_2 - \lambda^{-1}(f(x_1) - f(x_2)) \\ &= \lambda^{-1}(\lambda(x_1 - x_2) - f(x_1) + f(x_2)) \end{aligned}$$

Since U is open there exists an open ball at a contained in U, and then any closed ball $\overline{B}_{\delta_1}(a)$ with smaller radius is compact, convex and contained in U. Then by exercise 8.6 $|Df(x)|$ is bounded for $x \in \overline{B}_{\delta_1}(a)$, and using Corollary 10.12, for $x_1, x_2 \in \overline{B}_{\delta_1}(a)$

$$\|F_{y_0}(x_1) - F_{y_0}(x_2)\| \leq |\lambda^{-1}| \sup_{x \in B_{\delta_1}(a)} |Df(x) - Df(a)| \|x_1 - x_2\|$$

By continuity of Df, there exists $\delta_2 > 0$ such that if $|x - a| < \delta_2$ then $|Df(x) - Df(a)| < \frac{1}{2|\lambda^{-1}|}$ (the reason for this particular choice of bound will become clear later). Therefore if $x_1, x_2 \in B_{\delta_1}(a) \cap B_{\delta_2}(a)$ then

$$\|F_{y_0}(x_1) - F_{y_0}(x_2)\| < \tfrac{1}{2}\|x_1 - x_2\| \tag{18}$$

i.e. F_{y_0} is Lipschitz with constant < 1. Let $\delta < \min\{\delta_1, \delta_2\}$, then F_{y_0} is Lipschitz on the closed ball $\overline{B}_\delta(a)$, which is complete. In order to apply the Banach fixed point theorem it remains to show $F_{y_0}(\overline{B}_\delta(a)) \subset \overline{B}_\delta(a)$. For this note that

$$F_{y_0}(a) = a - \lambda^{-1}(f(a) - f(a)) = a$$

and therefore from (18) $\|F_{y_0}(x) - a\| < \frac{1}{2}\|x - a\|$, which shows that if $x \in \overline{B}_\delta(a)$ then $F_{y_0}(x) \in \overline{B}_\delta(a)$. Now by the Banach fixed point theorem F_{y_0} has a unique fixed point in $\overline{B}_\delta(a)$. Let $\gamma = \frac{\delta}{2|\lambda^{-1}|}$ and for any $y \in B_\gamma(y_0)$ define

$$F_y(x) := x - \lambda^{-1}(f(x) - y).$$

Applying the same argument as for y_0 we have the Lipschitz property

$$\|F_y(x_1) - F_y(x_2)\| < \tfrac{1}{2}\|x_1 - x_2\|$$

for all $x_1, x_2 \in \overline{B}_\delta(a)$. To show $F_y(\overline{B}_\delta(a)) \subset \overline{B}_\delta(a)$ note that

$$F_y(a) - a = a - \lambda^{-1}(f(a) - y) - a = \lambda^{-1}(y - y_0)$$

hence

$$\begin{aligned} F_y(x) - a &= F_y(x) - F_y(a) + F_y(a) - a \\ &= F_y(x) - F_y(a) + \lambda^{-1}(y - y_0) \end{aligned}$$

and therefore for $x \in \overline{B}_\delta(a)$ and $y \in B_\gamma(y_0)$

$$\begin{aligned}\|F_y(x) - a\| &\leq \|F_y(x) - F_y(a)\| + |\lambda^{-1}|\|y - y_0\| \\ &\leq \frac{1}{2}\|x - a\| + \frac{\delta}{2} \leq \delta\end{aligned}$$

by the Lipschitz property for F_y and the choice of γ. So $F_y(x)$ is a contraction mapping and has a unique fixed point. We define $g : B_\gamma(y_0) \to B_\delta(a)$ by $y \mapsto$ (unique fixed point of F_y), i.e. $F_y(g(y)) = g(y)$. It follows that $g(y) = g(y) - \lambda^{-1}(f(g(y)) - y)$, and then $f(g(y)) = y$. Moreover

$$F_{f(x)}(x) = x - \lambda^{-1}(f(x) - f(x)) = x$$

so x is the fixed point of $F_{f(x)}$, which means $g(f(x)) = x$. Thus g is the inverse of $f|B_\delta(a)$. Continuity of g:

$$\begin{aligned}\|g(y_1) - g(y_2)\| &= \|F_{y_1}(g(y_1)) - F_{y_2}(g(y_2))\| \\ &\leq \|F_{y_1}(g(y_1)) - F_{y_1}(g(y_2))\| + \|F_{y_1}(g(y_2)) - F_{y_2}(g(y_2))\| \\ &\leq \frac{1}{2}\|g(y_1) - g(y_2)\| + \|\lambda^{-1}(y_1 - y_2)\|\end{aligned}$$

rearranging gives:

$$\|g(y_1) - g(y_2)\| \leq 2|\lambda^{-1}|\|y_1 - y_2\|$$

so g is Lipschitz and therefore continuous.

Now we prove g is differentiable. Note that since Df is continuous, $Df(a) \in GL(n, \mathbb{R})$, and $GL(n, \mathbb{R})$ is open in $L(\mathbb{R}^n, \mathbb{R}^n) = M(n \times n, \mathbb{R})$ (see Week 10 worksheet), there is an open set $U' \subset U$ such that $Df(x) \in GL(n, \mathbb{R})$ for all $x \in U'$. So we set $V = B_\delta(a) \cap U'$ and $W = f(V) \subset B_\gamma(y_0)$, and then for all $y \in W$ we have that $Df(g(y))$ is invertible.

Now for $y \in W$ let $k = g(y+h) - g(y)$ with h sufficiently small so that k is well defined. Then $\lim_{h \to 0} k = 0$ (continuity of g) and $h = f(k + g(y)) - y = f(k + g(y)) - f(g(y))$. Letting $\Lambda = Df(g(y))$

$$\begin{aligned}\frac{\|g(y+h) - g(y) - \Lambda^{-1}h\|}{\|h\|} &= \frac{\|\Lambda^{-1}(\Lambda k - h)\|}{\|h\|} \\ &= \frac{\|\Lambda^{-1}(\Lambda k - f(k + g(y)) - f(g(y))\|}{\|h\|} \\ &\leq 2|\lambda^{-1}||\Lambda^{-1}|\frac{\|f(k + g(y)) - f(g(y)) - \Lambda k\|}{\|k\|}\end{aligned}$$

where we have used $\|k\| = \|g(y+h) - g(y) \leq K\|h\|$ (Lipschitz property for g). The squeeze theorem gives

$$\lim_{h\to 0} \frac{\|g(y+h) - g(y) - \Lambda^{-1}h\|}{\|h\|} = 0$$

which shows that $Dg(y) = \Lambda^{-1} = Df(g(y))^{-1}$. This means $Dg : B_\gamma(y_0) \to L(\mathbb{R}^n, \mathbb{R}^n)$ can be decomposed as

$$Dg = \text{Inv} \circ Df \circ g : B_\gamma(y_0) \to B_\delta(a) \to GL(n, \mathbb{R}) \to GL(n, \mathbb{R}).$$

Here $\text{Inv} : GL(n,\mathbb{R}) \to GL(n,\mathbb{R})$ denotes matrix inversion, which we claim is a smooth map. So Dg is a composition of continuous functions, therefore continuous, and g is C^1. By repeated applications of the chain rule it follows that g is C^k. □

✍ 10.14. Prove that matrix inversion $GL(n,\mathbb{R}) \to GL(n,\mathbb{R})$ is a smooth map.

Remarks.

The identification $L(\mathbb{R}^n, \mathbb{R}^n) = M(n \times n, \mathbb{R})$ depends on a choice of basis for $\mathbb{R}^n$. Usually we mean to use the standard basis on $\mathbb{R}^n$, but it doesn't actually matter for the above proof.

The inverse function theorem can be extended to maps between Banach spaces, but it takes a bit more work. For example, in order to generalise the mean value inequality we need to know how to integrate over a path in Banach space.

11 WEEK 12

✍ 11.1. Show that a linear map A is injective if and only if $\ker A := \{x : Ax = 0\} = \{0\}$.

Lemma 11.2. Let V be an r-dimensional subspace of $\mathbb{R}^n$. Then $\mathbb{R}^n \cong V \oplus V^\perp$, the dimension of $V^\perp$ is $n-r$, and for any $x \in \mathbb{R}^n$ there exist unique $y \in V$ and $z \in V^\perp$ such that $x = y + z$.

Proof. V is a finite dimensional vector space so it has a basis $\{v_1, \dots v_r\}$ where r is the dimension of V. We can assume this basis is orthonormal (by Gram-Schmidt 6.5), and can extend it to an orthonormal basis $\{v_1, \dots v_r, w_{r+1}, \dots w_n\}$ for $\mathbb{R}^n$ (choose a vector not in the span of v_i, orthonormalise, repeat if possible). So any $x \in \mathbb{R}^n$ can be written as a linear combination

$$x = \sum_{i=1}^{r} x^i v_i + \sum_{j=r+1}^{n} x^j w_j \tag{19}$$

where $x^i = x \cdot v_i$ for $i = 1, \dots, r$ and $x^j = x \cdot w_j$ (we emphasize that here x^i does *not* mean the ith component of x in the standard basis).

Suppose $z = \sum_{i=1}^{r} z^i v_i + \sum_{j=r+1}^{n} z^j w_j$ is in $V^\perp$, then $z^i = z \cdot v_i = 0$ for $i = 1, \dots r$, so $\{w_{r+1} \dots w_n\}$ is a basis for $V^\perp$, which therefore has dimension $n - r$. Define projections $\pi_V : \mathbb{R}^n \to V$ and $\pi_\perp : \mathbb{R}^n \to V^\perp$ by

$$\pi_V(x) := \sum_{i=1}^{r} x^i v_i \qquad \pi_\perp(x) = \sum_{j=r+1}^{n} x^j w_j$$

Then $(\pi_V, \pi_\perp) : \mathbb{R}^n \to V \oplus V^\perp$ is an isomorphism with inverse $(y, z) \mapsto y + z$ (for example, check injectivity: $(\pi_V, \pi_\perp)x = (0, 0)$ iff $x^i = 0$ for all $i = 1, \dots, n$ iff $x = 0$, we leave the other properties as an exercise). Moreover, observe that setting $y = \pi_V x$ and $z = \pi_\perp x$ gives the decomposition $x = y + z$. If this decomposition is not unique, then (19) is not unique and $\{v_1, \dots v_r, w_{r+1}, \dots w_n\}$ is not a basis. □

There is an isomorphism $V \cong \mathbb{R}^r$ given by mapping y to its coefficients $(y^1, \dots, y^r)$ with respect to the basis $\{v_1, \dots, v_r\}$, or perhaps more precisely: mapping each $v_i \mapsto e_i$ to the standard basis vector and extending linearly (compare exercise 1.12). Similarly $V^\perp \cong \mathbb{R}^{n-r}$, and so $V \oplus V^\perp \cong \mathbb{R}^n \oplus \mathbb{R}^{n-r} = \mathbb{R}^n$. From this point of view the above lemma can be rephrased as:

Corollary 11.3. There is an isomorphism (a.k.a. change of basis) $T : \mathbb{R}^n \to \mathbb{R}^n = \mathbb{R}^{n-m} \oplus \mathbb{R}^m$ such that $T(V) = \mathbb{R}^{n-m} \oplus \{\mathbf{0}\}$ and $T(V^\perp) = \{\mathbf{0}\} \oplus \mathbb{R}^m$.

Lemma 11.4. (Splitting lemma) Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a surjective linear map and let $K = \ker A$. Then there exists an isomorphism Φ such that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^m \\ \Big\downarrow \Phi & \nearrow_{\mathrm{pr}_2} & \\ K \oplus \mathbb{R}^m & & \end{array}$$

meaning $A = \mathrm{pr}_2 \circ \Phi$, where $\mathrm{pr}_2(a, b) := b$.

Proof. The kernel of a linear map is a subspace so by the previous lemma there exist projections $\pi_K, \pi_\perp$ such that $(\pi_K, \pi_\perp) : \mathbb{R}^n \cong K \oplus K^\perp$. We will prove that $A|K^\perp \to \mathbb{R}^m$ is an isomorphism. Suppose $v \in K^\perp$ and $Av = 0$. Then $V \in K^\perp \cap K = \{0\}$ and $v = 0$, so $A|K^\perp$ is injective. For surjectivity: let $w \in \mathbb{R}^m$, then since A is surjective there exists $x \in \mathbb{R}^n$ such that $Ax = w$. By the previous lemma $x = \pi_K x + \pi_\perp x$ so

$$w = Ax = A(\pi_K x + \pi_\perp x) = 0 + A\pi_\perp x$$

and $\pi_\perp x \in K^\perp$ so $A|K^\perp$ is surjective. We therefore define $\Phi : \mathbb{R}^n \to K \oplus \mathbb{R}^m$ by

$$\Phi := (\mathrm{Id}, A|K^\perp) \circ (\pi_K, \pi_\perp) = (\pi_K, A)$$

Φ is a composition of isomorphisms, therefore itself an isomorphism, and $Ax = A(\pi_K x + \pi_\perp x) = A\pi_\perp x = \mathrm{pr}_2\, \Phi x$, so the diagram commutes. □

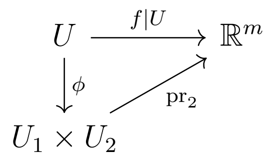
Notice that we have also proved the rank-nullity theorem, because by Lemma 11.2 the dimension of K is $n - m$.

Suppose now that $f : \mathbb{R}^n \to \mathbb{R}^m$ is a smooth map such that $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$ is surjective. Then by the splitting lemma there is an isomorphism Φ such that

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{Df(x)} & \mathbb{R}^m \\ \Big\downarrow \Phi & \nearrow_{\mathrm{pr}_2} & \\ \ker Df(x) \oplus \mathbb{R}^m & & \end{array}$$

By the Linearisation Dogma we expect to be able to prove something similar for f, at least locally.

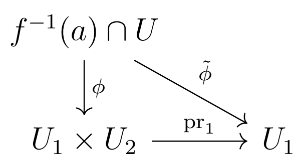
Proposition 11.5. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a smooth map and $Df(x)$ is surjective, then there exists an open subset $U \subset \mathbb{R}^n$ containing x, open subsets $U_1 \subset \mathbb{R}^{n-m}, U_2 \subset \mathbb{R}^m$, and a diffeomorphism $\phi : U \to U_1 \times U_2$ such that the following diagram commutes:



Proof. Let $K = \ker Df(x)$ and Define $F := (\pi_K, f) : \mathbb{R}^n \to K \oplus \mathbb{R}^m$. Then by properties 9.6 of the derivative $DF(x) = (\pi_K, Df(x))$, which is an isomorphism by the splitting lemma. By the inverse function theorem 10.13 there exists an open set $U \subset \mathbb{R}^n$ such that $F|U \to F(U)$ is a diffeomorphism. By definition of the product topology $F(U)$ contains an open set of the form $U_1 \times U_2$, so we let $U = F^{-1}(U_1 \times U_2)$ and define $\phi := F|U = (\pi_K, f)|U$, so $\mathrm{pr}_2 \circ \phi = f|U$. □

Theorem 11.6. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be smooth and suppose $a \in \mathbb{R}^m$ is a regular value of f, i.e. $Df(x)$ is surjective for every $x \in f^{-1}(a)$. Then $f^{-1}(a)$ is an $(n-m)$-dimensional submanifold of $\mathbb{R}^n$.

Proof. Given $x \in f^{-1}(a)$ we have $Df(x)$ surjective and therefore by the previous proposition there are open sets U, U_1, U_2 and a diffeomorphism $\phi : U \to U_1 \times U_2$ such that $\mathrm{pr}_2 \circ \phi = f$. We give $f^{-1}(a)$ the subspace topology so that $f^{-1}(a) \cap U$ is open in $f^{-1}(a)$. Moreover, for all $x \in f^{-1}(a) \cap U$ we have $a = f(x) = \mathrm{pr}_2 \circ \phi(x)$ and therefore $\phi(x) = (x_1, a)$ for some $x_1 \in U_1$. We claim that $\tilde{\phi} := \mathrm{pr}_1 \circ \phi | f^{-1}(a) \cap U \to U_1$ is a homemorphism:



- both ϕ, pr_1 are continuous, so the restriction is continuous in the subspace topology
- the inverse $u \mapsto \phi^{-1}(u, a)$ is also continuous: suppose $W \subset f^{-1}(a) \cap U$ is open, then $\phi(W) = W_1 \times \{a\}$ is open in $U_1 \times \{a\}$ because ϕ^{-1} is continuous, and therefore W_1 is open in U_1.

So for any $x \in f^{-1}(a)$ we have a chart $(\tilde{\phi}, f^{-1}(a) \cap U)$ for $f^{-1}(a)$ which contains x. It remains to show that overlapping charts are compatible. Suppose we have another chart $(\tilde{\gamma}, f^{-1}(a) \cap V)$ by

$$\begin{array}{ccc} f^{-1}(a) \cap V & & \\ \downarrow{\scriptstyle \gamma} & \searrow^{\tilde{\gamma}} & \\ V_1 \times V_2 & \xrightarrow{\mathrm{pr}_1} & V_1 \end{array}$$

which is such that $W := f^{-1}(a) \cap U \cap V \neq \emptyset$. Consider $\tilde{\gamma} \circ \tilde{\phi}^{-1} : \tilde{\phi}(W) \subset U_1 \to V_1$. For $u \in \tilde{\phi}(W)$ we have

$$\tilde{\gamma} \circ \tilde{\phi}^{-1}(u) = \mathrm{pr}_1 \circ \gamma \circ \phi^{-1}(u, a)$$

so $D(\tilde{\gamma} \circ \tilde{\phi})(u)$ is a partial derivative of $\mathrm{pr}_1 \circ \gamma \circ \phi^{-1}$. But γ and ϕ are diffeomorphisms and pr_1 is linear so the composition is smooth, and so is the partial derivative.

$\square$

Example 11.7. The n-sphere $S^n := \{x \in \mathbb{R}^n + 1 : \|x\| = 1\}$ is a submanifold of $\mathbb{R}^{n+1}$. The function $f : \mathbb{R}^{n+1} \backslash \{0\} \to \mathbb{R}$ defined by $f(x) := \|x\|^2$ has 1 as a regular value and $S^n = f^{-1}(1)$. To see that 1 is a regular value we calculate the derivative of f:

$$\begin{aligned} Df(x)v = \left.\tfrac{d}{dt}\right|_{t=0} f(x + tv) &= \left.\tfrac{d}{dt}\right|_{t=0} \|x + tv\|^2 \\ &= \left.\tfrac{d}{dt}\right|_{t=0} (x + tv) \cdot (x + tv) \\ &= \left.\tfrac{d}{dt}\right|_{t=0} (x \cdot x + 2tx \cdot v + t^2 v \cdot v) \\ &= 2x \cdot v \end{aligned}$$

Observe that for $x \neq 0$ we have $Df(x)$ surjective: given $w \in \mathbb{R}$ let $v = wx/2\|x\|^2$. So in fact every positive real number is a regular value of f.