University of
Cambridge

Mathematics Tripos

Part IB

Analysis II

Michaelmas, 2017

Lectures by
J. Rasmussen

Notes by
Qiangru Kuang

Contents

0 Introduction

In IA Analysis I, the primary space we are interested in is $\mathbb{R}$ and we studied notions such as continuity, convergence, differentiation, integration and solving equaiton through, for example, Intermediate Value Theorem. In Analysis II, we moved to the study general function space.

	$\mathbb{R}^m$	Function space
Convergence & continuity	✓	✓
Differentiation	✓	Calculus of variations
Integration	Probability and measure	??? (ask physicists)
Solving equations	inverse function theorem	existence of solutions for ODEs

Table 1: Comparison of Euclidean space and function space

1 Normed Vector Spaces

1.1 Definitions

A motivating example: if (a_n) is a sequence of real numbers, then $(a_n) \to a$ if

$$\forall \varepsilon > 0, \exists N s.t. \forall n > N, |a_n - a| < \varepsilon.$$

Now if I replace $\mathbb{R}$ by a real vector space V, what do I replace $|\cdot|$ with?

> **Definition** (Norm)**.** If V is a real vector space, a *norm* on V is a function $\|\cdot\| : V \to \mathbb{R}$ satisfying
>
> 1. $\forall \mathbf{v} \in V, \|\mathbf{v}\| \geq 0$ with equality if and only if $\mathbf{v} = \mathbf{0}$.
> 2. $\forall \mathbf{v}, \forall \lambda \in \mathbb{R}, \|\lambda \mathbf{v}\| = |\lambda| \|\mathbf{v}\|$
> 3. $\forall \mathbf{v}, \mathbf{w} \in V, \|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ (triangle inequality).

Example.

1. $V = \mathbb{R}^m, \mathbf{v} = (v_1, \dots, v_m)$,
   (a) $\|\mathbf{v}\| = (\sum_{i=1}^m v_i^2)^{1/2}$, the Euclidean norm,
   (b) $\|\mathbf{v}\|_\infty = \max |v_i|$, the max norm,
   (c) $\|\mathbf{v}\|_1 = \sum_{i=1}^m |v_i|$.
2. $V = C[0, 1]$,
   (a) $\|f\|_\infty = \max_{x \in [0,1]} |f(x)|$,
   (b) $\|f\|_2 = (\int_0^1 f(x)^2 dx)^{1/2}$, which comes from $\langle f, g \rangle = \int_0^1 f(x) g(x) dx$,
   (c) $\|f\|_1 = \int_0^1 |f(x)| dx$, the L^1 norm.

> **Definition** (Convergence)**.** Suppose $(V, \|\cdot\|)$ is a normed vector space and $(\mathbf{v}_n)$ is a sequence of elements of V. We say $(\mathbf{v}_n)$ *converges* to $\mathbf{v} \in V$, denoted $(\mathbf{v}_n) \to \mathbf{v}$, if $\forall \varepsilon > 0, \exists N s.t. \forall n > N, \|\mathbf{v}_n - \mathbf{v}\| < \varepsilon$. Equivalently, $(\mathbf{v}_n) \to \mathbf{v}$ if $(\|\mathbf{v}_n - \mathbf{v}\|) \to 0$.

Exercise. Suppose $V = \mathbb{R}^m, (\mathbf{v}_n) = (v_{n,1}, \dots, v_{n,m})$. Then $(\mathbf{v}_n) \to \mathbf{v}$ with respect to $\|\cdot\|_\infty$ means

$$\begin{aligned} &\left(\max_{1 \leq i \leq m} |v_{n,i} - v_i| \right) \to 0 \\ \iff &(|v_{n,i} - v_i|) \to 0 \text{ for all } 1 \leq i \leq m \\ \iff &(v_{n,i}) \to v_i \text{ for all } 1 \leq i \leq m \end{aligned}$$

The convergence with respect to $\|\cdot\|_1$ means

$$\begin{aligned} &\left(\sum_{i=1}^m |v_{n,i} - v_i| \right) \to 0 \\ \iff &\left(|v_{n,i} - v_i| \right) \to 0 \text{ for all } 1 \leq i \leq m \\ \iff &(v_{n,i}) \to v_i \text{ for all } 1 \leq i \leq m. \end{aligned}$$

Remark. Two different norms, same notion of convergence.

Convention. If I say $(\mathbf{v}_n) \to \mathbf{v}$ where $(\mathbf{v}_n)$ is a sequence in $\mathbb{R}^m$ without specifying the norm, I mean w.r.t. $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$. (all give the same notion for convergence).

Example. $V = C[0,1]$,

$$f_n(x) = \begin{cases} 1 - nx & x \in [0, 1/n] \\ 0 & x \geq 1/n \end{cases}$$

Then $\|f_n\|_1 = \int_0^1 |f_n(x)| dx = \frac{1}{2n} \to 0$ so $(f_n) \to 0$ w.r.t. $\|\cdot\|_1$.
But $\|f_n\|_\infty = 1$ so $(\|f_n\|_\infty) \nrightarrow 0$ i.e. $(f_n) \nrightarrow 0$ w.r.t. $\|\cdot\|_\infty$.

Remark. Two different norms, two different notions of convergence.

1.2 Continuity

Definition (Continuity)**.** Suppose V and W are normed vector spaces. We say a function $f : V \to W$ is *continuous* if

$$(f(\mathbf{v}_n)) \to f(\mathbf{v}) \text{ in } W \text{ whenever } (\mathbf{v}_n) \to \mathbf{v} \text{ in } V.$$

Example.

1. $f : V \to \mathbb{R}^m, f(\mathbf{v}) = (f_1(\mathbf{v}), \dots, f_m(\mathbf{v}))$ is continuous if and only if $f_i : V \to \mathbb{R}$ is continuous for all $1 \leq i \leq m$.
2. $\rho_i : \mathbb{R}^m \to \mathbb{R}, \rho_i(\mathbf{v}) = v_i$ is continuous.
3. $F : C[0,1] \to \mathbb{R}, F(f) = f(0)$,
   (a) If (f_n) is the sequence from the example on page 4, then $F(f_n) = 1$. Now $(f_n) \to \mathbf{0}$ w.r.t. $\|\cdot\|_1$. But $(F(f_n)) \nrightarrow 0 = F(\mathbf{0})$. So F is not continuous w.r.t. $\|\cdot\|_1$.
   (b) If $(g_n) \to g$ w.r.t. $\|\cdot\|_\infty$, then $(\max |g_n(x) - g(x)|) \to 0$ so $(|g_n(0) - g(0))| \to 0$, $(|F(g_n) - F(g)|) \to 0$, so $F(g_n) \to F(g)$.

   F is continuous w.r.t. $\|\cdot\|_\infty$ but not w.r.t. $\|\cdot\|_1$.
4. If $f : V_1 \to V_2$ and $g : V_2 \to V_3$ are continuous then $g \circ f : V_1 \to V_3$ are continuous.

   Proof. If $(\mathbf{v}_n) \to (\mathbf{v})$ in V_1 , then as f is continuous, $(f(\mathbf{v}_n)) \to (f(\mathbf{v}))$, then as g is continuous, $(g(f(\mathbf{v}_n))) \to (g(f(\mathbf{v})))$ in V_3. □
5. $\|\cdot\| : V \to \mathbb{R}$ is continuous.

   Proof. If $(\mathbf{v}_n) \to \mathbf{v}$, then $(\|\mathbf{v}_n - \mathbf{v}\|) \to 0$. Now

   $$0 \leq |\|\mathbf{v}_n\| - \|\mathbf{v}\|| \leq \|\mathbf{v}_n - \mathbf{v}\|$$

   by Reverse triangle inequality. So $(|\|\mathbf{v}_n - \mathbf{v}\||) \to 0$ by squeeze rule, i.e. $\|\mathbf{v}_n\| \to \|\mathbf{v}\|$. □

Lemma 1.1 (Reverse triangle inequality)**.** $\|\mathbf{v}-\mathbf{w}\| \geq |\|\mathbf{v}\| - \|\mathbf{w}\||$ *for all* $\mathbf{v}, \mathbf{w} \in V$.

Proof. By triangle inequality, $\|\mathbf{v}-\mathbf{w}\| + \|\mathbf{w}\| \geq \|\mathbf{v}\|$ so $\|\mathbf{v}-\mathbf{w}\| \geq \|\mathbf{v}\| - \|\mathbf{w}\|$ and $\|\mathbf{v}-\mathbf{w}\| \geq -\|\mathbf{v}\| + \|\mathbf{w}\|$ □

More generally, if $X \subseteq V$ is a subset, we say $f : X \to W$ is continuous if

$$(f(\mathbf{x}_n)) \to f(\mathbf{x})$$

in W whenever $(\mathbf{x}_n) \to \mathbf{x}$ in V for $\mathbf{x}$ and all $\mathbf{x}_n \in X$.

Example. $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}, x \mapsto \frac{1}{x}$ is continuous.

1.3 Open and Closed Subsets

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

Definition (Open and closed ball)**.** If $\mathbf{v}_0 \in V$ and $r \in \mathbb{R}$,

$$B_r(\mathbf{v}_0) = \{\mathbf{v} \in V : \|\mathbf{v} - \mathbf{v}_0\| < r\}$$

is the *open ball* of radius r centred at $\mathbf{v}_0$, and

$$\overline{B}_r(\mathbf{v}_0) = \{\mathbf{v} \in V : \|\mathbf{v} - \mathbf{v}_0\| \leq r\}$$

is the *closed ball* of radius r centred at $\mathbf{v}_0$.

Example.

1. $(V, \|\cdot\|) = (\mathbb{R}, |\cdot|)$, then

$$\begin{aligned} B_r(a) &= (a-r, a+r) \\ \overline{B}_r(a) &= [a-r, a+r] \end{aligned}$$

2. $V = \mathbb{R}^2$, then $\overline{B}_1(\mathbf{0})$ with respect to `to be filled in`.
3. $V = \mathbb{R}^3, \|\cdot\|_2$ is the "three-dimensional ball".
4. $(V, \|\cdot\|) = (C[0,1], \|\cdot\|_\infty)$,

$$\overline{B}_r(f) = \{g \in C[0,1] : f(x) - r \leq g(x) \leq f(x) + r \; \forall x \in [0,1]\}.$$

Proposition 1.2 (Alternate characterisation of continuity)**.** $f : V \to W$ *is continuous if and only if*

$$\forall \mathbf{v}_0 \in V, \forall \varepsilon > 0, \exists \delta > 0 s.t. \|\mathbf{v} - \mathbf{v}_0\| < \delta \Rightarrow \|f(\mathbf{v}) - f(\mathbf{v}_0)\| < \varepsilon \qquad (*)$$

i.e.

$$f(B_\delta(\mathbf{v}_0)) \subseteq B_\varepsilon(f(\mathbf{v}_0)).$$

Proof. Suppose ($*$) holds. Given $(v_n) \to v$, must show $(f(v_n)) \to f(v)$. Given $\varepsilon > 0$, pick δ such that $(f(B_\delta(v)) \subseteq B_\epsilon(f(v))$. Since $(v_n) \to (v)$, exists N such that whenever $n > N$, $\|v_n - v\| < \delta$, i.e. $v_n \in B_\delta(v)$, so $f(v_n) \in B_\epsilon(f(v))$. In other words, whenever $n > N$, $\|f(v_n) - f(v)\| < \varepsilon$.

Suppose ($*$) does not hold. Then exists some $v \in V$ and $\varepsilon > 0$ such that there is no $\delta > 0$ with $f(B_\delta(v)) \subseteq B_\varepsilon(f(v))$. In particular $f(B_{1/n}(v)) \nsubseteq B_\varepsilon(f(v))$ for all n. Pick $v_n \in B_{1/n}(v)$ with $f(v_n) \notin B_\varepsilon(f(v))$. Then $(v_n) \to v$, but $(f(v_n)) \nrightarrow f(v)$, since $\|f(v_n) - f(v)\| \geq \varepsilon$ for all n. f is not continuous. □

Definition (Open subset)**.** $U \subseteq V$ is an *open subset* of V if for every $u \in U$ there is some $\varepsilon > 0$ with $B_\varepsilon(u) \subseteq U$.

Proposition 1.3. *If $f : V \to W$ is continuous and $U \subseteq W$ is open then*

$$f^{-1}(U) = \{v \in V : f(v) \in V\}$$

is an open subset of V.

Proof. If $v \in f^{-1}(U)$, then $f(v) \in U$. U is open in W so exists $\varepsilon > 0$ such that $B_\varepsilon(f(v)) \subseteq U$. F is continuous so exist $\delta > 0$ such that $f(B_\delta(v)) \subseteq B_\varepsilon(f(v)) \subseteq U$. So $B_\delta(v) \subseteq f^{-1}(U)$. $f^{-1}(U)$ is open. □

Remark. The converse statement is also true: if for any $U \subseteq W$ open $f^{-1}(U)$ open in V, then f is continuous.

Example.

1. $(0, 1)$ is open in $\mathbb{R}$.
2. The function $h(v) = \|v - v_0\|$ is continuous: $h(v) = g \circ f(v)$ where $f(v) = v - v_0$, and $g(v) = \|v\|$. so $B_r(r) = h^{-1}((-r, r))$ is open in V.

Definition (Closed subset)**.** $C \subseteq V$ is a *closed subset* of V if $V \setminus C$ is open in V.

Corollary 1.4. *If $f : V \to W$ is continuous and $C \subseteq W$ is closed, then $f^{-1}(C)$ is closed in V.*

Proof.

$$f^{-1}(W \setminus C) = V \setminus f^{-1}(C)$$

so if $C \subseteq W$ is closed, $W \setminus C$ is open, so $f^{-1}(W \setminus C) = V \setminus f^{-1}(C)$ is open. Thus $f^{-1}(C) \subseteq V$ is closed. □

Example.

1. $[a, b]$ is closed in $\mathbb{R}$.
2. $h(v) = \|v - v_0\|, \overline{B}_r(v_0) = h^{-1}([0, r])$ so closed ball is closed.
3. $V, \emptyset$ are both open and closed in V.
4. $\mathbb{Q} \subseteq \mathbb{R}$ is neither open nor closed.

Proposition 1.5. *$C \subseteq V$ is closed if and only if for every sequence $(v_n) \to v$ with all $v_n \in C$, $v \in C$.*

Proof. Suppose C is closed and $(v_n) \to v \notin C$. Then $v \in V \setminus C$ is open, so $\exists \varepsilon > 0$ with $B_\varepsilon(v) \subseteq V \setminus C$, i.e. $B_\varepsilon(v) \cap C = \emptyset$. $(v_n) \to v$ so $\exists N$ such that $v_n \in B_\varepsilon(v)$ for all $n > N$. Thus $v_n \notin C$ for all $n > N$. In other word, if $(v_n) \to v$, all but finitely many of $v_n \notin C$.

Conversely, suppose C is not closed. Then $V \setminus C$ is not open so $\exists c \in V \setminus C$ such that there is no $\varepsilon > 0$ with $B_\varepsilon(v) \subseteq V \setminus C$. In other words, $B_\varepsilon(v) \cap C \neq \emptyset$ for all $\varepsilon > 0$. Pick $v_n \in B_{1/n}(v) \cap C$ for all $n > 0$. Then $\|v_n - v\| < 1/n$ so $(v_n) \to v$. All $v_n \in C$ but $v \in V \setminus C$. □

Example. The set $X = \{f \in C[0,1] : \forall x, f(x) > 0\}$ is not closed with respect to $\|\cdot\|_1$ or $\|\cdot\|_\infty$ since $f_n(x) = \frac{1}{n} \in X$, $(f_n) \to 0$ with respect to either norm but $0 \notin X$.

For future use, suppose for all $\alpha \in A$, $U_\alpha \subseteq V$ is open. Given $U = \bigcup_{\alpha \in A} U_\alpha$ and $f : U \to W$,

Proposition 1.6. *If $f|_{U_\alpha} : U_\alpha \to W$ is continuous for all $\alpha \in A$, then $f : U \to W$ is continuous.*

Note. The hypothesis that U_α is open is important. For example, let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 1$ if $x \in \mathbb{Q}$, $f(x) = 0$ otherwise, then $f|_\mathbb{Q}$ and $f|_{\mathbb{R}\setminus\mathbb{Q}}$ are both continuous but f is not.

Proof. Suppose $v_n, v \in U$ with $(v_n) \to v$. Must show $(f(v_n)) \to f(v)$. $v \in U$ so $v \in U_\alpha$ for some α. U_α is open so $\exists \varepsilon > 0$ with $B_\varepsilon(v) \subseteq U_\alpha$, $(v_n) \to v$ so $\exists N$ with $v_n \in B_\varepsilon(v)$ for all $n > N$. Let $u_i = v_{N+1}$, then $u_i \in U_\alpha$ and $(u_i) \to v$. Since $f|_{U_\alpha}$ is continuous, $(f(u_i)) \to f(v)$ which implies that $(f(v_n)) \to f(v)$. □

1.4 Lipschitz Equivalence

Recall from the introduction of norms that $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$ on $\mathbb{R}^n$ all induce the same notion of convergence. We want to generalise this idea.

Suppose $\|\cdot\|$ and $\|\cdot\|'$ are two norms on V. Consider

$$\begin{aligned} \mathrm{id}_V : (V, \|\cdot\|) &\to (V, \|\cdot\|') \\ v &\mapsto v \end{aligned}$$

Proposition 1.7. id_V *as above is continuous if and only if $\exists C \in \mathbb{R}$ with $\|v\|' \leq C\|v\|$ for all $v \in V$.*

Proof. Suppose $\|v\|' \leq C\|v\|$ for all v. To show id_V is continuous, must show $(v_n) \to v$ with respect to $\|\cdot\|'$ whenever $(v_n) \to v$ with respect to $\|\cdot\|$.

If $(v_n) \to v$ with respect ot $\|\cdot\|$, then $(\|v_n - v\|) \to 0$ so $(C\|v_n - v\|) \to 0$. We know

$$0 \leq \|v_n - v\|' \leq C\|v_n - v\|,$$

so by squeeze rule $\|v_n - v\| \to 0$. Thus $(v_n) \to v$ with respect to $\|\cdot\|'$.

Conversely, suppose id_V is continuous. There exist $\delta > 0$ such that

$$\mathrm{id}_V(B_\delta(0, \|\cdot\|) \subseteq B_1(0, \|\cdot\|')$$

Given $v \in V, v \neq 0$, exists $K \in \mathbb{R}$ with $\|Kv\| = \delta/2$ (take $K = \frac{\delta}{2\|v\|}$). Then $Kv \in B_\delta(0, \|\cdot\|) \Rightarrow Kv \in B_1(0, \|\cdot\|')$, i.e. $\|Kv\| = \delta/2, \|Kv\|' < 1$, so $\|Kv\|' \leq \frac{2}{\delta}\|Kv\| \Rightarrow K\|v\|' \leq \frac{2}{\delta}K\|v\|$. Let $C = \frac{2}{\delta}$. □

Joke. `The joke about a mathematician going for a firefighter interview... Well you should know it by now if you are a mathematician.`

Definition (Lipschitz equivalence)**.** If $\|\cdot\|$ and $\|\cdot\|'$ are two norms on V, they are said to be *Lipschitz equivalent* if

$$\begin{aligned} &\exists C > 0 s.t. \forall v \in V, \frac{1}{C}\|v\| \leq \|v\|' \leq C\|v\| \\ \Longleftrightarrow &\exists C_1, C_2 s.t. \|v\|' \leq C_1\|v\|, \|v\| \leq C_2\|v\|' \\ \Longleftrightarrow &\ \mathrm{id}_V : (V, \|\cdot\|) \to (V, \|\cdot\|') \text{ and } \mathrm{id}_V : (V, \|\cdot\|') \to (V, \|\cdot\|) \\ &\text{are both continuous.} \end{aligned}$$

Corollary 1.8. *If $(V, \|\cdot\|)$ and $(V, \|\cdot\|')$ are Lipschitz equivalent, then*

1. *$(v_n) \to v$ with respect to $\|\cdot\|$ if and only if with respect to $\|\cdot\|'$,*
2. *$f : V \to W$ is continuous with respect to $\|\cdot\|$ if and only if with respect to $\|\cdot\|'$,*
3. *$F : W \to V$ is continuous with respect to $\|\cdot\|$ if and only if with respect to $\|\cdot\|'$.*

Proof. Example proof: if $f : (V, \|\cdot\|) \to W$ is continuous, then $f' : (V, \|\cdot\|') \to W$ is the composition

$$(V, \|\cdot\|') \xrightarrow{\mathrm{id}_V} (V, \|\cdot\|) \xrightarrow{f} W$$

so continuous. □

Example.

1. $V = \mathbb{R}^n$, $\|v\|_\infty \leq \|v\|_2 \leq \|v\|_1 \leq n\|v\|_\infty$ so all three are Lipschitz equivalent.
2. $V = C[0, 1]$, $\mathrm{id}_V : (V, \|\cdot\|_\infty) \to (V, \|\cdot\|_1)$ is continuous but $\mathrm{id}_V : (V, \|\cdot\|_1) \to (V, \|\cdot\|_\infty)$ is not so not Lipschitz equivalent.

2 Uniform Convergence

2.1 Notions of Convergence

Suppose $A \subseteq \mathbb{R}, f, f_n : A \to \mathbb{R}$. We say f is *continuous* if given $x \in A$ and $\varepsilon > 0$, $\exists \delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever $y \in A$ and $|x - y| < \delta$. We say f is *bounded* if exists M such that $|f(x)| \leq M$ for all $x \in A$. Define

- $C(A) = \{f : A \to \mathbb{R} : f \text{ is continuous}\}$,
- $B(A) = \{f : A \to \mathbb{R} : f \text{ is bounded}\}$,

which are both vector space.

Example. $C[0,1] \subseteq B[0,1]$ by Maximum Value Theorem. $g(x) = \frac{1}{x} \in C(0,1]$ so $C(0,1] \nsubseteq B(0,1]$.

Definition (Pointwise Convergence)**.** $(f_n) \to f$ *pointwise* if

$$(f_n(x)) \to f(x) \text{ for all } x \in \mathbb{R}.$$

Definition (Uniform norm)**.** The *uniform norm* on $B(A)$ is given by

$$\|f\|_\infty = \sup_{x \in A} |f(x)|.$$

Definition (Uniform convergence)**.** If $f, f_n : A \to \mathbb{R}$, we say $f(x) \to f$ *uniformly on* A if $(f_n - f) \in B(A)$ for all n and $(\|f_n - f\|_\infty) \to 0$.

In other words,

- $(f_n) \to f$ pointwise means: you give me $x \in A$ and $\varepsilon > 0$, I have to find N such that $|f_n(x) - f(x)| < \varepsilon$ whenever $n > N$. *This N only has to work for that particular value of x.*
- $(f_n) \to f$ uniformly means: you give me $\varepsilon > 0$, I have to find N such that $|f_n(x) - f(x)| < \varepsilon$ for all $x \in A$ and $n > N$. *Same N works for all $x \in A$.*

Exercise. If $(f_n) \to f$ uniformly, then $(f_n) \to f$ pointwise. The converse is false.

Example.

- Suppose $A = \mathbb{R}, f_n(x) = x + \frac{1}{n}, f(x) = x$. Then $f_n(x) - f(x) = \frac{1}{n}$ so $(f_n) \to f$ uniformly.
- Let $g_n(x) = (x + \frac{1}{n})^2, g(x) = x^2$. Then $(g_n) \to g$ pointwise but $g_n(x) - g(x) = \frac{2x}{n} + \frac{1}{n^2}$ is not even bounded. So $(g_n) \nrightarrow g$ uniformly on $\mathbb{R}$.

Theorem 2.1. *Suppose $f_n \in C(A)$ for all n and $(f_n) \to f$ uniformly on A. Then $f \in C(A)$ as well.*

Slogan. The uniform limit of continuous functions is continuous.

Proof. Suppose f_n are continuous and $(f_n) \to f$ uniformly. Given $x \in A$ and $\varepsilon > 0$, must find $\delta > 0$ such that $|f(x)-f(y)| < \varepsilon$ whenever $y \in A$ and $|x-y| < \delta$.

Since $(f_n) \to f$ uniformly, there exists N such that $|f_n(x) - f(x)| < \varepsilon/4$ for all $x \in A$ and $n \geq N$. Since f_N is continuous, exists $\delta > 0$ such that $|f_N(x) - f_n(y)| < \varepsilon/2$ whenever $y \in A$ and $|x - y| < \delta$. Then if $|x - y| < \delta$,

$$\begin{aligned}|f(x) - f(y)| &\leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| \\ &\leq \varepsilon/4 + \varepsilon/2 + \varepsilon/4 \\ &= \varepsilon\end{aligned}$$

□

Example. Take $A = [0, 1]$,

- $f_n(x) = x^n, f(x) = 1$ if $x = 1, f(x) = 0$ if $x \neq 1$. Then $(f_n) \to f$ pointwise on $[0, 1]$ but $f_n \in C[0, 1], f \notin C[0, 1]$ so $(f_n) \not\to f$ uniformly on $[0, 1]$.
- $g_n(x) = x^n(1 - x), g(x) = 0$. Then $(g_n) \to g$ uniformly.

  Proof. Given $\varepsilon > 0, 1-\varepsilon < 1$ so $(1-\varepsilon)^n \to 0$. Pick N such that $(1-\varepsilon)^n < \varepsilon$ for all $n > N$. Then $|f_n(x)| = |(1-x)x^n| \leq 1 \cdot (1-\varepsilon)^n < \varepsilon$ for $x \in [0, 1-\varepsilon]$ and $|f_n(x)| = |(1 - x)x^n| < \varepsilon \cdot 1^n = \varepsilon$ for $x \in (1 - \varepsilon, 1]$. Thus $|f_n(x)| < \varepsilon$ for all $x \in [0, 1]$. □

Note. The converse, at least when taken literally, is false. See example sheet 1 Q11.

Remark. Everything I have said so far works fine for $A \subseteq V, f : A \to W$, where V, W are normed vector spaces.

Joke. A mathematician named Cliff measured his room for painting. His wife went off to the paint store and told the counter how much paint she needed. The counter said: "Thats a lot of paint. Are you sure you want that much?" To which the wife answered: "Well my husband is a mathematician. I'm sure he gets the numbers correct."

She arrived back home with really a lot of paint. Cliff moved all the paint in the house and suddenly said:

"Oh, damn! I measured the volumn instead of the area!"

Recall that if $f \in C[a, b]$ then $\|f\|_1 = \int_a^b |f(x)|dx$.

Definition (Convergence in measure)**.** f_n converges *in measure* to f if $(f_n) \to f$ with respect to $\| \cdot \|_1$,

Lemma 2.2. *If $(f_n) \in C[a,b]$ and $(f_n) \to f$ uniformly then $(f_n) \to f$ in measure.*

Proof. Given $\varepsilon > 0$, pick N such that $|f_n(x) - f(x)| < \varepsilon/2(b-a)$ for all $x \in [a,b]$. Then

$$\begin{aligned}\|f_n - f\| &= \int_a^b |f_n(x) - f(x)| dx \\ &\leq \int_a^b \varepsilon/2(b-a) dx \\ &= (\varepsilon/2(b-a))(b-a) \\ &= \varepsilon/2.\end{aligned}$$

□

Equivalently, the map $\mathrm{id} : (C[a,b], \|\cdot\|_\infty) \to (C[a,b], \|\cdot\|_1)$ is continuous.

Example. Let $A = [0,1]$,

1. $f(x) = \begin{cases} nx & x \in [0,1/n] \\ 2 - nx & x \in [1/n, 2/n] \\ 0 & x \geq 2/n \end{cases}$ Then $(f_n) \to 0$ pointwise and in measure but not uniformly.
2. $g_n(x) = \begin{cases} n^2 x & x \in [0,1/n] \\ 2n - n^2 x & x \in [1/n, 2/n] \\ 0 & x \geq 2/n \end{cases}$ Then $(g_n) \to 0$ pointwise but $(g_n) \not\to 0$ in measure or uniformly.

2.2 Power Series

Question. Given

$$f(x) = \sum_{i=0}^{\infty} \frac{x^i}{i!},$$

how do I know if $f(x)$ is continuous or differentiable?

Recall some facts about series from IA Analysis I:

1. The series $\sum_{i=0}^{\infty} c_i = c \in \mathbb{C}$ means that $(\sum_{i=0}^{\infty}) \to c$, as real vector space $(\mathbb{C}, \|\cdot\|) \cong (\mathbb{R}^2, \|\cdot\|)$.
2. $\sum_{i=0}^{\infty} c_i$ converges if and only if there exists $N \in \mathbb{N}$ such that $\sum_{i=N}^{\infty} c_i$ converges.
3. Geometric series: $\sum_{i=k}^{\infty} \alpha^i = \frac{\alpha^k}{1-\alpha}$ for $|\alpha| < 1$.
4. If $\sum_{i=0}^{\infty} c_i$ converges then $(c_i) \to 0$.
5. Comparison test: if $0 \leq a_i \leq b_i$ for all i and $\sum_{i=0}^{\infty} b_i$ converges then $\sum_{i=0}^{\infty} a_i$ converges and $\sum_{i=0}^{\infty} a_i \leq \sum_{i=0}^{\infty} b_i$.

6. Absolute convergence: if $\sum_{i=0}^{\infty}|c_i|$ converges then $\sum_{i=0}^{\infty}c_i$ converges and $|\sum_{i=0}^{\infty}c_i| \leq \sum_{i=0}^{\infty}|c_i|$.

Lemma 2.3. *If $0 \leq |c_i| \leq b_i$ for all i and $\sum_{i=0}^{\infty} b_i$ converges then $\sum_{i=0}^{\infty} c_i$ converges.*

Proof. Combine property 5 and 6. □

Definition (Power series)**.** A series of the form

$$\sum_{i=0}^{\infty} a_i(z-c)^i,$$

where $a_i, z, c \in C$ is called a *power series.* C is the *centre.*

Proposition 2.4. *If*

$$\sum_{i=0}^{\infty} a_i(z_0-c)^i$$

converges then

$$\sum_{i=0}^{\infty} a_i(z-c)^i$$

converges whenever

$$|z-c| < |z_0-c|.$$

Proof. By property 4 $(a_i(z_0-c)^i) \to 0$. Pick N such that $|a_i(z_0-c)^i| < 1$ for all $i \geq N$. By Property 2 it suffices to show that

$$\sum_{i=N}^{\infty} a_i(z-c)^i$$

converges. Now for $i \geq N$,

$$\boxed{|a_i(z-c)^i| = |a_i(z_0-c)^i| \cdot \left|\frac{z-c}{z_0-c}\right|^i < 1 \cdot \alpha^i}$$

Fundamental Estimate for Power Series

where $\alpha = |\frac{z-c}{z_0-c}|$. So if $|z-c| < |z_0-c|, \alpha < 1$, $\sum_{i=N}^{\infty}\alpha^i$ converges by property 3.

In summary, we have

$$|a_i(z-c)^i| < \alpha^i$$

for all $i \geq N$ and $\sum_{i=N}^{\infty}\alpha^i$ converges. By the lemma $\sum_{i=N}^{\infty} a_i(z-c)^i$ converges. □

Definition (Radius of convergence)**.**

$$R := \sup\{|z-c| : \sum_{i=0}^{\infty} a_i(z-c)^i \text{ converges}\}$$

is the *radius of convergence* of $\sum_{i=0}^{\infty} a_i(z-c)^i$.

Proposition 2.4 implies that if $z \in B_R(c)$ then $\sum_{i=0}^{\infty} a_i(z-c)^i$ converges. In other words, if we define

$$\begin{aligned} f : B_R(c) &\to \mathbb{C} \\ z &\mapsto \sum_{i=0}^{\infty} a_i(z-c)^i \\ P_n : B_R(c) &\to \mathbb{C} \\ z &\mapsto \sum_{i=0}^{n} a_i(z-c)^i \end{aligned}$$

Proposition 2.4 says that $(P_n) \to f$ pointwise on $B_R(c)$. As P_n are polynomials so they are continuous. A natural question is, is f continuous as well? We know this answer will be yes if we can prove that the convergence is uniform.

Theorem 2.5. *With notations as above,*

$$(P_n) \to f$$

uniformly on $\overline{B}_r(c)$ *whenever* $r < R$.

Note. Equivalently, we can say $(P_n) \to f$ uniformly on $B_r(c)$ for $r < R$. The closed ball $\overline{B}_r(c)$ is just a convention when talking about uniform convergence on a compact set.

Proof. Define

$$E_n(z) = f(z) - P_n(z) = \sum_{i=n+1}^{\infty} a_i(z-c)^i.$$

Fix $r < R$. Given $\varepsilon > 0$, need to find N such that $|E_n(z)| < \varepsilon$ whenever $n \geq N$ and $z \in \overline{B}_r(c)$.

Choose z_0 with $r < |z_0 - c| < R$ as in the proof of Proposition 2.4, pick N_0 such that $|a_i(z_0 - c)|^i < 1$ for $i \geq N_0$. Now we use Fundamental Estimate for Power Series. For $i \geq N_0$, we have $|a_i(z-c)^i| < \alpha(z)^i$ where $\alpha(z) = |\frac{z-c}{z_0-c}|$. For $z \in \overline{B}_r(c)$,

$$\alpha(z) = \left|\frac{z-c}{z_0-c}\right| \leq \frac{r}{|z_0-c|} = \alpha_0 < 1$$

since $r < |z_0 - c|$. Hence for $n > N_0$,

$$|E_n(z)| \leq \sum_{i=n+1}^{\infty} |a_i(z-c)^i| \leq \sum_{i=n+1}^{\infty} \alpha_0^i = \frac{\alpha_0^{n+1}}{1-\alpha_0}.$$

As $\alpha_0 < 1$, $(\alpha_0^i) \to 0$. Pick $N \geq N_0$ such that

$$\alpha_0^i < \varepsilon(1-\alpha_0)$$

for $i \geq N$. So for $n > N$,

$$|E_n(z)| < \frac{\varepsilon(1-\alpha_0)}{1-\alpha_0} = \varepsilon$$

for all $z \in \overline{B}_r(c)$. This is what we wanted. $\square$

Note. It need not be true that $(P_n) \to f$ uniformly on $B_R(c)$. For example,

$$\sum_{i=0}^{\infty} z^i$$

does not converge uniformly on $B_1(0)$.

Corollary 2.6. *f as above is continuous on $B_R(c)$.*

Proof. Let $U_r = B_r(c), r < R$. Then U_r is open in $\mathbb{C}$. $(P_n) \to f$ uniformly on U_r for $r < R$. Since the P_n are continuous $f|_{U_r}$ is continuous. By gluing lemma f is continuous on

$$U = \bigcup_{r<R} U_r = B_R(c).$$

□

To summarise, power series are locally uniformly convergent and thus continuous on its domain of convergence $B_R(c)$.

2.3 Integration & Differentiation

Recall from IA Analysis I that

Theorem 2.7 (Fundamental Theorem of Calculus)**.** *Suppose $f \in C[a,b]$, $c \in [a,b]$, then*

$$F(x) = \int_c^x f(y)dy$$

is well-defined for $x \in [a,b]$ and

$$F'(x) = f(x).$$

and the following properties of (Riemann) integral:

1. If $f(x) \le g(x)$ for $x \in [a,b]$, $\int_a^b f(x)dx \le \int_a^b g(x)dx$.
2. $\left|\int_a^b f(x)dx\right| \le \int_a^b |f(x)|dx$.
3. If $b < a$, $\int_a^b f(x)dx = -\int_b^a f(x)dx$.

Lemma 2.8. *If $|f(x)| \le C$ for all $x \in [a,b]$ then*

$$\left|\int_c^x f(t)dt\right| \le C|x-c|.$$

Proof. If $x \ge c$ then

$$\left|\int_c^x f(t)dt\right| \le \int_c^x |f(t)|dt \le \int_c^x Cdt = C(x-c).$$

If $x \le c$ then

$$\left|\int_c^x f(t)dt\right| \le \int_c^x |-f(t)|dt \le C|x-c|.$$

□

Now suppose $f_n \in C[a,b]$ and $(f_n) \to f$ uniformly on $[a,b]$ so $f \in C[a,b]$. Define

$$F_n(x) = \int_c^x f_n(t)dt$$
$$F(x) = \int_c^x f(t)dt$$

Proposition 2.9. *With notations above,*

$$(F_n) \to F$$

uniformly on $[a,b]$.

Proof. Given $\varepsilon > 0$, there exists N such that $|f_n(x) - f(x)| < \varepsilon/(b-a)$ for all $n \geq N$ and $x \in [a,b]$. Then

$$\begin{aligned} |F_n(x) - F(x)| &= \Big| \int_c^x f_n(t)dt - \int_c^x f(t)dt \Big| \\ &\leq \Big| \int_c^x (f_n(t) - f(t))dt \Big| \\ &\leq |x-c| \cdot \frac{\varepsilon}{b-a} \text{ by lemma} \\ &\leq \varepsilon \end{aligned}$$

since $x, c \in [a,b]$, $|x-c| \leq b-a$. Thus $(F_n) \to F$ uniformly on $[a,b]$. □

Now suppose $f(x) = \sum_{i=0}^{\infty} a_i(x-c)^i$ is a real power series with radius of convergence R. Then for $r < R$ and $P_n(x) = \sum_{i=0}^{n} a_i(x-c)^i$, $(P_n) \to f$ uniformly on $\overline{B}_r(c) = [c-r, c+r]$.

Corollary 2.10.

$$\int_c^x f(t)dt = \sum_{i=0}^{\infty} \frac{a_i}{i+1}(x-c)^{i+1}$$

for $x \in (c-R, c+R)$.

Proof. Given x, choose r with $|x-c| < r < R$. Then $(P_n) \to f$ on $[c-r, c+r]$ so by Proposition 2.9

$$(\mathbf{P}_n) \to \int_c^x f(t)dt$$

uniformly on $[c-r, c+r]$ where

$$\mathbf{P}_n(x) = \int_c^x \sum_{i=0}^{n} a_i(t-c)^i dt = \sum_{i=0}^{n} \frac{a_i}{i+1}(x-c)^{i+1}$$

Since uniform convergence implies pointwise convergence,

$$\sum_{i=0}^{\infty} \frac{a_i}{i+1}(x-c)^{i+1} = \int_c^x f(t)dt.$$

□

Question. If $(f_n) \to f$ uniformly on $[a, b]$ and f_n are differentiable, what can we say about (f'_n)?

The answer is, surprisingly, *absolutely nothing.*

Example. Let $f(x) = \frac{1}{n} \sin nx$. Then $(f_n) \to \mathbf{0}$ uniformly on $[0, \pi]$. But $f'_n(x) = \cos nx$ does not even converge for any $x \neq 0$.

Nevertheless, if $f(x) = \sum_{i=0}^{\infty} a_i (x-c)^i$ has radius of convergence R, we still have

Proposition 2.11. *f is differentiable on $(c - R, c + R)$ and*

$$f'(x) = \sum_{i=1}^{\infty} i a_i (x-c)^{i-1}.$$

In other words, power series can be differentiated term-by-term.

Lemma 2.12. *$g(x) = \sum_{i=1}^{\infty} i a_i (x-c)^{i-1}$ converges for all $y \in (c-R, c+R)$.*

Proof. Given $x \in (c - R, c + R)$, pick x_0 with $|x - c| < |x_0 - c| < R$. Then $\sum_{i=0}^{\infty} a_i (x_0 - c)^i$ converges, so by Fundamental Estimate for Power Series, there exists N such that

$$|a_i (x-c)^i| \leq \alpha^i$$

for all $i \geq N$, where $\alpha = \frac{|x-c|}{|x_0-c|} < 1$. Then

$$b_i := |i a_i (x-c)^{i-1}| \leq \left| \frac{i a_i}{x-c} \cdot (x-c)^i \right| \leq \frac{i}{|x-c|} \alpha^i$$

where we assume $y \neq c$. Now

$$\lim_{i \to \infty} \frac{i+1}{i} \cdot \alpha = \alpha < 1$$

so $\sum_{i=1}^{\infty} \frac{i}{|x-c|} \alpha^i$ converges by ratio test. Since

$$|i a_i (x-c)^{i-1}| \leq \frac{i}{|x-c|} \alpha^i,$$

$\sum_{i=1}^{\infty} i a_i (x-c)^{i-1}$ converges by comparison test. If $x = c$ then the convergence is obvious. □

Proof of proposition. $g(x) = \sum_{i=1}^{\infty} i a_i (x-c)^{i-1}$ converges on $(c - R, c + R)$, so by term-by-term integration

$$\int_c^x g(t) dt = \sum_{i=1}^{\infty} a_i (x-c)^i = f(x) - f(c).$$

Now $g(x)$ is continuous on $(c - R, c + R)$ so we can apply Fundamental Theorem of Calculus so $f'(x) = g(x)$ for all $x \in (c - R, c + R)$. □

Application. Power series solutions of ODEs are legit as long as you check the radius of convergence.

3 Compactness & Completeness

3.1 Compact subsets of $\mathbb{R}^n$

Let $(V, \|\cdot\|)$ be a normed vector space. If (v_n) is a sequence in V and (n_j) is an increasing sequence of positive integers (i.e. $n_{j+1} > n_j$) then (v_{n_j}) is a *subsequence* of (v_n).

Exercise. if $(v_n) \to v$ in v then any subsequence (v_{n_j}) converges to v as well.

Definition (Boundedness)**.** $A \subseteq V$ is bounded if there exists m such that $\|v\| \leq m$ for all $v \in A$.

Remark. If $\|\cdot\|$ and $\|\cdot\|'$ are Lipschitz equivalent then A is bounded with respect to $\|\cdot\|$ if and only if with respect to $\|\cdot\|'$. It follows that boundedness in $\mathbb{R}^n$ means with respect to any one of $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$.

Recall from IA Analysis I:

Theorem 3.1 (Bolzano-Weierstrass)**.** *A bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

Corollary 3.2 (Bolzano-Weierstrass for $\mathbb{R}^m$)**.** *A bounded sequence in $\mathbb{R}^m$ has a convergent subsequence.*

Proof. Inducton on m: if $m = 1$ then done by Bolzano-Weierstrass. Suppose it holds for $\mathbb{R}^{m-1}$ and let (v_n) be a bounded sequence in $\mathbb{R}^m$. Write $v_n = (v_{n,1}, \dots, v_{n,n}) = (w_n, v_{n,m})$ for some $w_n \in \mathbb{R}^{m-1}$. $\|w_n\|$ and $|v_{n,m}| \leq \|v_n\|$ so (v_n) is bounded implies that (w_n) and $(v_{n,m})$ are bounded. By induction (w_n) has a subsequence $(w_{n_j}) \to w \in \mathbb{R}^{m-1}$. Now consider $(v_{n_j,m})$. This is a bounded sequence in $\mathbb{R}$ so by Bolzano-Weierstrass there is a subsequence $(v_{n_{j_k},m}) \to v \in \mathbb{R}$. By Exercise $(w_{n_{j_k}}) \to w$ so

$$(v_{n_{j_k}}) = ((w_{n_{j_k}}, v_{n_{j_k},m})) \to (w, v) \in \mathbb{R}^m.$$

□

Definition (Sequential compactness)**.** $C \subseteq V$ is *sequentially compact* if any sequence (v_n) in C has a convergent subsequence $(v_{n_j}) \to v \in C$.

Remark. There is another (topological) definition of compactness using open covers. For metric spaces, in particular subspaces of normed spaces, these two are equivalent.

Example.

1. $\mathbb{R}$ is not compact as (n) has no convergent subsequence.
2. $(0, 1]$ is not compact as $(1/n) \to 0$ but $0 \notin A$.

Theorem 3.3 (Heine-Borel)**.** *$A \subseteq \mathbb{R}^m$ is compact if and only if A is closed and bounded.*

Proof.

- $\Leftarrow$: Suppose A is closed and bounded. Given a sequence (v_n) with $v_n \in A$, must find a convergent subsequence. Since A is bounded, (v_n) is bounded so by Bolzano-Weierstrass there is a convergent subsequence $(v_{n_j}) \to v \in \mathbb{R}^m$. As A is closed and $v_{n_j} \in A$, $(v_{n_j}) \to v$ implies that $v \in A$.
- $\Rightarrow$: If A is not closed or not bounded, we will find a sequence (v_n) with $v_n \in A$ with no convergent subsequence:
  - if A is not closed, there is a sequence $(v_n) \to v$ with $v_n \in A$ but $v \notin A$. Suppose $(v_{n_j}) \to w$ is a convergent subsequence. Then by Exercise $(v_{n_j}) \to v$. By uniqueness of limits $v = w \notin A$.
  - if A is not bounded, then for each $n > 0$ there exists $v_n \in A$ with $\|v_n\| \geq n$. Consider the sequence (v_n). Suppose there is a subsequence $(v_{n_j}) \to v$. Then we can find J such that $\|v_{n_j} - v\| < 1$ for all $j \geq J$. Pick $K \geq \max(J, \|v\| + 1)$. Then for $j \geq K$,

    $$\|v_{n_j}\| \leq \|v_{n_j} - v\| + \|v\| \leq 1 + \|v\| \leq K \leq n_j$$

    since $n_j \geq j \geq K$. So $\|v_{n_j}\| < n_j$ for $j \geq K$, contradiction.

□

Example. $(V, \|\cdot\|) = (C[0,1], \|\cdot\|_\infty)$. Consider $f_n(x) = \begin{cases} 1 - nx & x \in [0, 1/n] \\ 0 & x \geq 1/n \end{cases}$. Note if $f(x) = \begin{cases} 1 & x = 0 \\ 0 & x > 0 \end{cases}$ then $(f_n) \to f$ pointwise. Claim (f_n) has no convergent subsequence with respect to $\|\cdot\|_\infty$.

Proof. Suppose $(f_{n_j}) \to g$ uniformly. Then $(f_{n_j}) \to g$ pointwise. But we know from Exercise $(f_{n_j}) \to f$ pointwise so $f = g$. But f_n is continuous so g is continuous. Contradition. □

Note. $f_n \in \overline{B}_1(0) \subseteq V$.

Corollary 3.4. *$\overline{B}_1(0)$ is closed and bounded in $(C[0,1], \|\cdot\|_\infty)$ but is not compact.*

Proposition 3.5 (Continuous image of compact set)**.** *Suppose $C \subseteq V$ is compact and $f : C \to W$ is continuous then $f(C)$ is also compact.*

Proof. Suppose (w_n) is a sequence in $f(C)$. Pick $v_n \in C$ with $f(v_n) = w_n$. C is compact so (v_n) has a convergent subsequence $(v_{n_j}) \to v \in C$. f is continuous so

$$(w_{n_j}) = (f(v_{n_j})) \to f(v) \in f(C).$$

□

Joke. `There is no joke today.`

3.1.1 Application I: Maximum Value Theorem

Lemma 3.6. *If* $\emptyset \neq A \subseteq \mathbb{R}$ *is compact then* $\sup A \in A$.

Proof. A is closed and bounded so $\alpha = \sup A \in \mathbb{R}$. For each $n > 0$, exists $a_n \in A$ such that $\alpha - 1/n \leq a_n \leq \alpha$. Then $(a_n) \to \alpha$. Since A is closed and $a_n \in A$, $\alpha \in A$ as well. □

Theorem 3.7. *Suppose* $f : C \to \mathbb{R}$ *is continuous and* C *is compact and nonempty. Then exists* $v \in C$ *such that* $f(v) \geq f(v')$ *for all* $v' \in C$.

Proof. $f(C)$ is compact and nonempty by the Proposition, so Lemma implies that $\alpha = \sup f(C)$ exists and $\alpha \in f(C)$. Pick $v \in C$ with $f(v) = \alpha$. If $v' \in C$ then $f(v') \in f(C)$ so $f(v') \leq \alpha = f(v)$. □

Corollary 3.8. *Let* f *and* C *be as above. Then there exists* $v_{\_} \in C$ *with* $f(v_{\_}) \leq f(v')$ *for all* $v' \in C$.

Proof. Apply Theorem to $-f$. □

3.1.2 Application II: Equivalence of Norms on $\mathbb{R}^n$

Let $\|\cdot\|$ be some norm on $\mathbb{R}^m$.

Lemma 3.9. *The map* $\mathrm{id} : (\mathbb{R}^m, \|\cdot\|_1) \to (\mathbb{R}^m, \|\cdot\|)$ *is continuous.*

Proof. By the criterion in Proposition 1.7 it suffices to show that there is a constant C such that $\|v\| \leq C\|v\|_1$ for all $v \in \mathbb{R}^m$. Let $v = (v_1, \dots, v_m) = \sum_{i=1}^m v_i e_i$ where e_i is the standard basis vector. Take $C = \max_{1\leq i\leq m}\|e_i\|$. Then

$$\|v\| \leq \sum_{i=1}^{m} \|v_i e_i\| = \sum_{i=1}^{m} |v_i|\|e_i\| \leq C \sum_{i=1}^{m} |v_i| = C\|v\|_1$$

□

Corollary 3.10. *The map* $f : (\mathbb{R}^m, \|\cdot\|_1) \to (\mathbb{R}, |\cdot|)$ *given by* $f(v) = \|v\|$ *is continuous.*

Proof. $f = g \circ \mathrm{id}$ where g is the continuous map from $\mathbb{R}^m$ to $\mathbb{R}$ in Example 5 on page 4. □

Recall that two norms $\|\cdot\|$ and $\|\cdot\|'$ on V are Lipschitz equivalent if there exists C such that

$$\frac{1}{C}\|v\| \leq \|v\|' \leq C\|v\|$$

for all $v \in V$.

Remark. This is trivially true if $v = 0$ so suffices to check for $v \neq 0$.

Theorem 3.11. *If $\|\cdot\|$ is a norm on $\mathbb{R}^m$ then it is Lipschitz equivalent to $\|\cdot\|_1$.*

Proof. Let $S = \{v \in \mathbb{R}^m : \|v\|_1 = 1\}$. Claim S is compact with respect to $\|\cdot\|_1$: S is clearly bounded. Consider $g : (\mathbb{R}^m, \|\cdot\|_1) \to (\mathbb{R}, |\cdot|), v \mapsto \|v\|_1$. g is continuous and $S = g^{-1}(\{1\})$. As $\{1\} \subset \mathbb{R}$ is closed S is closed. So by Heine-Borel S is compact.

By Corollary $f : (S, \|\cdot\|_1) \to (\mathbb{R}, |\cdot|)$ given by $f(v) = \|v\|$ is continuous. By the Maximum Value Theorem there exists $v_\pm \in S$ with

$$f(v_-) \leq f(v') \leq f(v_+)$$

for all $v' \in S$. Let $C_\pm = f(v_\pm)$. Notice that

$$v_\pm \in S \Rightarrow \|v_\pm\|_1 = 1 \Rightarrow v_\pm \neq 0 \Rightarrow C_- = \|v_\pm\| \neq 0.$$

Let $C = \max(C_+, 1/C_-)$. Then

$$\frac{1}{C} \leq C_- \leq f(v) = \|v\| \leq C_+ \leq C$$

for all $v \in S$.

Finally, if $v \in \mathbb{R}^m \setminus \{0\}$ then $v/\|v\|_1 \in S$ so

$$\frac{1}{C} \leq \left\| \frac{v}{\|v\|_1} \right\| \leq C$$

and Lipschitz equivalence condition follows. □

Corollary 3.12. *Any two norms on $\mathbb{R}^m$ are Lipschitz equivalent.*

Proof. Lipschitz equivalence is an equivalence relation. □

3.2 Completeness

Let V be a normed vector space.

Definition (Cauchy sequence)**.** A sequence (v_n) in V is *Cauchy* if for any $\varepsilon > 0$, there exists N such that for all $n, m \geq N$, $\|v_n - v_m\| < \varepsilon$.

Example.

1. If $(v_n) \to v$ then (v_n) is Cauchy.

   Proof. Given $\varepsilon > 0$, pick N such that $\|v_n - v\| < \varepsilon/2$ for $n \geq N$. Then for $n, m \geq N$,
   $$\|v_n - v_m\| \leq \|v_n - v\| + \|v - v_m\| < \varepsilon.$$
   □

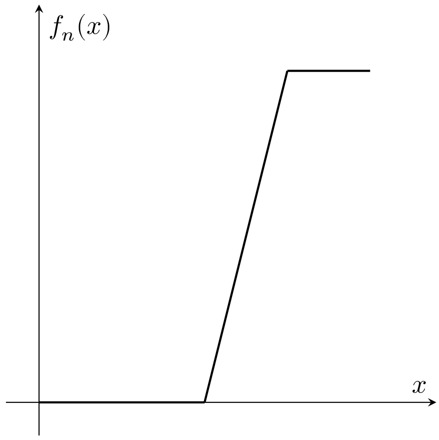
2. If $(\sum_{i=1}^n 1/i)$ is not Cauchy since for any fixed N, $\sum_{i=N}^m 1/i \to \infty$ as $m \to \infty$.

Informally, a Cauchy sequence *wants* to converge: given $\varepsilon > 0$, pick N such that $\|v_n - v_m\| < \varepsilon$ for all $n, m \geq N$. Then $\|v_n - v_N\| < \varepsilon$ for all $n \geq N$ so $v_n \in B_\varepsilon(v_N)$.

But there may *not* be anything for it to converge to!

Example. $V = C[0,1]$ with $\|\cdot\|_1$. Let $f_n(x) = \begin{cases} 0 & x \in [0, 1/2] \\ n(x - 1/2) & x \in [1/2, 1/2 + 1/n] \\ 1 & x \geq 1/2 + 1/n \end{cases}$

Then (f_n) is Cauchy but does not converge to any $f \in C[0,1]$.



Example ((Not quite an) Example)**.** Let $(V, \|\cdot\|) = (\mathbb{Q}, \|\cdot\|)$, let v_n be the first n-decimal place expansion of π. Then (v_n) is Cauchy but does not converge to any $v \in \mathbb{Q}$.

Definition (Completeness)**.** A normed vector space $(V, \|\cdot\|)$ is *complete* if whenever (v_n) is a Cauchy sequence in V, there exists $v \in V$ such that $(v_n) \to v$.

Example. $(C[0,1], \|\cdot\|_1)$ is not complete.

Theorem 3.13 (Completeness of Euclidean Space)**.** *$\mathbb{R}^m$ is complete.*

The proof uses two lemmas about Cauchy sequences: suppose V is a normed vector space and (v_n) is Cauchy, then

Lemma 3.14 (Boundedness of Cauchy sequence)**.** *(v_n) is bounded.*

Proof. Example sheet. □

Lemma 3.15. *If there exists a subsequence $(v_{n_j}) \to v \in V$ then $(v_n) \to v$.*

Proof. Given $\varepsilon > 0$, pick N such that for all $n, m \geq N$, $\|v_n - v_m\| < \varepsilon/2$. Since $(v_{n_j}) \to v$, exists J such that whenever $j \geq J$, $\|v_{n_j} - v\| < \varepsilon/2$. Pick $j \geq J$ such that $n_j \geq N$. Then for all $n \geq N$,

$$\|v_n - v\| \leq \|v_n - v_{n_j}\| + \|v_{n_j} - v\| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

□

Proof of Theorem 3.13. Suppose (v_n) is a Cauchy sequence in $\mathbb{R}^n$. By Lemma 3.14, (v_n) is bounded. By Bolzano-Weierstrass, it has a convergent subsequence $(v_{n_j}) \to v \in \mathbb{R}^n$. By Lemma 3.15 $(v_n) \to v \in \mathbb{R}^n$. So any Cauchy sequence in $\mathbb{R}^n$ converges to $v \in \mathbb{R}^n$. $\mathbb{R}^n$ is complete. □

We saw before that $(C[0,1], \|\cdot\|_\infty)$ does not have Bolzano-Weierstrass property. Nevertheless, it is complete:

> **Theorem 3.16.** *Let $[a, b]$ be a bounded interval in $\mathbb{R}$, then $C[a, b]$ is complete with respect to $\|\cdot\|_\infty$.*

Proof. Suppose (f_n) is Cauchy in $C[a, b]$. We must find $f \in C[a, b]$ such that $(f_n) \to f$ uniformly on $[a, b]$.

Give $\varepsilon > 0$, pick N such that whenever $n, m \geq N$,

$$|f_n(x) - f_m(x)| \leq \max_{x \in [a,b]} |f_n(x) - f_m(x)| = \|f_n - f_m\|_\infty < \varepsilon/2 \qquad (*)$$

i.e. for all $x \in [a, b]$, the sequence $(f_n(x))$ is Cauchy in $\mathbb{R}$. Since $\mathbb{R}$ is complete, $\lim_{n\to\infty} f_n(x)$ exists. Now define $f(x) = \lim_{n\to\infty} f_n(x)$. Claim $(f_n) \to f$ uniformly on $[a, b]$: given $\varepsilon > 0$, choose N as in ($*$). Given x, choose M such that $|f_n(x) - f(x)| < \varepsilon/2$ for $n \geq M$. Let $m' = \max\{M, N\}$. Then for $n \geq N$,

$$|f_n(x) - f(x)| \leq |f_n(x) - f_{m'}(x)| + |f_{m'}(x) - f(x)| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Since the uniform limit of continuous functions is continuous, $(f_n) \to f \in C[a, b]$. $(C[a, b], \|\cdot\|_\infty)$ is complete. □

> **Definition** (Normed subspace)**.** Let $(V, \|\cdot\|)$ and $(V', \|\cdot\|')$ be two normed vector spaces. We say V is a *normed vector subspace* of V' if
>
> 1. $V \leq V'$ as a vector space,
> 2. $\|v\| = \|v\|'$ for all $v \in V$.

Remark.

1. If $V \subseteq V'$, V is complete then V is a closed subset of V'.

   Proof. Suppose (v_n) is a sequence in V and $(v_n) \to v' \in V'$. Then (v_n) is Cauchy in V so $(v_n) \to v \in V$. By uniqueness of limits $v = v' \in V$ so $V \subseteq V$ is closed. □

2. If $(V, \|\cdot\|)$ is a normed space then there exists a complete normed space $(\overline{V}, \|\cdot\|)$ which contains V as a dense subspace (dense means that any $v \in V$ is a limit of sequence in V). This $\overline{V}$ is the *completion* of V and is unique up to isomorphism of normed space.

Example.

1. $\mathbb{R}$ is the completion of $\mathbb{Q}$ with $|\cdot|$.

2. The completion of $(C[0,1], \|\cdot\|_1)$ is the space of Lebesgue integrable functions on $[0,1]$.

3.3 Uniform Continuity

Suppose V and W are normed spaces and $X \subseteq V$,

Definition (Uniform continuity)**.** $f : X \to W$ is *uniformly continuous* if

$$\forall \varepsilon > 0, \exists \delta > 0 \, \forall v, w \in X, \, \|v - w\| < \delta \Rightarrow \|f(v) - f(w)\| < \varepsilon.$$

Note.

- f is continuous means that if I give you $\varepsilon > 0$ and $v \in X$, you find $\delta > 0$ such that $\|f(v) - f(w)\| < \varepsilon$ whenever $\|v - w\| < \delta$.

- f is uniformly means that if I give you $\varepsilon > 0$, you find $\delta > 0$ which works for all $v \in X$.

Clearly uniform convergence imples convergence but the converse is false.

Example.

1. $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ is not uniformly continuous since $f(x+\delta/2) - f(x) = x\delta + \delta^2/4 > 1$ if $x > 1/\delta$.

2. Suppose $a, b \in \mathbb{R}$, $f : (a,b) \to \mathbb{R}$ is uniformly continuous, then f is bounded on (a,b):

   Proof. Suppose $|f(x) - f(y)| < 1$ whenever $|x - y| < 2\delta$, then

   $$|f(x) - f(y)| \leq \left|f(x) - f(\frac{x+Y}{2})\right| + \left|f(\frac{x+y}{2}) - f(y)\right| < 1 + 1 = 2$$

   Similarly $|x - y| < n\delta$ implies that $|f(x) - f(y)| < n$ so f is bounded on (a,b). □

3. $f : (0,1) \to \mathbb{R}, f(x) = 1/x$ is continuous but not bounded so not uniformly continuous.

4. $f : (0,1) \to \mathbb{R}, f(x) = \sin(1/x)$ is not unifornly continuous since for any $\delta > 0$, there exists $x, x' \in (0, \delta)$ with $f(x) = 1, f(x') = -1$.

Theorem 3.17. *If $C \subseteq V$ is compact and $f : C \to W$ is continuous then f is uniformly continuous.*

Proof. By contradiction. Suppose f satisfies the hypotheses but is not uniformly continuous, i.e. $\exists a > 0$ such that there is no $\delta > 0$ with $\|f(v) - f(w)\| < a$ whenever $\|v - w\| < \delta$. Thus $\forall n > 0$, $\exists v_n, w_n$ such that $\|v_n - w_n\| < 1/n$ but $\|f(v_n) - f(w_n)\| \geq a$. Since C is compact, (v_n) has a convergent subsequence $(v_{n_j}) \to v \in C$. Claim $(w_{n_j}) \to v$:

Proof. $(v_{n_j}) \to v$ so $\lim_{j\to\infty}\|v_{n_j} - v\| = 0$. $\|v_n - w_n\| < 1/n$ so $\lim_{j\to\infty}\|v_{n_j} - w_{n_j}\| = 0$. Then

$$0 \leq \|w_{n_j} - v\| \leq \|w_{n_j} - v_{n_j}\| + \|v_{n_j} - v\|$$

so by squeeze rule $\lim_{j\to\infty}\|w_{n_j} - v\| = 0$. □

Now we claim $\lim_{j\to\infty}\|f(v_{n_j}) - f(w_{n_j})\| = 0$:

Proof. $(v_{n_j}) \to v$, f is continuous so $(f(v_{n_j})) \to f(v)$. Similarly $(w_{n_j}) \to v$, so $(f(w_{n_j})) \to f(v)$ as well. Then

$$0 \leq \|f(v_{n_j}) - f(w_{n_j})\| \leq \|f(v_{n_j}) - f(v)\| + \|f(w_{n_j}) - f(v)\|$$

and the result follows by squeeze rule. □

Now choose J such that $\|f(v_{n_j}) - f(w_{n_j})\| < a$ whenver $j \geq J$. Then let $N = n_J$, $\|f(v_n) - f(w_n)\| < a$ whenever $n \geq N$. However, this contradicts the fact that $\|f(v_n) - f(w_n)\| \geq a > 0$ for all n. □

3.3.1 Application: Riemann Integral

Recall from IA Analysis I: $g : [a, b] \to \mathbb{R}$ is *piecewise constant* if $\exists a = a_0 < a_1 < \cdots < a_n = b$ such that $g(x) = c_i$ for all $x \in (a_{i-1}, a_i)$. Let

$$P[a, b] = \{g : [a, b] \to \mathbb{R} : g \text{ is piecewise constant}\}$$

Notation. $f \leq g$ means for all $x \in [a, b]$, $f(x) \leq g(x)$.

If $g \in P[a, b]$, define

$$I(g) = \sum_{i=1}^{n} c_i(a_i - a_{i-1})$$

to be the signed area under the graph of g.

Lemma 3.18. *If $f, g \in P[a, b]$, then*

1. *$f + \lambda g \in P[a, b]$ and $I(f + \lambda g) = I(f) + \lambda I(g)$.*
2. *If $f \geq 0$ then $I(f) \geq 0$.*
3. *If $f \geq g$ then $I(f) \geq I(g)$.*

If $f \in B[a,b]$, define

$$\mathcal{U}(f) = \{g \in P[a,b] : g \geq f\}$$
$$\mathcal{L}(f) = \{g \in P[a,b] : g \leq f\}$$

Now define

$$U(f) = \{I(g) : g \in \mathcal{U}(f)\}$$
$$L(f) = \{I(g) : g \in \mathcal{L}(f)\}$$

so given $g_+ \in \mathcal{U}(f), g_- \in \mathcal{L}(f)$, $g_- \leq f \leq g_+$ so $I(g_-) \leq I(g_+)$. Now define

$$u(f) = \inf U(f)$$
$$l(f) = \sup L(f)$$

Definition (Riemann integral)**.** f is *Riemann integrable* if $u(f) = l(f)$. In this case define

$$\int_a^b f(x)dx = u(f) = l(f).$$

Theorem 3.19. *If $f \in C[a,b]$ then f is Riemann integrable.*

Proof. By Maximum Value Theorem $f \in B[a,b]$. Claim given $\varepsilon > 0$, $\exists g_+ \in \mathcal{U}(f), g_- \in \mathcal{L}(f)$ such that $I(g_+) - I(g_-) < \varepsilon$:

Proof. Since $[a,b]$ is compact, f is uniformly continuous. Choose $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon/(b-a)$ whenever $|x - y| < \delta$.

Choose $a = a_0 < \cdots < a_n = b$ such that $a_i - a_{i-1} < \delta$ for all i. Define

$$c_i^+ = \max_{x \in [a_{i-1}, a_i]} f(x)$$
$$c_i^- = \min_{x \in [a_{i-1}, a_i]} f(x)$$

If $x, y \in [a_{i-1}, a_i]$, $|x - y| < \delta$ so $c_i^+ - c_i^- < \varepsilon/(a_{i-1}, a_i)$. Now take $g_\pm(x) = c_i^\pm$ if $x \in [a_{i-1}, a_i]$. By construction for $x \in [a_{i-1}, a_i]$,

$$g_+(x) \geq c_i^+ = \max_{x \in [a_{i-1}, a_i]} f(x) \geq f(x)$$

so $g_+ \geq f$. Similar for g_-. It follows that $I(g_+) - I(g_-) < \varepsilon$. □

With this result, the proof of the theorem is almost apparent: since

$$I(g_-) \leq l(f) \leq u(f) \leq I(g_+),$$

$0 \leq u(f) - l(f) \leq I(g_+) - I(g_-) \leq \varepsilon$ for all $\varepsilon > 0$. Thus $u(f) = l(f)$ and f is Riemann integrable. □

4 Differential Calculus on $\mathbb{R}^n$

4.1 Derivative, First Attempt

It is natural to ask what is the generalisation of differentiation on $\mathbb{R}$ to higher dimension. More specifically,

Question.

1. What does it mean for $F : \mathbb{R}^2 \to \mathbb{R}$ to be differentiable?
2. If F is differentiable, what is its derivative?

Recall from multivariate calculus the partial derivative

$$\begin{aligned}\frac{\partial F}{\partial x_1}\Big|_{(a_1,a_2)} &= \mathcal{D}_1 F|_{(a_1,a_2)} \\ &= \left(\frac{d}{dt}F(a_1+t,a_2)\right)\Big|_{t=0} \\ &= \lim_{h\to 0}\frac{F(a_1+h,a_2)-F(a_1,a_2)}{h}\end{aligned}$$

Similar for $\frac{\partial F}{\partial x_2}$.

Note. Just becuase $\frac{\partial F}{\partial x_i}$ exists does not mean that F is differentiable at $a = (a_1, a_2)$:

Example.

- Let
$$F(x_1,x_2) = \begin{cases} 0 & \text{if } x_1 = 0 \text{ or } x_2 = 0 \\ 1 & \text{otherwise} \end{cases}$$
Then $\frac{\partial F}{\partial x_1} = \frac{\partial F}{\partial x_2} = 0$ at $(0,0)$ but F is not even continuous. Heuristically, a good definition of differentiability should not allow this as this is not compatible with the definition of differentiability on $\mathbb{R}$.
- If one insists that differentiability does not imply continuity (by whatever means) and declare F as differentiable whenever $\frac{\partial F}{\partial x_i}$ exists, then consider $G(x_1,x_2) = F(x_1+x_2, x_1-x_2)$, whose derivative $\frac{\partial G}{\partial x_1}$ does not even exist. This reveals a bigger problem with this definition of "differentiability": the composition of differentiable functions is not differentiable!

The failure in the above attempt to define differentiability is because the existence of partial derivatives is not strong enough as $\frac{\partial F}{\partial x_i}$ merely tells us about the behaviour of when F is restricted to coordinate axes.

To answer Q1, recall that $f : \mathbb{R} \to \mathbb{R}$, $f'(a) = c$ means that

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

Graphically, $y = T(x)$ is the tangent line to $y = f(x)$ at $(a, f(a))$ where $T(x) = f(a) + c(x - a)$. In other words,

$$\begin{aligned}\lim_{h\to 0} \frac{f(a+h) - T(a+h)}{h} &= \lim_{h\to 0}\left(\frac{f(a+h) - f(a)}{h} - \frac{ch}{h}\right)\\ &= f'(a) - c\\ &= c - c\\ &= 0\end{aligned}$$

so f is differentiable at a if and only if f is well-approximately by its tangent line near a in the sense that

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

Answer (Answer to Q1)**.** $F : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at a if it is well-approximated by its tangent plane $z = T(x_1, x_2)$ at $(a, F(a))$ in the sense that

$$\lim_{h\to 0} \frac{F(a+h) - T(a+h)}{\|h\|} = 0 \tag{$*$}$$

where the limit will be made rigorous later.

Now we have to ask: what is the tangent plane? It is the equation

$$\begin{aligned}z &= T(x_1, x_2)\\ &= F(a_1, a_2) + c_1(x_1 - a_1) + c_2(x_2 - a_2)\end{aligned}$$

for some c_1 and c_2 so

$$T(a+h) = F(a_1, a_2) + c_1h_1 + c_2h_2 = F(a_1, a_2) + L(h)$$

where $h = (h_1, h_2)$ and $L : \mathbb{R}^2 \to \mathbb{R}$ is a linear map.

Answer (Answer to Q2)**.** The derivative $\mathcal{D}F|_a$ is the linear map L.

What should c_1 and c_2 that we used to parameterise T be? Equation ($*$) says that

$$\lim_{h\to 0} \frac{F(a+h) - F(a_1, a_2) - c_1h_1 - c_2h_2}{\|h\|} = 0$$

Taking $h = (h_1, 0)$ gives

$$\begin{aligned}0 &= \lim_{h\to 0} \frac{F(a+h) - F(a_1, a_2) - c_1h_1}{h_1}\\ &= \lim_{h_1\to 0}\left(\frac{F(a_1+h, a_2) - F(a_1, a_2)}{h_1} - c_1\right)\\ &= \frac{\partial F}{\partial x}\Big|_{(a_1,a_2)}\end{aligned}$$

so $c_1 = \frac{\partial F}{\partial x_1}|_{(a_1,a_2)}$. Similar for c_2. So

$$\begin{aligned}\mathcal{D}F|_a &= \frac{\partial F}{\partial x_1}h_1 + \frac{\partial F}{\partial x_2}h_2\\ &= \begin{pmatrix}\frac{\partial F}{\partial x_1} & \frac{\partial F}{\partial x_2}\end{pmatrix}\begin{pmatrix}h_1\\ h_2\end{pmatrix}\end{aligned}$$

i.e. as a map $\mathcal{D}F|_a$ is given by the matrix

$$(\mathcal{D}_1 F|_a \quad \mathcal{D}_2 F|_a).$$

Example. Let F be the same as in the previous example. $\frac{\partial F}{\partial x_i} = 0$ so if there were a tangent plane at 0 it would be given by

$$z = T(x_1, x_2) = 0 + 0 \cdot x_1 + 0 \cdot x_2 = 0$$

which is, to say the least, a terrible approximation of F.

More generally,

Definition (Differentiability)**.** Given $F : \mathbb{R}^n \to \mathbb{R}^m$, F is *differentiable* at a if there is a linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h\to 0} \frac{F(a+h) - F(a) - L(h)}{\|h\|} = 0. \tag{$*$}$$

If so, say $\mathcal{D}F|_a = L$ is the *derivative* of F at a.

If $F = (F_1, \dots, F_m)$ is differentiable at a, the derivative is given by the matrix

$$\mathcal{D}F|_a = \begin{pmatrix} \mathcal{D}_1 F_1|_a & \cdots & \mathcal{D}_n F_1|_a \\ \vdots & \ddots & \vdots \\ \mathcal{D}_1 F_m|_a & \cdots & \mathcal{D}_n F_m|_a \end{pmatrix}$$

Example. If $\gamma : \mathbb{R} \to \mathbb{R}^n$,

$$\mathcal{D}\gamma|_t = \begin{pmatrix} \gamma_1'(t) \\ \vdots \\ \gamma_m'(t) \end{pmatrix}$$

Recall the chain rule: given $F : \mathbb{R}^m \to \mathbb{R}$,

$$\begin{aligned} \frac{d}{dt} F(\gamma(t)) &= \sum_{i=1}^m \frac{\partial F}{\partial x_i}\Big|_{\gamma(t)} \gamma_i'(t) \\ &= \begin{pmatrix} \frac{\partial F}{\partial x_1} \cdots \frac{\partial F}{\partial x_m} \end{pmatrix} \begin{pmatrix} \gamma_1'(t) \\ \vdots \\ \gamma_m'(t) \end{pmatrix} \\ &= [C] \end{aligned}$$

As matrix multiplication is composition of linear maps

$$\mathcal{D}(F \circ \gamma)|_t = \mathcal{D}F|_{\gamma(t)} \circ \mathcal{D}\gamma|_t.$$

Similarly for $F : \mathbb{R}^n \to \mathbb{R}^m$, $G : \mathbb{R}^m \to \mathbb{R}^l$ differentiable, we will show that

$$\mathcal{D}(F \circ G) = \mathcal{D}F \circ \mathcal{D}G.$$

4.2 Limits

Suppose V and W are normed spaces, $U \subseteq V$ open, and $v_0 \in U$, $f : U \setminus \{v_0\} \to W$.

Definition (Limit)**.** We say the *limit* of f at v_0 is w, denoted by

$$\lim_{v\to v_0} f(v) = w,$$

if for every $\varepsilon > 0$, exists $\delta > 0$ such that $\|f(v) - w\| < \varepsilon$ whenever $0 < \|v - v_0\| < \delta$ and $v \in U$.

Note.

1. $\lim_{v\to v_0} f(v) = w$ if and only if $\lim_{v\to v_0}\|f(v) - w\| = 0$.
2. $\lim_{v\to v_0} f(v) = w$ if and only if for every sequence (x_i) in $U \setminus \{v_0\}$ which converges to v_0, the sequence $(f(x_i)) \to w$.
3. If $\|\cdot\|_V$ and $\|\cdot\|'_V$ are Lipschitz equivalent on V and $\|\cdot\|_W$ and $\|\cdot\|'_W$ are Lipschitz equivalent on W then $\lim_{v\to v_0} f(v) = w$ with respect to $\|\cdot\|_V$ and $\|\cdot\|_W$ if and only if with respect to $\|\cdot\|'_V$ and $\|\cdot\|'_W$. In particular if $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ then the limit is unambiguously defined without specifying any norms.

Proposition 4.1 (Properties of limits)**.**

1. *$\lim_{v\to v_0}(f(v) + \lambda g(v)) = \lim_{v\to v_0} f(v) + \lambda \lim_{v\to v_0} g(v)$.*
2. *If $\lambda : U \setminus \{v_0\} \to \mathbb{R}$ then*

   $$\lim_{v\to v_0} \lambda(v) f(v) = \lim_{v\to v_0} \lambda(v) \lim_{v\to v_0} f(v)$$

   if both limits on RHS exist.
3. *Squeeze rule: if $\|f(v) - w\| \le \|g(v) - w\|$ and $\lim_{v\to w_o} g(v) = w$ then $\lim_{v\to v_0} f(v) = w$.*
4. *If $f : U \setminus \{v_0\} \to \mathbb{R}^m, f(v) = (f_1(v), \dots, f_m(v))$, $w = (w_1, \dots, w_m)$ then $\lim_{v\to v_0} f(v) = w$ if and only if $\lim_{v\to v_0} f_i(v) = w_i$ for all $1 \le i \le m$.*
5. *If $\lim_{v\to 0} f(v) = w$ and $x \in U \setminus \{0\}$ then $\lim_{t\to 0} f(tx) = w$.*

Proof. The first three are the same as the one dimensional case in IA Analysis I.

For (4), $(f(x)) \to w$ if and only if $(f_i(x_n)) \to w_i$ for all i so the result follows from sequential characterisation of convergence.

For (5), given $\varepsilon > 0$, exists $\delta > 0$ such that $\|f(v) - w\| < \varepsilon$ whenever $0 < \|v\| < \delta$. Then for $0 < |t| < \delta/\|x\|$, $0 < \|tx\| < \delta$ so $\|f(tx) - w\| < \varepsilon$ so $\lim_{t\to 0} f(tx) = w$. □

Note. $\lim_{t\to 0} f(tx) = w$ for all $x \in U \setminus \{0\}$ does not imply that $\lim_{v\to 0} f(v) = w$. This is similar to the case the existence of $\frac{\partial f}{\partial x_i}$ for all i does not imply differentiability of f.

Example. Let

$$\begin{aligned} f &: \mathbb{R}^2 \setminus \{0\} \to \mathbb{R} \\ f(x_1, x_2) &= \begin{cases} \frac{x_1^2}{x_2^2}(x_1^2 + x_2^2) & x_2 \neq 0 \\ 0 & x_2 = 0 \end{cases} \end{aligned}$$

Then if $x = (x_1, x_2)$ with $x_2 \neq 0$,

$$\lim_{t\to 0} f(tx) = \lim_{t\to 0} \frac{x_1^2}{x_2^2}(x_1^2 + x_2^2)t^2 = 0$$

If $x = (x_1, 0)$ with $x_1 \neq 0$ then $\lim_{t\to 0} f(tx) = \lim_{t\to 0} 0 = 0$. But

$$f(t, t^2) = \frac{t^2}{t^4}(t^2 + t^4) = 1 + t^2 \geq 1$$

for $t \neq 0$ so taking $t = \frac{1}{n}$, we find points $(1/n, 1/n^2)$ arbitrarily close to $(0, 0)$ with $f(1/n, 1/n^2) \geq 1$ so $\lim_{v\to 0} f(v) \neq 0$.

Definition (Continuity)**.** $f : U \to W$ is *continuous* at v_0 if

$$\lim_{v\to v_0} f(v) = f(v_0).$$

Lemma 4.2. *f is continuous at v_0 if and only if $(f(x_i)) \to f(v_0)$ whenever (x_i) is a sequence in U converging to v_0.*

Proof. Sequential characterisation of convergence. □

Recall that previously we have defined continuity of a function *on a domain*. Here we define continuity of a function *at a point*. Fortunately, these two notions of continuity agree:

Corollary 4.3. *f is continuous if and only if f is continuous at every $v \in U$.*

Proof. f is continuous at every $v \in U$ if and only if $(x_i) \to v \in U \Rightarrow (f(x_i)) \to f(v)$ if and only if f is continuous. □

4.3 Derivative, Revisited

Suppose $U \subseteq \mathbb{R}^n$ is open, $f : U \to \mathbb{R}^m$ and $a \in U$. We recall

Definition (Differentiability)**.** Given $F : \mathbb{R}^n \to \mathbb{R}^m$, F is *differentiable* at a if there is a linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h\to 0} \frac{F(a + h) - F(a) - L(h)}{\|h\|} = 0. \qquad (*)$$

If so, say $\mathcal{D}F|_a = L$ is the *derivative* of F at a.

Note.

1. Equivalently, $f(a + h) = f(a) + L(h) + \|h\|\alpha(h)$ where $\lim_{h\to 0} \alpha(h) = 0$.
2. If $(*)$ holds for two linear functions $L, L' : \mathbb{R}^n \to \mathbb{R}^m$ then $L = L'$ so $\mathcal{D}f|_a$ is well-defined.

Proof. Subtraction gives

$$\lim_{h\to 0}\frac{L(h)-L'(h)}{\|h\|}=0$$

which implies that for any $v\in\mathbb{R}^n\setminus\{0\}$,

$$0=\lim_{t\to 0^+}\frac{L(tv)-L'(tv)}{\|tv\|}=\lim_{t\to 0^+}\frac{L(v)-L'(v)}{\|v\|}=\frac{L(v)-L'(v)}{\|v\|}$$

so $0=L(v)-L'(v)=(L-L')(v)$ for all $v\neq 0$. They certainly agree at 0. □

3. The definition of derivative does not depend on the norm used.

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

Notation. $L(\mathbb{R}^n,\mathbb{R}^m)$ is the space of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$, which is isomorphic to $\mathcal{M}_{m,n}(\mathbb{R})$.

Example. If $L\in L(\mathbb{R}^n,\mathbb{R}^m), y_0\in\mathbb{R}^n, f(x)=y_0+L(x)$ then

$$\begin{aligned}f(a+h)&=y_0+L(a+h)\\&=y_0+L(a)+L(h)\\&=f(a)+L(h)\end{aligned}$$

so

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

so f is differentiable and $\mathcal{D}f|_a=L$.

Definition (Differentiability)**.** f is differentiable if f is differentiable at all $a\in U$.

There are two ways to think about derivatives:

- analyst way: $\mathcal{D}f:U\to L(\mathbb{R}^n,\mathbb{R}^m), a\mapsto\mathcal{D}f|_a$.
- geometer way: define the tangent bundle $TU=U\times\mathbb{R}^n$ if $U\subseteq\mathbb{R}^n$ is open, then

$$\begin{aligned}df:TU&\to T\mathbb{R}^m\cong\mathbb{R}^m\times\mathbb{R}^m\\(a,v)&\mapsto(f(v),\mathcal{D}f|_a(v))\end{aligned}$$

The analyst way says that given f in the previous example, the derivative $\mathcal{D}f:\mathbb{R}^n\to L(\mathbb{R}^n,\mathbb{R}^m), a\mapsto L$ is the constant map.

4.4 Directional Derivative

Let $F:U\to\mathbb{R}^n$.

Definition (Directional derivative)**.** If $v \in \mathbb{R}^n \setminus \{0\}$, the *directional derivative* of F at a in the direction v is

$$\mathcal{D}_v F|_a = \lim_{t \to 0^+} \frac{F(a+tv) - F(a)}{t}$$

if RHS exist.

Example. If e_i is the standard basis vector of $\mathbb{R}^n$ then $\mathcal{D}_{e_i} F|_a = \mathcal{D}_i F|_a$ if $\mathcal{D}_i F|_a$ exists.

Proposition 4.4. *If F is differentiable at a then*

$$\mathcal{D}_v F|_a = \mathcal{D}F|_a(v).$$

Proof. Suppose

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

Then by the property of limits

$$\begin{aligned} 0 &= \lim_{h \to 0^+} \frac{F(a+tv) - F(a) - L(tv)}{\|tv\|} \\ &= \lim_{h \to 0^+} \frac{F(a+tv) - F(a) - tL(v)}{t} \end{aligned}$$

so

$$\lim_{t \to 0^+} \frac{F(a+tv) - F(a)}{t} = L(v).$$

Thus

$$\mathcal{D}_v F|_a = \mathcal{D}F|_a(v).$$

□

Example. In the example above,

$$\mathcal{D}F|_a(e_i) = \mathcal{D}_{e_i} F|_a = \mathcal{D}_i F|_a.$$

If F is differentiable, the derivative is representated by the $1 \times n$ matrix

$$(\mathcal{D}_1 F|_a \cdots \mathcal{D}_n F|_a).$$

Corollary 4.5. *If F is differentiable at a then $D_v F|_a$ is a linear function of v.*

Example. Let

$$F(x_1, x_2) = \begin{cases} \frac{x_1^3 + x_2^3}{x_1^2 + x_2^2} & (x_1, x_2) \neq (0,0) \\ 0 & (x_1, x_2) = (0,0) \end{cases}$$

Then

$$\begin{aligned}\mathcal{D}_v F|_0 &= \lim_{t\to 0^+} \frac{F(tv) - F(0)}{t} \\ &= \lim_{t\to 0^+} \frac{t^3(v_1^3 + v_2^3)}{t^2(v_1^2 + v_2^2)} \cdot \frac{1}{t} \\ &= \frac{v_1^3 + v_2^3}{v_1^2 + v_2^2}\end{aligned}$$

so $D_v F|_0$ exists for all $v \neq 0$. But $\mathcal{D}_{(1,0)} F|_0 = \mathcal{D}_{(0,1)} F|_0 = \mathcal{D}_{(1,1)} F|_0 = 1$,

$$\mathcal{D}_{(1,0)} F|_0 + \mathcal{D}_{(0,1)} F|_0 \neq \mathcal{D}_{(1,1)} F|_0.$$

$D_v F|_0$ is not a linear function of v so F is not differentiable at 0.

Remark. If $F : U \to \mathbb{R}$, the *gradient* is

$$\nabla F|_a = (\mathcal{D}_1 F|_a \cdots \mathcal{D}_n F|_a) \in \mathbb{R}^n$$

There is a vector space isomorphism

$$\begin{aligned}\mathbb{R}^n &\to (\mathbb{R}^n)^* \\ w &\mapsto w^* = w \cdot -\end{aligned}$$

sending a vector to its dual. In particular

$$\nabla F|_a \mapsto \mathcal{D} F|_a$$

so

$$\mathcal{D}_v F|_a = \mathcal{D} F|_a(v) = \nabla F|_a \cdot v.$$

Proposition 4.6. *$f = (f_1, \dots, f_m) : U \to \mathbb{R}^m$ is differentiable at a with derivative $L = (L_1, \dots L_m)$ where $L_i \in (\mathbb{R}^n)^*$ if and only if f_i's are differentiable at a with derivative L_i for all $1 \leq i \leq m$.*

Proof. By property of limits,

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

if and only if

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

for all i. □

Corollary 4.7. *If f is differentiable at a then $\mathcal{D} f|_a \in L(\mathbb{R}^n, \mathbb{R}^m)$ is given by multiplication by the $m \times n$ matrix*

$$\mathcal{D} f|_a = \begin{pmatrix} \mathcal{D}_1 f_1|_a & \cdots & \mathcal{D}_n f_1|_a \\ \vdots & \ddots & \vdots \\ \mathcal{D}_1 f_m|_a & \cdots & \mathcal{D}_n f_m|_a \end{pmatrix}$$

Sometimes it is better to differentiate by hand:

Example. Let

$$\begin{aligned} F : \mathbb{R} \times \mathbb{R} &\to \mathbb{R} \\ (a_1, a_2) &\mapsto a_1 \cdot a_2 \end{aligned}$$

Then

$$F(a+h) = (a_1 + h_1) \cdot (a_2 + h_2) = \underbrace{a_1 \cdot a_2}_{F(a)} + \underbrace{h_1 \cdot a_2 + a_1 \cdot h_2}_{L_a(h)} + h_1 \cdot h_2$$

Observe that $L_a(h) = a_1 \cdot h_2 + a_2 \cdot h_1$ is a linear function of h. Claim $\mathcal{D}F|_a = L_a$:

Proof.

$$\lim_{h \to 0} \frac{F(a+h) - F(a) - L_a(h)}{\|h\|} = \lim_{h \to 0} \frac{h_1 \cdot h_2}{\|h\|}$$

By Cauchy-Schwarz,

$$0 \leq \frac{|h_1 \cdot h_2|}{\|h\|} \leq \frac{\|h_1\| \|h_2\|}{(\|h_1\|^2 + \|h_2\|^2)^{1/2}} = \frac{\|h_1\|}{(\|h_1\|^2 + \|h_2\|^2)^{1/2}} \cdot \|h_2\| \leq \|h_2\|$$

so by squeeze rule

$$\lim_{h \to 0} \frac{h_1 \cdot h_2}{\|h\|} = 0$$

and it follows that $\mathcal{D}F|_a = L_a$. ☐

Recall from IB Linear Algebra that given V, W vector spaces,

$$L(V, W) = \{T : V \to W : T \text{ is linear}\}.$$

Lemma 4.8. *If $L \in L(\mathbb{R}^n, \mathbb{R}^m)$ then L is continuous.*

Proof. Suppose $L(x) = (L_1(x), \dots, L_m(x))$. Write $L_i(x) = \sum_{j=1}^n a_{ij} x_j$ and let $A = (a_{ij})$ be the matrix representation of L. Since $\pi_j : x \mapsto x_j$ is continuous, L_i is a linear combination of continuous functions so continuous. All components of L are continuous and so is L. ☐

Note. The lemma assumes that the domain and codomain are finite-dimensional. The result does not hold for general normed spaces. For example,

$$\begin{aligned} (C[0,1], \|\cdot\|_1) &\to \mathbb{R} \\ f &\mapsto f(0) \end{aligned}$$

is not continuous.

Proposition 4.9. *If $f : U \to \mathbb{R}^m$ is differentiable at a then f is continuous at a.*

Proof. Since f is differentiable at a, we may write

$$f(a+h) = f(a) + L(h) + \|h\|\alpha(h)$$

where $L(h) = \mathcal{D}f|_a \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $\lim_{h\to 0} \alpha(h) = 0$.

Since L is continuous,

$$\begin{aligned}\lim_{h\to 0} f(a+h) &= \lim_{h\to 0} f(a) + \lim_{h\to 0} L(h) + \lim_{h\to 0} \|h\|\alpha(h) \\ &= f(a) + L(0) + \lim_{h\to 0}\|h\| \cdot \lim_{h\to 0} \alpha(h) \\ &= f(a) + 0 + 0 \\ &= f(a)\end{aligned}$$

Thus f is continuous at a. □

Joke. `Unfortunately the author missed this lecture and could not reproduce the joke in its original form.`[1]

Definition (C^1 space)**.** $F : U \to \mathbb{R}$ is C^1 if all the partial derivatives $\mathcal{D}_i F$ exist and are continuous on U. Denote

$$C^1(U) = \{F : U \to \mathbb{R} : F \text{ is } C^1\}.$$

Theorem 4.10. *If F is C^1 then F is differentiable on U.*

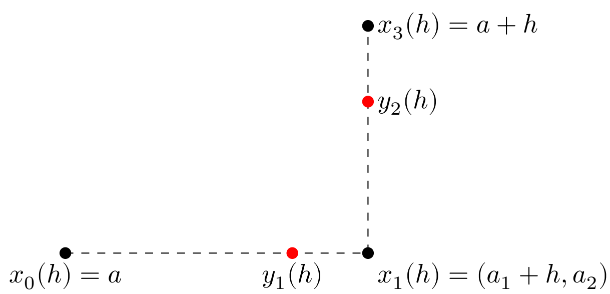
Joke. Lecturer: I'm not usually superstitious, but last time I lectured this course I messed up proving this theorem. The next day Trump was elected.

Proof. For the ease of notation we show the $n = 2$ case here. The proof for $n \geq 3$ is entirely analogous.

We need to show that $F(a+h) - F(a)$ is well-approximated by

$$L(h) = \mathcal{D}_1 F|_a \cdot h_1 + \mathcal{D}_2 F|_a \cdot h_2.$$

From now on assume $\|h\| < \varepsilon$ where $B_\varepsilon(a) \subseteq U$.



[1] The author would upon request produce a list of references from which an interested individual could hear the original joke.

The black dots in the above illustraction show that we can do a two step approximation

$$F(a+h) - F(a) = F(x_2) - F(x_1) + F(x_1) - F(x_0)$$

Recall from IA Analysis I

Theorem 4.11 (Mean Value Theorem)**.** *If $f : [a, b] \to \mathbb{R}$ is differentiable then there exists $c \in [a, b]$ such that*

$$f(b) - f(a) = (b - a)f'(c).$$

Now apply this to $f(t) = F(a_1 + t, a_2)$, $f'(t) = \mathcal{D}_1 F|_{(a_1+t,a_2)}$: there exists $y_1(h) = (a_1 + c_1(h), a_2)$ such that $c_1(h) \in [0, h_1]$ and

$$F(x_1) - F(x_0) = h_1 \cdot \mathcal{D}_1 F|_{y_1(h)}.$$

Similarly apply Mean Value Theorem to $g(t) = F(a_1 + h_1, a_2 + t)$: there exists $y_2(h) = (a_1 + h_1, a_2 + c_2(h))$ such that $c_2(h) \in [0, h_2]$ and

$$F(x_2) - F(x_1) = h_2 \cdot \mathcal{D}_2 F|_{y_2(h)}.$$

Notice that since $y_1(h) = (a_1 + c_1(h), a_2)$ where $c_1(h) \in [0, h_1]$, we have

$$\|y_1(h) - a\| \leq \|h_1\| \leq \|h\|.$$

Similarly $\|y_2(h) - a\| \leq \|h\|$. Claim that

$$\lim_{h \to 0} \mathcal{D}_1 F|_{y_1(h)} - \mathcal{D}_1 F|_a = 0$$

Proof. By hypothesis $\mathcal{D}_1 F$ is continuous so given $\varepsilon > 0$, there exists $\delta > 0$ such that $|\mathcal{D}_1 F|_z - \mathcal{D}_1 F|_a| < \varepsilon$ whenever $\|z - a\| < \delta$. Then if $\|h\| < \delta$, $\|y_1(h) - a\| \leq \|h\| < \delta$. Then

$$|\mathcal{D}_1 F|_{y_1(h)} - \mathcal{D}_1 F|_a| < \varepsilon.$$

□

Similarly we see that

$$\lim_{h \to 0} \mathcal{D}_2 F|_{y_1(h)} - \mathcal{D}_2 F|_a = 0.$$

Now $\frac{|h_1|}{\|h\|} \leq 1$ so by squeeze rule

$$\lim_{h \to 0} \frac{(\mathcal{D}_1 F|_{y_1(h)} - \mathcal{D}_1 F|_a)h}{\|h\|} = 0$$

and similary for $\mathcal{D}_2$. Then

$$\begin{aligned} F(a+h) - F(a) &= F(x_2) - F(x_1) + F(x_1) - F(x_0) \\ &= h_1 \cdot \mathcal{D}_1 F|_{y_1(h)} + h_2 \cdot \mathcal{D}_2 F|_{y_2(h)} \end{aligned}$$

Thus

$$F(a+h) - F(a) - L(h) = h_1 \cdot (\mathcal{D}_1 F|_{y_1(h)} - \mathcal{D}_1 F|_a) + h_2 \cdot (\mathcal{D}_2 F|_{y_2(h)} - \mathcal{D}_2 F|_a)$$

Divide by $\|h\|$ and take limits,

$$\begin{aligned}\lim_{h\to 0} \frac{F(a+h) - F(a) - L(h)}{\|h\|} &= \lim_{h\to 0} \frac{h_1 \cdot (\mathcal{D}_1 F|_{y_1(h)} - \mathcal{D}_1 F|_a)}{\|h\|} \\ &+ \lim_{h\to 0} \frac{h_2 \cdot (\mathcal{D}_2 F|_{y_2(h)} - \mathcal{D}_2 F|_a)}{\|h\|} \\ &= 0 + 0 \\ &= 0\end{aligned}$$

Thus F is differentiable at a. □

Definition (C^1 space)**.** $f = (f_1, \dots, f_m) : U \to \mathbb{R}^n$ is C^1 if f_i is C^1 for all i.

Corollary 4.12. *If $f \in C^1(U)$ then f is differentiable on U.*

Proof. f is differentiable if and only if f_i are differentiable for all i. □

4.5 Chain Rule

4.5.1 Limits and Compositions

Let V_1, V_2, V_3 be normed vector spaces, $U_1 \subseteq V_1, U_2 \subseteq V_2$ open, $v_0 \in U_1$, $f : U_1 \to U_2, g : U_2 \to V_3$.

Lemma 4.13. *If $\lim_{v\to v_0} f(v) = w$ and g is continuous at w then*

$$\lim_{v\to v_0} g(f(v)) = g(w).$$

Proof. g is continuous at w so given $\varepsilon > 0$, exists $\delta_1 > 0$ such that $\|g(w') - g(w)\| < \varepsilon$ whenever $\|w' - w\| < \delta_1$. As $\lim_{v\to v_0} f(v) = w$, exists $\delta_2 > 0$ such that $\|f(v) - w\| < \delta_1$ whenever $0 < \|v - v_0\| < \delta_2$. Taking $w' = f(v)$, we see that $\|g(f(v)) - g(w)\| < \varepsilon$ whenever $0 < \|v - v_0\| < \delta_2$. Thus

$$\lim_{v\to v_0} g(f(v)) = g(w).$$

□

4.5.2 Derivatives and Compositions

Let $U_1 \subseteq \mathbb{R}^n, U_2 \subseteq \mathbb{R}^m$ open, $F : U_1 \to U_2, g : U_2 \to \mathbb{R}^\ell$. If f is differentiable at $a \in U_1$, $\mathcal{D}f|_a \in L(\mathbb{R}^n, \mathbb{R}^m)$. If g is differentiable at $f(a) \in U_2$, $\mathcal{D}g|_{f(a)} \in L(\mathbb{R}^m, \mathbb{R}^\ell)$, so the composition

$$\mathcal{D}g|_{f(a)} \circ \mathcal{D}f|_a \in L(\mathbb{R}^n, \mathbb{R}^\ell).$$

Theorem 4.14 (Chain rule)**.** *Suppose f and g are as above. If f is differentiable at a and g is differentiable at $f(a)$, then $g \circ f$ is differentiable at a and*
$$\mathcal{D}(g \circ f)|_a = \mathcal{D}g|_{f(a)} \circ \mathcal{D}f|_a.$$

Proof. We first use alternate characterisation of differentiability to rephrase the problem. We know that
$$f(a+h) = f(a) + L_1(h) + \|h\|\alpha(h)$$
where $L_1 = \mathcal{D}f|_a \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $\lim_{h\to 0} \alpha(h) = 0$. Similarly
$$g(f(a)+r) = g(f(a)) + L_2(r) + \|r\|\beta(r)$$
where $L_2 = \mathcal{D}g|_{f(a)} \in L(\mathbb{R}^m, \mathbb{R}^\ell)$ and $\lim_{r\to 0} B(r) = 0$.

Joke. I had some trouble finding a good name for the infinitesimal variable. Certainly f or g can't be used although they are close to h. In the end I settled for r, although it does not have much h-ness.

Note. If I define $\beta(0) = 0$ then β is continuous at 0.

The objective is to show that
$$g(f(a+h)) = g(f(a)) + L_2(L_1(h)) + \|h\|\gamma(h)$$
where $\lim_{h\to 0} \gamma(h) = 0$.

So we want to find $\gamma(h)$. Using the estimates for $f(a+h)$ and $g(f(a)+r)$, we see that
$$\begin{aligned} g(f(a+h)) &= g(f(a) + \underbrace{L_1(h) + \|h\|\alpha(h)}_{E(h)}) \\ &= g(f(a)) + L_2(L_1(h) + \|h\|\alpha(h)) + \|E(h)\|\beta(E(h)) \\ &= g(f(\alpha)) + L_2(L_1(h)) + \|h\|L_2(\alpha(h)) + \|E(h)\|\beta(E(h)) \end{aligned}$$
i.e.
$$\gamma(h) = L_2(\alpha(h)) + \frac{\|E(h)\|}{\|h\|}\beta(E(h)).$$

Always remember divide and conquer. The first term is easy so we focus on $E(h)$ for now.

Lemma 4.15. *There exists $\delta > 0$ such that $\|E(h)\|/\|h\| \le M$ whenever $0 < \|h\| < \delta$.*

Proof. Recall that
$$\frac{E(h)}{\|h\|} = L_1\left(\frac{h}{\|h\|}\right) + \alpha(h)$$
where
$$\frac{h}{\|h\|} \in S = \{v \in \mathbb{R}^n : \|v\| = 1\}.$$
We have seen before in the proof of Theorem 3.11 that S is compact. By Max Value theorem there exists M' such that $\|L_1(v)\| \le M'$ for all $v \in S$.

Choose δ such that $\|\alpha(h)\| < 1$ whenever $0 < \|h\| < \delta$. Then for $0 < \|h\| < \delta$,

$$\frac{\|E(h)\|}{\|h\|} \le \left\| L\left(\frac{h}{\|h\|}\right) \right\| + \|\alpha(h)\| \le M' + 1 = M$$

□

Corollary 4.16.
$$\lim_{h \to 0} E(h) = 0.$$

Proof.
$$\lim_{h \to 0} E(h) = \lim_{h \to 0} \frac{E(h)}{\|h\|} \cdot \|h\| = 0$$
by squeeze rule. □

Finally we are ready to estimate $\gamma(h)$. L_2 is continuous so

$$\lim_{h \to 0} \alpha(h) = 0 \Rightarrow \lim_{h \to 0} L_2(\alpha(h)) = L_2(0) = 0.$$

β is continuous so

$$\lim_{h \to 0} E(h) = 0 \Rightarrow \lim_{h \to 0} \beta(E(h)) = \beta(0) = 0.$$

Applying Lemma, squeeze rule shows that

$$\lim_{h \to 0} \gamma(h) = 0.$$

□

Remark.

1. If $\gamma : \mathbb{R} \to \mathbb{R}^n, F : \mathbb{R}^n \to \mathbb{R}$, let $f = F \circ \gamma$. Then

$$\begin{aligned} \mathcal{D}(F \circ \gamma)|_t &= f'(t) \\ &= \mathcal{D}F|_{\gamma(t)} \circ \mathcal{D}\gamma|_t \\ &= (\mathcal{D}_1 F|_{\gamma(t)} \cdots \mathcal{D}_n F|_{\gamma(t)}) \begin{pmatrix} \gamma_1'(t) \\ \vdots \\ \gamma_n'(t) \end{pmatrix} \\ &= [\boldsymbol{\nabla} F|_{\gamma(t)} \cdot \gamma'(t)] \end{aligned}$$

i.e. $f'(t) = \boldsymbol{\nabla} F|_{\gamma(t)} \cdot \gamma'(t)$.

In IA Analysis I, we proved the Mean Value Theorem which is an important result in differential calculus. The result generalises to higher dimensions.

Definition (Convex set)**.** $U \subseteq \mathbb{R}^n$ is *convex* if whenever $x_0, x_1 \in U$ then

$$tx_1 + (1-t)x_0 \in U$$

for $t \in [0, 1]$.

Theorem 4.17 (Mean Value Inequality)**.** *Suppose $U \subseteq \mathbb{R}^n$ is open and convex. If $F : U \to \mathbb{R}$ is differentiable and satisfies*

$$\|\nabla F|_a\| \leq M$$

for all $a \in U$ then

$$|F(x_1) - F(x_0)| \leq M\|x_1 - x_0\|$$

for all $x_1, x_0 \in U$.

Proof. Given what we have so far, it takes more words to state the theorem than to prove it (well, maybe not).

For $t \in [0,1]$ define

$$\gamma(t) = tx_1 + (1-t)x_0.$$

Then

$$\gamma'(t) = x_1 - x_0$$

and $\operatorname{Im}\gamma \subseteq U$. Let $f(t) = F(\gamma(t))$, then

$$\begin{aligned} |f'(t)| &= |\nabla F|_{\gamma(t)} \cdot \gamma'(t)| \\ &\leq \left\|F|_{\gamma(t)}\right\| \cdot \|\gamma'(t)\| \text{ by Cauchy-Schwarz} \\ &\leq M\|x_1 - x_0\| \end{aligned}$$

Mean Value Theorem applied to $f(t)$ says that

$$\begin{aligned} |F(x_1) - F(x_0)| &= |f(1) - f(0)| \\ &= |f'(c)| \text{ for some } c \in [0,1] \\ &\leq M\|x_1 - x_0\| \end{aligned}$$

□

4.6 Higher Derivatives

Let $U \subseteq \mathbb{R}^n$ open, $a \in U$ and $f : U \to \mathbb{R}^m$ differentiable. Then $\mathcal{D}f_a \in L(\mathbb{R}^n, \mathbb{R}^m)$.

Question. How should we define the second derivative at a?

Given $v \in \mathbb{R}^n$, define

$$\begin{aligned} \alpha_v : U &\to \mathbb{R}^m \\ x &\mapsto \mathcal{D}f|_x(v) \end{aligned}$$

Note. α_v depends on f.

Example.

1. $\alpha_{e_i}(x) = \mathcal{D}f_x(e_i) = \mathcal{D}_i f(x)$.
2. $\alpha_{v_1+\lambda v_2}(x) = \mathcal{D}f|_x(v_1 + \lambda v_2) = \mathcal{D}f|_x(v_1) + \lambda\mathcal{D}f|_x(v_2) = \alpha_{v_1}(x) + \lambda\alpha_{v_2}(x)$.

Definition (Twice differentiability)**.** f is *twice differentiable* at $a \in U$ if α_v is differentiable at a for all $v \in \mathbb{R}^n$.

If so define

$$\begin{aligned} \mathcal{D}^2 f|_a : \mathbb{R}^n \times \mathbb{R}^n &\to \mathbb{R}^m \\ (v, w) &\mapsto \mathcal{D}\alpha_v|_a(w) \end{aligned}$$

Example.

$$\begin{aligned} \mathcal{D}^2 f|_a(e_i, e_j) &= \mathcal{D}\alpha_{e_i}|a(e_j) \\ &= \mathcal{D}(\mathcal{D}_i f)|_a(e_j) \\ &= \mathcal{D}_j(\mathcal{D}_i f)|_a \end{aligned}$$

Notation. Let

$$\mathcal{D}_{ij} f|_a := \mathcal{D}_j(\mathcal{D}_i f) = \frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i}\Big|_a.$$

Lemma 4.18. *If f is twice differentiable at a then*

$$\mathcal{D}^2 f|_a : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$$

is bilinear, i.e.

1. $\mathcal{D}^2 f|_a(v_1 + \lambda v_2, w) = \mathcal{D}^2 f|_a(v_1, w) + \lambda \mathcal{D}^2 f|_a(v_2, w)$,
2. $\mathcal{D}^2 f|_a(v, w_1 + \lambda w_2) = \mathcal{D}^2 f|_a(v, w_1) + \lambda \mathcal{D}^2 f|_a(v, w_2)$.

Proof.

1.

$$\begin{aligned} \mathcal{D}^2 f|_a(v_1 + \lambda v_2, w) &= \mathcal{D}\alpha_{v_1 + \lambda v_2}|_a(w) \\ &= \mathcal{D}(\alpha_{v_1} + \lambda \alpha_{v_2})|_a(w) \\ &= \mathcal{D}\alpha_{v_1}|_a(w) + \lambda \mathcal{D}\alpha_{v_2}|_a(w) \\ &= \mathcal{D}^2 f|_a(v_1, w) + \lambda \mathcal{D}^2 f|_a(v_2, w) \end{aligned}$$

2.

$$\begin{aligned} \mathcal{D}^2 f|_a(v, w_1 + \lambda w_2) &= \mathcal{D}\alpha_v|_a(w_1 + \lambda w_2) \\ &= \mathcal{D}\alpha_v|_a(w_1) + \lambda \mathcal{D}\alpha_v|_a(w_2) \\ &= \mathcal{D}^2 f|_a(v, w_1) + \lambda \mathcal{D}^2 f|_a(v, w_2) \end{aligned}$$

□

Now suppose $m = 1$, i.e. $F : U \to \mathbb{R}$. Then

$$\begin{aligned} \mathcal{D}^2 F|_a(v, w) &= \mathcal{D}^2 F|_a \left(\sum_{i=1}^n v_i e_i, \sum_{j=1}^n w_j e_j \right) \\ &= \sum_{i,j=1}^n v_i \mathcal{D}^2 F|_a(e_i, e_j) w_j \\ &= v^T H(a) w \end{aligned}$$

where

$$H(a) = (\mathcal{D}^2 F|_a(e_i, e_j)) = (\mathcal{D}_{ij} F|_a)$$

is the *Hessian matrix* of F at a.

Example. Let $F : \mathbb{R}^2 \to \mathbb{R}, F(x_1, x_2) = x_1^4 x_2^2$, then

$$\begin{array}{ll} \mathcal{D}_{11} F = 12 x_1^2 x_2^2 & \mathcal{D}_{12} F = 8 x_1^3 x_2 \\ \mathcal{D}_{21} F = 8 x_1^3 x_2 & \mathcal{D}_{22} F = 2 x_1^4 \end{array}$$

so

$$H(1,1) = \begin{pmatrix} 12 & 8 \\ 8 & 2 \end{pmatrix}$$

and

$$\mathcal{D}^2 F|_{(1,1)}((3,1),(1,2)) = \begin{pmatrix} 3 & 1 \end{pmatrix} \begin{pmatrix} 12 & 8 \\ 8 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Definition (C^2 space)**.** $F : U \to \mathbb{R}$ is C^2 if F is C^1 and all second partial derivatives $\mathcal{D}_{ij} F$ exist and are continuous on U. If so $F \in C^2(U)$.

Remark. We conduct a reality check that $F \in C^2(U)$ does imply that F is twice differentiable: if $F \in C^2(U)$ then $\mathcal{D}_i F \in C^1(U)$ for all i. Since $\alpha_v(x) = \mathcal{D} F|_x(v)$ it follows that

$$\alpha_v = \sum_{i=1}^n v_i \mathcal{D}_i F \in C^1(U).$$

By Theorem 4.10 α_v is differentiable for all v so F is twice differentiable on U.

Theorem 4.19 (Symmetry of Mixed Partials)**.** *If $U \subseteq \mathbb{R}^2$ is open and $F \in C^2(U)$ then*

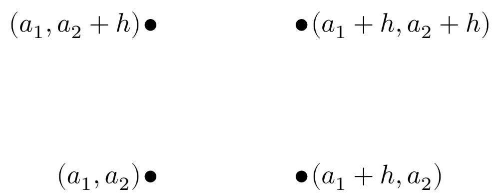
$$\mathcal{D}_{12} F = \mathcal{D}_{21} F.$$

Lemma 4.20.

$$\mathcal{D}_{21} F|_a = \lim_{h \to 0} \frac{1}{h^2} \left(F(a_1 + h, a_2 + h) - F(a_1 + h, a_2) - F(a_1, a_2 + h) + F(a_1, a_2) \right).$$

Note. $h \in \mathbb{R}$, not $\mathbb{R}^2$.

Proof. The proof is similar to the proof of Theorem 4.10 in that we apply Mean Value Theorem to a two step estimation.



Let

$$\begin{aligned}\Delta &= F(a_1 + h, a_2 + h) - F(a_1 + h, a_2) - F(a_1, a_2 + h) + F(a_1, a_2)\\ A(t) &= F(a_1 + h, t) - F(a_1, t)\end{aligned}$$

so that

$$\begin{aligned}\Delta &= A(a_2 + h) - A(a_2)\\ A'(t) &= \mathcal{D}_2 F|_{(a_1+h,t)} - \mathcal{D}_2 F|_{(a_1,t)}\end{aligned}$$

By Mean Value Theorem applied to A, we see that

$$\Delta = h \cdot A'(x_2(h))$$

where $x_2(h) \in [a_2, a_2 + h]$

$$\begin{aligned}&= h \cdot \left(\mathcal{D}_2 F|_{(a_1+h,x_2(h))} - \mathcal{D}_2 F|_{(a_1,x_2(h))}\right)\\ &= h \cdot (B(a_1 + h) - B(a_1))\end{aligned}$$

where

$$\begin{aligned}B(s) &= \mathcal{D}_2 F|_{(s,x_2(h))}\\ B'(s) &= \mathcal{D}_1(\mathcal{D}_2 F)|_{(s,x_2(h))} = \mathcal{D}_{21} F|_{(s,x_2(h))}\end{aligned}$$

so by Mean Value Theorem again

$$= h^2 \mathcal{D}_{21} F|_{(x_1(h),x_2(h))}$$

In summary, $\Delta = h^2 \mathcal{D}_{21} F|_{x_1(h),x_2(h)}$ so

$$\lim_{h\to 0} \frac{\Delta}{h^2} = \lim_{h\to 0} D_{21} F(x_1(h), x_2(h))$$

where $x_1(h) \in [a_1, a_1 + h], x_2(h) \in [a_2, a_2 + h]$. Thus we know

$$\lim_{h\to 0} (x_1(h), x_2(h)) = (a_1, a_2).$$

Since $F \in C^2(U)$, $\mathcal{D}_{21} F$ is continuous so

$$\lim_{h\to 0} \mathcal{D}_{21} F(x_1(h), x_2(h)) = \mathcal{D}_{21} F|_a = \lim_{h\to 0} \frac{\Delta}{h^2}.$$

□

Proof of Symmetry of Mixed Partials. This should be almost apparent from the symmetry of the expression on RHS above. Let $G(x_1, x_2) = F(x_2, x_1)$. Then

$$\begin{aligned}\mathcal{D}_{12} F|_{(a_1,a_2)} &= \mathcal{D}_{21} G|_{(a_2,a_2)}\\ &= \lim_{h\to 0} \frac{1}{h^2} (G(a_1 + h, a_2 + h) - G(a_1, a_2 + h) - G(a_1 + h, a_2) + G(a_1, a_2))\\ &= \lim_{h\to 0} \frac{1}{h^2} (F(a_1 + h, a_2 + h) - F(a_1 + h, a_2) - F(a_1, a_2 + h) + F(a_1, a_2))\\ &= \mathcal{D}_{21} F|_{(a_1,a_2)}\end{aligned}$$

□

Corollary 4.21. *Suppose $U \subseteq \mathbb{R}^n$ is open, $F \in C^2(U)$. Then*

$$D_{ij}F|_a = D_{ji}F|_a$$

for all $1 \leq i \leq j \leq n$ and all $a \in U$.

Proof. Apply Symmetry of Mixed Partials to

$$G(x_1, x_2) = F(a_1, \dots, a_{i-1}, x_1, a_{i+1}, \dots, a_{j-1}, x_2, a_{j+1}, \dots, a_n).$$

□

To summarise, if $F \in C^2(U)$ then the Hessian matrix

$$H|_a = \big(D_{ij}F|_a\big)$$

is symmetric. We have proved that the second derivative is given by

$$\begin{aligned} \mathcal{D}^2 F|_a : \mathbb{R}^n \times \mathbb{R}^n &\to \mathbb{R} \\ (v, w) &\mapsto v^T H|_a w \end{aligned}$$

so $\mathcal{D}^2 F|_a$ is a *symmetric bilinear form*:

$$\mathcal{D}^2 F|_a(v, w) = \mathcal{D}^2 F|_a(w, v).$$

We could rephrase our theory for second derivatives developed so far using the language of linear maps, which gives an alternative description from a slightly different point: if $f : U \to \mathbb{R}^m$ is C^2, then

$$\begin{aligned} \mathcal{D}f : U &\to L(\mathbb{R}^n, \mathbb{R}^m) \cong \mathcal{M}_{m,n}(\mathbb{R}) \cong \mathbb{R}^{mn} \\ \mathcal{D}(\mathcal{D}f)|_a &\in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)). \end{aligned}$$

i.e. if $w \in \mathbb{R}^n, v \in \mathbb{R}^n$,

$$\mathcal{D}(\mathcal{D}f)|_a(w) \in L(\mathbb{R}^n, \mathbb{R}^m).$$

Define a function

$$B(v, w) = (\mathcal{D}(\mathcal{D}f)|_a(w))(v)$$

which is a clearly bilinear map $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ and it is not hard too see that

$$B(v, w) = \mathcal{D}^2 f|_a(v, w)$$

as we defined it.

4.6.1 Third and Higher Derivatives

Definition (C^k space)**.** $F : U \to \mathbb{R}$ is C^k if the partial derivatives $\mathcal{D}_i F$ are C^{k-1} for all $1 \leq i \leq n$.

If $F \in C^k(U)$, define the kth derivative

$$\begin{aligned} \mathcal{D}^k F|_a : (\mathbb{R}^n)^k &\to \mathbb{R} \\ (v_1, \dots, v_k) &\mapsto \mathcal{D}(\mathcal{D}^{k-1} F(v_1, \dots v_{k-1}))|_a(v_k) \end{aligned}$$

which is a *symmetric multilinear form.*

Note. The above definition also applies to $F : U \to \mathbb{R}^m$.

4.7 Taylor's Formula

Let $U \subseteq \mathbb{R}^n$ be convex and open, $F \in C^k(U)$ and $x_0, x_0 + v \in U$. Define

$$f(t) = F(x_0 + tv).$$

Note that $f : [0,1] \to \mathbb{R}$.

> **Proposition 4.22.** *f is k-times differentiable and*
>
> $$f^{(k)}(t) = \mathcal{D}^k F|_{x_0+tv}(v, \dots, v).$$

Proof. If $G \in C^k(U)$, $g(t) = G(x_0 + tv)$ then

$$g'(t) = \mathcal{D}_v G|_{x_0+tv} = \mathcal{D}G|_{x_0+tv}(v). \qquad (*)$$

Proof is by induction on k. $k = 1$ is $(*)$ with $G = F$. In general define

$$\begin{aligned} h(t) &= f^{(k-1)}(t) \\ &= \mathcal{D}^{k-1} F|_{x_0+tv}(v, \dots, v) \\ &= H(x_0 + tv) \end{aligned}$$

where $H(x) = \mathcal{D}^{k-1} F|_x(v, \dots, v)$.

Applying $(*)$ to h gives

$$\begin{aligned} f^{(k)}(t) &= h'(t) \\ &= \mathcal{D}H|_{x_0+tv}(v) \\ &= \mathcal{D}(\mathcal{D}^{k-1} F(v, \dots, v))(v) \\ &= \mathcal{D}^k F(v, \dots, v) \end{aligned}$$

□

> **Corollary 4.23** (Multivariable Taylor's Formula)**.** *If U is open and convex, $x_0, x_0 + v \in U$ and $F \in C^k(U)$ then*
>
> $$F(x_0 + v) = \sum_{i=0}^{k-1} \frac{1}{i!} \mathcal{D}^i F|_{x_0}(v, \dots, v) + \frac{1}{k!} \mathcal{D}^k F|_{x_0+tv}(v, \dots, v)$$
>
> *for some $t \in [0,1]$.*

Proof. This seems like a horrible mess but, like many other things we have encountered in this course, its nothing more than ideas from IA Analysis I applied new (actually gneralised from old) definitions. Define

$$f(t) = F(x_0 + tv).$$

Then the single variable Taylor's formula says that

$$f(1) = \sum_{i=0}^{k-1} \frac{1}{i!} f^{(i)}(0) 1^i + \frac{1}{k!} f^{(k)}(t) 1^k$$

for some $t \in [0,1]$. Subsituting the formula for $f^{(i)}$ as in the proposition above gives the result required. □

4.8 Speed & Distance

Well the title says all. This a bewildering section that doesn't seem to go anywhere or belong to any part of this course. Nevertheless it is required by the faculty.[1]

All norms are Euclidean norms in this section since we require inner product.

Lemma 4.24. *If $\alpha : \mathbb{R} \to \mathbb{R}^n$ is C^1 then*

$$\frac{d}{dt}\|\alpha(t)\| \leq \|\alpha'(t)\|.$$

Proof. $\|\alpha(t)\| = (\alpha \cdot \alpha)^{1/2}$ so

$$\begin{aligned}\frac{d}{dt}(\alpha \cdot \alpha)^{1/2} &= \frac{1}{2}(\alpha \cdot \alpha)^{-1/2}(2\alpha' \cdot \alpha) \\ &= \frac{\alpha' \cdot \alpha}{(\alpha \cdot \alpha)^{1/2}} \\ &\leq \frac{\|\alpha'\|\|\alpha\|}{\|\alpha\|} \\ &= \|\alpha'\|\end{aligned}$$

by Cauchy-Schwarz. □

Corollary 4.25. *If $\gamma : \mathbb{R} \to \mathbb{R}^n$ is continuous then*

$$\left\|\int_0^1 \gamma(t)dt\right\| \leq \int_0^1 \|\gamma(t)\|dt$$

Note. If $\gamma(t) = v(t)$ is the velocity then this says displacement is smaller than distance on the odometer.

Proof. Let $\alpha(s) = \int_0^s \gamma(t)dt$. Then by the lemma

$$\frac{d}{ds}\|\alpha(s)\| \leq \|\alpha'(s)\| = \|\gamma(s)\|$$

where the equality comes from Fundamental Theorem of Calculus. Let $\beta(s) = \int_0^s \|\gamma(t)\|dt$ then $\beta'(s) = \|\gamma(s)\|$.

Since $\|\alpha(0)\| = \beta(0)$ and

$$\frac{d}{ds}\|\alpha(s)\| \leq \|\gamma(s)\| = \frac{d}{ds}\beta(s),$$

$$\|\alpha(s)\| \leq \beta(s)$$

for all $s \geq 0$. Take $s = 1$ to get the result required. □

And that marks the end of this vestigial section.

[1] "All right let's go ahead and get started."

5 Metric Spaces

In this chapter we take a short break from differential calculus (but don't forget them! We will need them shortly after).

5.1 Definitions

Definition (Metric space)**.** A *metric space* is a set X with a distance function $D : X \times X \to \mathbb{R}$ satisfying

1. positivity: $d(x, y) \geq 0, d(x, y) = 0$ if and only if $x = y$.
2. symmetry: $d(x, y) = d(y, x)$ for all $x, y \in X$.
3. triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

Example.

1. A normed space $(V, \|\cdot\|)$ is a metric space with

$$d(v, w) = \|v - w\|.$$

   Proof.

   (a) $d(v, w) = \|v - w\| \geq 0$ and $d(v, w) = 0$ if and only if $\|v, w\| = 0$ if and only if $v - w = 0$ if and only if $v = w$.

   (b) $d(v, w) = \|v - w\| = \|(-1)(w - v)\| = |-1| \cdot \|w - v\| = d(w, v)$.

   (c) $d(v_1, v_3) = \|v_1 - v_3\| \leq \|v_1 - v_2\| + \|v_2 - v_3\| = d(v_1, v_2) + d(v_2, v_3)$.

   □

2. If (X, d) is a metric space and $Y \subseteq X$ then $(Y, d|_{Y \times Y})$ is metric space. We say Y is a *subspace* of X.

3. For any set X, let

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

   which is the *discrete metric*.

Most of the definitions and theorems we gave about subsets of normed spaces apply equally well to metric spaces by replacing $\|v - w\|$ with $d(v, w)$. Actually metric is a more fundamental concept than norm: every norm induces a metric as outlined above but not vice versa. This means that we could have organised the contents in a more structured and formal way by introducing metric spaces and its properties upfront and subsequently allowing normed spaces to inherit these properties. However, we choose not to do so since

1. for most of the course up to this point, properties of metric spaces are in a sense add complexity but not richness to our theory because we work with $\mathbb{R}^n$ and function spaces, which come with a normed structure. Differential calculus in high dimension is already hard and we don't want to make things more complicated.

2. in fact, we don't use metric properties until the last chapter. It might be better to give an ad hoc definition here lest one forget if we do it at the very beginning.

That is enough digression about the structure of the course. As promised, here are some definitions and results that generalise easily those from normed space. You should find them at this point very familiar (and more so if you've taken IB Metric and Topological Spaces).

Definition (Convergence)**.** A sequence (x_n) in X *converges* to $x \in X$ if for every $\varepsilon > 0$, there exists N such that $d(x_n, x) < \varepsilon$ whenever $n > N$.

Definition (Continuity)**.** If (X, d_x) and (Y, d_Y) are metric spaces, $f : X \to Y$ is continuous if $(f(x_n)) \to f(x)$ with respect to d_Y whenever $(x_n) \to x \in X$ with respect to d_X.

Proposition 5.1 (Alternate characterisation of continuity)**.** *f is continuous if and only if for every $\varepsilon > 0$ and $x \in X$, there exists $\delta > 0$ such that $d(f(x'), f(x)) < \varepsilon$ whenever $d(x', x) < \delta$.*

Definition (Open ball)**.** The set

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

is the *open ball* of radius r centred at x.

Definition (Closed ball)**.** The set

$$B_r(x) = \{x' \in X : d(x', x) \leq r\}$$

is the *closed ball* of radius r centred at x.

Definition (Open subset)**.** $U \subseteq X$ is an *open* subset of X if for every $x \in U$, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$.

Proposition 5.2. *If $f : X \to Y$ is continuous and $U \subseteq Y$ is open then*

$$f^{-1}(U) \subseteq X$$

is open.

We have stressed this before but in case one has forgotten,

Note. Being open (and closed) is a property of a *subset*, not a space.

Example. Let $X = \mathbb{R}$ with metric $d(x, y) = |x - y|$. Then $[0, \frac{1}{2})$ is not an open subset of X. If $Y = [0, 1] \subseteq X$ with the subspace metric then $[0, \frac{1}{2})$ is an open subset of Y.

Definition (Closed subset)**.** C is a *closed subset* of X if $X \setminus C$ is an open subset.

Proposition 5.3. *$C \subseteq X$ is closed if and only if whenever $(x_n) \to x$ in X and x_n's are all in C then $x \in C$ as well.*

Definition (Cauchy sequence)**.** A sequence (x_n) in X is *Cauchy* if for every $\varepsilon > 0$ there exists N such that $d(x_n, x_m) < \varepsilon$ whenever $n, m \geq N$.

Definition (Completeness)**.** X is *complete* if whenever (x_n) is a Cauchy sequence in X, there exists $x \in X$ such that $(x_n) \to x$.

Proposition 5.4. *Suppose X is a complete metric space and $C \subseteq X$ is closed. Then C with the subspace metric is also complete.*

Proof. Suppose (x_n) is a Cauchy sequence in C. Then (x_n) is also a Cauchy sequence in X. Since X is complete there exists $x \in X$ such that $(x_n) \to x$. Since $C \subseteq X$ is closed $x \in C$ so C is complete. □

Joke. This is a story about John Conway. Before he moved to the U.S. he was a professor here in Cambridge. He was a very unusual guy and liked playing games. His office was full of toys, such as balls to study sphere packing. In fact he had two offices full of toys: the first one was filled up so he was given a second one.

One day he had his attic repainted. When the painters finished, they left behind this long roll of paper and an enormous pair of shears. They came back and collect hte shears but not didn't bother the paper.

Back then Conway was interested in finite simple group. Around that time someone suspected a new finite simple group and Conway proposed a way to build it. Other group theorists told Conway that, well, if you want to prove it then just write down the character table (which is enormous). But he was too busy to get started.

Just about then he thought it would a really good idea to use the paper at hand to do this. He cover the floor of attic with paper and started working on the character table. And indeed he found it, so now we have a finite simple group (actually three) called Conway group.[1]

5.2 Lipschitz Maps

Suppose (X, d_X) and (Y, d_Y) are metric spaces.

Definition (Lipschitz map)**.** $f : X \to Y$ is *K-Lipschitz*, where $K \in \mathbb{R}, K > 0$, if for every $x, x' \in X$,

$$d_Y(f(x), f(x')) \leq K d_X(x, x').$$

[1] Moral of the story: sometimes you just need a really big piece of paper!

Say f is *Lipschitz* if it is K-Lipschitz for some K.

Example.

1. If f is Lipschitz then it is uniformly continuous:

   Proof. Suppose f is K-Lipschitz. Given $\varepsilon > 0$, $d(f(x), f(x')) \leq Kd(x, x') < \varepsilon$ whenever $d(x, x') < \varepsilon / K$. □

2. Suppose $U \subseteq \mathbb{R}^n$ is open, $F \in C^1(U)$. If $K = \overline{B}_r(x_0) \subseteq U$ then $F|_K$ is Lipschitz:

   Proof. The function

   $$\begin{aligned} U &\to \mathbb{R}^n \to \mathbb{R} \\ x &\mapsto \nabla F|_x \mapsto \|\nabla F|_x\| \end{aligned}$$

   is continuous. K is closed and bounded and thus compact. Thus by Maximum Value Theorem there exists M such that $\|\nabla F|_x\| \leq M$ for all $x \in K$. $K = \overline{B}_r(x_0)$ is convex so by Mean Value Inequality

   $$|F(x) - F(x')| \leq M\|x - x'\|$$

   so f is M-Lipschitz. □

3. If $f : X \to Y$ is K_1-Lipschitz, $g : Y \to Z$ is K_2-Lipschitz then $g \circ f$ is K_1K_2-Lipschitz:

   Proof.

   $$\begin{aligned} d(g(f(x)), g(f(x'))) &\leq K_2 d(f(x), f(x')) \\ &\leq K_2 K_1 d(x, x') \end{aligned}$$

   □

4. Consequently, composition of Lipschitz maps is Lipschitz.

5. If $\|\cdot\|$ and $\|\cdot\|'$ are two norms on V. Then they are Lipschitz equivalent if and only if the maps

   $$\begin{aligned} \mathrm{id} &: (V, \|\cdot\|) \to (V, \|\cdot\|') \\ \mathrm{id} &: (V, \|\cdot\|') \to (V, \|\cdot\|) \end{aligned}$$

   are both Lipschitz.

5.2.1 Operator Norm

Definition (Operator norm)**.** Let V and W be normed spaces. Given $L \in L(V, W)$, the *operator norm* is

$$\|L\|_{\text{op}} = \sup_{v \in V \setminus \{0\}} \frac{\|L(v)\|_W}{\|v\|_V} = \sup_{v \in V \setminus \{0\}} \left\| L\left(\frac{v}{\|v\|}\right) \right\|$$

Remark. If V and W are finite-dimensional, $L \in L(V, W)$ is continuous and $S(V) = \{v \in V : \|v\| = 1\}$ is compact. By the Maximum Value Theorem,

$$\sup_{v \in S(V)} \|L(v)\| = \max_{v \in S(V)} \|L(v)\|.$$

so we can replace sup with *max* in the definition.

Observe that if $v \in V$ then $\|L(v)\| \leq \|L\|_{\text{op}} \cdot \|v\|$.

We call something a norm without checking whether it is a norm so we had better do it now:

> **Proposition 5.5.** $\|\cdot\|_{op}$ *is a norm on* $L(V, W)$.

Proof. Example sheet. ☐

Form here on let $V = (\mathbb{R}^n, \|\cdot\|_2)$ and $W = (\mathbb{R}^m, \|\cdot\|_2)$.

> **Proposition 5.6.** *Suppose* $U \subseteq \mathbb{R}^n$ *is open and convex,* $f : U \to \mathbb{R}^m$ *is differentiable and* $\|\mathcal{D}f|_x\|_{op} \leq M$ *for all* $x \in U$. *Then* f *is M-Lipschitz.*

Proof. The proof for the general case is similar to that of Mean Value Inequality and is left as an exercise. Here we want to draw our attention to the case where f is C^1 and show how it arises as a corollary of Mean Value Inequality.

Given $x_0, x_1 \in U$, define

$$\begin{aligned} x : [0,1] &\to U \\ t &\mapsto (1-t)x_0 + tx_1 \\ \gamma : [0,1] &\to \mathbb{R}^m \\ t &\mapsto f(x(t)) \end{aligned}$$

By Corollary 4.25,

$$\left\| \int_0^1 \gamma'(t) dt \right\| \leq \int_0^1 \|\gamma'(t)\| dt$$

By Fundamental Theorem of Calculus, LHS is

$$\left\| \int_0^1 \gamma'(t) dt \right\| = \|\gamma(1) - \gamma(0)\| = \|f(x_1) - f(x_0)\|,$$

and by chain rule the integrand on RHS is

$$\|\gamma'(t)\| = \left\| \mathcal{D}f|_{x(t)}(x'(t)) \right\| \leq \|\mathcal{D}f|_{x(t)}\| \cdot \|x'(t)\| \leq M\|x_1 - x_0\|.$$

Thus putting everyting together we get

$$\|f(x_1) - f(x_0)\| \leq M\|x_1 - x_0\|.$$

☐

5.3 Contraction Mapping Theorem

In this section we will learn a new way to solve equations.

Definition (Contraction map). Let X be a metric space. $f : X \to X$ is a *contraction map* if it is K-Lipschitz for some $K < 1$, i.e.

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

Intuitively, f shrinks distances, ergo the name.

Definition (Fixed point). $x \in X$ is a *fixed point* of $f : X \to X$ if $f(x) = x$.

Theorem 5.7 (Contraction Mapping Theorem). *Suppose X is a* complete *metric space. If $f : X \to X$ is a contraction map then f has a unique fixed point.*

Proof. Since f is a contraction, there is some $K < 1$ such that

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

We prove the uniqueness part first since it is short. Suppose x and x' are both fixed points of f, then

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

where $K < 1$. The only way for this to hold is $d(x, x') = 0$, i.e. $x = x'$.

Next we prove the more interesting part about existence. Heuristically, completeness appears in our hypothesis although it is not in any part of the definition of a contraction map or fixed point, so we better find a Cauchy sequence to which we can apply the condition. Pick $x_0 \in X$ and inductively define

$$x_{n+1} = f(x_n) = f^{n+1}(x_0).$$

Observe that

$$d(x_n, x_{n+1}) = d(f(x_{n-1}), f(x_n)) \leq K d(x_{n-1}, x_n)$$

so by induction we see that

$$d(x_n, x_{n-1}) \leq K^n d(x_0, x_1) = K^n R$$

where $R = d(x_0, x_1)$. Claim (x_n) is Cauchy:

Proof.

$$\begin{aligned} d(x_n, x_{n+r}) &\leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots + d(x_{n+r-1}, x_{n+r}) \\ &\leq K^n R + K^{n+1} R + \cdots + K^{n+r-1} R \\ &= K^n R \frac{1 - K^r}{1 - K} \\ &\leq K^n R \frac{1}{1 - K} \end{aligned}$$

As $K < 1$,

$$\lim_{n \to \infty} \frac{K^n R}{1 - K} = 0.$$

Given $\varepsilon > 0$, pick N such that $\frac{K^n R}{1-K} < \varepsilon$ whenever $n \geq N$. Then for $m \geq n \geq N$,

$$d(x_n, x_m) \leq \frac{K^n R}{1-K} < \varepsilon$$

so (x_n) is Cauchy. □

Since X is complete, there exists $x \in X$ such that $(x_n) \to x$. f is Lipschitz so it is continuous so $(f(x_n)) \to f(x)$, i.e. $(x_{n+1}) \to f(x)$. But $(x_{n+1}) \to x$ so by uniqueness of limits in metric space $f(x) = x$, i.e. x is a fixed point of f. □

Remark. How does this help us solve equations? The theorem says that the equation $f(x) = x$ has a unique solution and the proof shows that we can approximate the fixed point by starting with $x_0 \in X$ and repeatedly applying f.

In practice, not every map is contraction so we often have to restrict the domain of f in order to get a contraction map.

Example (Finding square roots using iteration)**.** An elementary method to find the square root of a non-negative number n is to let

$$\begin{aligned} f : (0, \infty) &\to (0, \infty) \\ x &\mapsto \frac{1}{2}\left(x + \frac{n}{x}\right) \end{aligned}$$

and iterate f.

Why does this work? We are essentially finding a fixed point of f. But then $x = f(x) = \frac{1}{2}(x + n/x)$ so $x^2 = n$. Therefore the fixed point of f is precisely $\sqrt{n}$.

That seems promising. Now we are left to show f is a contraction:

$$|f(x) - f(y)| = \frac{1}{2}\left|x + \frac{n}{x} - y - \frac{n}{y}\right| = \frac{1}{2}|x - y| \cdot \left|1 - \frac{n}{xy}\right|.$$

Unfortunately, this means that if x and y are small f is definitely not a contraction. To fix this, we restrict f to $I_K = [\sqrt{n}/K, K\sqrt{n}]$. Then $f(I_K) \subseteq I_K$ and if we choose $K = \sqrt{2}$, for example, then

$$\left|1 - \frac{n}{xy}\right| \leq |1 - 2| = 1$$

so

$$|f(x) - f(y)| \leq \frac{1}{2}|x - y| \cdot 1 = \frac{1}{2}|x - y|$$

for $x, y \in I_K$ so $f|_{I_{\sqrt{2}}}$ is a contraction map. Therefore if I start with $x_0 \in I_{\sqrt{2}}$ and iterate it will converge to $\sqrt{n}$.

6 Solving Equations

In this final chapter we are going to learn methods to solve equations, first in $\mathbb{R}^n$ and then in function space.

6.1 Inverse Function Theorem

> **Definition** (Diffeomorphism). Suppose $U_1, U_2 \subseteq \mathbb{R}^n$ is open. A map $f : U_1 \to U_2$ is a *diffeomorphism* if f is a bijection and both f and f^{-1} are C^1.

Example.

1. If $L \in L(\mathbb{R}^n, \mathbb{R}^n)$ is invertible (i.e. $\det L \neq 0$) then the *affine map*

$$\begin{aligned} A : \mathbb{R}^n &\to \mathbb{R}^n \\ x &\mapsto L(x) + y_0 \end{aligned}$$

   is a diffeomorphism with inverse $y \mapsto L^{-1}(y - y_0)$.

2. The polar coordinates map

$$\begin{aligned} f : (0,\infty) \times (0,\pi) &\to \mathbb{R} \times (0,\infty) \\ (r,\theta) &\mapsto (r\cos\theta, r\sin\theta) \end{aligned}$$

   is a diffeomorphism. It is easy to see that f is bijective and C^1 and its inverse is

$$f^{-1}(x,y) = \left(\sqrt{x^2+y^2}, \tan^{-1}\frac{y}{x}\right)$$

   also C^1.

The second example gives us another way to interpret diffeomorphism: if $f : U_1 \to U_2$ is a diffeomorphism then $(f_1(x), \dots, f_n(x))$ is a different coordinate system on U_1. This is the basis of differential geometry.

Remark.

1. If $f : U_1 \to U_2$ and $g : U_2 \to U_3$ are diffeomorphisms then so is the composition $g \circ f : U_1 \to U_3$.

2. If $U_1, U_2 \subseteq \mathbb{R}^n$ are open, write $U_1 \cong U_2$ if there is a diffeomorphism $f : U_1 \to U_2$. Then $\cong$ is an equivalence relation on open subsets of $\mathbb{R}^n$.

3. Differentiable functions are continuous so if $f : U_1 \to U_2$ is a diffeomorphism then it is also a homeomorphism. The converse is not true: $f : \mathbb{R} \to \mathbb{R}, x \mapsto x^3$ is a homeomorphism with $f^{-1}(x) = x^{1/3}$. However f is not a diffeomorphism since f^{-1} is not differentiable at 0.

4. If $f : U_1 \to U_2$ is a diffeomorphism then $f^{-1} \circ f = \mathrm{id}_{U_1}$. Applying the chain rule,

$$\mathcal{D}f^{-1}|_{f(x)} \circ \mathcal{D}f|_x = \mathcal{D}(\mathrm{id}_{U_1})|_x = \iota \in L(\mathbb{R}^n, \mathbb{R}^n).$$

This implies that $\mathcal{D}f|_x$ is invertible and

$$(\mathcal{D}f^{-1})|_{f(x)} = (\mathcal{D}f|_x)^{-1}.$$

We can use this to prove the above proposition: $f(x) = x^3$ is not a diffeomorphism since $\mathcal{D}f|_x = (3x^2)$ is not invertible at $x = 0$.

In the rest of the section we are going to prove a single theorem. This theorem gives a sufficient condition for the existence of a local diffeomorphism. Roughly it says that if we are interested in determining whether a map is a diffeomorphism near a point, we only have to consider its invertibility at that point.

This theorem is of enormous importance and is considered Theorem 0 of differential geometry:

Theorem 6.1 (Inverse Function Theorem)**.** *Suppose $U \subseteq \mathbb{R}^n$ is open, $f : U \to \mathbb{R}^n$ is C^1 and $x_0 \in U$ is such that $\mathcal{D}f|_{x_0}$ is invertible, then there exists $U_1 \subseteq U$, $U_2 \subseteq \mathbb{R}^n$ open with $x_0 \in U_1$ such that $f|_{U_1} : U_1 \to U_2$ is a diffeomorphism.*

As said, this is a huge theorem and we will take two lectures to prove it so before we start to prove it let us discuss its intuition and implications.

Remark.

1. We can choose $U_1 = B_\varepsilon(x_0)$ or $U_2 = B_\delta(f(x_0))$ but not both at once.
2. The theorem says that if $\det \mathcal{D}f|_{x_0} \neq 0$ then $(f_1(x), \dots, f_n(x))$ are local coordinates on $\mathbb{R}^n$ near x_0.
3. On $\mathbb{R}$, if $f : \mathbb{R} \to \mathbb{R}$ is C^1 and $f'(x) \neq 0$ then f is a diffeomorphism onto its image. But this is false in higher dimension: that $f : \mathbb{R}^n \to \mathbb{R}^n$ is C^1 with $\det \mathcal{D}f|_x \neq 0$ does not imply f is a diffeomorphism onto its image. For example, let

$$\begin{aligned} f : (0,\infty) \times \mathbb{R} &\to \mathbb{R}^2 \setminus \{0\} \\ (r,\theta) &\mapsto (r\cos\theta, r\sin\theta) \end{aligned}$$

has

$$\mathcal{D}f|_{r,\theta} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and $\det \mathcal{D}f|_{r,\theta} = r \neq 0$ (note that we excluded the origin in the definition, which is always a problem for polar coordinates). But f is not injective since $f(r,\theta) = f(r,\theta+2\pi)$. This shows that there is no global generalisation of the one-dimensional result and Inverse Function Theorem is, in a sense, the best alternative we can have in terms of the local generalisation thereof.

Proof of Inverse Function Theorem. As this proof is long, I will interleave it with comments and remarks for clarity.

First consider a special case. Let $f : U \to \mathbb{R}^n$, $0 \in U$, $f(0) = 0$ and $\mathcal{D}f|_0 = \iota \in L(\mathbb{R}^n, \mathbb{R}^n)$.

Remark. Before we prove the theorem for this special case, note that to pass from the special case to a general case where $\mathcal{D}g|_{x_0}$ is invertible, we consider the function
$$f(x) = L^{-1}(g(x + x_0) - g(x_0))$$
where $L = \mathcal{D}g|_{x_0} \in L(\mathbb{R}^n, \mathbb{R}^n)$.

The hard bit is to show that f is locally invertible near 0, i.e. there exist $\varepsilon > 0$, $\delta > 0$ such that if $y_0 \in B_\varepsilon(0)$ then there is a unique $x \in B_\delta(0)$with $f(x) = y_0$. If this holds, take $U_2 = B_\varepsilon(0)$, $U_1 = f^{-1}(U_2) \cap B_\delta(0)$, then $f : U_1 \to U_2$ is a bijection.

The way we solve $f(x) = y_0$ is to use the tool we have just acquired: Contraction Mapping theorem. If we could write down a contraction map N_{y_0} whose fixed point $N_{y_0}(x) = x$ solves $f(x) = y_0$ then we are done.

Let us write done the map first. Take
$$N_{y_0}(x) = x + y_0 - f(x).$$
Check that $N_{y_0}(x) = x$ if and only if $x = x + y_0 - f(x)$ if and only if $f(x) = y_0$.

Remark. Okay so this map works for us. The question is, how do I come up with this map? The answer (essentially) is Newton's method. Recall Newton's method in one-dimension: to solve $f(x) = y_0$ for $f : \mathbb{R} \to \mathbb{R}$, let $x_{m+1} = \widetilde{N}_{y_0}(x_n)$ where
$$\widetilde{N}_{y_0}(x) = x + \frac{y_0 - f(x)}{f'(x)}.$$
This easily generalises to a function $f : \mathbb{R}^n \to \mathbb{R}^n$: let
$$\widetilde{N}_{y_0}(x) = x + (\mathcal{D}f|_x)^{-1}(y_0 - f(x)).$$

It is now only one step from its final form. We simplify further by approximation $\mathcal{D}f|_x \approx \mathcal{D}f|_0$. By hypothesis $\mathcal{D}f|_0 = \iota$ so we get
$$N_{y_0}(x) = x + \iota^{-1}(y_0 - f(x)) = x + y_0 - f(x).$$

We want to find $\delta > 0$ such that $N_y|_{B_\delta(0)}$ is a contraction map. We first prove it is Lipschitz:

Lemma 6.2. *There exists $\delta > 0$ such that $N_y|_{B_\delta(0)}$ is $\frac{1}{2}$-Lipschitz.*

Proof. Notice that
$$\mathcal{D}N_y|_x = \mathcal{D}(x + y - f(x))|_x = \iota - \mathcal{D}f|_x.$$

Now f is C^1 so the map $\mathcal{D}f : U \to L(\mathbb{R}^n, \mathbb{R}^n) \cong \mathbb{R}^{n^2}$ is continuous. All norms on $\mathbb{R}^{n^2}$ are equivalent so $\mathcal{D}f$ is continuous with respect to $\|\cdot\|_{\text{op}}$. Thus there exists $\delta > 0$ such that
$$\|\mathcal{D}f|_x - \mathcal{D}f|_0\|_{\text{op}} < \frac{1}{2}$$
whenever $\|x\| < \delta$, i.e. $\left\|\mathcal{D}N_y|_x\right\|_{\text{op}} = \|\mathcal{D}f|_x - \iota\|_{\text{op}} < \frac{1}{2}$.

By Proposition 5.6, $N_y|_{B_\delta(0)}$ is $\frac{1}{2}$-Lipschitz. □

That is one step in the correct direction. However, we still haven't quite got a contraction map as $N_y(x)$ may end up not in the ball $B_\delta(0)$. Our objective is to make sure when y is small enough it is also small.

Lemma 6.3. *Let* $\varepsilon = \frac{\delta}{2}$. *Then if* $y \in B_\varepsilon(0)$,

$$N_y(\overline{B}_\delta(0)) \subseteq B_\delta(0).$$

Proof. Suppose $x \in \overline{B}_\delta(0)$. Then

$$\begin{aligned} \|N_y(x)\| &\le \|N_y(0)\| + \|N_y(x) - N_y(0)\| \\ &\le \|0 + y - f(0)\| + \frac{1}{2}\|x - 0\| \\ &= \|y\| + \frac{1}{2}\|x\| \\ &< \frac{\delta}{2} + \frac{\delta}{2} \\ &= \delta \end{aligned}$$

so $N_y(x) \in B_\delta(0)$. ☐

In summary, we get

$$N_y|_{\overline{B}_\delta(0)} : \overline{B}_\delta(0) \to B_\delta(0)$$

is $\frac{1}{2}$-Lipschitz and is thus a contraction map.

Proposition 6.4. *Let* ε *and* δ *be as above. If* $y \in B_\varepsilon(0)$ *there is a unique* $x \in B_\delta(0)$ *with* $f(x) = y$.

Proof. Check the hypotheses for Contraction Mapping Theorem: $\mathbb{R}^n$ is complete and $\overline{B}_\delta(0) \subseteq \mathbb{R}^n$ is closed so $\overline{B}_\delta(0)$ is a complete metric space. Contraction Mapping Theorem applied to $N_y|_{\overline{B}_\delta(0)} : \overline{B}_\delta(0) \to \overline{B}_\delta(0)$ shows that there is a unique $x \in \overline{B}_\delta(0)$ with $N_y(x) = x$, i.e. there is a unique $x \in \overline{B}_\delta(0)$ with $f(x) = y$. To remove the bar on top of the ball, apply N_y again so we get

$$x = N_y(x) \in B_\delta(0).$$

☐

Corollary 6.5. *There exists* $U_1 \subseteq U$ *open with* $0 \in U_1$ *and* $U_2 \subseteq \mathbb{R}^n$ *open with* $0 \in U_2$ *such that* $f|_{U_1} : U_1 \to U_2$ *is a bijection.*

Proof. Take $U_2 = B_\varepsilon(0)$ and $U_1 = f^{-1}(U_2) \cap B_\delta(0)$. f is continuous so $f^{-1}(U_2)$ is open so U_1 is open. By construction $f(U_1) \subseteq U_2$. The proposition above says that for every $y \in U_2$ then there exists a unique $x \in U_1$ with $f(x) = y$. ☐

So far we have got a well-define map $g = (f|_{U_1})^{-1} : U_2 \to U_1$. The next goal is to show g is C^1 so that $f|_{U_1} : U_1 \to U_2$ is a diffeomorphism.

Lemma 6.6. g *is* 2*-Lipschitz.*

Proof. Suppose $g(y_1) = x_1$ and $g(y_2) = x_2$. Then $N_{y_1}(x_1) = x_1$ and $N_{y_2}(x_2) = x_2$. Notice that

$$N_{y_1}(x) - N_{y_2}(x) = x + y_1 - f(x) - x - y_2 + f(x) = y_1 - y_2.$$

We have

$$\begin{aligned}\|x_1 - x_2\| &= \|N_{y_1}(x_1) - N_{y_2}(x_2)\| \\ &\le \|N_{y_1}(x_1) - N_{y_1}(x_2)\| + \|N_{y_1}(x_2) - N_{y_2}(x_2)\| \\ &\le \frac{1}{2}\|x_1 - x_2\| + \|y_1 - y_2\|\end{aligned}$$

so $\frac{1}{2}\|x_1 - x_2\| \le \|y_1 - y_2\|$, i.e. $\|g(y_1) - g(y_2)\| \le 2\|y_1 - y_2\|$. □

Corollary 6.7. *g is continuous.*

Proof. g is 2-Lipschitz so uniformly continuous. □

Suppose $y \in U_2, g(y) = x$. We want to show $\mathcal{D}g|_y = (\mathcal{D}f|_x)^{-1}$. Note that although $\mathcal{D}f|_0$ is invertible, we don't know about $\mathcal{D}f|_x$ yet.

Lemma 6.8. *$\mathcal{D}f|_x$ is invertible.*

Proof. In the proof of a previous lemma we know that $\|\mathcal{D}f|_x - \iota\|_{\text{op}} < 1/2$. Write $A = \iota - \mathcal{D}f|_x$ so $\|A\|_{\text{op}} < 1/2$. From an exercise on example sheet we know the series $B = \sum_{n=0}^{\infty} A^n$ converges and $B = (\iota - A)^{-1} = (\mathcal{D}f|_x)^{-1}$. □

We have set up everything properly and are ready for the final step. The ideas involved should be very reminiscent of those in the proof of (baby) Inverse Function Theorem in IA Analysis I.

Proposition 6.9. *$g : U_2 \to U_1$ is C^1.*

Proof. Fix $y \in U_2$, let $g(y) = x$. Since U_2 is open there exists $\eta > 0$ such that $B_\eta(y) \subseteq U_2$. For $\kappa \in B_\eta(0)$, define

$$h(\kappa) = g(y + \kappa) - g(y),$$

i.e. $g(y + \kappa) = g(y) + h(\kappa)$ so $f(x + h(\kappa)) = y + \kappa$. We know f is differentiable at $x \in U_1 \subseteq U$ so

$$\begin{aligned}y + \kappa &= f(x + h(\kappa)) \\ &= f(x) + L(h(\kappa)) + \|h(\kappa)\|\alpha(h(\kappa))\end{aligned}$$

where $L = \mathcal{D}f|_x$ and $\lim_{h\to 0} \alpha(h) = 0$ by the definition of differentiability

$$= y + L(g(y + \kappa) - g(y)) + \|h(k)\|\alpha(h(\kappa))$$

which implies that

$$g(y + \kappa) = g(y) + L^{-1}(\kappa) - \|h(k)\|L^{-1}(\alpha(h(\kappa)).$$

To show that g is differentiable, it suffices to show that

$$\lim_{\kappa\to 0}\frac{\|h(\kappa)\|}{\|\kappa\|}L^{-1}(\alpha(h(\kappa)))=0.$$

Now g is 2-Lipschitz so

$$\|h(\kappa)\|=\|g(y+\kappa)-g(y)\|\le 2\|\kappa\|.$$

Since h is continuous $\lim_{\kappa\to 0}h(\kappa)=0$. α and L^{-1} are continuous at 0 so

$$\lim_{\kappa\to 0}\frac{\|h(\kappa)\|}{\|\kappa\|}L^{-1}(\alpha(h(\kappa)))=0.$$

by squeeze rule. Therefore g is differentiable.

We are almost there. To show g is C^1, the derivative map

$$\begin{aligned} U_2 &\to L(\mathbb{R}^n,\mathbb{R}^n)\\ x &\mapsto (\mathcal{D}f|_x)^{-1}\end{aligned}$$

is continuous by composition. □

We have essentially done the proof. Now we have to prove the theorem in the form stated. Begin with an obseravtion:

Lemma 6.10. *Suppose $f: U_1\to U_2$ is C^1. If $U_3\subseteq U_1$ is open then $f|_{U_3}: U_3\to f(U_3)$ is C^1.*

Proof. $f(U_3)=(f^{-1})^{-1}(U_3)$ is open since f^{-1} is C^1 and thus continuous. By construction $f|_{U_3}: U_3\to f(U_3)$ is a bijection. As f and f^{-1} are C^1, $f|_{U_3}$ and $(f^{-1})|_{f(U_3)}$ are C^1. □

Now suppose $\mathcal{D}f|_{x_0}=L$ is invertible. Let

$$\tilde{f}(x)=L^{-1}(f(x+x_0)-f(x_0)),$$

which is to say $\tilde{f}=A_2\circ f\circ A_1$ where

$$\begin{aligned} A_1,A_2&:\mathbb{R}^n\to\mathbb{R}^n\\ A_1(x)&=x+x_0\\ A_2(y)&=L^{-1}(y-f(x_0))\end{aligned}$$

are both affine diffeomorphisms. Now $\tilde{f}(0)=0$ so

$$\begin{aligned}\mathcal{D}\tilde{f}|_0&=\mathcal{D}A_2|_{f(A_1(x))}\circ\mathcal{D}f|_{A_1(0)}\circ\mathcal{D}A_1|_0\\ &=L^{-1}\circ L\circ\iota\\ &=\iota\end{aligned}$$

$\tilde{f}: A_1^{-1}(U)\to\mathbb{R}^n$ satisfies all requirements we used above and thus there exists $\widetilde{U}_1\subseteq A_1^{-1}(U)$ such that $\tilde{f}|_{\widetilde{U}_1}:\widetilde{U}_1\to\tilde{f}(\widetilde{U}_1)$ is a diffeomorphism. All that's left is to write the correct subset on which the functions are defined.

Let $U_1=A_1(\widetilde{U}_1)\subseteq U$. Then

$$f|_{U_1}=A_2^{-1}|_{\tilde{f}(A_1^{-1}(U_1))}\circ\tilde{f}|_{A_1^{-1}(U_1)}\circ A_1^{-1}|_{U_1}.$$

Done! □

Remark. The proof of differentiability of g is non-examinable.

6.2 Implicit Function Theorem

Question. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is C^1 and $y_0 \in \mathbb{R}^m$. What can I say about

$$f^{-1}(y_0),$$

the set of solutions to $f(x) = y_0$?

Let's begin with a simple example. Let

$$F : \mathbb{R}^2 \to \mathbb{R}$$
$$(x_1, x_2) \mapsto x_1^2 - x_2^2$$

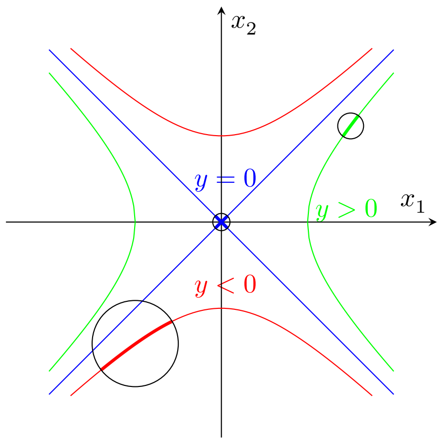


Figure 1: Level sets for different y's

For most $x \in \mathbb{R}^2$, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \cap F^{-1}(F(x))$ is the image of a parameterised curve. But something different happens at the origin: two curves intersect there and there is no neighbourhood of the origin in which it looks like a curve.

Question. What is different about $(0, 0)$?

The answer lies in the derivative of the map. $\mathcal{D}F|_{(x_1,x_2)} = (2x_1, -2x_2) \in L(\mathbb{R}^2, \mathbb{R})$ is surjective everywhere except $(0, 0)$.

Theorem 6.11. *Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is C^1, $x_0 \in \mathbb{R}^n$ and $\mathcal{D}f|_{x_0} \in L(\mathbb{R}^n, \mathbb{R}^m)$ is* surjective. *Then there is an open subset $U \subseteq \mathbb{R}^n$, $x_0 \in U$ and diffeomorphism $h : U \to U'$ such that*

$$f(x) = (h_1(x), \dots, h_m(x))$$

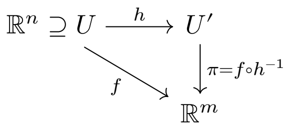
for $x \in U$.

The theorem says that there are local coordinates $(y_1, \dots, y_n)$ on $\mathbb{R}^n$ near x_0 such that $y_i = h_i(x)$, with respect to which f is the projection of the first m coordinates.

Proof. Let $L_1 = \mathcal{D}f|_{x_0} \in L(\mathbb{R}^n, \mathbb{R}^m)$. L_1 is surjective so $\dim \ker L_1 = n - m$ by Rank-nullity. Pick $L_2 \in L(\mathbb{R}^n, \mathbb{R}^{n-m})$ such that $\ker L_2 \cap \ker L_1 = 0$. There are many possibilities for L_2, for example the orthogonal projection onto $\ker L_1 \cong \mathbb{R}^{n-m}$. Define

$$\begin{aligned} h : \mathbb{R}^n &\to \mathbb{R}^m \times \mathbb{R}^{n-m} \cong \mathbb{R}^n \\ x &\mapsto (f(x), L_2(x)) \end{aligned}$$

Then the derivative $\mathcal{D}h|_{x_0} : \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^{n-m}$ is given by $(\mathcal{D}f|_{x_0}, \mathcal{D}L_2|_{x_0}) = (L_1, L_2)$.



If $v \in \ker \mathcal{D}h|_{x_0}$ then $v \in \ker L_1 \cap \ker L_2 = 0$ so $v = 0$. That is to say $\mathcal{D}h|_{x_0} \in L(\mathbb{R}^n, \mathbb{R}^n)$ is injective and thus invertible. Thus by Inverse Function Theorem there exists $U \subseteq \mathbb{R}^n$ open, $x_0 \in U$ such that $h|_U$ is a diffeomorphism. By construction

$$f(x) = (h_1(x), \dots, h_m(x)).$$

□

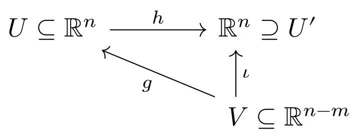
Corollary 6.12. *With f and x_0 as above, there is an open set $V \subseteq \mathbb{R}^{n-m}$ and an injective map $g : V \to \mathbb{R}^n$ such that*

$$f^{-1}(y_0) \cap U = g(V)$$

where $y_0 = f(x_0)$. That is to say $f^{-1}(y_0)$ is locally the image of a function $g : V \to \mathbb{R}^n$.

Proof. Consider the map

$$\begin{aligned} \iota : \mathbb{R}^{n-m} &\to \mathbb{R}^n \cong \mathbb{R}^m \times \mathbb{R}^{n-m} \\ z &\mapsto (y_0, z) \end{aligned}$$



Let $V = \iota^{-1}(U')$, which is open since ι is continuous. $g : V \to \mathbb{R}^n$ is given by $g = h^{-1} \circ \iota$. Then g is injective since both h^{-1} and ι are, and $f(x) = y_0$ if and only if $h(x) = (y_0, z)$, if and only if $x \in \operatorname{Im} g$. □

We have the tools ready for (stating) our theorem. Before that let's introduce some terminologies. Let $f : \mathbb{R}^n \to \mathbb{R}^m$.

Definition (Critical point)**.** $x \in \mathbb{R}^n$ is a *critical point* of f if $\mathcal{D}f|_x$ is not surjective.

Definition (Critical value). $y \in \mathbb{R}^m$ is a *critical value* of f if any $x \in f^{-1}(y)$ is a critical point.

Definition (Singular). If $x \in f^{-1}(y)$ is a critical point then x is a *singular point* of $f^{-1}(y)$.

If y is a critical value of f then $f^{-1}(y)$ is *singular*.

And the one final definition:

Definition (Manifold). An *n-manifold* is a metric space X such that any $p \in X$ has an open neighbourhood U homeomorphic to an open subset of $\mathbb{R}^n$.

With these fancy languages, our previous result could be succintly stated as follow:

Theorem 6.13 (Global Implicit Function Theorem). *If $y \in \mathbb{R}^m$ is not a critical value of f, $f^{-1}(y)$ is an $(n-m)$-dimensional manifold.*

Remark. In differential geometry, Sard's Theorem asserts that in a sense, most points in $\mathbb{R}^m$ are not critical values.

6.3 Solving ODEs

Given $\mathbf{V} : \mathbb{R}^n \to \mathbb{R}^n$ which is C^1 (think of it as a vector field on $\mathbb{R}^n$) and $\mathbf{x}_0 \in \mathbb{R}^n$, the main question in the section is to find a map $\mathbf{x} : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ for some $\varepsilon > 0$ satisfying

$$\frac{d\mathbf{x}}{dt} = \mathbf{V}(\mathbf{x}(t)),\ \mathbf{x}(0) = \mathbf{x}_0. \tag{$*$}$$

Example. Solve

$$\begin{cases} x_1'(t) = x_2 \sin(x_1 x_2^2) \\ x_2'(t) = e^{x_1^2 + x_2^3} \end{cases}$$

with initial conditions

$$\begin{cases} x_1(0) = 0 \\ x_2(0) = 0 \end{cases}$$

Let's set up a physical model for $n = 2$. Let $\mathbb{R}^2$ be the surface of the ocean. $\mathbf{V}(\mathbf{p})$ is the velocity of the current at position $\mathbf{p}$. If I drop a rubber duck in at position $\mathbf{x}_0$ at time $t = 0$, then $\mathbf{x}(t)$ is the position of the duck at time t. Then ($*$) says that $\frac{d\mathbf{x}}{dt}$, the velocity of the duck at time t, equals to $\mathbf{V}(\mathbf{x}(t))$, the velocity of the current at the duck's position, i.e. the duck moves at the same speed as the current. For this reason, solutions to ($*$) are often called *flowlines* of $\mathbf{V}$.

The tangent vector to $\mathbf{x}(t)$ at time t is $\mathbf{x}'(t) = \mathbf{V}(\mathbf{x}(t))$, i.e. the flowline $\mathbf{x}(t)$ is everywhere tangent to $\mathbf{V}$.

Remark.

1. If $\|\mathbf{V}(\mathbf{p})\|$ grows rapidly with $\|\mathbf{p}\|$ then the duck may escape to infinity in finite time. So, the physical model being a model, shouldn't be taken too literally. For example for $n = 1$, the problem

$$\frac{dx}{dt} = x^2,\ x(0) = 1$$

has a solution $x = \frac{1}{1-t}$. There is no solution on $[0, t)$ for any $t > 1$. This hints that we may not be able to get global solution, but only a local solution in an ε-neighbourhood of certain points.

2. Argurably a more realistic physical model would allow $\mathbf{V}$ to have time-dependence, i.e.

$$\mathbf{V} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$$

and solve

$$\frac{d\mathbf{x}}{dt} = \mathbf{V}(t, \mathbf{x}), \tag{$**$}$$

However, there is a cheap trick that reduces the problem to the previous form. Consider $y_0(t), y_1(t), \dots, y_n(t)$ which are components of $\mathbf{y} : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$, satisfying

$$\begin{aligned} y_0'(t) &= 1 \\ y_1'(t) &= V_1(y_0(t), \dots, y_n(t)) \\ &\vdots \\ y_n'(t) &= V_n(y_0(t), \dots, y_n(t)) \end{aligned}$$

with initial conditions

$$\begin{aligned} y_0(0) &= 0 \\ (y_1(0), \dots, y_n(0)) &= \mathbf{x}_0 \end{aligned}$$

This is an equation of type $(*)$ with $\widetilde{\mathbf{V}}(\mathbf{p}) = (1, \mathbf{V}(\mathbf{p}))$. Any solution satisfies $y_0(t) = t$ so $\mathbf{x}(t) = (y_1(t), \dots, y_n(t))$ is a solution to $(**)$. Therefore we can reduce the problem of solving equations of type $(**)$ to the problem of solving equations of type $(*)$.

To solve this type of equation, we will use Contraction Mapping Theorem. Recall that we need two things: a complete metric space X and a contraction $F : X \to X$.

Take

$$X_\varepsilon = \{\mathbf{x} : [-\varepsilon, \varepsilon] \to \mathbb{R}^n : \mathbf{x} \text{ is continuous}\}$$

where ε is to be determined. This is a normed vector space with

$$\|\mathbf{x}\| = \max_{t \in [-\varepsilon, \varepsilon]} \|\mathbf{x}(t)\|_2 = \big\|\|\mathbf{x}(\cdot)\|_2\big\|_\infty$$

We showed back in Theorem 3.16 that $C[a, b]$ is complete with respect to $\|\cdot\|_\infty$.

Corollary 6.14. *X_ε is complete.*

Proof. Suppose $(\mathbf{x}_k)$ is a Cauchy sequence in X_ε. Write $\mathbf{x}_k(t) = (x_{k,1}(t), \dots, x_{k,n}(t))$. Then for $1 \le i \le n$, $(x_{k,i})$ is Cauchy in $C[-\varepsilon, \varepsilon]$. Thus there exists $y_i \in C[-\varepsilon, \varepsilon]$ such that $(x_{k,i}) \to y_i$. Then $(\mathbf{x}_k) \to \mathbf{y} = (y_1(t), \dots, y_n(t))$ in X_ε so X_ε is complete. □

Before we look for a contraction map note that the equation $\mathbf{x}'(t) = \mathbf{V}(\mathbf{x}(t))$ is *not* good for iteration since differentiation is a map $C^r \to C^{r-1}$. Instead, consider the integral equation

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t \mathbf{V}(\mathbf{x}(s))ds \tag{†}$$

Proposition 6.15. *If* $\mathbf{x} \in X_\varepsilon$, $\mathbf{x}$ *satisfies* (†) *if and only if* $\mathbf{x}$ *satisfies* (∗).

Proof. Suppose $\mathbf{x} \in X_\varepsilon$. Then $\mathbf{V} \circ \mathbf{x} : [-\varepsilon, \varepsilon] \to \mathbb{R}^n$ is continuous so the function

$$\mathbf{y}(t) = \mathbf{x}_0 + \int_0^t \mathbf{V}(\mathbf{x}(s))ds$$

is well-defined. By Fundamental Theorem of Calculus

$$\mathbf{y}'(t) = \mathbf{V}(\mathbf{x}(t)).$$

Thus if $\mathbf{x}$ satisfies (†) then $\mathbf{x}'(t) = \mathbf{V}(\mathbf{x}(t))$ so $\mathbf{x}$ satisfies (∗).

Conversely, if $\mathbf{x}'(t) = \mathbf{V}(\mathbf{x}(t))$, $\mathbf{x}$ is differentiable and thus continuous so $\mathbf{V} \in X_\varepsilon$. Moreover $\mathbf{x}'(t) = \mathbf{V}(\mathbf{x}(t))$ is continuous (since $\mathbf{V} \circ \mathbf{x}$ is) so $\mathbf{x}(t)$ is C^1. Integrating both sides of (∗) gives (†). □

Consider the map

$$\begin{aligned} F &: X_\varepsilon \to X_\varepsilon \\ F(\mathbf{x})(t) &= \mathbf{x}_0 + \int_0^t \mathbf{V}(\mathbf{x}(s))ds \end{aligned}$$

then $\mathbf{x}$ solves (†) if and only if $\mathbf{x}$ is a fixed point of F. Now there is only one question left: when is F a contraction map?

Proposition 6.16. *If* $\mathbf{V}$ *is K-Lipschitz the F is $K\varepsilon$-Lipschitz.*

Thus if $\mathbf{V}$ is Lipschitz, taking ε small enough guarantees that F is a contraction.

Proof. For $t \in [-\varepsilon, \varepsilon]$,

$$\begin{aligned}
\|F(\mathbf{x}_1)(t) - F(\mathbf{x}_2)(t)\|_2 &= \left\| \int_0^t \mathbf{V}(\mathbf{x}_1(s)) - \mathbf{V}(\mathbf{x}_2(s)) ds \right\|_2 \\
&\le \int_0^t \|\mathbf{V}(\mathbf{x}_1(s)) - \mathbf{V}(\mathbf{x}_2(s))\|_2 ds \\
&\le \int_0^t K\|\mathbf{x}_1(s) - \mathbf{x}_2(s)\|_2 ds \\
&\le Kt \max_{s\in[0,t]} \|\mathbf{x}_1(s) - \mathbf{x}_2(s)\|_2 \\
&\le Kt \max_{s\in[-\varepsilon,\varepsilon]} \|\mathbf{x}_1(s) - \mathbf{x}_2(s)\|_2 \\
&= Kt\|\mathbf{x}_1 - \mathbf{x}_2\|
\end{aligned}$$

so

$$\|F(\mathbf{x}_1) - F(\mathbf{x}_2)\| = \max_{t\in[-\varepsilon,\varepsilon]} \|F(\mathbf{x}_1)(t) - F(\mathbf{x}_2)(t)\|_2 \le K\varepsilon\|\mathbf{x}_1 - \mathbf{x}_2\|$$

□

Corollary 6.17. *If* $\mathbf{V}$ *is K-Lipschitz then there exists a unique solution to* $(*)$ *on* $[-\varepsilon, \varepsilon]$ *for any* $\varepsilon < 1/K$.

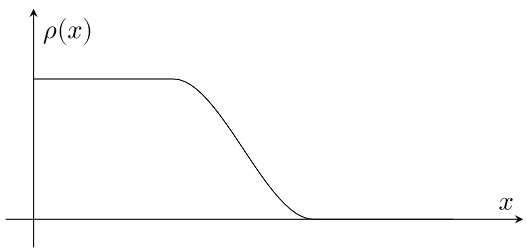
Proof. If $\varepsilon < 1/K$ then $F : X_\varepsilon \to X_\varepsilon$ is a contraction. X_ε is complete so F has a unique fixed point so $(*)$ has a unique solution. □

This is a nice result but not quite the answer to what we have asked at the beginning of the section since $\mathbf{V}$, being C^1, may not be Lipschitz in general.

Question. What if $\mathbf{V}$ is C^1 but not Lipschitz?

As a common trick, choose a *cut-off function* $\rho : [0, \infty) \to \mathbb{R}$ which is C^1 and satisfies

$$\rho(x) = \begin{cases} 1 & x \le 1 \\ 0 & x \ge 2 \end{cases}$$



Now let $\widetilde{\mathbf{V}}(\mathbf{p}) = \mathbf{V}(\mathbf{p})\rho(\|\mathbf{p} - \mathbf{x}_0\|)$. Intuitively this weight ρ localises $\mathbf{V}$ to the ball $\overline{B}_1(\mathbf{x}_0)$.

Lemma 6.18. $\widetilde{\mathbf{V}}$ *is Lipschitz.*

Proof. $\widetilde{\mathbf{V}}$ is C^1 so the map

$$\begin{aligned} \mathbb{R}^n &\to L(\mathbb{R}^n, \mathbb{R}^n) \to \mathbb{R}^n \\ \mathbf{p} &\mapsto \mathcal{D}\widetilde{\mathbf{V}}|_{\mathbf{p}} \mapsto \left\|\mathcal{D}\widetilde{\mathbf{V}}|_{\mathbf{p}}\right\|_{\text{op}} \end{aligned}$$

is continuous. $\overline{B}_2(\mathbf{x}_0)$ is compact so there exists M such that $\left\|\mathcal{D}\widetilde{\mathbf{V}}|_{\mathbf{p}}\right\|_{\text{op}} \leq M$ for all $\mathbf{p} \in \overline{B}_2(\mathbf{x}_0)$. On the other hand $\widetilde{\mathbf{V}}(\mathbf{p}) = 0$ for $\mathbf{p} \notin \overline{B}_2(\mathbf{x}_0)$ so $\mathcal{D}\widetilde{\mathbf{V}}|_{\mathbf{p}} = 0$. Thus $\left\|\mathcal{D}\widetilde{\mathbf{V}}|_{\mathbf{p}}\right\|_{\text{op}} \leq M$ for all $\mathbf{p} \in \mathbb{R}^n$ so $\widetilde{\mathbf{V}}$ is M-Lipschitz. □

By the corollary, the equation

$$\tilde{\mathbf{x}}'(t) = \widetilde{\mathbf{V}}(\tilde{\mathbf{x}}(t)),\ \tilde{\mathbf{x}}(0) = \mathbf{x}(0) = \mathbf{x}_0 \qquad (***)$$

has a unique solution on $[-\varepsilon, \varepsilon]$ when $\varepsilon < 1/M$. Since $\widetilde{\mathbf{V}}$ is continuous, there exists M' such that $\|\widetilde{\mathbf{V}}(\mathbf{p})\|_2 \leq M'$ for all $\mathbf{p} \in \overline{B}_2(\mathbf{x}_0)$, i.e. $\|\widetilde{\mathbf{V}}(\mathbf{p})\| \leq M'$ for all $\mathbf{p} \in \mathbb{R}^n$. Choose $\varepsilon < \min\{1/M, 1/M'\}$.

Lemma 6.19. *With ε and $\tilde{\mathbf{x}}$ as above, $\|\tilde{\mathbf{x}}(t) - \mathbf{x}_0\|_2 < 1$ for all $t \in [-\varepsilon, \varepsilon]$.*

Proof.

$$\begin{aligned} \|\tilde{\mathbf{x}}(t) - \mathbf{x}_0\|_2 &= \left\|\int_0^t \widetilde{\mathbf{V}}(\tilde{\mathbf{x}}(s)) ds\right\|_2 \\ &\leq \left|\int_0^t \|\widetilde{\mathbf{V}}(\tilde{\mathbf{x}}(s))\|_2 ds\right| \\ &\leq M'|t| \\ &\leq M'\varepsilon \\ &< 1 \end{aligned}$$

□

So for $t \in [-\varepsilon, \varepsilon]$,

$$\begin{aligned} \tilde{\mathbf{x}}'(t) &= \widetilde{\mathbf{V}}(\tilde{\mathbf{x}}(t)) \\ &= \mathbf{V}(\tilde{\mathbf{x}}(t))\rho(\|\tilde{\mathbf{x}}(t) - \mathbf{x}_0\|_2) \\ &= \mathbf{V}(\tilde{\mathbf{x}}(t)) \end{aligned}$$

i.e. $\tilde{\mathbf{x}}$ solves $(*)$.

Conversely, a solution to $(*)$ on $[-\varepsilon, \varepsilon]$ gives a solution to $(***)$.

In summary, we have proved that

Theorem 6.20. *If $\mathbf{V} : \mathbb{R}^n \to \mathbb{R}^n$ is C^1 and $\mathbf{x}_0 \in \mathbb{R}^n$, there exists $\varepsilon > 0$ such that there is a unique $\mathbf{x} : [-\varepsilon, \varepsilon] \to \mathbb{R}^n$ satisfying*

$$\mathbf{x}'(t) = \mathbf{V}(\mathbf{x}(t)),\ \mathbf{x}(0) = \mathbf{x}_0.$$

Remark.

1. If $\mathbf{V}$ is Lipschitz then there exists a unique solution $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$, which solves the equation for all time t. See example sheet.

2. If $\mathbf{V}$ is continuous it can still be shown that solutions exist.

3. If $\mathbf{V}$ is not Lipschitz, however, the solutions may not be unique. For example consider

$$x'(t) = \frac{3}{2}x^{1/3},\ x(0) = 0.$$

Then $x(t) = 0$ and $x(t) = t^{3/2}$ are both solutions. The reason is that $V'(y) = \frac{1}{3}y^{-2/3}$ is not bounded and thus V is not Lipschitz.

6.4 General ODEs

We end our course by a brief disussion on general ODEs. In general, an ODE may not have a solution, for (a stupid) example

$$(x'(t))^2 = -1,\ x(0) = 1$$

or have many solutions

$$(x'(t))^2 = 1,\ x(0) = 0.$$

But we can reduce a general ODE into a form we are familiar with, namely first order equation:

1. Eliminate t (as a term in the equation) at the cost of adding an extra variable. We have already shown how to do this before. For example,

$$t^2(y')^2 + y^3 = \sin t$$

can be rewritten as

$$\begin{cases} x_2(t)^2(x_1')^2 + x_1^3 = \sin x_2 \\ x_2' = 1 \end{cases}$$

2. Eliminate higher derivatives by adding extra variables. For example,

$$y''y + (y')^3 = 1$$

can be rewritten as

$$\begin{cases} x_2'x_1 + x_2^3 = 1 \\ x_1' = x_2 \end{cases}$$

3. After the previous two steps, we have reduced the problem to the form

$$F(\mathbf{y}, \mathbf{y}') = 0,\ \mathbf{y}(0) = \mathbf{y}_0.$$

At this stage, we can try to

(a) look for solutions to $F(\mathbf{y}_0, \mathbf{z}_0) = 0$, which is an algebraic problem,

(b) and then use Global Implicit Function Theorem to find a function $G : B_\varepsilon(\mathbf{y}_0) \to \mathbb{R}^n$ such that

$$\begin{cases} G(\mathbf{y}_0) = \mathbf{z}_0 \\ F(\mathbf{y}, G(\mathbf{y})) = 0 \end{cases}$$

If I find such a G, then solutions to

$$\mathbf{y}'(t) = G(\mathbf{y}(t)),\ \mathbf{y}(0) = \mathbf{y}_0$$

will be solutions to $F(\mathbf{y}, \mathbf{y}') = 0$.

Index