

# MA 200: Multivariable Calculus

Naman Mishra

August 2024

# Contents

<b>1</b>	<b>Linear algebra</b>	<b>4</b>
1.1	Matrix norms . . . . .	11
<b>2</b>	<b>Differentiation</b>	<b>27</b>

# Lectures

1	Fri, August 2	Norms and equivalence . . . . .	3
2	Mon, August 5	One norm to rule them all . . . . .	7
3	Wed, August 7	Big-Oh and matrix norms . . . . .	9
4	Fri, August 9	Topology? . . . . .	15
5	Mon, August 12	Continuity of the inverse; differentiation . . .	25
6	Mon, August 19	Differentiation in $\mathbb{R}^n$ . . . . .	28
7	Wed, August 21	. . . . .	29
8	Fri, August 23	. . . . .	31
9	Sun, August 25	. . . . .	33

# Assignments

1	due August 12 . . . . .	16
2	due August 12 . . . . .	36

# The course

## Grading

- Homework: 20%
- Quizzes: 20%
- Midterm: 20%
- Final: 40%

## Textbooks

- 

**Lecture 1.**  
Friday  
August 2

# Chapter 1

## Linear algebra

**Definition 1.1** (homogeneous function). Let  $V$  be a vector space over  $\mathbb{R}$ . A function  $f: V \setminus \{0\} \rightarrow \mathbb{R}$  is called a *homogeneous function* of degree  $k$  if

$$f(rx) = r^k f(x)$$

for each  $x \in V \setminus \{0\}$  and  $r > 0$ .

*Remarks.*

- If  $f$  and  $g$  are homogeneous functions of degree  $k$  and  $l$  respectively, then  $f \cdot g$  is homogeneous of degree  $k + l$  and  $f/g$  is homogeneous of degree  $k - l$  (provided  $g$  is never zero).
- $f \equiv 0$  is homogeneous of any degree.

**Definition 1.2** (norm). Let  $V$  be a vector space over  $\mathbb{R}$ . A norm  $\|\cdot\|$  on  $V$  is a function from  $V$  to  $\mathbb{R}$  that satisfies

(N1) (Positivity)  $\|x\| \geq 0$  for any  $x \in V$ .

(N2) (Definiteness)  $\|x\| = 0$  iff  $x = 0$ .

(N3) (Homogeneity)  $\|rx\| = |r|\|x\|$  for any  $x \in V$  and  $r \in \mathbb{R}$ .

(N4) (Triangle inequality)  $\|x + y\| \leq \|x\| + \|y\|$  for any  $x, y \in V$ .

**Definition 1.3** (normed linear space). A vector space  $V$  equipped with a norm  $\|\cdot\|$  is called a *normed linear space*.

*Remark.* Any normed linear space  $(V, \|\cdot\|)$  can be given a metric space structure by defining the distance  $d(x, y)$  between  $x, y \in V$  as  $\|x - y\|$ .

The set  $B(x, r) := \{y \in V \mid \|x - y\| < r\}$  is called the open ball of radius  $r$  centered at  $x$ .

The set  $S(x, r) := \{y \in V \mid \|x - y\| = r\}$  is called the sphere of radius  $r$  centered at  $x$ .

**Exercise 1.4** (reverse triangle inequality). *Let  $V$  be a normed linear space. Show that*

$$\left| \|x\| - \|y\| \right| \leq \|x - y\| \quad (1.1)$$

for any  $x, y \in V$ .

*Proof.* First observe from homogeneity (N3) that  $\|x\| = \|-x\|$  for any  $x \in V$ . Next, from the triangle inequality (N4) we have

$$\|x\| \leq \|x - y\| + \|y\|$$

so that

$$\|x\| - \|y\| \leq \|x - y\|.$$

Similarly,

$$\|y\| \leq \|y - x\| + \|x\|$$

so that

$$-\|x - y\| \leq \|x\| - \|y\|.$$

Combining these gives the result. ■

This shows that  $f = x \mapsto \|x\|$  is a (Lipschitz) continuous function on  $V$ .

**Definition 1.5** (metric space). A *metric space* is a set  $X$  equipped with a function  $d: X \times X \rightarrow \mathbb{R}$  called a *metric* that satisfies the following properties:

(M1)  $d(x, y) \geq 0$  for any  $x, y \in X$ .

(M2)  $d(x, y) = 0$  iff  $x = y$ .

(M3)  $d(x, y) = d(y, x)$  for any  $x, y \in X$ .

(M4)  $d(x, z) \leq d(x, y) + d(y, z)$  for any  $x, y, z \in X$ .

**Exercise 1.6** (self). *Show that any normed linear space  $(V, \|\cdot\|)$  is a metric space under the distance  $d(x, y) = \|x - y\|$ .*

*Proof.* (M1) and (M2) are immediate from (N1) and (N2). (N3) implies (M3) by scaling by  $-1$ . Triangle implies triangle. ■

**Definition 1.7** (continuity). Let  $(X, d)$  and  $(Y, \rho)$  be metric spaces. A function  $f: X \rightarrow Y$  is called *continuous* at  $a \in X$  iff

$$\begin{aligned} x_n \rightarrow a &\implies f(x_n) \rightarrow f(a), \text{ or} \\ d(x_n, a) \rightarrow 0 &\implies \rho(f(x_n), f(a)) \rightarrow 0 \end{aligned}$$

**Exercise 1.8** (product metric spaces). *Let  $(X_1, d_1)$  and  $(X_2, d_2)$  be metric spaces. Let  $d: X_1 \times X_2 \rightarrow \mathbb{R}$  be defined by*

$$d((x_1, x_2), (y_1, y_2)) := d_1(x_1, y_1) + d_2(x_2, y_2).$$

Show that  $d$  is a metric on  $X_1 \times X_2$ .

Let  $(z_n)_{n \in \mathbb{N}} = ((x_n, y_n))_{n \in \mathbb{N}}$  be a sequence in  $X_1 \times X_2$ . Show that  $z_n \rightarrow (x, y)$  iff  $x_n \rightarrow x$  and  $y_n \rightarrow y$ .

*Proof.* Suppose  $x_n \rightarrow x$  and  $y_n \rightarrow y$ . That is,  $d_1(x_n, x) \rightarrow 0$  and  $d_2(y_n, y) \rightarrow 0$ . Thus  $d_1(x_n, x) + d_2(y_n, y) \rightarrow 0$ .

Conversely if  $d_1(x_n, x) + d_2(y_n, y) \rightarrow 0$  and each is nonnegative, then  $d_1(x_n, x) \rightarrow 0$  and  $d_2(y_n, y) \rightarrow 0$ . ■

*Remark.*  $\tilde{d}$  given by

$$\tilde{d}((x_1, x_2), (y_1, y_2)) := \min\{d_1(x_1, y_1), d_2(x_2, y_2)\}$$

is *not* a metric on  $X_1 \times X_2$  as it fails definiteness.

However,  $\max\{d_1, d_2\}$  is a metric.

**Exercise 1.9.** Let  $(V, \|\cdot\|)$  be a normed linear space.

- The addition map  $(x, y) \mapsto x + y$  is a continuous map from  $V \times V$  to  $V$ .
- The scalar multiplication map  $(\alpha, x) \mapsto \alpha x$  is continuous from  $\mathbb{R} \times V$  to  $V$ .

*Solution.*

- $\|x' + y' - (x + y)\| \leq \|x' - x\| + \|y' - y\| = \|(x', y') - (x, y)\|$ .
- $\|\alpha' x' - \alpha x\| \leq \|\alpha' x' - \alpha x'\| + \|\alpha x' - \alpha x\| = |\alpha' - \alpha| \|x'\| + |\alpha| \|x' - x\|$ .

Thus choosing  $\delta = \varepsilon / \max\{|\alpha|, \|x\|\}$  gives

$$\|\alpha' x' - \alpha x\| \leq \max\{|\alpha|, \|x\|\} (|\alpha' - \alpha| + \|x' - x\|) < \varepsilon$$

whenever  $|\alpha' - \alpha| + \|x' - x\| < \delta$ .

Repeated in problem 1.1. ■

*Examples.*

- $(\ell^p \text{ norm}) \mathbb{R}^n$  with  $p \in [1, \infty]$  and

$$\|x\|_p := \left( |x_1|^p + \cdots + |x_n|^p \right)^{1/p}$$

where

$$\|x\|_\infty := \max\{|x_1|, \dots, |x_n|\}$$

is the limit of the  $\ell^p$  norms as  $p \rightarrow \infty$ .

**Exercise 1.10.** See problem 1.6.

**Definition 1.11** (norm equivalence). Let  $\|\cdot\|_a$  and  $\|\cdot\|_b$  be two norms on  $V$ . We say that  $\|\cdot\|_a$  and  $\|\cdot\|_b$  are *equivalent* if there exist  $c_1, c_2 > 0$  such that

$$c_1\|x\|_a \leq \|x\|_b \leq c_2\|x\|_a$$

for all  $x \in V$ . We write  $\|\cdot\|_a \sim \|\cdot\|_b$ .

**Exercise 1.12.** Check that  $\sim$  is an equivalence relation.

*Solution.* Reflexivity is obvious. Symmetry is since

$$c_1\|x\|_a \leq \|x\|_b \leq c_2\|x\|_a \implies \frac{1}{c_2}\|x\|_b \leq \|x\|_a \leq \frac{1}{c_1}\|x\|_b.$$

For transitivity, let

$$\begin{aligned} c_1\|x\|_a &\leq \|x\|_b \leq c_2\|x\|_a, \\ c_3\|x\|_b &\leq \|x\|_c \leq c_4\|x\|_b. \end{aligned}$$

Then

$$c_1c_3\|x\|_a \leq \|x\|_c \leq c_2c_4\|x\|_a. \quad \blacksquare$$

**Proposition 1.13.** Equivalent norms induce the same topology. That is, let  $\|\cdot\|_a \sim \|\cdot\|_b$ . Then a set is open (resp. compact) under  $\|\cdot\|_a$  iff it is open (resp. compact) under  $\|\cdot\|_b$ .

**Lecture 2.**

Monday

August 5

*Proof.* Suppose  $c_1\|x\|_a \leq \|x\|_b \leq c_2\|x\|_a$ .

Let  $U \subseteq V$  be open under  $\|\cdot\|_a$ . Let  $x \in U$ . There exists  $\varepsilon > 0$  such that  $\|y - x\|_a < \varepsilon \implies y \in U$ . But then  $\|y - x\|_b < c_1\varepsilon \implies y \in U$ . Thus  $U$  is open under  $\|\cdot\|_b$ .

Compactness follows from openness. ■

**Proposition 1.14.** Every  $\ell^p$  norm is equivalent to  $\ell^\infty$ .

*Proof.* Let  $x \in \mathbb{R}^n$ . Then  $\|x\|_\infty \leq \|x\|_p \leq n^{\frac{1}{p}}\|x\|_\infty$ . ■

The usual topology on  $\mathbb{R}^n$  is the one induced by the Euclidean norm. This norm itself is induced by the inner product  $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ . Using Cauchy-Schwarz, we can define the angle between two vectors  $x, y \in \mathbb{R}^n$  to be

$$\cos^{-1}\left(\frac{\langle x, y \rangle}{\|x\|\|y\|}\right).$$

**Proposition 1.15.** Let  $\|\cdot\|$  be any norm on  $\mathbb{R}^n$ . Then the function  $x \mapsto \|x\|$  is Lipschitz continuous with respect to the Euclidean topology.



*Proof.*

$$\begin{aligned}\|x\| &= \left\| \sum x_i e_i \right\| \\ &\leq \sum |x_i| \|e_i\| \\ &\leq M \|x\|_2\end{aligned}$$

where  $M = \sum \|e_i\|$ .

The reverse triangle inequality gives

$$\begin{aligned}|\|x\| - \|y\|| &\leq \|x - y\| \\ &\leq M \|x - y\|_2.\end{aligned}$$

■

**Theorem 1.16.** *Any two norms on  $\mathbb{R}^n$  are equivalent.*

*Proof.* Let  $\|\cdot\|$  be any norm on  $\mathbb{R}^n$ . Then  $x \mapsto \|x\|$  is continuous with respect to  $\|\cdot\|_2$ . Let

$$S(0, 1)_{\|\cdot\|_2} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\} = S^{n-1}.$$

$\|\cdot\|$  attains a minimum and a maximum on  $S^{n-1}$  by compactness. Thus there exist positive constants  $c_1, c_2$  such that

$$c_1 \leq \|x\| \leq c_2$$

for all  $x \in S^{n-1}$ .

Now for any  $x \in \mathbb{R}^n \setminus \{0\}$ , dividing by  $\|x\|_2$  gives a point that lies on  $S^{n-1}$ . Thus

$$c_1 \leq \left\| \frac{x}{\|x\|_2} \right\| \leq c_2.$$

By homogeneity (N3),

$$c_1 \|x\|_2 \leq \|x\| \leq c_2 \|x\|_2.$$

This is also trivially true for  $x = 0$ .

Thus  $\|\cdot\| \sim \|\cdot\|_2$ .

■

*Remark.* The idea of the proof is as follows.

Any homogenous function is determined by its value on the unit sphere. A homogenous function of degree *zero* is essentially nothing but a function on the unit sphere ( $f(v) = f(\hat{v})$ ).

The function  $x \mapsto \frac{\|x\|}{\|x\|_2}$  is a continuous homogenous function on degree 0. The unit sphere is known to be compact under the Euclidean norm (and every other, but not before we complete the proof). Thus

$$c_1 \leq \frac{\|x\|}{\|x\|_2} \leq c_2$$

for some positive constants  $c_1, c_2$ .

Definiteness and  $\triangle$  are required for the ratio to be continuous. Homogeneity is required for it to be homogenous. Is positivity required?

*Remark.* We technically only need to show  $c_1\|x\|_2 \leq \|x\|$ , since the other inequality is proven in the previous proof. It is nonetheless clearer to show both inequalities.

**Exercise 1.17** (Self). Show that (N1) follows from (N3) and (N4).

*Solution.* Let  $v \in V$ . By triangle inequality,  $\|v\| = \|-v + 2v\| \leq \|-v\| + \|2v\|$ . By homogeneity, this is  $3\|v\|$ . Thus  $\|v\| \leq 3\|v\|$ , so  $\|v\| \geq 0$ . ■

*Remarks* (Finite-dimensional vector spaces).

- Let  $V$  be a vector space over  $\mathbb{R}$  with dimension  $n < \infty$ . Using a basis for  $V$ , any norm on  $V$  induces a norm on  $\mathbb{R}^n$ , and vice versa. Norms on  $V$  are in a one-to-one correspondence with norms on  $\mathbb{R}^n$ .
- Thus any two norms on  $V$  are equivalent.
- Any two inner products on  $V$  will also be equivalent due to this.
- Any finite-dimensional vector space over  $\mathbb{R}$  is complete.

**Exercise 1.18.** Let  $f: \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$  be given by  $f(x) = \frac{1}{x}$ . Show that  $f$  is continuous. What is the key idea of your proof?

*Solution.* Let  $x_0 \in \mathbb{R} \setminus \{0\}$  and  $\varepsilon > 0$ . Choose  $\delta = \min\{\varepsilon \cdot \frac{1}{2}|x_0|^2, \frac{1}{2}|x_0|\}$ . Then for any  $x$  in the  $\delta$ -neighbourhood of  $x_0$ ,

$$\begin{aligned} |f(x) - f(x_0)| &= \left| \frac{1}{x} - \frac{1}{x_0} \right| \\ &= \frac{|x_0 - x|}{|x||x_0|} \\ &< \frac{\delta}{|x||x_0|} \\ &< \frac{2\delta}{|x_0|^2} \\ &\leq \varepsilon. \end{aligned}$$

*Remark.* The proof works by bounding  $\frac{1}{|x|}$ . The rest goes to zero as  $x \rightarrow a$ . We will do a similar proof in proposition 1.39.

On  $\mathbb{R}^n$ , we will always fix the  $\ell^2$ -norm

**Lecture 3.**  
Wednesday  
August 7

*Notation.*

$$L(\mathbb{R}^n, \mathbb{R}^m) = \{T: \mathbb{R}^n \rightarrow \mathbb{R}^m \mid T \text{ is linear}\}$$

and

$$M_{m \times n}(\mathbb{R}) \cong L(\mathbb{R}^n, \mathbb{R}^m)$$

using the isomorphism  $A \mapsto T_A$  where

$$\begin{aligned} T_A: \mathbb{R}^n &\rightarrow \mathbb{R}^m \\ v &\mapsto Av, \end{aligned}$$

where  $v$  is interpreted as a column vector. We will also write  $L(\mathbb{R}^n)$  for  $L(\mathbb{R}^n, \mathbb{R}^n)$ .

**Definition 1.19** (liminf and limsup). Let  $f: X \rightarrow \mathbb{R}$  be a function on a topological space  $X$ . We define the *limit inferior* and *limit superior* of  $f$  as

$$\begin{aligned} \liminf_{x \rightarrow a} f(x) &= \sup_V \inf_{x \in V} f(x) \\ \limsup_{x \rightarrow a} f(x) &= \inf_V \sup_{x \in V} f(x) \end{aligned}$$

where  $V$  ranges over all open neighbourhoods of  $a$  that contain at least one point other than  $a$ .

**Exercise 1.20** (self). Let  $(X, d)$  be a connected metric space with at least two points. Then

$$\begin{aligned} \liminf_{x \rightarrow a} f(x) &= \lim_{\varepsilon \searrow 0} \inf_{0 < d(x, a) < \varepsilon} f(x) \\ \limsup_{x \rightarrow a} f(x) &= \lim_{\varepsilon \searrow 0} \sup_{0 < d(x, a) < \varepsilon} f(x) \end{aligned}$$

*Proof.* We first need to show that  $\{x \in X \mid 0 < d(x, a) < \varepsilon\}$  is non-empty for each  $\varepsilon > 0$ . Suppose this were not the case for some  $\varepsilon$ . Then  $B(a, \varepsilon) = \{a\} = \overline{B(a, \varepsilon/2)}$  is clopen, contradicting the connectedness of  $X$ .

Notice that for any  $\varepsilon_1 > \varepsilon_2 > 0$ ,

$$\{x \in X \mid 0 < d(x, a) < \varepsilon_1\} \subseteq \{x \in X \mid 0 < d(x, a) < \varepsilon_2\},$$

so the infimum increases as  $\varepsilon \searrow 0$ . ■

**Definition 1.21** (*O* notation). Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $g: \mathbb{R}^n \rightarrow \mathbb{R}^k$ . We say that

(i)  $f(x) = o(g(x))$  as  $x \rightarrow a$  if

$$\lim_{x \rightarrow a} \frac{\|f(x)\|}{\|g(x)\|} = 0,$$

(ii)  $f(x) = O(g(x))$  as  $x \rightarrow a$  if

$$\limsup_{x \rightarrow a} \frac{\|f(x)\|}{\|g(x)\|} < \infty.$$

where the assumption is that  $g$  is non-zero in some neighbourhood of  $a$ .

**Exercise 1.22.** Show that the definition of  $O$  is equivalent to the following:

We say that  $f(x) = O(g(x))$  as  $x \rightarrow a$  if there exists an open neighbourhood  $V$  of  $a$  such that  $\frac{\|f(x)\|}{\|g(x)\|}$  is bounded on  $V$ .

*Solution.* Call the ratio  $h$ .

$$\begin{aligned} \inf_V \sup_{x \in V} h(x) \leq \infty &\iff \exists V (\sup_{x \in V} h(x) < \infty) \\ &\iff \exists V \exists M (\forall x \in V, h(x) \leq M) \end{aligned}$$

■

**Exercise 1.23.**

## 1.1 Matrix norms

**Definition 1.24** (Hilbert-Schmidt norm). For a matrix  $A \in M_{m \times n}(\mathbb{R})$ , we define the *Hilbert-Schmidt* or *Frobenius* norm by

$$\|A\|_{HS} = \left( \sum_{i,j} a_{ij}^2 \right)^{1/2}$$

**Exercise 1.25.** Show that  $\|A\|_{HS}^2 = \text{Tr}(A^\top A) = \text{Tr}(AA^\top)$ .

*Solution.*

$$\begin{aligned} (A^\top A)_{ii} &= \sum_k (A^\top)_{ik} A_{ki} \\ &= \sum_k a_{ki}^2 \\ \implies \text{Tr}(A^\top A) &= \sum_i \sum_k a_{ki}^2 \\ &= \|A\|_{HS}^2. \end{aligned}$$

Since  $\|A\|_{HS} = \|A^\top\|_{HS}$ , we also have  $\text{Tr}(AA^\top) = \|A\|_{HS}^2$ . ■

**Proposition 1.26.** *Any linear transformation is continuous.*

*Proof.* Let  $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ . Then

$$\begin{aligned} \|Tx\| &= \left\| T\left(\sum x_i e_i\right) \right\| \\ &= \left\| \sum x_i T e_i \right\| \\ &\leq \sum |x_i| \|T e_i\| \\ &\leq \|x\| \sum \|T e_i\| \\ &= M \|x\| \end{aligned} \tag{1.2}$$

where  $M = \|T e_1\| + \dots + \|T e_n\|$ .

Now

$$\|Tx - Ty\| = \|T(x - y)\| \leq M \|x - y\|$$

says that  $T$  is Lipschitz continuous with Lipschitz constant  $M$ . ■

We temporarily define two norms on  $M_{m \times n}(\mathbb{R})$ :

$$\begin{aligned} \|T\|_S &= \sup_{\|x\|=1} \|Tx\| \\ \|T\|_B &= \sup_{\|x\| \leq 1} \|Tx\| \end{aligned}$$

**Lemma 1.27.**  $\|T\|_S = \|T\|_B$ .

*Proof.* From the definition it is obvious that  $\|T\|_S \leq \|T\|_B$ . Now for any  $x \in \mathbb{R}^n \setminus \{0\}$ , let  $y = x/\|x\|$ .

$$\begin{aligned} \|Ty\| &\leq \|T\|_S \\ \frac{\|Tx\|}{\|x\|} &\leq \|T\|_S \\ \implies \|Tx\| &\leq \|T\|_S \|x\| \end{aligned}$$

Thus for  $\|x\| \leq 1$ , we have  $\|Tx\| \leq \|T\|_S$  (check 0 separately). So  $\|T\|_B \leq \|T\|_S$ . ■

**Definition 1.28** (Operator norm). For any  $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ , we define the *operator norm* by

$$\|T\| = \sup_{\|x\|=1} \|Tx\|$$

From the previous lemma, we can also write

$$\|T\| = \sup_{\|x\| \leq 1} \|Tx\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|}.$$

From equation (1.2), we have

$$\|T\| \leq \|Te_1\| + \cdots + \|Te_n\|.$$

So the operator norm is finite.

**Proposition 1.29.** *The operator norm is a norm on  $L(\mathbb{R}^n, \mathbb{R}^m)$ .*

*Proof.* Let  $T, S \in L(\mathbb{R}^n, \mathbb{R}^m)$ .

(N1) Positivity is by positivity of the vector norm.

(N2) Suppose  $T$  is not identically zero. Let  $v \neq 0$  be such that  $\|Tv\| \neq 0$ . Then

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|} \geq \frac{\|Tv\|}{\|v\|} > 0.$$

(N3)  $\|\lambda T\| = \sup_{\|x\|=1} \|\lambda Tx\| = |\lambda| \sup_{\|x\|=1} \|Tx\| = |\lambda| \|T\|.$

(N4)

$$\begin{aligned} \|T + S\| &= \sup_{\|x\|=1} \|(T + S)x\| \\ &\leq \sup_{\|x\|=1} \|Tx\| + \|Sx\| \\ &\leq \sup_{\|x\|=1} \|Tx\| + \sup_{\|x\|=1} \|Sx\| \\ &= \|T\| + \|S\|. \end{aligned}$$

**Proposition 1.30.** *Let  $T_2 \in L(\mathbb{R}^m, \mathbb{R}^n)$  and  $T_1 \in L(\mathbb{R}^n, \mathbb{R}^k)$ . Then*

$$\|T_1 \circ T_2\| \leq \|T_1\| \|T_2\|$$

*Proof.* Let  $x \in \mathbb{R}^m$  with  $\|x\| = 1$ . Then

$$\|T_1 T_2 x\| \leq \|T_1\| \|T_2 x\| \leq \|T_1\| \|T_2\|.$$

Since  $M_{m \times n}(\mathbb{R}) \cong \mathbb{R}^{mn}$ , we can conclude that the Hilbert-Schmidt norm and the operator norm are equivalent, as are any two norms on  $M_{m \times n}(\mathbb{R})$ . Thus we can talk about openness and continuity without specifying the norm. Problem 1.10 discusses their equivalence with specific bounds.

**Proposition 1.31.**  $\text{GL}_n(\mathbb{R})$  is open in  $M_n(\mathbb{R})$ .

*Proof.*  $\det: M_n(\mathbb{R}) \rightarrow \mathbb{R}$  is continuous because it is a polynomial in the entries of the matrix. Note that  $\text{GL}_n(\mathbb{R}) = \det^{-1}(\mathbb{R} \setminus \{0\})$ , so it is the preimage of an open set, which is open by proposition 1.37. ■

Determinants pose a problem in infinite dimensions. This also doesn't provide estimates on the size of the neighbourhood. We will go through Rudin's proof in proposition 1.39 which avoids determinants.

Meanwhile, here is a special case.

**Lemma 1.32.** *There is an open ball around  $I$  in  $M_n(\mathbb{R})$  that is contained in  $GL_n(\mathbb{R})$ .*

*Proof.* A reasonable guess for the radius is 1 (intuiting from the 1D case). Let  $X \in M_n(\mathbb{R})$  with  $\|X - I\| < 1$ .

Let  $v \in \mathbb{R}^n \setminus \{0\}$ . Then  $\|(X - I)v\| < \|v\|$  implies that  $(X - I)v \neq v$ . Thus  $Xv \neq 0$  and so  $X$  is invertible. ■

This will be useful in proposition 1.39. This can also be proven by borrowing the following result from  $\mathbb{C}$ .

$$\frac{1}{1-z} = 1 + z + z^2 + \dots \quad \text{for } |z| < 1.$$

(This was the first thought a student had when prompted.) We approach it this way as well, since this gives us an explicit inverse.

**Lemma 1.33.** *Let  $Z \in M_n(\mathbb{R})$  be such that  $\|Z\| < 1$ . Then*

- (i)  $\sum_{n=0}^{\infty} Z^n$  converges.
- (ii)  $I - Z$  is invertible.
- (iii)  $(I - Z)^{-1} = \sum_{n=0}^{\infty} Z^n$ .

*Proof.* By  $\|AB\| \leq \|A\|\|B\|$ ,  $\|Z^k\| \leq \|Z\|^k$ .

It is easy to see that the series converges by the Cauchy criterion. For any  $\varepsilon > 0$ , there is some  $n$  such that

$$\left\| \sum_{k=n}^m Z^k \right\| \leq \sum_{k=n}^m \|Z^k\| < \varepsilon$$

for all  $m > n$ .

To see that  $I - Z$  is invertible, note that

$$\|I - Z\| \geq \left| \|I\| - \|Z\| \right| = 1 - \|Z\| > 0.$$

Finally, let  $S_n = \sum_{k=0}^n Z^k$  and  $S_\infty = \lim_{n \rightarrow \infty} S_n$ . Then  $(I - Z)S_n = I - Z^{n+1}$  and so  $(I - Z)S_n \rightarrow I$  as  $n \rightarrow \infty$ . Since matrix multiplication is continuous, we can take the limit inside the product and get  $(I - Z)S_\infty = I$ . ■

*Remark.* For infinite-dimensional spaces, we also need to show  $S_\infty(I - Z) = I$ , which will be done in the exact same way.

**Proposition 1.34.**  $A \mapsto A^{-1}$  is continuous on  $\text{GL}_n(\mathbb{R})$ .

*Proof.* Let  $A \in \text{GL}_n(\mathbb{R})$ . Then  $A^{-1} = \frac{1}{\det A} \text{adj } A$ . Each entry of  $A^{-1}$  is a rational function in the entries of  $A$ , so  $A \mapsto A^{-1}$  is continuous by exercise 1.35. ■

**Exercise 1.35.** Let  $U \subseteq \mathbb{R}^n$  be an open set. Let  $f: U \rightarrow \mathbb{R}^m$  be such that

$$f(x) := (f_1(x), f_2(x), \dots, f_n(x)), \quad x \in U$$

Show that  $f$  is continuous at  $a \in U$  iff each  $f_i$  is continuous at  $a$ .

*Solution.* Consider the  $\ell^1$  norm on  $\mathbb{R}^m$ .

Suppose  $f$  is continuous. Since  $|f_1(x) - f_1(y)| \leq \|f(x) - f(y)\|$ , so is each  $f_i$ .

Suppose each  $f_i$  is continuous at  $a$ . For any  $\varepsilon > 0$ , there exists  $\delta_i > 0$  such that  $|f_i(x) - f_i(a)| < \frac{1}{m}\varepsilon$  in a  $\delta_i$ -neighbourhood of  $a$ . Let  $\delta = \min\{\delta_1, \delta_2, \dots, \delta_n\}$ . ■

**Exercise 1.36.** Let  $f(x) = o(g(x))$  and  $g(x) = O(h(x))$ . Then show that  $f(x) = o(h(x))$ .

*Solution.*

$$\limsup_{x \rightarrow a} \frac{\|f(x)\|}{\|h(x)\|} < c \quad \text{and} \quad \lim_{x \rightarrow a} \frac{\|g(x)\|}{\|h(x)\|} = 0$$

Thus

$$\limsup_{x \rightarrow a} \frac{\|f(x)\|}{\|h(x)\|} = 0. \quad \blacksquare$$

**Proposition 1.37.** Suppose  $X$  and  $Y$  are metric spaces. Then the following are equivalent.

(i)  $f$  is continuous.

(ii)  $f^{-1}(V)$  is open whenever  $V$  is open in  $Y$ .

*Solution.* Suppose  $f$  is continuous. Let  $V \subseteq Y$  be open.

Let  $x \in f^{-1}(V)$ . Then  $f(x) \in V$ . There is some  $\varepsilon > 0$  such that  $B(f(x), \varepsilon) \subseteq V$ . But by continuity, there is some  $\delta > 0$  such that  $f(B(x, \delta)) \subseteq B(f(x), \varepsilon) \subseteq V$ . Thus  $B(x, \delta) \subseteq f^{-1}(V)$ .

Conversely, suppose  $f^{-1}(V)$  is open whenever  $V$  is open. Then for any  $x \in X$  and  $\varepsilon > 0$ , we have that  $f^{-1}(B(f(x), \varepsilon))$  is open. So some  $\delta$ -neighbourhood of  $x$  in  $X$  that is contained in  $f^{-1}(B(f(x), \varepsilon))$ . ■



## Assignment 1

**Problem 1.1.** Let  $(V, \|\cdot\|)$  be a normed linear space.

up August 2  
due August 12  
quiz August 14

(i) Show that the addition map  $(u, v) \mapsto u + v$  is continuous.

(ii) Show that the scalar multiplication map  $(\alpha, u) \mapsto \alpha u$  is continuous.

*Proof.*

$$(i) \quad \|u_2 + v_2 - (u_1 + v_1)\| \leq \|u_2 - u_1\| + \|v_2 - v_1\|.$$

$$(ii) \quad \|\alpha_2 u_2 - \alpha_1 u_1\| = \|\alpha_2 u_2 - \alpha_1 u_2 + \alpha_1 u_2 - \alpha_1 u_1\| = \|(\alpha_2 - \alpha_1)u_2 + \alpha_1(u_2 - u_1)\| \leq |\alpha_2 - \alpha_1|\|u_2\| + |\alpha_1|\|u_2 - u_1\|. \quad \blacksquare$$

**Problem 1.2.** Let  $(V, \|\cdot\|)$  be a normed linear space. Prove that

$$|\|x\| - \|y\|| \leq \|x - y\|$$

for all  $x, y \in V$ . Show that the function  $x \mapsto \|x\|$  from  $V$  to  $\mathbb{R}$  is continuous.

*Proof.* By the  $\triangle$  inequality,

$$\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\| \implies \|x\| - \|y\| \leq \|x - y\|.$$

Similarly

$$\|y\| = \|y - x + x\| \leq \|y - x\| + \|x\| \implies \|x\| - \|y\| \geq -\|x - y\|.$$

Thus

$$|\|x\| - \|y\|| \leq \|x - y\|.$$

To show that  $\|\cdot\|$  is continuous, do what exactly? Notice

$$|\|x\| - \|y\|| \leq \|x - y\|? \quad \blacksquare$$

**Problem 1.3.** For  $x, y \in \mathbb{R}^n$ , show that

$$|\langle x, y \rangle| \leq \|x\|_2 \|y\|_2 \tag{1.3}$$

Also show that the two sides in equation (1.3) are equal if and only if  $x$  and  $y$  are linearly dependent over  $\mathbb{R}$ .

*Proof.* If either of  $x$  or  $y$  is 0, both sides are 0.

Suppose  $x, y \neq 0$ . Let  $\hat{x} = \frac{x}{\|x\|_2}$  and  $\hat{y} = \frac{y}{\|y\|_2}$ . Then proving equation (1.3) amounts to proving

$$|\langle \hat{x}, \hat{y} \rangle| \leq 1$$

because of homogeneity of the inner product.

$$\begin{aligned}
 0 &\leq \sum_{i=1}^n (\hat{x}_i - \hat{y}_i)^2 \\
 0 &\leq \sum_{i=1}^n \hat{x}_i^2 - 2\hat{x}_i\hat{y}_i + \hat{y}_i^2 \\
 2 \sum_{i=1}^n \hat{x}_i\hat{y}_i &\leq \sum_{i=1}^n \hat{x}_i^2 + \sum_{i=1}^n \hat{y}_i^2 \\
 \langle \hat{x}, \hat{y} \rangle &\leq 1.
 \end{aligned}$$

Similarly  $\langle -\hat{x}, \hat{y} \rangle \leq 1$ , which gives  $\langle \hat{x}, \hat{y} \rangle \geq -1$ . ■

**Problem 1.4.** Let  $\{x_k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n$  and  $x \in \mathbb{R}^n$ . Show that  $\{x_k\}_{k \in \mathbb{N}}$  converges to  $x$  if and only if  $\{\langle x_k, y \rangle\}$  converges to  $\langle x, y \rangle$  for all  $y \in \mathbb{R}^n$ .

*Proof.* Suppose  $x_k \rightarrow x$ . Let  $y \in \mathbb{R}^n$ . Then

$$|\langle x_k, y \rangle - \langle x, y \rangle| = |\langle x_k - x, y \rangle| \leq \|x_k - x\| \|y\| \rightarrow 0.$$

Conversely, suppose  $\langle x_k, y \rangle \rightarrow \langle x, y \rangle$  for all  $y \in \mathbb{R}^n$ . Then  $\langle x_k, e_i \rangle \rightarrow \langle x, e_i \rangle$  for all  $i$ . Thus  $x_k \rightarrow x$  componentwise. ■

**Problem 1.5.** Let  $1 < p, q < \infty$  be such that  $\frac{1}{p} + \frac{1}{q} = 1$ . Show that for any  $a \geq 0$  and  $x \geq 0$  the following holds:

$$xa \leq \frac{a^p}{p} + \frac{x^q}{q}. \quad (1.4)$$

Show that in equation (1.4) equality holds if and only if  $x^q = a^p$ .

*Proof.* Let  $a \geq 0$  be fixed. Define  $f(x) = xa - \frac{a^p}{p} - \frac{x^q}{q}$ . This is differentiable on  $[0, \infty)$  since  $q > 0$ .  $f'(x) = a - x^{q-1}$ . Thus

$$\begin{aligned}
 f'(x) \leq 0 &\iff x^{q-1} \leq a \\
 &\iff x^{q/p} \leq a \\
 &\iff x^q \leq a^p.
 \end{aligned}$$

Thus  $f$  is decreasing on  $[a^{p/q}, \infty)$  and increasing on  $[0, a^{p/q}]$ . Thus  $f(x) \geq f(a^{p/q}) = 0$ . Moreover, since  $f'(x) \neq 0$  for  $x^q \neq a^p$ , we have  $f(x) = 0 \iff x^q = a^p$ .

Thus  $xa \leq \frac{a^p}{p} + \frac{x^q}{q}$  with equality only if  $x^q = a^p$ . ■

**Problem 1.6.** For  $1 \leq p \leq \infty$  and  $x = (x_1, x_2, \dots, x_n)$ , we define

$$\|x\|_p = \begin{cases} \left( |x_1|^p + |x_2|^p + \dots + |x_n|^p \right)^{\frac{1}{p}} & 1 \leq p < \infty \\ \max_{1 \leq i \leq n} |x_i| & p = \infty \end{cases}$$

(i) Let  $1 \leq q \leq \infty$  be such that  $\frac{1}{p} + \frac{1}{q} = 1$ . For any  $x, y \in \mathbb{R}^n$ , show that

$$|\langle x, y \rangle| \leq \|x\|_p \|y\|_q \text{ and } \|x + y\|_p \leq \|x\|_p + \|y\|_p. \quad (1.5)$$

(ii) Show that  $\|\cdot\|_p$  defines a norm on  $\mathbb{R}^n$ .

(iii) Show that  $\|x\|_\infty = \lim_{p \rightarrow \infty} \|x\|_p$  for any  $x \in \mathbb{R}^n$ .

*Proof.* We first deal with the case  $p = \infty$  for parts (a) and (b).

(i)  $q = 1$ .

$$\begin{aligned} |\langle x, y \rangle| &= |x_1 y_1 + x_2 y_2 + \cdots + x_n y_n| \\ &\leq |x_1| |y_1| + |x_2| |y_2| + \cdots + |x_n| |y_n| \\ &= \max_{1 \leq i \leq n} |x_i| (|y_1| + |y_2| + \cdots + |y_n|) \\ &= \|x\|_\infty \|y\|_1 \end{aligned}$$

and

$$\begin{aligned} \|x + y\|_\infty &= \max_{1 \leq i \leq n} |x_i + y_i| \\ &\leq \max_{1 \leq i \leq n} (|x_i| + |y_i|) \\ &\leq \max_{1 \leq i, j \leq n} (|x_i| + |y_j|) \\ &= \max_{1 \leq i \leq n} |x_i| + \max_{1 \leq j \leq n} |y_j| \\ &= \|x\|_\infty + \|y\|_\infty. \end{aligned}$$

(ii) We have positivity by definition.  $\|x\|_p = 0 \iff \max_{1 \leq i \leq n} |x_i| = 0 \iff |x_1| = |x_2| = \cdots = |x_n| = 0 \iff x = 0$ , so definiteness holds. Homogeneity is since

$$\|\alpha x\|_\infty = \max_{1 \leq i \leq n} |\alpha x_i| = |\alpha| \max_{1 \leq i \leq n} |x_i| = |\alpha| \|x\|_\infty.$$

Triangle inequality is proven above.

Thus  $\|\cdot\|_\infty$  is a norm.

Now we deal with the case  $1 \leq p < \infty$ .

(i) For  $|\langle x, y \rangle| \leq \|x\|_p \|y\|_q$ , we only concern ourselves with  $1 < p, q < \infty$ . The case  $p = 1$  requires  $q = \infty$ , which is covered above with  $p$  and  $q$  interchanged. We will show that the ratio of the two sides is bounded

by 1.

$$\begin{aligned}
\frac{|\langle x, y \rangle|}{\|x\|_p \|y\|_q} &= \left| \frac{x_1 y_1 + x_2 y_2 + \cdots + x_n y_n}{\|x\|_p \|y\|_q} \right| \\
&\leq \sum_{i=1}^n \frac{|x_i| |y_i|}{\|x\|_p \|y\|_q} \\
&\leq \sum_{i=1}^n \left( \frac{1}{p} \frac{|x_i|^p}{\|x\|_p^p} + \frac{1}{q} \frac{|y_i|^q}{\|y\|_q^q} \right) \quad (\text{by equation (1.4)}) \\
&= \frac{1}{p} \frac{\sum_i |x_i|^p}{\|x\|_p^p} + \frac{1}{q} \frac{\sum_i |y_i|^q}{\|y\|_q^q} \\
&= \frac{1}{p} + \frac{1}{q} \\
&= 1.
\end{aligned}$$

We use this result to prove the triangle inequality. (We did this in a UM 204 assignment last semester, with ample of hints and time to spare.)

$$\begin{aligned}
\|x + y\|_p^p &= \sum_{i=1}^n |x_i + y_i|^p \\
&= \sum_{i=1}^n |x_i + y_i| |x_i + y_i|^{p-1} \\
&\leq \sum_{i=1}^n |x_i| |x_i + y_i|^{p-1} + \sum_{i=1}^n |y_i| |x_i + y_i|^{p-1}
\end{aligned}$$

Let  $X = (|x_1|, |x_2|, \dots, |x_n|)$  and  $Z = (|x_1 + y_1|^{p-1}, |x_2 + y_2|^{p-1}, \dots, |x_n + y_n|^{p-1})$ . Then by equation (1.4),

$$\begin{aligned}
\sum_{i=1}^n |x_i| |x_i + y_i|^{p-1} &= |\langle X, Z \rangle| \\
&\leq \|X\|_p \|Z\|_q
\end{aligned}$$

where  $q = \frac{p}{p-1}$

$$\begin{aligned}
&\leq \|x\|_p (|x_1 + y_1|^p + \cdots + |x_n + y_n|^p)^{\frac{p}{p-1}} \\
&= \|x\|_p \|x + y\|_p^{p-1}.
\end{aligned}$$

Similarly,

$$\sum_{i=1}^n |y_i| |x_i + y_i|^{p-1} \leq \|y\|_p \|x + y\|_p^{p-1}.$$

This gives

$$\begin{aligned}\|x + y\|_p^p &\leq (\|x\|_p + \|y\|_p)\|x + y\|_p^{p-1} \\ \|x + y\|_p &\leq \|x\|_p + \|y\|_p.\end{aligned}$$

- (ii) Positivity is again by definition.  $\|x\|_p = 0 \iff |x_i|^p = 0$  for all  $i$ , which is iff  $x = 0$ . Homogeneity is trivial to check.

$$\begin{aligned}\|\alpha x\|_p &= (|\alpha x_1|^p + |\alpha x_2|^p + \cdots + |\alpha x_n|^p)^{\frac{1}{p}} \\ &= (|\alpha|^p |x_1|^p + |\alpha|^p |x_2|^p + \cdots + |\alpha|^p |x_n|^p)^{\frac{1}{p}} \\ &= |\alpha| \|x\|_p.\end{aligned}$$

Triangle inequality is proven above.

Thus  $\|\cdot\|_p$  is a norm.

We now prove part (c). The case  $x = 0$  is trivial since  $\|x\|_p = \|x\|_\infty = 0$  for any  $p$ .

WLOG let  $\|x\|_\infty = |x_1| > 0$ . Then for  $1 \leq p < \infty$ ,

$$\begin{aligned}\|x\|_p &= |x_1| \left( 1 + \frac{|x_2|^p}{|x_1|^p} + \cdots + \frac{|x_n|^p}{|x_1|^p} \right)^{\frac{1}{p}} \\ &\leq |x_1| \cdot n^{\frac{1}{p}}\end{aligned}$$

Further,

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{\frac{1}{p}} \geq (|x_1|^p)^{\frac{1}{p}} = |x_1|.$$

Thus

$$|x_1| \leq \|x\|_p \leq n^{\frac{1}{p}} |x_1|.$$

As  $p \rightarrow \infty$ ,  $n^{\frac{1}{p}} \rightarrow 1$ . Thus by the squeeze theorem,  $\|x\|_p \rightarrow |x_1| = \|x\|_\infty$ . ■

**Problem 1.7.** Let  $C[a, b]$  be the set of all complex-valued continuous functions on  $[a, b]$ .

- (i) Let  $f \in C[a, b]$  be such that  $f$  is non-negative and  $\int_a^b f(x) dx = 0$ . Show that  $f \equiv 0$ .

- (ii) For  $f \in C[a, b]$ , define

$$\|f\|_\infty := \sup_{x \in [a, b]} |f(x)|, \quad \|f\|_1 := \int_a^b |f(x)| dx.$$

Show that  $\|\cdot\|_\infty$  and  $\|\cdot\|_1$  are norms on  $C[a, b]$ .

- (iii) Are the above two norms on  $C[a, b]$  equivalent? Are they comparable?

*Solution.*

- (i) Suppose  $f$  is non-zero at some point  $c \in [a, b]$ . By continuity,  $f(x) \geq \frac{f(c)}{2}$  in some neighbourhood  $[c - \delta, c + \delta]$ . Then  $f$  is lower bounded by the step function

$$g(x) = \begin{cases} \frac{f(c)}{2} & x \in [c - \delta, c + \delta] \\ 0 & \text{otherwise} \end{cases}$$

which has positive integral. This would force  $\int_a^b f(x) dx > 0$ . Contradiction! Such a  $c$  cannot exist.

- (ii) Clearly both are non-negative.  $\|f\|_\infty = 0 \iff |f(x)| \leq 0$  for all  $x \in [a, b]$ , which is iff  $f \equiv 0$ . Definiteness of  $\|\cdot\|_1$  is by the previous part. Homogeneity is obvious. Triangle inequality is an extension of the triangle inequality for complex numbers.
- (iii) They are *not* equivalent. Consider  $[a, b] = [0, 1]$  and  $f(x) = e^{-\lambda x}$ . Then  $\|f\|_\infty = 1$  and  $\|f\|_1 = \frac{1-e^{-\lambda}}{\lambda}$ . One can choose  $\lambda$  to make  $\|f\|_1$  arbitrarily close to 0. Thus there are no constants  $c_1, c_2 > 0$  such that

$$c_1 \|f\|_\infty \leq \|f\|_1 \leq c_2 \|f\|_\infty.$$

However, we *can* compare the norms as

$$\|f\|_1 \leq (b-a)\|f\|_\infty.$$

This is simply by noticing that the constant function  $x \mapsto \|f\|_\infty$  upper bounds  $|f(x)|$  and has integral  $(b-a)\|f\|_\infty$  over  $[a, b]$ . ■

**Problem 1.8.** For  $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ , let  $\|A\|$  denote the operator norm of  $A$ . Show that

$$\|A\| = \inf\{M : \|Ax\| \leq M\|x\| \text{ for all } x \in \mathbb{R}^n\}.$$

*Proof.*  $\|Ax\| \leq M\|x\|$  is trivially true for  $x = 0$  no matter what  $M$  is. Thus

$$\begin{aligned} & \inf\{M : \|Ax\| \leq M\|x\| \text{ for all } x \in \mathbb{R}^n\} \\ &= \inf\{M : \|Ax\| \leq M\|x\| \text{ for all } x \in \mathbb{R}^n \setminus \{0\}\} \\ &= \inf\{M : \left\|A \frac{x}{\|x\|}\right\| \leq M \text{ for all } x \in \mathbb{R}^n \setminus \{0\}\} \\ &= \inf\{M : \|Ay\| \leq M \text{ for all } y \in S^{n-1}\} \\ &= \inf\{\text{upper bounds of } \{\|Ay\| : y \in S^{n-1}\}\} \\ &= \sup\{\|Ay\| : y \in S^{n-1}\} \\ &= \|A\|. \end{aligned}$$

**Problem 1.9.** Let  $A$  be a real symmetric  $n \times n$  matrix.

- (i) Show that all eigenvalues of  $A$  are real.

(ii) For  $1 \leq i \leq n$ , let  $\lambda_i$  denote the eigenvalues of  $A$ . Show that

$$\|A\| = \max_{1 \leq i \leq n} |\lambda_i|.$$

*Solution.*

(i) View  $A$  as a linear operator on  $\mathbb{C}^n$ . Let  $\lambda$  be an eigenvalue of  $A$  and  $v$  be the corresponding eigenvector. Then

$$\lambda \langle v, v \rangle = \langle Av, v \rangle = \langle v, Av \rangle = \bar{\lambda} \langle v, v \rangle.$$

Thus  $\lambda = \bar{\lambda}$  is real.

(ii) (assuming spectral theorem) WLOG let  $\lambda_1 = \max_{1 \leq i \leq n} |\lambda_i|$ . Write any vector  $x \in \mathbb{R}^n$  as a linear combination of orthonormal eigenvectors  $x = \sum_{i=1}^n c_i v_i$ , where  $v_i$  is the eigenvector corresponding to  $\lambda_i$ . Then  $Ax = \sum_{i=1}^n c_i \lambda_i v_i$ .

$$\begin{aligned} \|Ax\|^2 &= \sum_{i=1}^n c_i^2 \lambda_i^2 \\ &\leq \lambda_1^2 \sum_{i=1}^n c_i^2 \\ &= \lambda_1^2 \|x\|^2. \end{aligned}$$

Thus  $\|A\| \leq \lambda_1$ . Moreover,  $\|Av_1\| = |\lambda_1| \|v_1\|$ . Thus  $\|A\| \geq \lambda_1$ . ■

**Problem 1.10.** Let  $A \in L(\mathbb{R}^n, \mathbb{R}^m)$  and  $B \in L(\mathbb{R}^k, \mathbb{R}^n)$ . Show that

$$\|A\| \leq \|A\|_{HS} \leq \sqrt{n} \|A\| \quad \text{and} \quad \|AB\|_{HS} \leq \|A\|_{HS} \|B\|_{HS}.$$

*Proof.*  $\|A\|_{HS} = \sqrt{\text{Tr}(A^\top A)}$ . Recall that the trace of a matrix is the sum of its eigenvalues.

Let  $v_1, v_2, \dots, v_n$  be orthonormal eigenvectors of  $A^\top A$  with eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  (spectral theorem). Each  $\lambda_i$  is non-negative, since  $\langle A^\top A x, x \rangle = \langle Ax, Ax \rangle \geq 0$ .

Then for any  $x = \sum_{i=1}^n c_i v_i$  with  $\|x\| = 1$ ,

$$\|Ax\|^2 = \langle Ax, Ax \rangle = \langle A^\top A x, x \rangle = \sum_{i=1}^n c_i^2 \lambda_i \leq \lambda_1$$

where the equality holds for  $x = v_1$ . Thus  $\|A\| = \sqrt{\lambda_1}$ . Since  $\|A\|_{HS}^2 = \sum_{i=1}^n \lambda_i$ , we have  $\lambda_1 \leq \|A\|_{HS}^2 \leq n \lambda_1$ . This gives  $\|A\| \leq \|A\|_{HS} \leq \sqrt{n} \|A\|$ .

For  $1 \leq i \leq m$  and  $1 \leq j \leq k$  let

$$a_i = (A_{i1} \ A_{i2} \ \cdots \ A_{in})^\top, \quad b_j = \begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{nj} \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} \langle a_1, b_1 \rangle & \langle a_1, b_2 \rangle & \cdots & \langle a_1, b_k \rangle \\ \langle a_2, b_1 \rangle & \langle a_2, b_2 \rangle & \cdots & \langle a_2, b_k \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle a_m, b_1 \rangle & \langle a_m, b_2 \rangle & \cdots & \langle a_m, b_k \rangle \end{pmatrix}$$

so by Cauchy-Schwarz,

$$\begin{aligned} \|AB\|_{HS}^2 &= \sum_{i=1}^m \sum_{j=1}^k \langle a_i, b_j \rangle^2 \\ &\leq \sum_{i=1}^m \sum_{j=1}^k \|a_i\|^2 \|b_j\|^2 \\ &= \left( \sum_{i=1}^m \|a_i\|^2 \right) \left( \sum_{j=1}^k \|b_j\|^2 \right) \\ &= \|A\|_{HS}^2 \|B\|_{HS}^2. \end{aligned}$$

■

*Remark.* A far simpler proof that I missed is the following.

$$\begin{aligned} \|Ax\|^2 &\leq \sum_i \langle a_i, x \rangle^2 & \|A\|_{HS}^2 &= \sum_j \sum_i a_{ij}^2 \\ &\leq \sum_i \|a_i\|^2 \|x\|^2 & &= \sum_j \|Ae_j\|^2 \\ &= \|A\|_{HS}^2 \|x\|^2 & &\leq \sum_j \|A\|^2 \\ & & &= n \|A\|^2. \end{aligned}$$

## Quiz

**Problem 1.11.** Recall the definition of a [homogeneous function](#). Let  $f: \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}$  be a continuous, non-vanishing homogenous function of degree  $k$  and  $\|\cdot\|$  be a fixed norm on  $\mathbb{R}^n$ . Show that there exist positive constants  $C_1, C_2 > 0$  such that

$$C_1 \|x\|^k \leq |f(x)| \leq C_2 \|x\|^k,$$

for every  $0 \neq x \in \mathbb{R}^n$ .

*Proof.* Choose  $C_1 = \min_{\|x\|=1} |f(x)|$  and  $C_2 = \max_{\|x\|=1} |f(x)|$ . They exist by compactness of the unit sphere, and are positive since  $f$  does not vanish.

Then for any  $x \neq 0$ ,

$$|f(x)| = \|x\|^k \left| f\left(\frac{x}{\|x\|}\right) \right|$$

is bounded between  $C_1 \|x\|^k$  and  $C_2 \|x\|^k$ .

■



**Problem 1.12.** Let  $V$  be a vector space over  $\mathbb{R}$ . Let  $d$  be the discrete metric on  $V$ . Is  $d$  induced by a norm on  $V$ ?

*Solution.* No. Suppose  $d(x, y) = \|x - y\|$  for some norm  $\|\cdot\|$ , for all  $x, y \in V$ . Let  $x \neq y$ . Then  $d(x, y) = 1 = \|x - y\|$ . But  $d(2x, 2y) = 1 = \|2x - 2y\| = 2\|x - y\| = 2$ . Contradiction! ■

**Problem 1.13.** For  $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ , show that  $\|A\| = \|A^\top\|$ .

*Proof.* Notice by Cauchy-Schwarz that for any vector  $v$  in a real inner product space,

$$\|v\| = \sup_{\|w\|=1} \langle w, v \rangle.$$

(The supremum is achieved at  $v/\|v\|$  for  $v \neq 0$ .) Then

$$\begin{aligned} \|A\| &= \sup_{x \in S^{n-1}} \|Ax\| \\ &= \sup_{x \in S^{n-1}} \sup_{y \in S^{m-1}} \langle y, Ax \rangle \\ &= \sup_{y \in S^{m-1}} \sup_{x \in S^{n-1}} \langle A^\top y, x \rangle \\ &= \sup_{y \in S^{m-1}} \|A^\top y\| \\ &= \|A^\top\|. \end{aligned}$$

**Problem 1.14.** Find maximum of  $x + 2y + 3z$  subject to the condition  $x^2 + y^2 + z^2 = 1$ .

*Solution.* The function is continuous and the constraint is compact. Thus a maximum exists.

Let  $r = (x, y, z)$  and  $n = (1, 2, 3)$ . As discussed in the previous problem,

$$\max_{\|r\|=1} \langle n, r \rangle = \|n\|.$$

Thus the maximum is  $\sqrt{14}$ . ■

**Problem 1.15.** See problem 1.9.

**Exercise 1.38.** Let  $Z$  be as in  $(I - Z)^{-1} = I + Z + O(Z^2)$  and also  $(I - Z)^{-1} = I + Z + o(Z^2)$ .

**Lecture 5.**  
Monday  
August 12

The proof of proposition 1.34 is nice and sweet. However, the proof in Rudin generalises better to infinite dimensions. We thus prove it again.

**Proposition 1.39.**

(i) Let  $A \in M_n(\mathbb{R})$  be such that  $\|I - A\| < 1$ . Then  $A \in \text{GL}_n(\mathbb{R})$ .

(ii) Let  $A \in \text{GL}_n(\mathbb{R})$  be fixed and let  $B \in M_n(\mathbb{R})$  be such that

$$\|B - A\| < \|A^{-1}\|^{-1}.$$

Then  $B \in \text{GL}_n(\mathbb{R})$ .

(iii)  $A \mapsto A^{-1}$  is continuous on  $\text{GL}_n(\mathbb{R})$ .

*Remark.* The second part shows that  $\text{GL}_n(\mathbb{R})$  is open in  $M_n(\mathbb{R})$ .

*Proof.* We proved the first part earlier in lemma 1.32 and again in lemma 1.33 (let  $Z = I - A$ , then  $I - Z = A$ ).

For the second part, let  $A \in \text{GL}_n(\mathbb{R})$  be fixed and let  $\|B - A\| < \|A^{-1}\|^{-1}$ . We can write  $B - A$  as  $A(A^{-1}B - I)$ . Now

$$\begin{aligned} \|A^{-1}B - I\| &= \|A^{-1}(B - A)\| \\ &\leq \|A^{-1}\| \|B - A\| \\ &< 1. \end{aligned}$$

Then by the first part,  $A^{-1}B \in \text{GL}_n(\mathbb{R})$ , so that  $B \in \text{GL}_n(\mathbb{R})$ .

For the last part, we want  $B^{-1} \rightarrow A^{-1}$  as  $B \rightarrow A$ .

$$B^{-1} - A^{-1} = A^{-1}(A - B)B^{-1} \quad (1.6)$$

We need to bound  $\|B^{-1}\|$ . Let  $W$  be an open neighbourhood of  $A$  of radius  $\frac{1}{2}\|A^{-1}\|^{-1}$ . Then  $W \subseteq \text{GL}_n(\mathbb{R})$ .

For any  $B \in W$ ,  $\|A - B\| \|A^{-1}\| < \frac{1}{2}$  and

$$\begin{aligned} \|B^{-1}\| - \|A^{-1}\| &\leq \|B^{-1} - A^{-1}\| \\ &\leq \|B^{-1}\| \|A - B\| \|A^{-1}\| \quad (\text{by equation (1.6)}) \\ &\leq \frac{1}{2} \|B^{-1}\|. \end{aligned}$$

This bounds  $\|B^{-1}\|$  above by  $2\|A^{-1}\|$ . Using equation (1.6) again, we have

$$\begin{aligned} \|B^{-1} - A^{-1}\| &\leq \|A^{-1}\| \|A - B\| \|B^{-1}\| \\ &\leq 2\|A^{-1}\|^2 \cdot \|A - B\|. \end{aligned}$$

As  $B \rightarrow A$ ,  $B^{-1} \rightarrow A^{-1}$ . ■

*Idea.* This is similar in spirit to exercise [1.18](#).

- Equation [\(1.6\)](#) is similar to taking the common denominator in  $\frac{1}{x} - \frac{1}{a}$ .
- The choice of  $W$  is similar to choosing  $\delta \leq \frac{1}{2}|a|$ , and leads to an identical bound.

# Chapter 2

## Differentiation

**Definition 2.1.** Let  $f: \mathbb{R} \rightarrow \mathbb{R}$  be a function. We say that  $f$  is *differentiable* at  $a \in \mathbb{R}$  if

$$\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$$

exists. We denote this limit by  $f'(a)$  and call it the *derivative* of  $f$  at  $a$ .

This doesn't make sense for  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  when  $n > 2$  (for  $n = 2$  we can identify  $\mathbb{R}^2$  with  $\mathbb{C}$ ).

**Theorem 2.2** (Hurwitz' theorem).  $\mathbb{R}^n$  is a

We will redefine differentiability for real functions.

**Proposition 2.3.** Let  $U$  be an open subset of  $\mathbb{R}$  and  $f: U \rightarrow \mathbb{R}$ . Let  $a \in U$ . Then  $f$  is differentiable at  $a$  if and only if there exists a linear map  $T \in L(\mathbb{R}, \mathbb{R})$  such that

$$f(a+h) - f(a) = Th + o(h).$$

*Proof.* Suppose  $f$  is differentiable at  $a \in U$ .

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

We can rewrite this as

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{f(a+h) - f(a) - f'(a)h}{h} &= 0 \\ \implies \lim_{h \rightarrow 0} \frac{|f(a+h) - f(a) - T_{f'(a)}h|}{|h|} &= 0 \end{aligned}$$

where  $T_\alpha \in L(\mathbb{R}, \mathbb{R})$  is the linear map  $x \mapsto \alpha x$ .

Conversely, suppose there exists a linear map  $T$  such that  $f(a+h) -$

$f(a) - Th = o(h)$ . Then

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{|f(a+h) - f(a) - Th|}{|h|} &= 0 \\ \implies \lim_{h \rightarrow 0} \left| \frac{f(a+h) - f(a)}{h} - T(1) \right| &= 0 \\ \implies \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} &= T(1). \end{aligned}$$

■

**Definition 2.4.** Let  $U \subseteq \mathbb{R}^n$  be an open set containing  $a$ . Let  $f: U \rightarrow \mathbb{R}^m$ . We say that  $f$  is *differentiable* at  $a$  if there exists a linear map  $T \in L(\mathbb{R}^n, \mathbb{R}^m)$  such that

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

We say that  $T$  is the *derivative* of  $f$  at  $a$  and write  $f'(a) = T$ .

If  $f$  is differentiable at every point in  $U$ , we say that  $f$  is differentiable on  $U$ .

## Lecture 6.

Monday

August 19

Writing  $f'(a)$  requires the derivative to be unique.

**Proposition 2.5.** Let  $T_1, T_2 \in L(\mathbb{R}^n, \mathbb{R}^m)$  be satisfying the definition of differentiability at  $a$  for  $f: U \rightarrow \mathbb{R}^m$ . Then  $T_1 = T_2$ .

*Proof.* Let  $T = T_1 - T_2$ . Then

$$\begin{aligned} Th &= T_1h - T_2h \\ &= (f(a+h) - f(a) - T_2h) - (f(a+h) - f(a) - T_1h) \\ &= o(h) - o(h) = o(h). \end{aligned}$$

We have  $\lim_{h \rightarrow 0} \frac{\|Th\|}{\|h\|} = 0$ . Let  $v \in \mathbb{R}^n \setminus \{0\}$ . As  $t \rightarrow 0$ ,  $tv \rightarrow 0$ . Thus

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{\|T(tv)\|}{\|tv\|} &= \lim_{t \rightarrow 0} \frac{|t| \|Tv\|}{|t| \|v\|} \\ &= \frac{\|Tv\|}{\|v\|} = 0. \end{aligned}$$

Thus  $Tv = 0$  for all  $v \in \mathbb{R}^n$ . ■

**Proposition 2.6.** Differentiability at a point implies continuity at that point.

*Proof.* Suppose  $f$  is differentiable at  $a$  with  $f'(a) = T$ . Let

$$q(h) = f(a+h) - f(a) - Th.$$

$$\begin{aligned}
\|f(a+h) - f(a)\| &= \|f(a+h) - f(a) - Th + Th\| \\
&\leq \|q(h)\| + \|Th\| \\
&\leq \frac{\|q(h)\|}{\|h\|} \|h\| + \|T\| \|h\|.
\end{aligned}$$

As  $h \rightarrow 0$ , each term goes to 0. ■

For *finding* the derivative, it is helpful to do the following:

- Use little- $o$  notation.
- Identify the linear map  $T$ .
- Ignore the little- $o$  terms.

If  $f(a+h) = f(a) + Th + o(h)$ , then  $f'(a) = T$ .

*Examples.*

- Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  be given by  $f(x) = c$  for some constant  $c \in \mathbb{R}^m$ . For any  $a \in \mathbb{R}^n$ , we can write  $f(a+h) = f(a) + 0 + 0$ . Thus  $f'(a) = 0$ .
- Let  $f \in L(\mathbb{R}^n, \mathbb{R}^m)$ . Then  $f(a+h) = f(a) + f(h) + 0$ . Thus  $f'(a) = f$ .
- Let  $f: \mathbb{R} \rightarrow \mathbb{R}$  be given by  $f(x) = x$ . This is a special case of the previous example.  $f'(a) = \text{id}$ . Thus  $f'(A)(H) = A^2H + AHA + HA^2$ .

Even though we are developing calculus on  $\mathbb{R}^n$ , it is trivially extended to all finite-dimensional normed linear spaces over  $\mathbb{R}$  via the natural identification with  $\mathbb{R}^n$ .

We will continue our examples.

*Examples.*

- Let  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  be given by  $f(x, y) = xy$ . Write  $a = (a_1, a_2)$ ,  $h = (h_1, h_2)$ . Then

$$\begin{aligned}
f(a+h) &= f(a_1+h_1, a_2+h_2) \\
&= (a_1+h_1)(a_2+h_2) \\
&= a_1a_2 + a_1h_2 + a_2h_1 + h_1h_2 \\
&= f(a) + \begin{pmatrix} a_2 & a_1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + o(h).
\end{aligned}$$

Let us show  $h_1h_2 = o(h)$ .

$$\frac{|h_1h_2|}{\|h\|} = |h_1| \frac{|h_2|}{\|h\|} \leq |h_1| \rightarrow 0.$$

Thus  $f'(a)$  is the map  $(h_1, h_2) \mapsto a_2h_1 + a_1h_2$ . As a matrix, this is  $\begin{pmatrix} a_2 & a_1 \end{pmatrix}$ .

**Lecture 7.**  
Wednesday  
August 21

- Let  $f: M_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$  be given by  $f(X) = X$ .

We could identify  $M_n(\mathbb{R})$  with  $\mathbb{R}^{n^2}$  and construct a linear map from  $\mathbb{R}^{n^2}$  to  $\mathbb{R}^{n^2}$ . It is however advisable to construct a linear map from  $M_n(\mathbb{R})$  to  $M_n(\mathbb{R})$ . This is again a special case of the second example. Thus  $f'(A) = f = \text{id}$ .

- Let  $f: M_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$  be given by  $f(X) = X^2$ . Then

$$\begin{aligned} f(A + H) &= (A + H)^2 \\ &= A^2 + AH + HA + H^2 \\ &= f(A) + AH + HA + o(H) \end{aligned}$$

since

$$\frac{\|H^2\|}{\|H\|} \leq \frac{\|H\|^2}{\|H\|} = \|H\| \rightarrow 0.$$

Thus  $f'(A)(H) = AH + HA$ .

- Let  $f: M_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$  be given by  $f(X) = X^3$ . Then

$$\begin{aligned} f(A + H) &= (A + H)^3 \\ &= A^3 + A^2H + AHA + HA^2 \\ &\quad + AH^2 + HAH + H^2A + H^3 \\ &= A^3 + (A^2H + AHA + HA^2) + o(H). \end{aligned}$$

- Let  $f: \text{GL}_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$  be given by  $f(X) = X^{-1}$ . Recall that if  $\|Z\| < 1$ , then

$$(I - Z)^{-1} = I + Z + O(Z^2) = I + Z + o(Z).$$

Thus for small enough  $\|H\|$ ,

$$\begin{aligned} (A + H)^{-1} &= \left( A(I + A^{-1}H) \right)^{-1} \\ &= \left( I - A^{-1}H + o(-A^{-1}H) \right) A^{-1} \\ &= A^{-1} - A^{-1}HA^{-1} + o(H). \end{aligned} \tag{2.1}$$

Let us do the  $o(H)$  term more carefully. Let  $(I + A^{-1}H)^{-1} = I - A^{-1}H + u(H)$  where

$$\lim_{H \rightarrow 0} \frac{\|u(H)\|}{\| -A^{-1}H \|} = 0.$$

Then

$$\begin{aligned} (A + H)^{-1} &= (I + A^{-1}H)^{-1} A^{-1} \\ &= (I - A^{-1}H + u(H)) A^{-1} \\ &= A^{-1} - A^{-1}HA^{-1} + u(H)A^{-1}. \end{aligned}$$

But

$$\begin{aligned}
 \frac{\|u(H)A^{-1}\|}{\|H\|} &\leq \frac{\|u(H)\|}{\|H\|} \|A^{-1}\| \\
 &\leq \frac{\|u(H)\|}{\| -A^{-1}H \|} \frac{\| -A^{-1}H \|}{\|H\|} \|A^{-1}\| \\
 &\leq \frac{\|u(H)\|}{\| -A^{-1}H \|} \|A^{-1}\|^2 \\
 &\rightarrow 0.
 \end{aligned}$$

Thus from equation (2.1),  $f'(A)(H) = -A^{-1}HA^{-1}$ .

However, we can simply use ?? as follows:

$$u(H) = o(-A^{-1}H) = o(O(H)) = o(H).$$

- Let  $f: M_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$  be given by  $f(X) = X^{-2}$ . Then

$$\begin{aligned}
 f(A+H) &= (A+H)^{-2} \\
 &= (A^{-1} - A^{-1}HA^{-1} + o(H))(A^{-1} - A^{-1}HA^{-1} + o(H)) \\
 &= A^{-2} - A^{-2}HA^{-1} - A^{-1}HA^{-2} + o(H).
 \end{aligned}$$

Thus  $f'(A)(H) = -A^{-1}(A^{-1}H + HA^{-1})A^{-1}$ .

*Remarks.*

- $$\frac{1}{x+h} - \frac{1}{x} = \frac{1}{x+h}(x - (x+h))\frac{1}{x} \rightarrow -\frac{1}{x}h\frac{1}{x}$$
- $$\frac{1}{(x+h)^2} - \frac{1}{x^2} = \frac{1}{(x+h)^2}(x^2 - (x+h)^2)\frac{1}{x^2}$$

**Exercise 2.7** (Sum). Let  $U \subseteq \mathbb{R}^n$  be open and  $a \in U$ . Let  $f, g: U \rightarrow \mathbb{R}^m$  both be differentiable at  $U$ . Then  $f + g$  is differentiable at  $a$  with  $(f + g)'(a) = f'(a) + g'(a)$ .

**Lecture 8.**  
Friday  
August 23

**Lemma 2.8.** Let  $f, g, h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Suppose  $g = O(f)$  and  $h = o(g)$ . Then  $h = o(f)$ .

*Proof.*

$$\frac{\|h(x)\|}{\|f(x)\|} = \frac{\|h(x)\|}{\|g(x)\|} \frac{\|g(x)\|}{\|f(x)\|}$$

The first term goes to 0 and the second term is bounded. Thus  $\frac{\|h(x)\|}{\|f(x)\|} \rightarrow 0$ . ■



**Proposition 2.9** (chain rule). *Let  $U \subseteq \mathbb{R}^n$  be open and  $a \in U$ . Let  $f: U \rightarrow \mathbb{R}^m$  be differentiable at  $a$ . Let  $V \subseteq \mathbb{R}^m$  be an open set containing  $f(a)$  and  $g: \mathbb{R}^m \rightarrow \mathbb{R}^k$  be differentiable at  $f(a)$ . Then  $g \circ f$  is differentiable at  $a$  with*

$$(g \circ f)'(a) = g'(f(a)) \circ f'(a).$$

*Proof.* For small enough  $h$ ,

$$\begin{aligned} g(f(a+h)) - g(f(a)) &= g(f(a) + f'(a)h + u(h)) - g(f(a)) \\ &= g'(f(a))[f'(a)h + u(h)] + v(f'(a)h + u(h)) \end{aligned}$$

where  $\frac{\|u(h)\|}{\|h\|} \rightarrow 0$  and  $\frac{\|v(f'(a)h + u(h))\|}{\|f'(a)h + u(h)\|} \rightarrow 0$ .

Call  $f(a) = b$  and  $f'(a)h + u(h) = k(h)$  for convenience. We have

$$\frac{\|u(h)\|}{\|h\|} \rightarrow 0 \quad \text{and} \quad \frac{\|v(k(h))\|}{\|k(h)\|} \rightarrow 0$$

and

$$g(f(a+h)) - g(f(a)) = g'(b)f'(a)h + g'(b)u(h) + v(k(h)).$$

We need to show

$$g'(b)u(h) + v(k(h)) = o(h).$$

The first term is easy, since  $\|g'(b)u(h)\| \leq \|g'(b)\|\|u(h)\|$ .

For the second term, we write

$$\begin{aligned} \frac{\|v(k(h))\|}{\|h\|} &= \frac{\|v(k(h))\|}{\|k(h)\|} \frac{\|k(h)\|}{\|h\|} \\ &\leq \frac{\|v(k(h))\|}{\|k(h)\|} \left( \|f'(a)\| + \frac{\|u(h)\|}{\|h\|} \right) \end{aligned}$$

which goes to 0 as the second term is bounded.

Let  $\varepsilon > 0$  be given. Then there exists a neighbourhood  $W_1$  of 0 in  $V$  on which

$$\|v(k)\| \leq \varepsilon \|k\|.$$

Since  $f$  is continuous, there exists a neighbourhood  $W_2$  of 0 in  $U$  such that

$$k(h) = f(a+h) - f(a) \in W_1$$

for each  $h \in W_2$ . Thus on this neighbourhood  $W_2$ , we have

$$\frac{\|v(k(h))\|}{\|h\|} \leq \varepsilon \left( \|f'(a)\| + \frac{\|u(h)\|}{\|h\|} \right).$$

Thus

$$\lim_{h \rightarrow 0} \frac{v(k(h))}{h} = 0. \quad \blacksquare$$

How does Rudin deal with the vanishing of  $k(h)$ ?

*Examples.*

- Consider  $f: \text{GL}_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$  and  $g: M_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$  given by

$$f(X) = X^{-1} \quad \text{and} \quad g(X) = X^2.$$

These are differentiable with  $f'(X)(H) = -X^{-1}HX^{-1}$  and  $g'(X)(H) = XH + HX$ . By the chain rule, we have

$$\begin{aligned} (g \circ f)'(X)(H) &= g'(f(X))(f'(X)(H)) \\ &= X^{-1}(-X^{-1}HX^{-1}) + (-X^{-1}HX^{-1})X^{-1} \\ &= -X^{-1}(X^{-1}H + HX^{-1})X^{-1}. \end{aligned}$$

**Definition 2.10** (directional derivative). Let  $U \subseteq \mathbb{R}^n$  be open and  $a \in U$ . Let  $f: U \rightarrow \mathbb{R}$ . For  $v \in \mathbb{R}^n$ , we define

$$(D_v f)(a) := \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t}$$

to be the *directional derivative* of  $f$  at  $a$  in the direction of  $v$ .

**Exercise 2.11.** Show that  $D_{\lambda v} = \lambda D_v$ .

*Solution.* If  $\lambda = 0$ , both sides are 0. Otherwise,

$$\begin{aligned} (D_{\lambda v} f)(a) &= \lim_{t \rightarrow 0} \frac{f(a + t\lambda v) - f(a)}{t} \\ &= \lambda \lim_{t \rightarrow 0} \frac{f(a + (\lambda t)v) - f(a)}{\lambda t} \\ &= \lambda \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t} \\ &= \lambda (D_v f)(a) = (\lambda D_v f)(a). \end{aligned} \quad \blacksquare$$

**Definition 2.12** (partial derivative). Let  $U \subseteq \mathbb{R}^n$  be open and  $a \in U$ . Let  $f: U \rightarrow \mathbb{R}$ . We define

$$\frac{\partial f}{\partial x_i}(a) = D_i f(a) = \partial_i f(a) := (D_{e_i} f)(a)$$

to be the *partial derivative* of  $f$  at  $a$  with respect to the  $i$ -th coordinate.

Observe that if  $g = x \mapsto f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$ , then  $D_i f(a) = g'(a_i)$ .

**Definition 2.13** (gradient). Let  $U \subseteq \mathbb{R}^n$  be open and  $f: U \rightarrow \mathbb{R}$ . We define the *gradient*  $\nabla f: U \rightarrow \mathbb{R}^n$  of  $f$  by

$$(\nabla f)(a) := \left( \frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a) \right)$$

for all  $a \in U$ .

Throughout this lecture,  $U \subseteq \mathbb{R}^n$  is open and  $a \in U$ , and  $f: U \rightarrow \mathbb{R}^m$  is differentiable at  $a$ .

**Lecture 9.**  
Sunday  
August 25

**Proposition 2.14.** Let  $f = (f_1, \dots, f_m)$ .

Then  $\frac{\partial f_j}{\partial x_i}(a)$  exists for each  $i \in [n]$  and  $j \in [m]$ . Moreover,

$$f'(a)(e_i) = \sum_{j=1}^m \frac{\partial f_j}{\partial x_i}(a) e_j$$

for all  $i \in [n]$ . Equivalently,

$$f'(a) = \begin{pmatrix} \nabla f_1(a) \\ \vdots \\ \nabla f_m(a) \end{pmatrix}.$$

We have abused notation in the above statement to let  $e_1$  be in  $\mathbb{R}^n$  or  $\mathbb{R}^m$  depending on the context.

*Proof.* Let  $f'(a) = T \in L(\mathbb{R}^n, \mathbb{R}^m)$ . For  $i \in [n]$ ,

$$\begin{aligned} \frac{\partial f_j}{\partial x_i}(a) &= \lim_{t \rightarrow 0} \frac{f_j(a + te_i) - f_j(a)}{t} \\ \implies \frac{\partial f_j}{\partial x_i}(a) - T(e_i)_j &= \lim_{t \rightarrow 0} \frac{f_j(a + te_i) - f_j(a) - tT(e_i)_j}{t} \\ &= \lim_{t \rightarrow 0} \left( \frac{f(a + te_i) - T(te_i) - f(a)}{t} \right)_j \\ &= 0. \end{aligned}$$

Thus

$$T(e_i) = \sum_{j=1}^m \frac{\partial f_j}{\partial x_i}(a) e_j \quad \blacksquare$$

**Definition 2.15** (curve). A map  $\gamma: (a, b) \rightarrow \mathbb{R}^n$  is a *curve* in  $\mathbb{R}^n$ . If  $\gamma$  is differentiable, it is a *differentiable curve*.

*Remark.* Chain rule remains valid for curves. Suppose  $f: U \rightarrow \mathbb{R}$  and  $\gamma: (a, b) \rightarrow U$ . Then

$$\begin{aligned} (f \circ \gamma)'(t) &= f'(\gamma(t))(\gamma'(t)) \\ &= f'(\gamma(t)) \left( \sum \gamma'_i(t) e_i \right) \\ &= \sum \gamma'_i(t) f'(\gamma(t))(e_i) \\ &= \sum \gamma'_i(t) \frac{\partial f}{\partial x_i}(\gamma(t)) \\ &= \nabla f(\gamma(t)) \cdot \gamma'(t). \end{aligned}$$

**Definition 2.16.** Let  $\gamma: (-\varepsilon, \varepsilon) \rightarrow U$  be such that  $\gamma(0) = a$ .  $\gamma'(0)$  is the *tangent vector* to  $\gamma$  at  $a$ , and  $(f \circ \gamma)'(0)$  is the *derivative of  $f$  along  $\gamma$  at  $a$* .

**Proposition 2.17.** For each  $v \in \mathbb{R}^n$ ,  $D_v f(a)$  exists and equals  $\nabla f(a) \cdot v$ .

*Proof.* Let  $\gamma(t) = a + tv$ . There exist some  $\varepsilon > 0$  such that  $\gamma(-\varepsilon, \varepsilon) \subseteq U$  and  $\gamma'(0) = v$ . Then

$$D_v f(a) = \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t} = (f \circ \gamma)'(0) = \nabla f(a) \cdot v. \quad \blacksquare$$

## Assignment 2

up August 2  
due August 12  
quiz August 14