MA 200: Multivariable Calculus

Naman Mishra

August 2024

Contents

Lectures

Assignments

The course

Lecture 1.
Friday
August 2

Grading

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

Textbooks

-

Chapter 1

Linear algebra

1.1 Normed linear spaces

Definition 1.1 (homogeneous function)**.** Let V be a vector space over $\mathbb{R}$. A function $f\colon V \setminus \{0\} \to \mathbb{R}$ 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 $B(x, r) \coloneqq \{y \in V \mid \|x - y\| < r\}$ is called the open ball of radius r centered at x.

The set $S(x, r) \coloneqq \{y \in V \mid \|x - y\| = r\}$ 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*

$$\big|\|x\| - \|y\|\big| \le \|x - y\| \tag{1.1}$$

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

$$\|x\| \le \|x - y\| + \|y\|$$

so that

$$\|x\| - \|y\| \le \|x - y\|.$$

Similarly,

$$\|y\| \le \|y - x\| + \|x\|$$

so that

$$-\|x - y\| \le \|x\| - \|y\|.$$

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 $d \colon X \times X \to \mathbb{R}$ called a *metric* that satisfies the following properties:

(M1) $d(x, y) \ge 0$ 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) $d(x, z) \le d(x, y) + d(y, z)$ 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, ρ) be metric spaces. A function $f \colon X \to Y$ is called *continuous* at $a \in X$ iff

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

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

$$d((x_1, x_2), (y_1, y_2)) \coloneqq d_1(x_1, y_1) + d_2(x_2, y_2).$$

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 $z_n \to (x, y)$ iff $x_n \to x$ and $y_n \to y$.

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

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

Remark. $\widetilde{d}$ given by

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

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.

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

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

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

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

Repeated in problem 2.1. ■

Examples.

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

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

  where

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

  is the limit of the l^p norms as $p \to \infty$.

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 these 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

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

Lecture 2. Monday August 5

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 $\|\cdot\|_b$.*

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. $\blacksquare$

Proposition 1.14. *Every ℓ^p norm is equivalent to ℓ^∞.*

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

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

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

Lemma 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.

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

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

The reverse triangle inequality gives

$$\begin{aligned}\Big|\|x\| - \|y\|\Big| &\le \|x - y\| \\ &\le M\|x - y\|_2.\end{aligned}$$ ∎

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

$$c_1 \le \|x\| \le c_2$$

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

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

By homogeneity (N3),

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

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 ($f(v) = f(\widehat{v})$).

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

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

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 $f\colon \mathbb{R}\setminus\{0\} \to \mathbb{R}$ 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 δ-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 $x \to a$. We will do a similar proof in theorem 1.39.

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

Lecture 3. Wednesday August 7

Lecture 3: Big-Oh and matrix norms

Notation.

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

and

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

using the isomorphism $A \mapsto T_A$ where

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

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 $f\colon X \to \mathbb{R}$ be a function on a topological space X. We define the *limit inferior* and *limit superior* of f as

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

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*

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

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 ε. 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 $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and $g\colon \mathbb{R}^n \to \mathbb{R}^k$. We say that

(1) $f(x) = o(g(x))$ as $x \to a$ if

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

(2) $f(x) = O(g(x))$ as $x \to a$ if

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

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 $x \to a$ 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.

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

Exercise 1.23. *If $f_1, f_2 = O(g)$ then $f_1 \pm f_2 = O(g)$. If $f_1, f_2 = o(g)$ then $f_1 \pm f_2 = o(g)$.*

Solution. It do be a self-evident truth. ■

1.2 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

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

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

Solution.

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

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

Proposition 1.26. *Any linear transformation is continuous.*

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

$$\begin{aligned}
\|Tx\| &= \left\|T\Big(\sum x_i e_i\Big)\right\| \\
&= \left\|\sum x_i T e_i\right\| \\
&\le \sum |x_i| \|Te_i\| \\
&\le \|x\| \sum \|Te_i\| \\
&= M\|x\| \qquad (1.2)
\end{aligned}$$

where $M = \|Te_1\| + \cdots + \|Te_n\|$.

Now

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

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

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

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

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

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

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

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

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

$$\|T_1T_2x\| \leq \|T_1\|\|T_2x\| \leq \|T_1\|\|T_2\|.$$

■

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. $\mathrm{GL}_n(\mathbb{R})$ *is open in* $\mathrm{M}_n(\mathbb{R})$.

Proof. $\det\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ is continuous because it is a polynomial in the entries of the matrix. Note that $\mathrm{GL}_n(\mathbb{R}) = \det^{-1}(\mathbb{R} \setminus \{0\})$, 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 theorem 1.39 which avoids determinants.

We need to figure out a special case first: what ball around the identity matrix is fully invertible? A reasonable guess for the radius is 1 (intuiting from the 1D case).

Lemma 1.32. *The open ball of radius* 1 *around* I *in* $\mathrm{M}_n(\mathbb{R})$ *is contained in* $\mathrm{GL}_n(\mathbb{R})$.

Proof. Let $X \in \mathrm{M}_n(\mathbb{R})$ 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 also be useful in theorem 1.39. This can also be proven by borrowing the following result from $\mathbb{C}$.

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

(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* $Z \in \mathrm{M}_n(\mathbb{R})$ *be such that* $\|Z\| < 1$. *Then*

(1) $\sum_{n=0}^{\infty} Z^n$ *converges.*

(2) $I - Z$ *is invertible.*

(3) $(I - Z)^{-1} = \sum_{n=0}^{\infty} Z^n$.

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 $m > n$.

Now let $S_n = \sum_{k=0}^{n} Z^k$ and $S_\infty = \lim_{n\to\infty} S_n$. Then $(I - Z)S_n = I - Z^{n+1}$ and so $(I - Z)S_n \to I$ as $n \to \infty$. 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* $\mathrm{GL}_n(\mathbb{R})$.

Proof. Let $A \in \mathrm{GL}_n(\mathbb{R})$. Then $A^{-1} = \frac{1}{\det A} \operatorname{adj} A$. 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. ■

Lecture 4. Friday August 9

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

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

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

Solution. Consider the ℓ^∞ 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 $|f_i(x) - f_i(a)| < \varepsilon$ in a δ_i-neighbourhood of a. Let $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_n\}$. ■

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

Solution.

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

Thus

$$\limsup_{x\to a} \frac{\|f(x)\|}{\|h(x)\|} = 0.$$ ■

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

(1) f is continuous.

(2) $f^{-1}(V)$ 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 δ-neighbourhood of x in X that is contained in $f^{-1}(B(f(x), \varepsilon))$. ■

Assignment 1

up August 2
due August 12
quiz August 14

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

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

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

Proof.

(1) $\|u_2 + v_2 - (u_1 + v_1)\| \le \|u_2 - u_1\| + \|v_2 - v_1\|$.

(2) $\|\alpha_2 u_2 - \alpha_1 u_1\| = \|\alpha_2 u_2 - \alpha_1 u_2 + \alpha_1 u_2 - \alpha_1 u_1\| = \|(\alpha_2 - \alpha_1)u_2 + \alpha_1(u_2 - u_1)\| \le |\alpha_2 - \alpha_1|\|u_2\| + |\alpha_1|\|u_2 - u_1\|$. ■

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

$$\big|\|x\| - \|y\|\big| \le \|x - y\|$$

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,

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

Similarly

$$\|y\| = \|y - x + x\| \le \|y - x\| + \|x\| \implies \|x\| - \|y\| \ge -\|x - y\|.$$

Thus

$$\big|\|x\| - \|y\|\big| \le \|x - y\|.$$

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

$$\big|\|x\| - \|y\|\big| \le \|x - y\|?$$ ■

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

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

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 $x, y \ne 0$. Let $\widehat{x} = \dfrac{x}{\|x\|_2}$ and $\widehat{y} = \dfrac{y}{\|y\|_2}$. Then proving equation (1.3) amounts to proving

$$|\langle \widehat{x},\, \widehat{y}\rangle| \le 1$$

because of homogeneity of the inner product.

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

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

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 $x_k \to x$. Let $y \in \mathbb{R}^n$. Then

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

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

$$xa \leq \frac{a^p}{p} + \frac{x^q}{q}. \tag{1.4}$$

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 $x = (x_1, x_2, \ldots, x_n)$, we define

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

(1) 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

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

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

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

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

(1) $q = 1$.

$$\begin{aligned}
|\langle x,\, y\rangle| &= |x_1y_1 + x_2y_2 + \cdots + x_ny_n| \\
&\leq |x_1||y_1| + |x_2||y_2| + \cdots + |x_n||y_n| \\
&= \max_{1\leq i\leq n} |x_i|(|y_1| + |y_2| + \cdots + |y_n|) \\
&= \|x\|_\infty \|y\|_1
\end{aligned}$$

and

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

(2) 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 $1 \leq p < \infty$.

(1) 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.

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

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.)

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

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

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

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

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

Similarly,

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

This gives

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

(2) 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 $1 \le p < \infty$,

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

Further,

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

Thus

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

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

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

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

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

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

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

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

Solution.

(1) Suppose f is non-zero at some point $c \in [a, b]$. By continuity, $f(x) \ge \frac{f(c)}{2}$ 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 $\int_a^b f(x)\,\mathrm{d}x > 0$. Contradiction! Such a c cannot exist.

(2) 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.

(3) 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 λ 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

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

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

(1) Show that all eigenvalues of A are real.

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

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

Solution.

(1) View A as a linear operator on $\mathbb{C}^n$. Let λ 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.

(2) (assuming spectral theorem) WLOG let $\lambda_1 = \max_{1\le i\le n}|\lambda_i|$. 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 λ_i. Then $Ax = \sum_{i=1}^n c_i \lambda_i v_i$.

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

Thus $\|A\| \le \lambda_1$. Moreover, $\|Av_1\| = |\lambda_1| \|v_1\|$. Thus $\|A\| \ge \lambda_1$. ■

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

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

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

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

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

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

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 $\lambda_1 \le \|A\|_{HS}^2 \le n\lambda_1$. This gives $\|A\| \le \|A\|_{HS} \le \sqrt{n}\|A\|$.

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

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

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,

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

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

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

Quiz

Problem 1.11. Recall the definition of a homogeneous function. Let $f\colon \mathbb{R}^n\setminus\{0\} \to \mathbb{R}$ 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

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

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

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 $x \ne 0$,

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

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 $x \ne y$. 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.

Lecture 5. Monday August 12

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

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.

Theorem 1.39.

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

(2) Let $A \in \mathrm{GL}_n(\mathbb{R})$ be fixed and let $B \in \mathrm{M}_n(\mathbb{R})$ be such that

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

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

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

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

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

$$\begin{aligned} \|A^{-1}B-I\| &= \|A^{-1}(B-A)\| \\ &\le \|A^{-1}\|\|B-A\| \\ &< 1. \end{aligned} \tag{1.6}$$

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

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

$$B^{-1} - A^{-1} = A^{-1}(A-B)B^{-1} \tag{1.7}$$

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

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

$$\begin{aligned} \|B^{-1}\| - \|A^{-1}\| &\le \|B^{-1} - A^{-1}\| \\ &\le \|B^{-1}\|\|A-B\|\|A^{-1}\| && \text{(by equation (1.7))} \\ &\le \frac{1}{2}\|B^{-1}\|. \end{aligned}$$

This bounds $\|B^{-1}\|$ above by $2\|A^{-1}\|$. Using equation (1.7) again, we have

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

As $B \to A$, $B^{-1} \to A^{-1}$. ■

Idea. This is similar in spirit to exercise 1.18.

- Equation (1.7) 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

2.1 The derivative

Definition 2.1. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function. We say that f is *differentiable* at $a \in \mathbb{R}$ if
$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$
exists. We denote this limit by $f'(a)$ and call it the *derivative* of f at a.

This doesn't make sense for $f\colon \mathbb{R}^n \to \mathbb{R}^m$ 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 $f\colon U \to \mathbb{R}$. 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 $a \in U$.
$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = f'(a)$$
We can rewrite this as
$$\begin{aligned} &\lim_{h \to 0} \frac{f(a+h) - f(a) - f'(a)h}{h} = 0 \\ \implies &\lim_{h \to 0} \frac{|f(a+h) - f(a) - T_{f'(a)}h|}{|h|} = 0 \end{aligned}$$
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

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

■

Lecture 6. Monday August 19

Definition 2.4. Let $U \subseteq \mathbb{R}^n$ be an open set containing a. Let $f\colon U \to \mathbb{R}^m$. 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

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

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.

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 satisftying the definition of differentiability at a for $f\colon U \to \mathbb{R}^m$. Then $T_1 = T_2$.*

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

$$\begin{aligned}
Th &= T_1 h - T_2 h \\
&= (f(a+h) - f(a) - T_2 h) - (f(a+h) - f(a) - T_1 h) \\
&= o(h) - o(h) = o(h).
\end{aligned}$$

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

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

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

Remark. Since T is linear, $h \mapsto \frac{\|Th\|}{\|h\|}$ is homogenous of degree 0. Thus it is defined by its value on the unit sphere. It's limit at 0 can exist only when it is constant on the unit sphere, which implies that it is constant everywhere and equals the limit.

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

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

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

We know that $\frac{\|q(h)\|}{\|h\|} \to 0$ as $h \to 0$.

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

As $h \to 0$, each term goes to 0. ∎

For *finding* the derivative, it is helpful to follow these steps:

- 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 $f\colon \mathbb{R}^n \to \mathbb{R}^m$ 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 $f\colon \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x$. This is a special case of the previous example. $f'(a) = \mathrm{id}$. 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 $\mathbb{R}^n$.

Lecture 7.
Wednesday
August 21

We will continue our examples.

Examples.

- Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ 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 $h_1 h_2 = o(h)$.

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

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

- Let $f\colon \mathrm{M}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ be given by $f(X) = X$.

  We could identify $\mathrm{M}_n(\mathbb{R})$ 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 $\mathrm{M}_n(\mathbb{R})$ to $\mathrm{M}_n(\mathbb{R})$. This is again a specical case of the second example. Thus $f'(A) = f = \mathrm{id}$.

- Let $f\colon \mathrm{M}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ 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

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

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

- Let $f\colon \mathrm{M}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ be given by $f(X) = X^3$. Then

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

- Let $f\colon \mathrm{GL}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ 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 $\|H\|$,

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

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

$A^{-1}H + u(H)$ where

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

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

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

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 $f\colon \mathrm{M}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ 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.

- $$\frac{1}{x+h} - \frac{1}{x} = \frac{1}{x+h}(x - (x+h))\frac{1}{x} \to -\frac{1}{x}h\frac{1}{x}$$

- $$\begin{aligned}\frac{1}{(x+h)^2}-\frac{1}{x^2} &= \frac{1}{(x+h)^2}(x^2-(x+h)^2)\frac{1}{x^2}\\ &= -\frac{1}{(x+h)^2}(hx+xh+h^2)\frac{1}{x^2}\\ &= -\frac{1}{(x+h)^2}hx\frac{1}{x^2}-\frac{1}{(x+h)^2}(x+h)h\frac{1}{x^2}\\ &= -\frac{1}{(x+h)^2}h\frac{1}{x}-\frac{1}{x+h}h\frac{1}{x^2}\end{aligned}$$

Lecture 8. Friday August 23

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

Proposition 2.8 (chain rule)**.** *Let $U \subseteq \mathbb{R}^n$ be open and $a \in U$. Let $f\colon U \to \mathbb{R}^m$ be differentiable at a. Let $V \subseteq \mathbb{R}^m$ be an open set containing $f(a)$ and $g\colon V \to \mathbb{R}^k$ be differentiable at $f(a)$. Then $g \circ f$ is differentiable at a with*

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

Perhaps more intuitive notation is

$$D_a(g\circ f) = D_{f(a)}g \circ D_a f,$$

so that

$$D_a(g\circ f)(h) = D_{f(a)}g(D_a f(h)).$$

This is exactly what the chain rule says.

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

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

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

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

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. This is problematic if $k(h)$ vanishes at some points arbitrarily close to 0.

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

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

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

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

Thus

$$\lim_{h\to 0}\frac{v(k(h))}{h} = 0. \qquad \blacksquare$$

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

Example. Consider $f\colon \mathrm{GL}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ and $g\colon \mathrm{M}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ 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

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

2.2 Directions and partials

Definition 2.9 (directional derivative)**.** Let $U \subseteq \mathbb{R}^n$ be open and $a \in U$. Let $f\colon U \to \mathbb{R}$. For $v \in \mathbb{R}^n$, we define

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

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

Exercise 2.10. *Show that $D_{\lambda v} = \lambda D_v$.*

Lecture 8: The chain rule; directional derivatives

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

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

■

Definition 2.11 (partial derivative)**.** Let $U \subseteq \mathbb{R}^n$ be open and $a \in U$. Let $f\colon U \to \mathbb{R}$. We define

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

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

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

Lecture 9.
Monday
August 26

Definition 2.12 (gradient)**.** Let $U \subseteq \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}$. We define the *gradient* $\nabla f\colon U \to \mathbb{R}^n$ of f by

$$(\nabla f)(a) \coloneqq \left(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\right)$$

for all $a \in U$.

Throughout this lecture, $U \subseteq \mathbb{R}^n$ is open and $a \in U$, and $f\colon U \to \mathbb{R}^m$.

Proposition 2.13. *Suppose $f = (f_1, \ldots, f_m)$ is differentiable at a. 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 $i \in [n]$,

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

Thus

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

■

Definition 2.14 (curve)**.** A map $\gamma\colon (a,b) \to \mathbb{R}^n$ is a *curve* in $\mathbb{R}^n$. If γ is differentiable, it is a *differentiable curve.*

Remark. Chain rule remains valid for curves.

Let $m = 1$ so that $f\colon U \to \mathbb{R}$, and let $\gamma\colon (a,b) \to U$. We will treat $f'(a)$ to be a scalar, instead of a map from $\mathbb{R}^n$ to $\mathbb{R}$. We will also treat $\gamma'(t)$ to be a vector in $\mathbb{R}^n$, not a map from $\mathbb{R}$ to $\mathbb{R}^n$. Another way to think about this is that we are identifying $\mathbb{R}^{1\times 1}$ with $\mathbb{R}$ and $\mathbb{R}^{n\times 1}$ with $\mathbb{R}^n$.

Then

$$\begin{aligned}
(f\circ\gamma)'(t) &= f'(\gamma(t))(\gamma'(t)) \\
&= f'(\gamma(t))\Big(\sum \gamma_i'(t)e_i\Big) \\
&= \sum \gamma_i'(t) f'(\gamma(t))(e_i) \\
&= \sum \gamma_i'(t) \frac{\partial f}{\partial x_i}(\gamma(t)) \\
&= \nabla f(\gamma(t)) \bullet \gamma'(t).
\end{aligned}$$

If we continue to treat $f'(a)$ and $\gamma'(t)$ as maps, we would write

$$\begin{aligned}
(f\circ\gamma)'(t)(h) &= f'(\gamma(t))(\gamma'(t)(h)) \\
&= \text{what now?}
\end{aligned}$$

Definition 2.15. Let $\gamma\colon (-\varepsilon, \varepsilon) \to U$ be such that $\gamma(0) = a$. $\gamma'(0)$ is the *tangent vector* to γ at a, and $(f\circ\gamma)'(0)$ is the *derivative of f along γ at a.*

Proposition 2.16. *If f is differentiable at a, then for each $v \in \mathbb{R}^n$, $D_v f(a)$ exists and equals $\nabla f(a) \bullet v$.*

Proof. Let $\gamma(t) = a + tv$. There exist some $\varepsilon > 0$ such that $\gamma(-\varepsilon, \varepsilon) \subseteq U$

and $\gamma'(0) = v$. Then

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

Lecture 10.
Wednesday
August 28

Examples.

- Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

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

  Note that $|f(x,y)| = |x| \frac{y^2}{x^2+y^2} \leq |x|$ for $(x,y) \neq 0$, so f is continuous at $(0,0)$. Suppose f were differentiable at 0. Then the derivative of f at 0 would be

$$\begin{pmatrix} D_1 f(0) & D_2 f(0) \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}.$$

  Then

$$\begin{aligned} 0 &= \lim_{(h,k) \to 0} \frac{|f(h,k) - f(0,0) - f'(0)(h,k)|}{\|(h,k)\|} \\ &= \lim_{(h,k) \to 0} \frac{f(h,k)}{\|(h,k)\|} \\ &= \lim_{(h,k) \to 0} \frac{hk^2}{(h^2+k^2)^{3/2}}. \end{aligned}$$

  This is homogenous, and thus constant along lines through the origin.

$$\begin{aligned} 0 &= \lim_{t \to 0} \frac{t^3}{(2t)^{3/2}} \\ &= \frac{1}{2\sqrt{2}}. \end{aligned}$$

  This is essentially showing that $D_{(1,1)} f(0) \neq f'(0)(1,1)$.

- Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = [(x,y) \neq 0](x^2+y^2) \sin\left(\frac{1}{\sqrt{x^2+y^2}}\right).$$

  f is continuous at 0. We first find the gradient.

$$\begin{aligned} D_1 f(x,y) &= \lim_{x \to 0} \frac{f(x,0)}{x} \\ &= \lim_{x \to 0} x \sin\left(\frac{1}{|x|}\right) \\ &= 0. \end{aligned}$$

  Similarly,

$$D_2 f(x,y) = 0.$$

Thus if the derivative exists, it must be 0.

$$\begin{aligned}\lim_{(h,k)\to 0} &\frac{|f(h,k) - f(0,0) - f'(0)(h,k)|}{\|(h,k)\|} \\ &= \lim_{(h,k)\to 0} \|(h,k)\| \sin\left(\frac{1}{\|(h,k)\|}\right) \\ &= 0.\end{aligned}$$

Thus the derivative is indeed 0.

- Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

  $$f(x,y) = [(x,y) \neq 0]\sqrt{x^2+y^2}\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right).$$

  Then $f(x,0) = x\sin\left(\frac{1}{x}\right)$ is not differentiable at 0. Thus f is not differentiable at 0.

2.3 The mean value theorem

Recall these theorems from real analysis.

Theorem 2.17 (Rolle's theorem)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). If $f(a) = f(b)$, then there exists $c \in (a,b)$ such that $f'(c) = 0$.*

Theorem 2.18 (mean value theorem)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). Then there exists $c \in (a,b)$ such that*

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

Does an analogue of the mean value theorem hold for several variables? Certainly not fully. Consider

$$f(t) = (\cos t, \sin t).$$

Then $f(0) = f(2\pi)$, but $|f'(t)| = 1$ is never 0.

Proposition 2.19. *Let $\gamma\colon [a,b] \to \mathbb{R}^n$ be continuous on $[a,b]$ and differentiable on (a,b). Then there exists $c \in (a,b)$ such that*

$$\|\gamma(b) - \gamma(a)\| \leq (b-a)\|\gamma'(c)\|.$$

We would love to use the mean value theorem. Looking at the projection on $\gamma(b) - \gamma(a)$ allows us to do this.

Proof. WLOG let $\gamma(a) = 0$. Define $f\colon [a,b] \to \mathbb{R}$ as $f(t) = \langle \gamma(t), \gamma(b)\rangle$. This is differentiable, with

$$f'(t) = \langle \gamma'(t), \gamma(b)\rangle \leq \|\gamma'(t)\|\|\gamma(b)\|.$$

By the mean value theorem, there exists $c \in (a, b)$ such that

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

Thus

$$\begin{aligned} &\langle \gamma(b), \gamma(b) \rangle \leq \|\gamma'(c)\| \|\gamma(b)\| (b - a) \\ \implies &\|\gamma(b)\| \leq (b - a) \|\gamma'(c)\|. \end{aligned}$$

■

Assignment 2

up August 16
due August 30
quiz August 29

Problem 2.1. Determine whether in each case the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ is continuous or not.

(1) $f(x,y) = [(x,y) \neq 0]\ \frac{x \sin^2 y}{x^2+y^2}$

(2) $f(x,y) = [(x,y) \neq 0]\ \frac{\sin(x^2+y^2)}{x^2+y^2} + [(x,y) = 0]\ 1$

(3) $f(x,y) = [(x,y) \neq 0]\ \frac{xy}{\sqrt{x^2+y^2}}$

(4) $f(x,y) = [(x,y) \neq 0]\ \frac{xy}{x^2+y^2}$

(5) $f(x,y) = [(x,y) \neq 0]\ \frac{xy^2}{x^2+y^4}$

Solution. All of these are sums, products, quotients, and compositions of continuous functions in $\mathbb{R}^2 \setminus \{(0,0)\}$. Thus we only need to check continuity at $(0,0)$.

(1) Continuous.

$$|f(x,y)| \leq |x| \frac{\sin^2 y}{y^2} \leq |x| \to 0$$

(2) Continuous.

(3) Continuous.

$$|f(x,y)| \leq \frac{|xy|}{|y|} = |x| \to 0$$

(4) Not continuous. This is homogenous of degree 0, but not constant on the unit circle.

(5) Not continuous.

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

Problem 2.2. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be as in problem 2.1(4). Show that $D_1 f$ and $D_2 f$ exist at every point in $\mathbb{R}^2$, although the function is not continuous at $(0,0)$.

Solution. They obviously exist in $\mathbb{R}^2 \setminus \{(0,0)\}$. At $(0,0)$, we have

$$D_1 f(0,0) = \lim_{h \to 0} \frac{f(h,0) - f(0,0)}{h} = 0.$$

Similarly $D_2 f(0,0) = 0$. ■

Problem 2.3. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \frac{xy(x^2-y^2)}{x^4+y^4} \cdot [(x,y) \neq 0].$$

Show that $D_1 f$ and $D_2 f$ exist at every point in $\mathbb{R}^2$, although the function is not continuous at $(0,0)$.

Solution. Obvious except at $(0,0)$. Also obvious at $(0,0)$.

For continuity, notice again that this is homogenous of degree 0, but its value at $(1,0)$ is different from at $(\cos 1, \sin 1)$. ■

Problem 2.4. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be as in problem 2.1(5). Show that for every $v \in \mathbb{R}^2$, the directional derivative $D_v f$ exists at every point of $\mathbb{R}^2$, although the function is not continuous at $(0,0)$.

Solution. Let $v = (a,b) \neq 0$. Obvious but at the origin.

$$\begin{aligned} D_v f(0,0) &= \lim_{t \to 0} \frac{f(t)}{t} \\ &= \lim_{t \to 0} \frac{ab^2}{a^2 + b^4 t^2} \\ &= \frac{ab^2}{a^2 + b^4}. \end{aligned}$$

■

Problem 2.5. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \frac{xy^3}{x^2 + y^6} \cdot [(x,y) \neq 0].$$

Show that for every $v \in \mathbb{R}^2$, the directional derivative $D_v f$ exists at every point of $\mathbb{R}^2$, although the function is not continuous at $(0,0)$.

Solution. It's the same thing.

$$\begin{aligned} D_v f(0,0) &= \lim_{t \to 0} \frac{ab^3 t}{a^2 + b^6 t^4} \\ &= 0. \end{aligned}$$

Continuity is similarly disproven by (t^3, t). ■

Problem 2.6. Let U be an open subset of $\mathbb{R}^n$ and $f\colon U \to \mathbb{R}^m$ be such that

$$f(x) \coloneqq (f_1(x), \dots, f_m(x)) \quad \text{for } x \in U.$$

(1) Suppose that f is differentiable at $a \in U$. Show that each f_k is differentiable at a for $k \in [m]$, with

$$f_k'(a)(v) = \langle f'(a)(v),\, e_k \rangle \quad \text{for } v \in \mathbb{R}^n.$$

(2) Suppose that each $f_k\colon U \to \mathbb{R}$ is differentiable at $a \in U$ for $k \in [m]$. Prove that f is differentiable at a with

$$f'(a)(v) = (f_1'(a)(v), \dots, f_m'(a)(v)) \quad \text{for } v \in \mathbb{R}^n.$$

Solution.

(1) Let $f'(a) = T \in L(\mathbb{R}^n, \mathbb{R}^m)$. Let $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$. Then

$$\begin{aligned} f_k'(a)(v) &= \lim_{t\to 0} \frac{f_k(a+tv) - f_k(a)}{t} \\ \implies f_k'(a)(v) - \langle T(v), e_k\rangle &= \lim_{t\to 0} \frac{f_k(a+tv) - f_k(a) - T(v)_k}{t} \\ &= \lim_{t\to 0} \frac{f(a+tv) - f(a) - T(v)}{t} \cdot e_k \\ &= 0. \end{aligned}$$

(2) Let $f_k(a+v) = f_k(a) + f_k'(a)(v) + \|v\|\varepsilon_k(v)$ where $\varepsilon_k(v) \to 0$ as $v \to 0$. Then

$$\begin{aligned} f(a+v) - f(a) &= \sum_{k=1}^m (f_k(a+v) - f_k(a))e_k \\ &= \sum_{k=1}^m f_k'(a)(v)e_k + \|v\| \sum_{k=1}^m \varepsilon_k(v)e_k. \end{aligned}$$

Since $\varepsilon_k(v) \to 0$ for each k, $(\varepsilon_1(v), \ldots, \varepsilon_m(v)) \to 0$. Thus

$$f'(a)(v) = \sum_{k=1}^m f_k'(a)(v)e_k.$$ ■

Problem 2.7. Let U be an open subset of $\mathbb{R}^n$. Let $h\colon U \to \mathbb{R}^{k+m}$ be given by

$$h(x) = (f(x), g(x)) \quad \text{for } x \in U,$$

where $f\colon U \to \mathbb{R}^k$ and $g\colon U \to \mathbb{R}^m$.

(1) Suppose that h is differentiable at $a \in U$. Show that both f and g are differentiable at a, with

$$h'(a)(v) = (f'(a)(v), g'(a)(v)) \quad \text{for } v \in \mathbb{R}^n.$$

(2) Suppose that both f and g are differentiable at $a \in U$. Prove that h is differentiable at a with

$$h'(a)(v) = (f'(a)(v), g'(a)(v)) \quad \text{for } v \in \mathbb{R}^n.$$

Solution. Use the previous problem. ■

Problem 2.8. Calculate the total derivative of the following maps.

(1) Let $f\colon M_n(\mathbb{R}) \to M_n(\mathbb{R})$ be given by

$$f(X) \coloneqq X^\top.$$

(2) Let $f\colon M_n(\mathbb{R}) \to \mathbb{R}$ be given by

$$f(X) \coloneqq \operatorname{Tr}(X).$$

(3) Let $f\colon M_n(\mathbb{R}) \to M_n(\mathbb{R})$ be given by

$$f(X) \coloneqq XX^\top.$$

(4) Let $B \in M_n(\mathbb{R})$ be fixed. Let $f\colon M_n(\mathbb{R}) \to M_n(\mathbb{R})$ be given by

$$f(X) \coloneqq X^\top BX.$$

(5) Let $A \in M_n(\mathbb{R})$ be fixed. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be given by

$$f(x) = \langle Ax,\, x\rangle.$$

Solution.

(1) $f(X+H) = X^\top + H^\top + 0$ where $H \mapsto H^\top$ is linear and $0 = o(H)$. Thus $f'(X)$ is given by

$$f'(X)(H) = H^\top.$$

(2) $f(X+H) = \mathrm{Tr}(X) + \mathrm{Tr}(H) + 0$ where $H \mapsto \mathrm{Tr}(H)$ is linear and $0 = o(H)$. Thus $f'(X)$ is given by

$$f'(X)(H) = \mathrm{Tr}(H).$$

(3) $f(X+H) = f(X) + XH^\top + HX^\top + HH^\top$ where $H \mapsto XH^\top + HX^\top$ is linear and $HH^\top = o(H)$, since $\|HH^\top\| \leq \|H\|\|H^\top\| = \|H\|^2$. Thus $f'(X)$ is given by

$$f'(X)(H) = XH^\top + HX^\top.$$

(4) $f(X+H) = f(X)+X^\top BH+H^\top BX+H^\top BH$ where $H \mapsto X^\top BH+H^\top BX$ is linear and $H^\top BH = o(H)$ since $\|H^\top BH\| \leq \|B\|\|H\|^2$. Thus $f'(X)$ is given by

$$f'(X)(H) = X^\top BH + H^\top BX.$$

(5) $f(x+h) = f(x) + \langle Ax,\, h\rangle + \langle Ah,\, x\rangle + \langle Ah,\, h\rangle$ where $h \mapsto \langle Ah,\, x\rangle + \langle Ax,\, h\rangle$ is linear and $\langle Ah,\, h\rangle = o(h)$ since $|\langle Ah,\, h\rangle| \leq \|Ah\|\|h\| \leq \|A\|\|h\|^2$. ∎

Problem 2.9. Let U be an open subset of $M_n(\mathbb{R})$. Let $f\colon U \to M_n(\mathbb{R})$ and $g\colon U \to M_n(\mathbb{R})$ be two maps that are both differentiable at $A \in U$.

(1) Show that the map $\psi\colon U \to M_n(\mathbb{R})$ given by

$$\psi(X) \coloneqq f(X)g(X)$$

is differentiable at A and $\psi'\colon M_n(\mathbb{R}) \to M_n(\mathbb{R})$ is given by

$$\psi'(A)(H) = (f'(A)(H))g(A) + f(A)(g'(A)(H)).$$

(2) Using problem 2.9(1), answer the following questions (Don't use power series).

(a) Let $f\colon M_n(\mathbb{R}) \to M_n(\mathbb{R})$ be given by

$$f(A) := A^2.$$

Calculate $f'(A)$.

(b) Let $f\colon \mathrm{GL}_n(\mathbb{R}) \to M_n(\mathbb{R})$ be given by

$$f(A) := A^{-1}.$$

Calculate $f'(A)$.

(c) Let $f\colon \mathrm{GL}_n(\mathbb{R}) \to M_n(\mathbb{R})$ be given by

$$f(A) := A^{-2}.$$

Calculate $f'(A)$.

(d) Let $f\colon M_n(\mathbb{R}) \to M_n(\mathbb{R})$ be given by

$$f(A) := A^3.$$

Calculate $f'(A)$.

(e) Let $f\colon \mathrm{GL}_n(\mathbb{R}) \to M_n(\mathbb{R})$ be given by

$$f(A) := A^{-3}.$$

Calculate $f'(A)$.

Solution.

(1) We have

$$\begin{aligned}\psi(A+H) &= f(A+H)g(A+H)\\ &= (f(A) + f'(A)(H) + o(H))(g(A) + g'(A)(H) + o(H))\\ &= \psi(A) + f(A)g'(A)(H) + f'(A)(H)g(A) + o(H).\end{aligned}$$

The remaining terms are $o(H)$ because:

- $o(H) \cdot \text{constant} = o(H)$
- $o(H) \cdot \text{linear} = o(H)$ since the linear term is bounded in a neighbourhood.
- $o(H) \cdot o(H) = o(H)$ for the same reason (and because obviously).
- linear $\cdot$ linear $= o(H)$ because

$$\frac{\|T_1(H)T_2(H)\|}{\|H\|} \le \frac{\|T_1\|\|H\|\|T_2\|\|H\|}{\|H\|} = \|T_1\|\|T_2\|\|H\| \to 0$$

for any linear maps T_1, T_2.

Thus ψ is differentiable at A with

$$\psi'(A)(H) = f(A)g'(A)(H) + f'(A)(H)g(A). \tag{$*$}$$

(2) (a) We know $X \mapsto X$ is differentiable everywhere, with the derivative being the identity map. Thus using ($*$) yields that f is differentiable everywhere, and

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

(b) We know that $X \mapsto X^{-1}$ is continuous on $\mathrm{GL}_n(\mathbb{R})$ (theorem 1.39). Write $(A+H)^{-1} - A^{-1} = -(A+H)^{-1}HA^{-1}$ in

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

since

$$\frac{\|(A^{-1} - (A+H)^{-1})HA^{-1}\|}{\|H\|} \le \left\|A^{-1}\right\| \left\|A^{-1} - (A+H)^{-1}\right\| \to 0$$

by the continuity of $X \mapsto X^{-1}$. Thus f is differentiable everywhere, and

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

(c) Using the previous part and ($*$), we have that f is differentiable everywhere with

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

Alternatively, we can conclude that f is differentiable as before, but compute the derivative using the product rule on $I = A^2A^{-2}$. Any constant map has derivative 0, so

$$\begin{aligned} & (AH + HA)A^{-2} + A^2 f'(A)(H) = 0 \\ \implies & f'(A)(H) = -A^{-2}(AH + HA)A^{-2}. \end{aligned}$$

(d) Since $X \mapsto X^2$ and $X \mapsto X$ are differentiable everywhere, ($*$) gives that f is differentiable everywhere with

$$f'(A)(H) = A^2H + (AH + HA)A = A^2H + AHA + HA^2.$$

(e) Since $X \mapsto X^{-1}$ is differentiable everywhere, ($*$) twice gives that f is also differentiable everywhere. Then the product rule on $A^3A^{-3} = I$ gives

$$\begin{aligned} & (A^2H + AHA + HA^2)A^{-3} + A^3 f'(A)(H) = 0 \\ \implies & f'(A)(H) = -A^{-3}(A^2H + AHA + HA^2)A^{-3}. \end{aligned}$$ ∎

Problem 2.10. Let U be an open subset of $M_n(\mathbb{R}) \times M_n(\mathbb{R})$. Let $F\colon U \to M_n(\mathbb{R})$ be given by

$$F(X, Y) := XY.$$

Then for any $(A, B) \in U$, calculate $F'(A, B)$. Using problem 2.7 and composition rule, give an alternate proof of problem 2.9(1).

Assignment 2 — due August 30

Solution. We have

$$\begin{aligned} F(X+H, Y+K) &= (X+H)(Y+K) \\ &= XY + HY + XK + HK \\ &= F(X,Y) + HY + XK + o(H,K), \end{aligned}$$

since HK is the product of two linear maps from (H, K) (projections). Thus F is differentiable everywhere, and

$$F'(X,Y)(H,K) = HY + XK.$$

For problem 2.9(1), consider

$$\psi(X) = F(f(X), g(X)) = (F \circ h)(X),$$

where $h(X) := (f(X), g(X))$. By problem 2.7, h is differentiable at A with

$$h'(A)(H) = (f'(A)(H), g'(A)(H)).$$

Thus by the composition rule, ψ is differentiable at A with

$$\begin{aligned} \psi'(A)(H) &= F'(h(A))(h'(A)(H)) \\ &= F'(f(A), g(A))(f'(A)(H), g'(A)(H)) \\ &= f'(A)(H)g(A) + f(A)g'(A)(H). \end{aligned}$$

■

Lecture 11. Friday August 30

Proposition 2.20. *Let U be a convex open subset of $\mathbb{R}^n$. Let $f\colon U \to \mathbb{R}^m$ be differentiable and let*

$$\sup_{x\in U}\|f'(a)\| = M.$$

Then for any $a, b \in U$,

$$\|f(a) - f(b)\| \leq M\|b - a\|.$$

Proof. Consider the function $\gamma\colon [0,1] \to \mathbb{R}^n$ given by $\gamma(t) = (1-t)a + tb$. Since U is convex and f is differentiable, $g = f \circ \gamma$ is differentiable with

$$g'(t) = f'(\gamma(t))\gamma'(t) = f'(\gamma(t))(b - a).$$

Then by proposition 2.19 applied on g,

$$\|f(b) - f(a)\| = \|g(1) - g(0)\| \leq \|g'(c)\| \leq M\|b - a\|. \qquad \blacksquare$$

Proposition 2.21. *Let $U \subseteq \mathbb{R}^n$ be open and connected and $f\colon U \to \mathbb{R}^m$ be differentiable. If $f' \equiv 0$ on U, then f is constant on U.*

Proof. Fix $a \in U$. Let $W = \{x \in U \mid f(x) = f(a)\}$. We will show that W is clopen in the subset topology on U.

Let $x \in U$ be contained in an open ball $B \subseteq U$. By proposition 2.20, f is constant on B. Thus W is open in U. But $W = f^{-1}(f(a))$ is closed in U. Thus W is all of U. $\blacksquare$

Proposition 2.22. *Let $U \subseteq \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}^m$ be such that $D_j f_i$ exist and are bounded on U. Then f is continuous.*

Proof. Fix an $i \in [m]$. Let each D_j be bounded by M on all of U. Let $x \in U$ and $\varepsilon > 0$ be such that $B = B(x, \varepsilon) \subseteq U$. Let $(y_1, \ldots, y_n) \in B$. Consider $x^{(j)} = x^{(j-1)} + (y_j - x_j)e_i$ with $x^{(0)} = x$. Then

$$\|f_i(x^{(j)}) - f_i(x^{(j-1)})\| \leq M|y_j - x_j|$$

by the mean value theorem. By the triangle inequality,

$$\|f_i(y) - f_i(x)\| \leq M\sum_{j=1}^{n}|y_j - x_j| \leq Mn\varepsilon. \qquad \blacksquare$$

Definition 2.23. Let $U \subseteq \mathbb{R}^n$ be open. $C^1(U, \mathbb{R}^m)$ is the set of all functions $f\colon U \to \mathbb{R}^m$ such that $f'\colon U \to L(\mathbb{R}^n, \mathbb{R}^m)$ exists and is continuous.

Lecture 12. Monday September 2

Proposition 2.24. *Let $U \subseteq \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}^m$. Then $f \in C^1(U)$ iff $D_j f_i$ exist and are continuous on U.*

Lecture 12: C^1 from partials

Proof. Suppose $f \in C^1(U)$. The derivative at any point exists, hence so do the partial derivatives. Moreover, the partial derivatives are continuous functions of the derivative, hence they are continuous.

$$D_j f_i(x) = \langle e_i \mid f'(x) \mid e_j \rangle.$$

More directly,

$$\begin{aligned} |D_j f_i(x) - D_j f_i(y)| &= |\langle e_i \mid f'(x) - f'(y) \mid e_j \rangle| \\ &\le \|f'(x) - f'(y)\|. \end{aligned}$$

Conversely, suppose the partial derivatives are continuous. Let $T\colon U \to L(\mathbb{R}^n, \mathbb{R}^m)$ be given by

$$T(x) = \begin{pmatrix} D_1 f_1(x) & D_2 f_1(x) & \dots & D_n f_1(x) \\ D_1 f_2(x) & D_2 f_2(x) & \dots & D_n f_2(x) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f_m(x) & D_2 f_m(x) & \dots & D_n f_m(x) \end{pmatrix}.$$

For any $x, x + h \in U$ close enough, we have (similar to the proof of proposition 2.22)

$$f_i(x+h) - f_i(x) = \sum_{j=1}^n D_j f_i(c_j) h_j$$

by the mean value theorem, where $\|c_j - x\| \le \|h\|$. Thus

$$\begin{aligned} |f_i(x+h) - f_i(x) - (T(x)h)_i| &= \left| \sum_{j=1}^n (D_j f_i(c_j) - D_j f_i(x)) h_j \right| \\ &\le \sum_{j=1}^n |D_j f_i(c_j) - D_j f_i(x)| |h_j| \\ &\le \|h\| \sum_{j=1}^n |D_j f_i(c_j) - D_j f_i(x)|. \end{aligned}$$

The second term goes to zero by continuity, thus T is the derivative of f. T is continuous as each entry is continuous. ■

2.4 The inverse function theorem

2.4.1 Contraction maps

Lecture 13.
Wednesday
September 4

Lecture 13: Contractions and the Banach fixed point

Definition 2.25 (contraction map)**.** Let X, Y be two metric spaces. Then $f\colon X \to Y$ is a *contraction map* if $d(f(x_1), f(x_2)) < d(x_1, x_2)$ for each $x_1 \neq x_2 \in X$.

If there is a $k \in [0, 1)$ such that

$$d(f(x_1), f(x_2)) \leq kd(x_1, x_2) \text{ for all } x_1, x_2 \in X,$$

f is a *strict contraction map*.

This is a special case of a Lipschitz map, with Lipschitz constant less than (for strict contractions) or equal to (for contractions) 1.

Theorem 2.26 (Banach fixed-point theorem)**.** *Let X be a complete metric space and $f\colon X \to X$ be a strict contraction map. Then f has a unique fixed point z.*

Furthermore, for any $x_0 \in X$, the sequence $x_{n+1} = f(x_n)$ converges to z.

Proof. Let $k \in [0, 1)$ be the contraction factor, and let $x_0 \in X$. Then $d(x_{n+1}, x_{n+2}) \leq kd(x_n, x_{n+1})$. By induction, $d(x_n, x_{n+1}) \leq k^n d(x_0, x_1)$. Thus $(x_n)_n$ is Cauchy, and since X is complete, it converges to some $z \in X$.

Since f is Lipschitz,

$$f(z) = f\left(\lim_{n\to\infty} x_n\right) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = z.$$

Suppose there were another fixed point $z' \neq z$. Then

$$d(z, z') = d(f(z), f(z')) \leq kd(z, z'),$$

a contradiction. ■

Examples.

- $x \mapsto x/2$ is a contraction map on $(0, 1)$, but has no fixed point. This does not contradict the theorem, since $(0, 1)$ is incomplete.
- $\cos$ on $[0, \pi/2]$ is a (weak) contraction map, but $\cos$ on $[0, \pi/3]$ is a strict contraction map.
- $f\colon \mathbb{R} \to \mathbb{R}$ given by $x \mapsto \sqrt{x^2+1}$ is a (weak) contraction map but not a strict one.

  $$f'(x) = \frac{x}{\sqrt{x^2+1}} \implies |f'(x)| < 1$$

  It is not a strict contraction since there are no fixed points.

Tao states a consequence immediately after the Banach fixed-point theorem, which we cover as part of the inverse function theorem.

Lecture 13: Contractions and the Banach fixed point

Lemma 2.27. *Let $B(0;r)$ be a ball in $\mathbb{R}^n$ centered at the origin, and let $g\colon B(0;r) \to \mathbb{R}^n$ be a map such that $g(0) = 0$ and*

$$\|g(x) - g(y)\| \leq k\|x - y\|$$

for all $x, y \in B(0;r)$, for some $k < 1$. Then the function $f\colon B(0;r) \to \mathbb{R}^n$ defined by $f(x) = x + g(x)$ is injective, and furthermore the image $f(B(0;r))$ contains the ball $B(0;r - kr)$.

Proof. For $x, y \in B(0;r)$, $x + g(x) = y + g(y)$ implies $\|x - y\| = \|g(y) - g(x)\| \leq k\|x - y\|$, so that $x - y = 0$.

Fix a $y \in B(0;r - kr)$. We need an x such that $f(x) = y$, or $x = y - g(x)$. Thus, we need a fixed point of the map $x \mapsto y - g(x)$. Denote this map by ϕ_y. Then for $x \in B(0;r)$,

$$\|\phi_y(x)\| \leq \|y\| + \|g(x)\| \leq r - kr + kr = r. \tag{2.2}$$

Thus ϕ_y maps $B(0;r)$ to itself. However, this is not complete, so we cannot apply the Banach fixed-point theorem directly. We must first restrict the domain to a smaller closed ball. In fact, this also follows from equation (2.2), where $\|\phi_y(x)\| \leq kr + \|y\| < r$.

Thus ϕ_y is a strict contraction map on $B(0; kr + \|y\|)$. By the Banach fixed-point theorem, there is a unique fixed point x such that $x = \phi_y(x)$. This x is the desired preimage of y. ■

Remark. Let $f\colon X \to Y$ be a contraction map. Let $A \subseteq X$. Then $f|_A\colon A \to Y$ is also a contraction map.

Lecture 14.
Friday
September 6

2.4.2 Single variable

Theorem 2.28 (1D inverse function theorem)**.** *Let $U \subseteq \mathbb{R}$ be open $f\colon U \to \mathbb{R}$ be C^1, and $f'(a) \neq 0$ for some $a \in U$. Then there exists an open interval $J \ni a$ such that*

(1) f is injective on J,

(2) $f(J)$ is an open interval in $\mathbb{R}$.

(3) Let $g\colon f(J) \to J$ be the inverse of $f|_J$. Then $g \in C^1(f(J))$.

Proof. WLOG assume $f'(a) > 0$. Since $f \in C^1(\mathbb{R})$, there is an open interval $J \ni a$ on which $f' > \frac{f'(a)}{2}$. Then $f|_J$ is strictly increasing and hence injective.

$f(J)$ is connected since J is. Choosing J small enough makes $f(J)$ bounded. Since f' is never zero on J, f does not attain a maximum or minimum on J. Thus $f(J)$ is of the form (d_1, d_2) and hence open.

Now let g be as in the statement.

Claim. *g is continuous.*

Proof of claim. Let $y_0 = f(x_0) \in f(J)$ and $y = f(x) \in f(J)$. By the mean value theorem, $y - y_0 = f'(c)(x - x_0)$ for some $c \in (x_0, x)$. Thus

$$|g(y) - g(y_0)| = \frac{1}{f'(c)}|y - y_0| \leq \frac{2}{f'(a)}|y - y_0|.$$

This proves that g is Lipschitz. □

Claim. *g is differentiable.*

Proof of claim. Let $y_0 = f(x_0)$ and $y = f(x) \in f(J)$. Then

$$\begin{aligned}\lim_{y \to y_0} \frac{g(y) - g(y_0)}{y - y_0} &= \lim_{y \to y_0} \frac{1}{\frac{f(g(y)) - f(g(y_0))}{g(y) - g(y_0)}} \\ &= \frac{1}{f'(g(y_0))}.\end{aligned}$$

Thus g is differentiable with

$$g'(y) = \frac{1}{f'(g(y))}.$$ □

Now g' is the composition of continuous functions

$$f(J) \xrightarrow{g} J \xrightarrow{f'} \mathbb{R}^{\times} \xrightarrow{1/\cdot} \mathbb{R}^{\times}.$$

Thus g' is continuous. ■

Lecture 15. Monday September 9

2.4.3 Several variables

Theorem 2.29 (inverse function theorem)**.** *Let $U \subseteq \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}^n$. Suppose $f \in C^1(U, \mathbb{R}^n)$ and $f'(a)$ is invertible for some $a \in U$. Then*

(1) there exists an open set $V \subseteq U$ containing a such that f is injective on V.

(2) $f(V) \subseteq_{\text{op}} \mathbb{R}^n$.

(3) let $g\colon f(V) \to V$ be the inverse of f. Then $g \in C^1(f(V), \mathbb{R}^n)$ and $g'(y) = f'(g(y))^{-1}$.

Proof. Let $A = f'(a)$. There is an open ball V containing a such that $\|f'(x) - A\| < \frac{1}{2}\|A^{-1}\|^{-1}$ for $x \in U$. Call $f'(x)$ as B. For each $y \in \mathbb{R}^n$ define the map

$$\begin{aligned}\phi_y \colon U &\to \mathbb{R}^n \\ x &\mapsto x + A^{-1}(y - f(x)).\end{aligned}$$

Note that $f(x) = y \iff \phi_y(x) = x$. Since

$$\phi_y'(x) = I - A^{-1}f'(x) = I - A^{-1}B,$$

we have $\|\phi_y'(x)\| < \frac{1}{2}$ by equation (1.6). By proposition 2.20, this is a contraction.

$$\|\phi_y(x_1) - \phi_y(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|. \tag{2.3}$$

Thus ϕ_y has at most one fixed point, and so f is injective on V.

We need to show that $f(V)$ is open. Let $y_0 = f(x_0) \in f(V)$. Choose an $r > 0$ such that $\overline{B} = \overline{B(x_0, r)} \subseteq V$.

Claim. *Whenever* $|y - y_0| < \frac{1}{2}\|A^{-1}\|^{-1}r$, $y \in f(V)$.

Proof of claim. We will show that ϕ_y is a contraction map on $\overline{B}$. Let $x \in \overline{B}$. Then

$$\begin{aligned}\|\phi_y(x) - x_0\| &\leq \|\phi_y(x) - \phi_y(x_0)\| + \|\phi_y(x_0) - x_0\| \\ &\leq \frac{1}{2}\|x - x_0\| + \|A^{-1}\|\|y - y_0\| \\ &< \frac{1}{2}r + \frac{1}{2}r = r.\end{aligned}$$

That is, ϕ_y maps $\overline{B}$ to $\overline{B}$, and we already showed it is strictly contracting with factor $\frac{1}{2}$. □

By the Banach fixed point theorem, ϕ_y has a unique fixed point in $\overline{B}$, so that $y \in f(V)$. Thus $f(V)$ is open (y_0 was arbitrary).

Lecture 16.
Wednesday
September 11

Similarly,

$$\begin{aligned}\|\phi_y(x_1) - \phi_y(x_2)\| &= \|x_1 - x_2 - A^{-1}(f(x_1) - f(x_2))\| \\ &\geq \|x_1 - x_2\| - \|A^{-1}(f(x_1) - f(x_2))\| \\ &\geq \|x_1 - x_2\| - \|A^{-1}\|\|f(x_1) - f(x_2)\|.\end{aligned}$$

Combining this with equation (2.3) gives

$$\|x_1 - x_2\| \leq 2\|A^{-1}\|\|f(x_1) - f(x_2)\|. \tag{2.4}$$

This shows that g is (Lipschitz) continuous.

Let $y_0 = f(x_0)$ and $y_0 + k = f(x_0 + h)$. Let $r(k) = g(y_0 + k) - g(y_0) -$

$f'(g(y_0))^{-1}(k)$. Then

$$\begin{aligned} r(k) &= h - f'(x_0)^{-1}(k) \\ &= -f'(x_0)^{-1}(k - f'(x_0)h) \\ &= -f'(x_0)^{-1}(f(x_0+h) - f(x_0) - f'(x_0)h) \\ &= -f'(x_0)^{-1}o(h). \end{aligned}$$

Let $T = -f'(x_0)^{-1}$. Then

$$\begin{aligned} \frac{\|r(k)\|}{\|k\|} &= \frac{\|T(o(h))\|}{\|k\|} \\ &\leq \|T\| \frac{\|o(h)\|}{\|k\|} \end{aligned}$$

Equation (2.4) gives $\|h\| \leq 2\|A^{-1}\|\|k\|$, so

$$\frac{\|r(k)\|}{\|k\|} \leq 2\|T\|\|A^{-1}\| \frac{\|o(h)\|}{\|h\|} \to 0$$

since $h \to 0$ as $k \to 0$ (by (2.4)). Thus $r(k) = o(k)$, proving that g is differentiable at y_0 with

$$g'(y_0) = f'(g(y_0))^{-1}.$$

Since g and f' are continuous, g is C^1.

$$f(V) \xrightarrow{g} V \xrightarrow{f'} \mathrm{GL}_n(\mathbb{R}) \xrightarrow{(\cdot)^{-1}} \mathrm{GL}_n(\mathbb{R}).$$ ■

Corollary 2.30. *Let f be as in theorem 2.29, but suppose that $f'(x)$ is invertible for every $x \in U$. Then*

(1) f is locally injective.

(2) f is an open map.

Proof. (1) is only a restatement of theorem 2.29 (1).

For (2), let $E \subseteq U$ be open. For each $a \in E$, there exists an open set $V_a \subseteq E$ containing a such that $f(V_a)$ is open. Then $f(E) = \bigcup_{a \in E} f(V_a)$ is open. ■

Let $U \subseteq_{\text{op}} \mathbb{R}^n$ and $f \colon U \to \mathbb{R}^m$ be differentiable. Let $f = (f_1, \ldots, f_m)$ and write

$$[f'(x)]_{i,j} = [D_j f_i(x)]_{i,j}.$$

$f' \colon U \to L(\mathbb{R}^n, \mathbb{R}^m) \cong M_{m \times n}(\mathbb{R})$. This has a normed linear space structure itself. Thus we can talk about differentiability of f'.

Definition 2.31. Let $U \subseteq_{\text{op}} \mathbb{R}^n$. Write $f^{(k)}$ for the kth derivative of f, with $f^{(0)} = f$. Then

$$C^k(U, \mathbb{R}^m) = \{g\colon U \to \mathbb{R}^m \mid g, g', g'', \dots, g^{(k)} \text{ exist and are continuous}\}.$$

Note that for $k \geq 1$, $g \in C^k(U, \mathbb{R}^m)$ iff $g' \in C^{k-1}(U, L(\mathbb{R}^n, \mathbb{R}^m))$.

Also recall that $f \in C^1(U, \mathbb{R}^k)$ iff $D_j f_i \in C^1(U, \mathbb{R})$ for all i, j.

Thus $f' \in C^1(U, L(\mathbb{R}^n, \mathbb{R}^m))$ iff $D_k D_j f_i \in C^1(U, \mathbb{R})$ for all $i \in [m]$ and $j, k \in [n]$.

Thus $f \in C^2(U, \mathbb{R}^m)$ iff all second order partial derivatives of f exist and are continuous.

Exercise 2.32. *Compute the second derivative of $X \mapsto X^{-1}$ on $\mathrm{GL}_n(\mathbb{R})$.*

Solution. End of the next lecture. ∎

Lecture 16: Inverse function theorem – the beginning

Assignment 3

up September 6
due
September 14

Problem 3.5. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

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

(1) Show directly that f is continuous.

(2) Prove that $D_1 f$ and $D_2 f$ are bounded functions in $\mathbb{R}^2$ (hence f is continuous by problem 4)

(3) Let u be any vector in $\mathbb{R}^2$. Show that the directional derivative $D_u f(0,0)$ exists, and that its absolute value is at most one.

(4) Show that f is not differentiable at $(0,0)$.

Solution.

(1) Let $(x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$ and let $(h,k) \in \mathbb{R}^2$ be such that $\|(h,k)\| < \|(x,y)\|$. Then

$$f(x+h, y+k) = \frac{(x+h)^3}{(x+h)^2 + (y+k)^2}.$$

As $(h,k) \to (0,0)$, the numerator goes to x^3 and the denominator goes to $x^2 + y^2 > 0$. Therefore $f(x+h,y+k) \to f(x,y)$ as $(h,k) \to (0,0)$.

For continuity at $(0,0)$, note that

$$f(h,k) = \frac{h^3}{h^2+k^2} \leq \frac{h^3}{h^2} = h,$$

so that $f(h,k) \to 0$ as $(h,k) \to (0,0)$.

(2) For each $y \in \mathbb{R}$, let $g_y = x \mapsto f(x,y)$, and for each $x \in \mathbb{R}$, let $h_x = y \mapsto f(x,y)$. Then for $(x,y) \neq (0,0)$, we have

$$\begin{aligned} D_1 f(x,y) = g_y'(x) &= \frac{3x^2(x^2+y^2) - x^3(2x)}{(x^2+y^2)^2} = \frac{x^4 + 3x^2y^2}{(x^2+y^2)^2} \\ D_2 f(x,y) = h_x'(y) &= \frac{-2x^3y}{(x^2+y^2)^2}. \end{aligned}$$

Write

$$\begin{aligned} |D_1 f(x,y)| &= \frac{x^2}{x^2+y^2}\frac{x^2+3y^2}{x^2+y^2} \leq 1 \\ |D_2 f(x,y)| &= \frac{x^2}{x^2+y^2}\frac{2|xy|}{x^2+y^2} \leq 1. && \text{(AM-GM)} \end{aligned}$$

For $(0,0)$, we have $g_0(x) = x$ and $h_0(y) = 0$. Thus

$$D_1 f(0,0) = 1, \quad D_2 f(0,0) = 0.$$

Thus $D_1 f$ and $D_2 f$ are both bounded by 1 in all of $\mathbb{R}^2$.

(3) Let $u = (u_1, u_2) \neq 0$. ($D_{(0,0)} f$ is trivially 0.) Then

$$\begin{aligned} D_u f(0,0) &= \lim_{t \to 0} \frac{f(tu_1, tu_2)}{t} \\ &= \lim_{t \to 0} \frac{t^3 u_1^3}{t^2(u_1^2 + u_2^2) \cdot t} &= \frac{u_1^3}{u_1^2 + u_2^2}. \end{aligned}$$

(4) If f were differentiable at $(0,0)$, then $f'(0,0)$ would be the given by the matrix $\begin{bmatrix} 1 & 0 \end{bmatrix}$. Then $D_u f(0,0) = u_1$ for all $u \neq 0$. By the previous part, this is not the case (for example, when $u = (1,1)$). Thus f is not differentiable at $(0,0)$. ■

Problem 3.14. Let $f\colon \mathbb{R} \to \mathbb{R}$ be given by

$$f(x) = \begin{cases} x + 2x^2 \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Show that

(1) f is differentiable on $\mathbb{R}$ and $f'(0) = 1$.

(2) f' is continuous except at 0.

(3) f' is bounded on $(-1, 1)$.

(4) On any open interval around 0, there are points x with $f'(x) > 0$ and also points x with $f'(x) < 0$.

(5) f is not one-to-one in any neighborhood of 0.

Thus the continuity of f' cannot be eliminated from the hypothesis of the inverse function theorem even for the case $n = 1$.

Solution.

(1) Sum, product and chain rules give that f is differentiable everywhere on $\mathbb{R}^\times$. We can compute it as

$$f'(x) = 1 + 4x \sin \frac{1}{x} - 2 \cos \frac{1}{x}.$$

For differentiability at 0, we have

$$f'(0) = \lim_{h \to 0} 1 + 2h \sin \frac{1}{h} = 1.$$

(2) As the sum, product, composition of continuous functions, f' is continuous everywhere on $\mathbb{R}^\times$. For continuity at 0, consider the sequence $x_n = \frac{1}{2n\pi}$. As $n \to \infty$, $x_n \to 0$ but $f'(x_n) \to -1 \neq f(0)$. Thus f' is not continuous at 0.

(3) We know $f'(0) = 1$, and for $|x| < 1$, we have
$$|f'(x)| \leq 1 + 4|x| + 2 \leq 7.$$

(4) We have already produced a sequence $x_n \to 0$ such that $f'(x_n) = -1$ for each n. Consider the sequence $y_n = \frac{1}{(2n+1)\pi}$. Then $y_n \to 0$ and $f'(y_n) = 3$ for each n. Thus there exist x arbitrarily close to 0 with $f'(x) > 0$, and x arbitrarily close to 0 with $f'(x) < 0$.

(5) Let U be any open neighborhood of 0.

f is continuous since it is differentiable. By problem 13, f is one-to-one on U only if it is either strictly increasing or strictly decreasing on U.

If f is increasing on U, then $f'(x) \geq 0$ for all $x \in U$. If f is decreasing on U, then $f'(x) \leq 0$ for all $x \in U$. However, from the previous part, we know that there are points in U where f' is positive and points where f' is negative. Thus neither of these cases can hold, so f is not one-to-one in any neighborhood of 0. ■

Problem 3.16. Suppose X, Y are normed linear spaces over $\mathbb{R}$ (need not be of finite dimension) and $T\colon X \to Y$ is linear. Prove that the following are equivalent.

(1) T is continuous at some point of X.

(2) T is continuous at 0.

(3) T is continuous.

(4) There exists a constant $C > 0$ such that $\|T(x)\| \leq C\|x\|$ for all $x \in X$.

(5) For every bounded subset V of X, $T(V)$ is bounded in Y.

(6) $T(U)$ is a bounded subset of Y, where $U \coloneqq \{x \in X : \|x\| \leq 1\}$.

(7) T is uniformly continuous.

Solution. Each implication in
$$(7) \implies (3) \implies (2) \implies (1)$$
is obvious. We prove (1) $\implies$ (7) to close this chain.

Claim. *If T is continuous at some $x_0 \in X$, then T is uniformly continuous.*

Proof of claim. Let $\varepsilon > 0$ and $\delta > 0$ be such that
$$\|T(x) - T(x_0)\| < \varepsilon \quad \text{whenever } \|x - x_0\| < \delta.$$
Let $x, x' \in X$ be such that $\|x - x'\| < \delta$. Then
$$\begin{aligned}\|T(x) - T(x')\| &= \|T(x_0 + (x - x')) - T(x_0)\| && \text{(linearity)}\\ &< \varepsilon\end{aligned}$$
since $(x_0 + (x - x')) - x_0 = x - x'$. Thus T is uniformly continuous. □

We now show
$$(4) \implies (5) \implies (6) \implies (4).$$

((4) $\implies$ (5)) Let V be such that $\|v\| \leq M$ for all $v \in V$. Then for any $v \in V$, we have
$$\|T(v)\| \leq C\|v\| \leq CM$$
so that $T(V)$ is bounded.

((5) $\implies$ (6)) U is bounded, so $T(U)$ is bounded.

((6) $\implies$ (4)) Let $T(u) \leq C$ for all $u \in U$. Let $x \in X$ be arbitrary. Since T is linear,
$$\begin{aligned} T(x) &= \|x\| T\left(\frac{x}{\|x\|}\right) \\ &\leq C\|x\| \end{aligned}$$
since $\frac{x}{\|x\|} \in U$.

We have proven
$$(7) \iff (3) \iff (2) \iff (1) \quad \text{and} \quad (4) \iff (5) \iff (6).$$
We prove (2) $\iff$ (6) to prove equivalence of all statements.

Suppose T is continuous at 0. Then there exists $\delta > 0$ such that
$$\|T(x)\| < 1 \quad \text{whenever } \|x\| < \delta.$$
Plugging $x = \delta u$ gives
$$\|T(\delta u)\| < 1 \quad \text{whenever } \|\delta u\| < \delta.$$
By linearity of T and absolute homogeneity of the norm, we have
$$\|T(u)\| < \frac{1}{\delta} \quad \text{whenever } \|u\| < 1.$$
Thus $T(U)$ is bounded.

Conversely, suppose $T(U)$ is bounded. Then for any $x \in X$, $\|T(x)\| \leq C\|x\|$ for some $C > 0$ (since (6) $\implies$ (4)). For any $\varepsilon > 0$, choose $\delta = \frac{\varepsilon}{C}$. Then if $\|x\| < \delta$, we have
$$\|T(x)\| \leq C\|x\| < C\delta = \varepsilon.$$
Thus T is continuous at 0. ∎

Lecture 17.
Friday
September 13

Theorem 2.33. *Let $U \subseteq_{\mathrm{op}} \mathbb{R}^n$ and $f = (f_1, \dots, f_m) \colon U \to \mathbb{R}^m$. Then*

(1) for $k \geq 1$, $f \in C^k(U, \mathbb{R}^m)$ iff f' exists and $f' \in C^{k-1}(U, L(\mathbb{R}^n, \mathbb{R}^m))$.

(2) $f \in C^2(U, \mathbb{R}^m)$ iff all partial derivatives of f_i upto second order exist and are continuous.

(3) $f \in C^k(U, \mathbb{R}^m)$ iff all partial derivatives of f_i upto kth order exist and are continuous.

Definition 2.34 (smoothness)**.** $f \colon U \subseteq_{\mathrm{op}} \mathbb{R}^n \to \mathbb{R}^m$ is *smooth* if $f \in C^\infty(U, \mathbb{R}^m)$, where

$$C^\infty(U, \mathbb{R}^m) = \bigcap_{k=1}^{\infty} C^k(U, \mathbb{R}^m).$$

Corollary 2.35. *From theorem 2.33, f is smooth iff all partial derivatives of f_i of all orders exist and are continuous.*

Example. The function

$$\begin{aligned} i \colon \mathrm{GL}_n(\mathbb{R}) &\to X^{-1} \in \mathrm{M}_n(\mathbb{R}) \\ X &\mapsto X^{-1} \end{aligned}$$

is smooth, since

$$i(X) = \frac{1}{\det X} \operatorname{adj} X$$

is rational in the entries of X. That is, each entry of $i(X)$ is a rational function of the entries of X. Thus the partial derivatives of each coordinate function of i exist and are continuous.

Definition 2.36 (diffeomorphism)**.** Let $U, V \subseteq_{\mathrm{op}} \mathbb{R}^n$. A bijective differentiable map $f \colon U \to V$ is a *diffeomorphism* if $f^{-1} \colon V \to U$ is also differentiable.

More generally, a bijective C^k map $f \colon U \to V$ is a C^k-diffeomorphism if f^{-1} is also C^k.

We will now state and prove a generalised version of the inverse function theorem.

Theorem 2.37 (generalized inverse function theorem)**.** *Let $U \subseteq_{\mathrm{op}} \mathbb{R}^n$ and $f \colon U \to \mathbb{R}^n$ be C^k with $k \geq 1$. Suppose $f'(a)$ is invertible for some $a \in U$. Then there exist open sets $V \ni a, W \subseteq_{\mathrm{op}} \mathbb{R}^n$ such that $f \colon V \to W$ is a C^k-diffeomorphism.*

The case $k = 1$ is the usual inverse function theorem.

Exercise 2.38. *Let $U \subseteq_{\text{op}} \mathbb{R}^n$ and $V \subseteq_{\text{op}} \mathbb{R}^m$. Suppose $f\colon U \to V$ is a differentiable bijection, whose inverse is also differentiable. Then $m = n$.*

Solution. $f^{-1} \circ f = \mathrm{id}_V$ and $f \circ f^{-1} = \mathrm{id}_U$. Both of these are differentiable, with the derivative being the identity map everywhere. By the composition rule,

$$\begin{aligned} \mathrm{id}_U'(x) &= (f^{-1} \circ f)'(x) = f^{-1\prime}(f(x)) \circ f'(x) \\ \mathrm{id}_V'(y) &= (f \circ f^{-1})'(y) = f'(f^{-1}(y)) \circ f^{-1\prime}(y) \end{aligned}$$

requires $f'(x) \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $f^{-1\prime}(y) \in L(\mathbb{R}^m, \mathbb{R}^n)$ to be injective. This requires $m = n$. ∎

Fact 2.39. *Let $U \subseteq_{\text{op}} \mathbb{R}^n$ and $V \subseteq_{\text{op}} \mathbb{R}^m$. Suppose $f\colon U \to V$ is a homeomorphism. Then $m = n$.*

The proof is hard.

For the midterm, you should be comfortable with finding derivatives without any computation. You will not be required to prove the derivatives. For example, recall the product rule: If $f(x) = u(x)v(x)$, then $f'(a)(h) = u'(a)(h)v(a) + u(a)v'(a)(h)$.

Examples.

- $f(X) = X^2$ has derivative $f'(A)(H) = HA + AH$.
- $f(X) = XBX$ has derivative $f'(A)(H) = HBA + ABH$.
- $f(X) = X^{-1}BX$ has derivative $f'(A)(H) = -A^{-1}HA^{-1}BA + A^{-1}BH$.
- $f(X) = X^{-1}$ has derivative $f'(A)(K) = -A^{-1}KA^{-1}$. What is f''?

$$\begin{aligned} f\colon &\ \mathrm{GL}_n \to \mathrm{M}_n \\ f'\colon &\ \mathrm{GL}_n \to L(\mathrm{M}_n, \mathrm{M}_n) \\ f''\colon &\ \mathrm{GL}_n \to L(\mathrm{M}_n, L(\mathrm{M}_n, \mathrm{M}_n)) \\ f''(A)\colon &\ \mathrm{M}_n \to L(\mathrm{M}_n, \mathrm{M}_n) \\ f''(A)(H)\colon &\ \mathrm{M}_n \to \mathrm{M}_n \\ f''(A)(H)(K)\colon &\ \mathrm{M}_n \end{aligned}$$

We know $f'(A)(H) = A^{-1}HA^{-1}$. We want $f'(A+K) - f'(A)$. We can evaluate this pointwise. We will see why in **??**.

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

and we use $(A+K)^{-1} = A^{-1} - A^{-1}KA^{-1} + o(K)$ to write

$$\begin{aligned} f'(A+K)(H) - f'(A)(H) &= A^{-1}HA^{-1} - A^{-1}KA^{-1}HA^{-1} \\ &\quad - A^{-1}HA^{-1}KA^{-1} - A^{-1}HA^{-1} \\ &= -A^{-1}KA^{-1}HA^{-1} - A^{-1}HA^{-1}KA^{-1} \\ &= -\frac{1}{A}K\frac{1}{A}H\frac{1}{A} - \frac{1}{A}H\frac{1}{A}K\frac{1}{A}. \end{aligned}$$

This is the second derivative of $X \mapsto X^{-1}$.

$$\begin{aligned} T\colon \mathrm{GL}_n &\to L(\mathrm{M}_n, L(\mathrm{M}_n, \mathrm{M}_n)) \\ T(A)(K)(H) &= -A^{-1}KA^{-1}HA^{-1} - A^{-1}HA^{-1}KA^{-1}. \end{aligned}$$

What we have shown is that

$$\lim_{K\to 0} \frac{\|f'(A+K)(H) - f'(A)(H) - T(A)(K)(H)\|}{\|K\|} = 0$$

for each $H \in \mathrm{M}_n$. Taking the supremum over all $\|H\| = 1$ gives

$$\lim_{K\to 0} \frac{\|f'(A+K) - f'(A) - T(A)(K)\|}{\|K\|} = 0.$$

Let $v \in \mathbb{R}^2$. Let $R_v = \{tv \mid t \in \mathbb{R}^+\} \subseteq \mathbb{R}^2 \setminus \{0\}$, and $L_v = R_v \cup R_{-v}$. If $v = 0$, $R_v = L_v = \{0\}$. **Lecture 18.** Monday September 23

What functions are constant on each R_v? f satisfying $f(tv) = f(v)$ for all $t > 0$, $v \in \mathbb{R}^2$. These are precisely homogenous functions of degree 0.

We can identify $\mathbb{R}^2 \setminus \{0\}$ with $\mathbb{C}^\times$. Define the argument function $\arg\colon \mathbb{C}^\times \to [0, 2\pi)$ by declaring $z = |z|e^{i\arg(z)}$ for all $z \in \mathbb{C}^\times$. Note that any homogenous function of degree 0 on $\mathbb{R}^2 \setminus \{0\}$ is precisely a real valued function of $\arg z$.

Example. Any homogenous function on $\mathbb{R}^2 \setminus \{0\}$ can be extended to a continuous function on $\mathbb{R}^2$ iff it is constant. Thus $f(x,y) = xy \odot \frac{1}{x^2+y^2}$ is not continuous at 0.

Question 2.40. *Give an example of $f\colon \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$ which is constant on each L_v.*

Solution. $(x,y) \mapsto |\arg(x+iy)|$ works obviously, where $\arg(x+iy) \in [-\pi, \pi)$. Being constant on each L_v implies being a function of $|\arg|$. ■

Definition 2.41 (evenness)**.** A map $f\colon X \to Y$ between two vector spaces is said to be even (resp. odd) if $f(-x) = f(x)$ (resp. $-f(x)$) for all $x \in X$.

As a special case, functions constant on each L_v are even homogenous functions of degree 0.

Question 2.42. *Give an example of a function $f\colon \mathbb{R}^2 \to \mathbb{R}$ such that $(D_v f)(0)$ exists for any $v \in \mathbb{R}^2$ but f is not even continuous at 0.*

Solution. $D_v f(0)$ exists for all v implies that the restriction of f to any line L passing through the origin is differentiable and hence continuous at the origin.

If we cook up a null-homogenous function, then

$$\lim_{t\to 0} \frac{f(tv) - f(0)}{t} = \lim_{t\to 0} \frac{f(v)}{t}$$

does not exist. If we choose an odd homogenous function of degree 1 and set $f(0) = 0$, we would have

$$\lim_{t\to 0} \frac{f(tv) - f(0)}{t} = \lim_{t\to 0} f(v) = f(v).$$

Let us try $f(x, y) = x \odot \frac{x^2}{x^2+y^2}$. This fails to be differentiable since

$$D_v f(0) = \frac{v_1^3}{v_1^2 + v_2^2}$$

is not a linear transform of v. But it is continuous! In fact, any homogenous function of degree ≥ 1 will have limit 0 at 0. It helps to make it non-homogenous by considering

$$\begin{aligned} f(x, y) &= xy^2 \odot \frac{1}{x^2 + y^4} \\ \implies D_v f(0) &= \lim_{t\to 0} \frac{v_1 v_2^2 t^3}{v_1^2 t^3 + v_2 t^5} \\ &= \lim_{t\to 0} \frac{v_1 v_2^2}{v_1^2 + v_2 t^2} \\ &= v_1 \odot \frac{v_2^2}{v_1^2} \end{aligned}$$

and

$$\begin{aligned} \lim_{t\to 0} f(t^2, t) &= \lim_{t\to 0} \frac{t^4}{t^4 + t^4} \\ &= 1/2 \end{aligned} \qquad\qquad \begin{aligned} \lim_{t\to 0} f(0, t) &= \lim_{t\to 0} 0 \\ &= 0 \end{aligned}$$

so f is not continuous at 0. ■

Recall theorem 2.33 and corollary 2.35. Can we make the stronger statement

Lecture 19. Wednesday September 25

Proposition 2.35′. *f is smooth iff all partial derivatives of f of all orders exist?*

No! A counterexample is

Example. the function

$$f(x, y) = \mathbf{1}_{\{xy\neq 0\}} \odot e^{-\frac{x^2}{y^2} - \frac{y^2}{x^2}}.$$

This is not even continuous!

2.5 Generalizations

Proposition 2.44. *Let X and Y be (possibly infinite-dimensional) normed spaces, Let T be a linear map from X to Y. Then the following are equivalent:*

(1) T is continuous.

(2) T is bounded (i.e., T maps bounded sets to bounded sets).

(3) there is a constant M such that $\|Tx\| \leq M\|x\|$ for all $x \in X$.

Proof. (1) $\implies$ (2) holds since

$$\lim_{n\to\infty} T(x_n) = T\Big(\lim_{n\to\infty} x_n\Big).$$

(2) $\implies$ (3) since the unit ball is bounded. If (3) holds, then

$$T(y) = T(x) + T(y-x) \leq T(x) + M\|y-x\| \to T(x)$$

as $y \to x$. ■

Lecture 20. Friday September 27

Example. $f\colon x \in \mathbb{R} \mapsto x^2 \in \mathbb{R}$. Set-theoretically,

$$f = \{(x, x^2) \mid x \in \mathbb{R}\}$$

(ignoring the codomain discussion). The graph of f is

$$G(f) = \{(x, f(x)) \mid x \in \mathbb{R}\}.$$

2.6 The implicit function theorem

Let $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ be a differentiable function. Define the relation

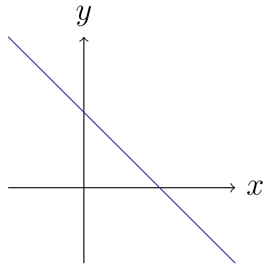
$$S \coloneqq \{(x,y) \mid f(x,y) = 0\}$$

from $\mathbb{R}^n$ to $\mathbb{R}^m$.

The impicit function theorem gives sufficient conditions on f such that S is a function (graph of a function) on $D(S)$ (domain of S). In other words, y can be solved in terms of x. In other other words, $y = \phi(x)$ for some function $\phi\colon D(S) \to \mathbb{R}^m$.

Examples.

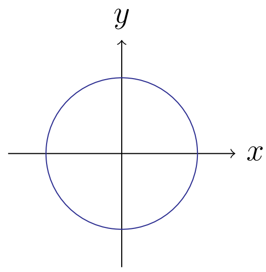
- $f\colon \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ defined by $f(x,y) = x+y-1$. Then $S = \{(x,y) \mid x+y = 0\}$ is the straight line



$D(S) = \mathbb{R}$, and S is a function ($y = 1 - x$).

- Let $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be given by
$$f(x, y) = y^5 + y^3 + y + x.$$
Fix an $x_0 \in \mathbb{R}$ and let $p = f(x_0)$. Then p is a strictly increasing odd degree polynomial, so there is exactly one $y_0 \in \mathbb{R}$ such that $f(x_0, y_0) = 0$.

- Let $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be given by
$$f(x, y) = x^2 + y^2 - 1.$$
The graph of f is the unit circle in $\mathbb{R}^2$.



Let $(a, b) \in S$. The implicit function theorem tells us when S can be *locally* a function around (a, b). That is, there is a neighbourhood V of (a, b) such that $S_1 \coloneqq V \cap S$ is a function on $D(S_1)$.

In this case, $D(S) = [-1, 1]$, and S is locally a function around each point except $(-1, 0)$ and $(1, 0)$. For $b > 0$, $y = \sqrt{1 - x^2}$ locally, and for $b < 0$, $y = -\sqrt{1 - x^2}$ locally.

Theorem 2.45 (implicit function theorem)**.** *Let $k \geq 1$, $U \subseteq_{\text{op}} \mathbb{R}^m \times \mathbb{R}^n$ and $f \in C^k(U, \mathbb{R}^m)$. Suppose $(a, b) \in U$ is such that $f(a, b) = 0$ and $\frac{\partial f}{\partial y}(a, b)$ is invertible. Then there exists a neighbourhood V of (a, b) and W of a, and a function $\phi \colon W \to \mathbb{R}^n$ such that*

(1) $(x, \phi(x)) \in V$ for all $x \in W$,

(2) $\phi(a) = b$,

(3) $f(x, \phi(x)) = 0$ for all $x \in W$.

(4) If $f(x, y) = 0$ for $x \in W$ and $(x, y) \in V$, then $y = \phi(x)$.

(5) $\phi'(a) = -\left(\frac{\partial f}{\partial y}(a, b)\right)^{-1} \frac{\partial f}{\partial x}(a, b)$.

Proof. Define $F \colon U \to \mathbb{R}^{n+m}$ by $F(x, y) = (x, f(x, y))$. Obviously F is C^k on U. Then $F'(x, y)(h, k) = (h, f'(x, y)(h, k))$ is zero iff $h = 0$ and $f'(x, y)(0, k) = \frac{\partial f}{\partial y}(x, y)(k) = 0$. Since $\frac{\partial f}{\partial y}(a, b)$ is invertible, $F'(a, b)(h, k) = 0$ iff $h = 0$ and $k = 0$. Thus $F'(a, b)$ is invertible.

By the inverse function theorem, there exist open sets $V_1 \ni (a, b)$ and $W_1 \ni (a, 0)$ such that $F \colon V_1 \to W_1$ is a C^k-diffeomorphism. Let $W \ni a$ be an open set such that $(x, 0) \in W_1$ for all $x \in W$. Then there exists a function $\phi \colon W \to \mathbb{R}^n$ given by $\phi(x) = \pi_2(F^{-1}(x, 0))$ such that

(1) $(x, \phi(x)) \in V_1$ for all $x \in W$ (since $\pi_1(F^{-1}(x, 0)) = x$),

(2) $\phi(a) = \pi_2(F^{-1}(a, 0)) = \pi_2(a, b) = b$,

(3) $F(x, \phi(x)) = F(x, \pi_2(F^{-1}(x, 0))) = (x, 0)$, so that $f(x, \phi(x)) = 0$ for all $x \in W$.

(4) yada yada ∎

Lecture 21. Wednesday October 2 / **Lecture 22.** Friday October 4

Example. Let $f(x, y) = x^3 e^y + 2x \cos(xy) - 3$. Can we write $y = \phi(x)$ in an open neighborhood V of 1 such that $f(x, \phi(x)) = 0$ for all $x \in V$? What is $\phi'(1)$? Then

$$\frac{\partial f}{\partial y}(x, y) = x^3 e^y - 2x^2 \sin(xy)$$
$$\frac{\partial f}{\partial x}(x, y) = 3x^2 e^y + 2\cos(xy) - 2xy \sin(xy)$$

and $\frac{\partial f}{\partial y}(1, 0) = 1 \neq 0$.

Let $U \subseteq \mathbb{R}^n$ be open and $f \colon U \to \mathbb{R}$ Then $f \in C^k(U)$ iff all partial derivatives of f upto order k exist on U and are continuous on U, which is iff $D_i f \in C^{k-1}(U)$ for all $i \in [n]$.

Notation.

- $D_i f(a) = \frac{\partial f}{\partial x_i}(a)$.
- $D_{ij} f(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a) = (D_i(D_j f))(a)$. This only makes sense if $D_j f$ exists in a neighbourhood of a, and is called a *mixed* partial derivative.
- In general, $D_{i_1,\ldots,i_k} f(a) = \frac{\partial^3 f}{\partial x_{i_1} \partial \ldots \partial x_{i_k}}(a) = D_{i_1}(D_{i_2}(\ldots(D_{i_k} f)\ldots))(a)$.

Question 2.46. *Suppose $D_{ij} f(a)$ and $D_{ji} f(a)$ exist. Can we say that they are equal?*

Solution. No, not necessarily. TODO ■

Theorem 2.47. *Let $U \subseteq_{\text{op}} \mathbb{R}^2$ and $f : U \to \mathbb{R}$. Suppose*

(1) $D_1 f$, $D_2 f$ and $D_{21} f$ exist on U.

(2) $D_{21} f$ is continuous at $(a, b) \in U$.

Then $D_{12} f(a, b)$ exists and

$$D_{12} f(a, b) = D_{21} f(a, b).$$

Proof. Define the function

$$\begin{aligned}\Delta(h, k) &:= f(a + h, b + k) - f(a + h, b) + f(a, b) - f(a, b + k)\\ &= [f(a + h, b + k) - f(a + h, b)] - [f(a, b + k) - f(a, b)].\end{aligned}$$

Define $f(t, b + k) - f(t, b)$ to be $u(t)$. Then

$$\begin{aligned}\Delta(h, k) &= u(a + h) - u(a)\\ &= hu'(a + \theta_1 h) \quad \text{for some } \theta_1 \in (0, 1)\\ &= h[D_1 f(a + \theta_1 h, b + k) - D_1 f(a + \theta_1 h, b)]\\ &= hk D_{21} f(a + \theta_1 h, b + \theta_2 k) \quad \text{for some } \theta_2 \in (0, 1).\end{aligned}$$

Since $D_{21} f$ is continuous at (a, b), we have that

$$\begin{aligned}D_2 f(a + h, b) - D_2 f(a, b) &= \lim_{k \to 0} \frac{\Delta(h, k)}{k}\\ &= h D_{21} f(a + \theta_1 h, b),\end{aligned}$$

so that

$$\begin{aligned}D_{12} f(a, b) &= \lim_{h \to 0} \frac{D_2 f(a + h, b) - D_2 f(a, b)}{h}\\ &= D_{21} f(a, b).\end{aligned}$$

■

Lecture 23. Monday October 7

Corollary 2.48. *Let $U \subseteq_{\text{op}} \mathbb{R}^n$ and $f : U \to \mathbb{R}$. Suppose $f \in C^k(U)$. Tgeb $D_{i_1 \ldots i_m} f = D_{i_{\sigma(1)} \cdots i_{\sigma(m)}} f$ for all $1 \le m \le k$, $i_j \in [n]$ and $\sigma \in S_m$.*

Proof. Exercise ∎

Definition 2.49. Let $A \in M_n(\mathbb{R})$ be symmetric. We define $B_A \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$B_A(x, y) = \langle Ax,\, y\rangle = x^T A y$$

and $q_A \colon \mathbb{R}^n \to \mathbb{R}$ by

$$q_A(x) = B_A(x, x) = \langle Ax,\, x\rangle = x^T A x.$$

Theorem 2.50 (Taylor's theorem)**.** *Let $U \subseteq_{\mathrm{op}} \mathbb{R}^n$ be convex and $f \in C^m(U; \mathbb{R})$. Then for any $a, a+h \in U$,*

$$f(a+h) = \sum_{k=0}^{m} \sum_{i_1,\dots,i_k} \frac{1}{k!}(D_{i_1\dots i_k} f(a)) h_{i_1} \dots h_{i_k} + r(h)$$

where

$$r(h) = \frac{1}{m!} \sum_{i_1,\dots,i_m} (D_{i_1\dots i_m} f(c) - D_{i_1\dots i_m} f(a)) h_{i_1} \dots h_{i_m}$$

for some c in the segment a–$(a+h)$ is $o(\|h\|^m)$.

Proof. Let $\gamma(t) = a + th$ and $g = f \circ \gamma$. If f is C^m, then so is g. By the 1-dimensional Taylor's theorem,

$$g(1) = \sum_{k=0}^{m-1} \frac{1}{k!} g^{(k)}(0) + \frac{1}{m!} g^{(m)}(\xi)$$

for some $\xi \in (0, 1)$. By the composition rule,

$$\begin{aligned} g'(t) &= f'(\gamma(t))\gamma'(t) = \sum_{i=1}^{n} D_i f(\gamma(t)) h_i \\ g''(t) &= \sum_{j=1}^{n} \sum_{i=1}^{n} D_{ji} f(\gamma(t)) h_i h_j \\ &\vdots \\ g^{(m)}(t) &= \sum_{i_1,\dots,i_m} D_{i_1\dots i_m} f(\gamma(t)) h_{i_1} \dots h_{i_m} \end{aligned}$$

∎

Assignment 4

up October 22
due October 30

Problem 4.4. Find first, second and third derivative of the following maps. (NOTE: Only write the final answer as linear map. Do not give any calculation.)

(1) Let $B \in M_n(\mathbb{R})$ be fixed. Let $f\colon M_n(\mathbb{R}) \to M_n(\mathbb{R})$ be given by

$$f(X) := XBX.$$

(2) Let $f\colon \mathrm{GL}_n(\mathbb{R}) \to M_n(\mathbb{R})$ be given by

$$f(X) := X^\top X^{-1}.$$

Solution.

(1)

$$\begin{aligned} f'(X)(H_1) &= H_1BX + XBH_1 \\ f''(X)(H_2)(H_1) &= H_1BH_2 + H_2BH_1 \\ f'''(X) &= 0 \end{aligned}$$

(2)

$$\begin{aligned} f'(X)(H_1) &= H_1^\top X^{-1} - X^\top X^{-1} H_1 X^{-1} \\ f''(X)(H_2)(H_1) &= -\sum_{\sigma \in S_2} \left(H_{\sigma(1)}^\top - X^\top X^{-1} H_{\sigma(1)}\right) X^{-1} H_{\sigma(2)} X^{-1} \\ f'''(X)(H_3)(H_2)(H_1) &= \sum_{\sigma \in S_3} \left(H_{\sigma(1)}^\top - X^\top X^{-1} H_{\sigma(1)}\right) X^{-1} H_{\sigma(2)} X^{-1} H_{\sigma(3)} X^{-1} \end{aligned}$$

Expanded out, this is (where $\overline{X} = X^{-1}$ for readability)

$$\begin{aligned} f'(X)(H) &= H^\top \overline{X} - X^\top \overline{X} H \overline{X} \\ f'(X)(K)(H) &= -H^\top \overline{X} K \overline{X} - K^\top \overline{X} H \overline{X} + X^\top \overline{X}(H \overline{X} K \overline{X} + K \overline{X} H \overline{X}) \\ f'''(X)(L)(K)(H) &= H^\top \overline{X}(K \overline{X} L + L \overline{X} K)\overline{X} \\ &\quad + K^\top \overline{X}(L \overline{X} H + H \overline{X} L)\overline{X} \\ &\quad + L^\top \overline{X}(H \overline{X} K + K \overline{X} H)\overline{X} \\ &\quad - X^\top \overline{X} H \overline{X}(K \overline{X} L + L \overline{X} K)\overline{X} \\ &\quad - X^\top \overline{X} K \overline{X}(L \overline{X} H + H \overline{X} L)\overline{X} \\ &\quad - X^\top \overline{X} L \overline{X}(H \overline{X} K + K \overline{X} H)\overline{X} \end{aligned}$$

■

Problem 4.8.

(1) Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a continuously differentiable function. Show that f is not one-one. (**Hint: Implicit function theorem**)

Solution.

(1) Assume $D_1 f \equiv 0$ everywhere. Then by the mean value theorem, $f(0,0) = f(1,0)$, so f is not one-one. Otherwise, assume $D_1 f(a,b) \neq 0$ for some $(a,b) \in \mathbb{R}^2$. Then by the implicit function theorem, there exists a neighbourhood $U \subseteq \mathbb{R}$ of a and a neighbourhood $V \subseteq \mathbb{R}$ of b such that for each $y \in V$, there exists an $x \in U$ such that $f(x,y) - f(a,b) = 0$. f is not one-one. ■

Problem 4.12. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} \dfrac{xy(x^2-y^2)}{x^2+y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

(1) Show that f is continuous on $\mathbb{R}^2$.

(2) Show that

$$D_1 f(x,y) = \begin{cases} \dfrac{y(x^4+4x^2y^2-y^4)}{(x^2+y^2)^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

and

$$D_2 f(x,y) = \begin{cases} \dfrac{x(x^4-4x^2y^2-y^4)}{(x^2+y^2)^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

(3) Show that $D_1 f$, $D_2 f$ are continuous on $\mathbb{R}^2$.

(4) Show that $D_1 f(0,y) = -y$ and $D_2 f(x,0) = x$ for any $x, y \in \mathbb{R}$.

(5) Show that $D_{12} f(0,0) = 1$ and $D_{21} f(0,0) = -1$.

(6) Show that $D_{12} f$ and $D_{21} f$ exist at every point of $\mathbb{R}^2$ and are continuous except at $(0,0)$.

(7) Does the function f belong to $C^1(\mathbb{R}^2)$ or $C^2(\mathbb{R}^2)$?

Solution.

(1) f is rational outside the origin, and hence continuous there. For the limit at the origin,

$$\begin{aligned} f(x,y) &= \frac{xy}{x^2+y^2}(x^2-y^2) \\ &\leq \frac{1}{2}(x^2-y^2) \\ &\to 0. \end{aligned}$$

Thus f is continuous on $\mathbb{R}^2$.

(2) Let $(x, y) \neq 0$. Then

$$\begin{aligned}
D_1 f(x, y) &= \frac{(3x^2y - y^3)(x^2 + y^2) - 2x^2y(x^2 - y^2)}{(x^2 + y^2)^2} \\
&= \frac{y(3x^4 + 3x^2y^2 - x^2y^2 - y^4 - 2x^4 + 2x^2y^2)}{(x^2 + y^2)^2} \\
&= \frac{y(x^4 + 4x^2y^2 - y^4)}{(x^2 + y^2)^2}. \\
D_2 f(x, y) &= \frac{(x^3 - 3xy^2)(x^2 + y^2) - 2xy^2(x^2 - y^2)}{(x^2 + y^2)^2} \\
&= \frac{x(x^4 + x^2y^2 - 3x^2y^2 - 3y^4 - 2x^2y^2 + 2y^4)}{(x^2 + y^2)^2} \\
&= \frac{x(x^4 - 4x^2y^2 - y^4)}{(x^2 + y^2)^2}.
\end{aligned}$$

Finally,

$$\begin{aligned}
D_1 f(0, 0) &= \lim_{x \to 0} \frac{f(x, 0) - f(0, 0)}{x} = \lim_{x \to 0} 0 = 0. \\
D_2 f(0, 0) &= \lim_{y \to 0} \frac{f(0, y) - f(0, 0)}{y} = \lim_{y \to 0} 0 = 0.
\end{aligned}$$

(3) Continuity outside the origin is again clear. For the limit at the origin, note that $D_1 f$ and $D_2 f$ are homogeneous of degree 1, and that $|D_1 f|$ and $|D_2 f|$ attain some maximum value M_1 and M_2 respectively on the unit sphere (compactness and continuity). Thus

$$\begin{aligned}
\lim_{u \to 0} D_1 f(u) &= \lim_{u \to 0} \|u\| D_1 f(u) \leq \lim_{u \to 0} M_1 \|u\| = 0. \\
\lim_{u \to 0} D_2 f(u) &= \lim_{u \to 0} \|u\| D_2 f(u) \leq \lim_{u \to 0} M_2 \|u\| = 0.
\end{aligned}$$

(4) Since $D_1 f(0, 0) = D_2 f(0, 0) = 0$, these are true at the origin. For $y \neq 0$,

$$D_1 f(0, y) = \lim_{x \to 0} \frac{f(x, y) - f(0, y)}{x} = \lim_{x \to 0} \frac{y(x^2 - y^2)}{x^2 + y^2} = -y.$$

For $x \neq 0$,

$$D_2 f(x, 0) = \lim_{y \to 0} \frac{f(x, y) - f(x, 0)}{y} = \lim_{y \to 0} \frac{x(x^2 - y^2)}{x^2 + y^2} = x.$$

(5)

$$\begin{aligned}
D_{12} f(0, 0) &= \lim_{x \to 0} \frac{D_2 f(x, 0)}{x} = 1. \\
D_{21} f(0, 0) &= \lim_{y \to 0} \frac{D_1 f(0, y)}{y} = -1.
\end{aligned}$$

(6) Outside the origin, f is a rational function and hence smooth. Thus $D_{12}f$ and $D_{21}f$ exist and are continuous everywhere except at $(0,0)$. If $D_{21}f$ was continuous at the origin, then by the equality of mixed partials, $D_{12}f(0,0) = D_{21}f(0,0)$ (since $D_1 f$ and $D_2 f$ are continuous). This is not the case, as shown in the previous part.

(7) We saw that $D_{21}f$ is not continuous at the origin. Thus $f \notin C^2(\mathbb{R}^2)$. By part (3), $f \in C^1(\mathbb{R}^2)$. ■

Lecture 24. Wednesday October 9

Definition 2.51 (multiindex)**.** Let $n \geq 1$ be fixed. $\alpha \in \mathbb{N}^n$ is called a *multiindex.* For $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$, we define

$$\begin{aligned} \alpha! &:= \alpha_1!\alpha_2!\dots\alpha_n! \\ |\alpha| &:= \alpha_1 + \alpha_2 + \dots + \alpha_n \\ h^\alpha &:= h_1^{\alpha_1} h_2^{\alpha_2} \dots h_n^{\alpha_n} \\ D^\alpha &:= D_1^{\alpha_1} D_2^{\alpha_2} \dots D_n^{\alpha_n} \end{aligned}$$

Example. Let $n = 5$ and $\alpha = (1, 2, 0, 1, 2)$. Then

$$\begin{aligned} \alpha! &= 1!2!0!1!2! = 2 \\ |\alpha| &= 1 + 2 + 0 + 1 + 2 = 6 \\ h^\alpha &= h_1 h_2^2 h_4 h_5^2 \\ D^\alpha f &= D_1 D_2^2 D_4 D_5^2 f \\ &= \frac{\partial^6 f}{\partial x_1\, \partial x_2^2\, \partial x_4\, \partial x_5^2} \end{aligned}$$

If f is smooth,

$$D_{i_1 \dots i_k} f = D_{i_{\sigma(1)}} \dots D_{i_{\sigma(k)}} f$$

for any $\sigma \in S_n$. Thus we can rewrite Taylor's theorem as

$$\boxed{f(a+h) = \sum_{k=0}^{m} \frac{1}{k!} \sum_{|\alpha|=k} \frac{k!}{\alpha!} D^\alpha f(a) h^\alpha + r(h) = \sum_{|\alpha| \leq m} \frac{1}{\alpha!} D^\alpha f(a) h^\alpha + r(h)}$$

since the number of sequences $(i_1, \dots, i_k)$ with α_1 occurrences of 1, α_2 occurrences of 2, etc. is $\binom{k}{\alpha_1, \alpha_2, \dots, \alpha_n} = \frac{k!}{\alpha!}$. We can write $r(h)$ as

$$r(h) = \sum_{|\alpha|=m} \frac{1}{\alpha!} (D^\alpha f(c) - D^\alpha f(a)) h^\alpha.$$

Definition 2.52 (definiteness)**.** Let $A \in M_n(\mathbb{R})$ be a symmetric matrix. We say that A or q_A is

(1) *positive* (resp. *negative*) *definite* if $q_A(x) > 0$ (resp. $q_A(x) < 0$) for all $x \in \mathbb{R}^n \setminus \{0\}$;

(2) *positive* (resp. *negative*) *semidefinite* if it is *not* definite and $q_A(x) \geq 0$ (resp. $q_A(x) \leq 0$) for all $x \in \mathbb{R}^n$;

(3) *indefinite* if it is neither definite not semidefinite.

Remark. A is positive (semi)definite iff $-A$ is negative (semi)definite.

Proposition 2.53. *If A is a positive definite matrix, then there exist $c_1, c_2 > 0$ such that*

$$c_1 \|x\|^2 \leq q_A(x) \leq c_2 \|x\|^2$$

for all $x \in \mathbb{R}^n$.

Proof. q_A is a continuous positive function on the compact set S^{n-1}. Let c_1 and c_2 be the minimum and maximum values of q_A on S^{n-1}. For any $x \in \mathbb{R}^n$, $q_A(x) = \|x\|^2 q_A(\widehat{x})$, so

$$c_1\|x\|^2 \leq q_A(x) \leq c_2\|x\|^2.$$

Alternatively, note that q_A is a norm on $\mathbb{R}^n$, and any two norms are equivalent. ■

Throughout this lecture, let $S \subseteq \mathbb{R}^n$ and $f\colon S \to \mathbb{R}$.

Lecture 25. Friday October 11

Definition 2.54 (critical point)**.** Let $a \in S^\circ$ be such that f is differentiable at a. We say that a is a *critical point* of f if $f'(a) = 0$.

Definition 2.55 (extrema)**.** We say that $a \in S$ is a

(1) *local maximum* (resp. *local minimum*) of f if there exists an $r > 0$ such that $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$) for all $x \in S \cap B(a,r)$.

(2) *global maximum* (resp. *global minimum*) of f if $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$) for all $x \in S$.

(3) *local extremum* if it is a local maximum or minimum.

(4) *saddle point* if it is neither a local maximum nor a local minimum. Confirm equality signs

Proposition 2.56. *Let $a \in S^\circ$ be such that f is differentiable at a, and a is a local extremum of f. Then a is a critical point of f.*

Proof. For $i \in [n]$ let $g_i\colon (-\varepsilon,\varepsilon) \to \mathbb{R}$ be given by $g_i(t) = f(a+te_i)$, where ε is chosen so that everything is defined. Then g_i is differentiable at 0 and $g_i'(0) = D_i f(a)$. Moreover, 0 is a local extremum of g_i. Thus by the 1-dimensional result, $g_i'(0) = 0$, so that $f'(a) = \begin{bmatrix} D_1 f(a) & \cdots & D_n f(a) \end{bmatrix} = 0$. ■

Examples.

- Find the maximum volume of an axes-parallel cuboid that can be inscribed in the ellipsoid

$$E = \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + \frac{y^2}{2} + \frac{z^2}{4} = 1\}.$$

  Geometrically, it only makes sense to keep the cuboid symmetric about the origin in each direction, and to keep its vertices on the ellipsoid. Thus let its vertices be $(\pm x, \pm y, \pm z)$, where $x,y,z > 0$. The volume is $V = 8xyz$, and $z = 2\sqrt{1 - x^2 - \frac{y^2}{2}}$. Thus $V(x,y) = 16xy\sqrt{1-x^2-\frac{y^2}{2}}$. Let $S = \{(x,y) \in \mathbb{R}^2 \mid x^2 + \frac{y^2}{2} \leq 1, x \geq 0, y \geq 0\}$.

S is compact, and $f \equiv 0$ on ∂S. The only possible local extrema are the critical points. In the interior,

$$\begin{aligned}
D_1V(x,y) &= 16y \cdot \frac{1}{2\sqrt{x^2 - x^4 - \frac{x^2y^2}{2}}} \cdot (2x - 4x^3 - xy^2) = 0 \\
&\implies x = \frac{1}{2}\sqrt{2-y^2}. \\
D_2V(x,y) &= 16x \cdot \frac{1}{2\sqrt{y^2 - x^2y^2 - \frac{y^4}{2}}} \cdot (2y - 2x^2y - 2y^3) = 0 \\
&\implies y = \sqrt{1-x^2}.
\end{aligned}$$

Together, these imply $x = \frac{1}{2}\sqrt{2-1+x^2}$, so $3x^2 = 1$ or $x = \frac{1}{\sqrt{3}}$. This gives $y = \frac{\sqrt{2}}{\sqrt{3}}$, and $z = \frac{2}{\sqrt{3}}$. The corresponding volume is $V = \frac{16}{3\sqrt{3}}$.

Lecture 26. Monday October 14

Examples.

- Find the maximum volume of an open rectangular garbage box which has surface area $3\,\mathrm{m}^3$.

  Let the dimensions of the box be x, y, z, with volume xyz and surface area $xy + 2yz + 2zx = 3$. Then $z = \frac{3-xy}{2(x+y)}$, so that the volume

  $$V(x,y) = \frac{xy(3-xy)}{2(x+y)}.$$

  The domain of interest is $S = \{(x,y) \in \mathbb{R}^2 \mid x > 0, y > 0, xy < 3\}$. This is open and unbounded. We'll fix that later.

  For critical points, notice that $V(x,y) = V(y,x)$, and

  $$\begin{aligned}
  D_1V(x,y) = 0 &\implies (3y - 2xy^2)(x+y) - xy(3-xy) = 0 \\
  &\implies 3y^2 - x^2y^2 - 2xy^3 = 0 \\
  &\implies 3 - x^2 - 2xy = 0.
  \end{aligned}$$

  Symmetrically, $D_2V(x,y) = 0 \implies 3 - y^2 - 2xy = 0$. Thus $x = y$, so $x = y = 1$. The volume at $(1,1)$ is $\frac{1}{2}$. We claim that this is the maximum volume.

  To see this, observe that

  $$V(x,y) = \frac{3xy}{2(x+y)} < \frac{3}{2}\min(x,y).$$

  Thus whenever $x < \frac{1}{3}$ or $y < \frac{1}{3}$, we have $V(x,y) < \frac{1}{2}$. Thus we can restrict our view to the compact set

  $$K = \{(x,y) \in \mathbb{R}^2 \mid x \geq \frac{1}{3}, y \geq \frac{1}{3}, xy \leq 3\}.$$

For $xy = 3$, $z = 0$ so $V(x, y) = 0$. For $x = \frac{1}{3}$ or $y = \frac{1}{3}$, $V(x, y) < \frac{1}{2}$. Thus the maximum volume is attained at $(x, y) = (1, 1)$ as computed above.

Definition 2.57. We will denote the set of symmetric $n \times n$ matrices by $S(n)$. Alternatively, $S(n)$ is the disjoint union of the sets of positive definite, negative definite, indefinite and semi-definite matrices.

Exercise 2.58. *Let $A \in S(n)$. Then*

(1) A is positive (resp. negative) definite iff all eigenvalues of A are strictly positive (resp. negative).

(2) A is indefinite iff A has both positive and negative eigenvalues.

(3) A is positive (resp. negative) semi-definite iff all eigenvalues of A are non-negative (resp. non-positive) and A is singular.

Exercise 2.59. *Let $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in S(2)$. Under what conditions is A positive definite, negative definite, indefinite and semi-definite?*

Lecture 27. Wednesday October 16

Examples.

- Let $f \colon \begin{array}{c} \mathbb{R}^2 \to \mathbb{R} \\ (x, y) \mapsto x^2 + y^2 \end{array}$. Then $(0, 0)$ is the only critical point. For any $(x, y) \neq (0, 0)$, $f(x, y) > f(0, 0)$. Thus it is a global minimum.
- Let $f \colon \begin{array}{c} \mathbb{R}^2 \to \mathbb{R} \\ (x, y) \mapsto -(x^2 + y^2) \end{array}$. Then $(0, 0)$ is the only critical point. For any $(x, y) \neq (0, 0)$, $f(x, y) < f(0, 0)$. Thus it is a global maximum.
- Let $f \colon \begin{array}{c} \mathbb{R}^2 \to \mathbb{R} \\ (x, y) \mapsto x^2 - y^2 \end{array}$. Then $(0, 0)$ is the only critical point. But this is neither a local minimum nor a local maximum. To see this, let $\varepsilon > 0$ be arbitrary. Then $f(\varepsilon, 0) > f(0, 0) > f(0, \varepsilon)$. Thus, $(0, 0)$ is a saddle point.

Let $f \colon S \subseteq \mathbb{R}^n \to \mathbb{R}$ and $a \in S^\circ$. Suppose a is a critical point of f. We want a criterion to determine whether a is a local minimum, local maximum, or neither. Recall that for single-variable functions, the second derivative provides such a criterion.

- If $f''(a) > 0$, then a is a local minimum.
- If $f''(a) < 0$, then a is a local maximum.

- If $f''(a) = 0$, the test is inconclusive.

Proposition 2.60. *Let $f\colon B(a,r) \to \mathbb{R}$ be C^2. Suppose $f'(a) = 0$. If $H_f(a)$ is* $\begin{cases} \textit{positive definite} \\ \textit{negative definite} \\ \textit{indefinite} \end{cases}$ *then a is a* $\begin{cases} \textit{local minimum} \\ \textit{local maximum} \\ \textit{saddle point} \end{cases}$*.*

If H_f is semi-definite, then the test is inconclusive.

Proof. Let $|h| < r$. Since $B(a,r)$ is open and convex, Taylor's theorem yields

$$f(a+h) = f(a) + f'(a)(h) + \frac{1}{2}q_{f,a}(h) + r(h),$$

where $r(h) = o(\|h\|^2)$ and $q_{f,a} = q_{H_f(a)}$ is the quadratic form associated to the Hessian of f evaluated at a.

If $H_f(a) \succ 0$ then $q_{f,a}(h) \geq \lambda\|h\|^2$ for some $\lambda > 0$. If $f'(a) = 0$, then $\frac{f(a+h)-f(a)}{\|h\|^2} \geq \lambda + \frac{o(\|h\|^2)}{\|h\|^2} \xrightarrow{h\to 0} \lambda > 0$. Thus for small enough h, $f(a+h) > f(a)$ and a is a local minimum. Symmetrically, if $H_f(a) \prec 0$, then a is a local maximum.

If $H_f(a)$ is indefinite, then there exist h_1, h_2 of unit norm such that $q_{f,a}(h_1) = \lambda_1 > 0$ and $q_{f,a}(h_2) = \lambda_2 < 0$. Then

$$\begin{aligned} \frac{f(a+th_1)-f(a)}{t^2} &= \lambda_1 + \frac{o(t^2)}{t^2} \xrightarrow{t\to 0} \lambda_1 > 0, \\ \frac{f(a+th_2)-f(a)}{t^2} &= \lambda_2 + \frac{o(t^2)}{t^2} \xrightarrow{t\to 0} \lambda_2 < 0. \end{aligned}$$

Thus there are points arbitrarily close to a where f is greater than $f(a)$, as well as points where f is less than $f(a)$. ∎

Examples.

- $f(x,y) = x^2 + y^4$. The only critical point is $(0,0)$. The Hessian is $\begin{pmatrix} 2 & 0 \\ 0 & 12y^2 \end{pmatrix}$, which when evaluated at $(0,0)$ is $\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$. This is positive semi-definite, so our test is inconclusive. It is easy to see that $(0,0)$ is the global minimum.

- For $f(x,y) = -(x^2+y^4)$, the Hessian at $(0,0)$ is negative semi-definite. Again the test is inconclusive, but $(0,0)$ is the global maximum.

- For $f(x,y) = x^2 - y^4$, the critical point is $(0,0)$ and the Hessian there is the same as for x^2+y^4. The test is inconclusive, but this time $(0,0)$ is a saddle point.

- For $f(x,y) = x^4 - y^2$, the Hessian is negative semi-definite, and again $(0,0)$ is a saddle point.

Lecture 28

Lecture 28.
Friday
October 18

Definition 2.61 (level set)**.** Let $U \subseteq \mathbb{R}^n$ and $f \colon U \to \mathbb{R}$. The set

$$L_c(f) = \{x \in U \mid f(x) = c\}$$

is the *level set* of f with level c.

Remark. For $n = 2$ and 3, we call these level curves and level surfaces. Note that they need not be curves or surfaces, as in the case of constant functions.

Examples.

- $f(x, y) = x^2 + y^2$ has level sets $L_c(f) = \varnothing$ if $c < 0$, $L_0(f) = \{(0, 0)\}$, and $L_c(f) =$ the circle of radius $\sqrt{c}$ centered at the origin when $c > 0$.
- $f(x, y) = xy$ has level sets $L_c(f) = \{(x, y) \in \mathbb{R}^2 \mid xy = c\}$ which are axis-aligned rectangular hyperbolas.

Suppose $U \subseteq \mathbb{R}^n$ and $f \colon U \to \mathbb{R}$ is C^1. Let $a \in L_c$ and $f'(a) \neq 0$ (that is, $D_i f(a) \neq 0$ for some i). Let $g = f - f(a)$, so that $g(a) = 0$. $D_i g(a) = D_i f(a) \neq 0$.

By the implicit function theorem, L_c can be parameterized by $n - 1$ coordinates in an open neighborhood of a. If $\nabla f \neq 0$ at any point of L_c, then L_c can be parameterized locally by $n - 1$ coordinates at each point. In other words, L_c is an $(n - 1)$-dimensional manifold.

Definition 2.62 (regularity)**.** Given a C^1-function $f \colon U \to \mathbb{R}$, a value $c \in \mathbb{R}$ is *regular* if $L_c(f) \neq \varnothing$ and $\nabla f(x) \neq 0$ for every $x \in L_c(f)$.

Lecture 29.
Monday
October 21

Exercise 2.63. *Compute* $\det'$*.*

Solution. We know that $\frac{\mathrm{d}}{\mathrm{d}t} \det(I + At) = \operatorname{Tr}(A)$. Why?

Thus $\det(I + At) = 1 + \operatorname{Tr}(A)t + o(t)$. So

$$\begin{aligned}
\det(X + H) &= \det(X) \det(I + X^{-1}H) \\
&= \det(X)(1 + \operatorname{Tr}(X^{-1}H) + o(\|H\|)) \\
&= \det(X) + \operatorname{Tr}(X^{-1}H) \det(X) + o(\|H\|) \\
\implies \det{}'(X)(H) &= \operatorname{Tr}(X^{-1}H) \det(X)
\end{aligned}$$

What about non-invertible matrices? ■

Theorem 2.64 (Lagrange multipliers)**.** *Let $U \subseteq \mathbb{R}^n$ be open and $f \colon U \to \mathbb{R}$ be differentiable. Let $g \colon U \to \mathbb{R}^m$ be C^1 and S be the level zero set of g. Let $a \in S$ be a local extremum of f in S and $g'(a)$ have rank m. Then there exist scalars $\lambda_1, \lambda_2, \ldots, \lambda_m$ such that*

$$\nabla f(a) = \sum_{i=1}^{m} \lambda_i \nabla g_i(a).$$

Remarks.

- $[f'(a)] = \nabla f(a)$ and $[g_i'(a)] = \nabla g_i(a)$. Thus the theorem can be written as
$$\begin{aligned}[f'(a)] &= \sum_{i=1}^{m} \lambda_i [g_i'(a)] \\ \implies f'(a) &= \sum_{i=1}^{m} \lambda_i g_i'(a).\end{aligned}$$
- $[g'(a)] = \begin{pmatrix} \nabla g_1(a) \\ \nabla g_2(a) \\ \vdots \\ \nabla g_m(a) \end{pmatrix}$. Then $g'(a)$ has rank m iff $\nabla g_1(a)$, $\nabla g_2(a)$, $\dots$, $\nabla g_m(a)$ are linearly independent.
- $\lambda_1, \lambda_2, \dots, \lambda_m$ are called the Lagrange multipliers.

Examples.

- Find the maximum value of $x_1^2 x_2^2 \dots x_n^2$ subject to the condition $x_1^2 + x_2^2 + \dots + x_n^2 = \alpha^2$.

  Here, $g\colon \mathbb{R}^n \to \mathbb{R}$ and $g(x) = x_1^2 + x_2^2 + \dots + x_n^2 - \alpha^2$. f is obviously $\mathbb{R}^n \to \mathbb{R}$ with $f(x) = x_1^2 x_2^2 \dots x_n^2$.
$$\begin{aligned} f'(x) &= \begin{pmatrix} 2x_1 x_2^2 \dots x_n^2 & 2x_1^2 x_2 \dots x_n^2 & \dots & 2x_1^2 x_2^2 \dots x_{n-1} \end{pmatrix} \\ g'(x) &= \begin{pmatrix} 2x_1 & 2x_2 & \dots & 2x_n \end{pmatrix} \end{aligned}$$
  By the Lagrange multipliers theorem, there exists a λ such that
$$\begin{aligned} x_1 x_2^2 \dots x_n^2 &= \lambda x_1 \\ x_1^2 x_2 \dots x_n^2 &= \lambda x_2 \\ &\vdots \\ x_1^2 x_2^2 \dots x_{n-1} &= \lambda x_n \end{aligned}$$
  In other words,
$$x_1^2 x_2^2 \dots x_{n-1}^2 = \lambda x_1^2 = \lambda x_2^2 = \dots = \lambda x_n^2.$$
  If any x_i is zero, then λ is zero (since some x_i must be non-zero everywhere on the sphere). This is one possible solution. Indeed, any point where any x_i is zero is a (global) minimum.

  Otherwise, we must have
$$|x|_1 = |x|_2 = \dots = |x|_n = \frac{\alpha}{\sqrt{n}},$$

for which $f(x) = \frac{\alpha^{2n}}{n^n}$. This is obviously the global maximum by the AM-GM inequality.

Let $W \subseteq \mathbb{R}^n$. Then $W^\perp = \{x \in \mathbb{R}^n \mid \forall y \in W, \langle x, y \rangle = 0\}$.

Exercise 2.65.

(1) $W^\perp$ is a subspace of $\mathbb{R}^n$.

(2) If W is a subspace of $\mathbb{R}^n$, then $(W^\perp)^\perp = W$ and $\mathbb{R}^n = W \oplus W^\perp$ (orthogonal direct sum).

(3) If $V \subseteq W$ then $W^\perp \subseteq V^\perp$.

Lecture 30. Wednesday October 23

2.7 The method of Lagrange multipliers

We will first consider the case of two variables.

Proposition 2.66. *Let $f, g \in C^1(\mathbb{R}^2, \mathbb{R})$. Let (x_0, y_0) be an extremum of f subject to the constraint $g(x, y) = 0$. Assume also that $\nabla g \neq 0$. Then there exists $\lambda \in \mathbb{R}$ such that*

$$\nabla(f + \lambda g)(x_0, y_0) = 0.$$

Remark. This is NOT a sufficient condition.

The method is then to find x, y that satisfy

$$\begin{cases} f_x + \lambda g_x = 0, \\ f_y + \lambda g_y = 0, \\ g = 0. \end{cases}$$

These are 3 equations in 3 unknowns.

Example. Find the extrema of $f(x, y) = x^2 + y^2 - x - y - xy$ in the domain $D = \{(x, y) \mid x + y \leq 3, x \geq 0, y \geq 0\}$.

Step 1 Find the critical points in the interior of D.

$$f_x(x, y) = 2x - y - 1 \stackrel{!}{=} 0,$$
$$f_y(x, y) = 2y - x - 1 \stackrel{!}{=} 0.$$

This has unique solution $x = y = 1$.

Step 2 Find the critical points on the boundary of D.

$(x = 0)$ $f(0, y) = y^2 - y$, with critical point $y = \frac{1}{2}$.

$(y = 0)$ $f(x, 0) = x^2 - x$, with critical point $x = \frac{1}{2}$.

$(x+y=3)$ Here we will use Lagrange multipliers. (I think this is overkill.)

$$\begin{aligned}(2x-y-1)+\lambda(1)&=0,\\(2y-x-1)+\lambda(1)&=0,\\x+y-3&=0.\end{aligned}$$

Again by symmetry, $x=y$, so $x=y=\frac{3}{2}$ and $\lambda=-\frac{1}{2}$ is the unique solution.

endpoints $(0,0)$, $(0,3)$, $(3,0)$.

Step 3 Compute all critical values.

$$\begin{aligned}f(1,1)&=-1\\f(0,\frac{1}{2})&=-\frac{1}{4}\\f(\frac{1}{2},0)&=-\frac{1}{4}\\f(\frac{3}{2},\frac{3}{2})&=-\frac{3}{4}\\f(0,0)&=0\\f(0,3)&=6\\f(3,0)&=6\end{aligned}$$

Thus f attains a global minimum at $(1,1)$, and global maxima at $(0,3)$ and $(3,0)$.

Example. Find the highest and lowest point of the ellipse that lies in the intersection of the cylinder $x^2+y^2=1$ and the plane $x+y+2z=2$.

We wish to extremize z over the given constraints. Let

$$\begin{aligned}f(x,y,z)&=z,\\g(x,y,z)&=\begin{pmatrix}x^2+y^2-1\\x+y+2z-2\end{pmatrix}.\end{aligned}$$

Computing the derivatives,

$$\begin{aligned}f'(x,y,z)&=\begin{pmatrix}0&0&1\end{pmatrix},\\g'(x,y,z)&=\begin{pmatrix}2x&2y&0\\1&1&2\end{pmatrix}.\end{aligned}$$

Clearly g' has full rank everywhere (except the origin, which we are not concerned with). We wish to solve

$$\begin{aligned} f'(x,y,z) &= \begin{pmatrix}\lambda_1 & \lambda_2\end{pmatrix} g'(x,y,z) \\ \begin{pmatrix}0 & 0 & 1\end{pmatrix} &= \begin{pmatrix}2x\lambda_1+\lambda_2 & 2y\lambda_1+\lambda_2 & 2\lambda_2\end{pmatrix} \\ \implies \lambda_2 &= \frac{1}{2} \\ \implies x = y &= -\frac{1}{4\lambda_1}. \end{aligned}$$

Since (x,y,z) must satisfy the constraints, we have

$$\frac{1}{16\lambda_1^2}+\frac{1}{16\lambda_1^2}=1, \text{ so } \lambda_1=\pm\frac{1}{2\sqrt{2}}.$$

The corresponding x and y are $\mp\frac{1}{\sqrt{2}}$. The highest point is $(-\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}},1+\frac{1}{\sqrt{2}})$ and the lowest is $(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},1-\frac{1}{\sqrt{2}})$.

Example. Find the operator norm of $A=\begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix}$.

We wish to maximize $\|Ax\|^2$ subject to $\|x\|^2=1$. Let

$$\begin{aligned} f(x,y) &= \left\|\begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}\right\|^2 = \left\|\begin{pmatrix}x+2y\\ y\end{pmatrix}\right\|^2 = x^2+4xy+5y^2, \\ g(x,y) &= x^2+y^2-1. \end{aligned}$$

Following the method,

$$\begin{aligned} f'(x,y) &= \begin{pmatrix}2x+4y & 4x+10y\end{pmatrix}, \\ g'(x,y) &= \begin{pmatrix}2x & 2y\end{pmatrix} \neq 0 \\ 2x+4y &\overset{!}{=} 2\lambda x, \\ 4x+10y &\overset{!}{=} 2\lambda y, \end{aligned}$$

Rearranging gives

$$\begin{aligned} (1-\lambda)x+2y &= 0, \\ 2x+(5-\lambda)y &= 0. \end{aligned}$$

For these to have non-trivial solutions, we must have

$$(1-\lambda)(5-\lambda)-4=0,$$

which gives $\lambda=3\pm2\sqrt{2}$. Then the equations simplify to $x=-(1\pm\sqrt{2})y$. Let $u=-(1\pm\sqrt{2})$. The corresponding x and y on the unit circle are given by

$$\begin{aligned}
&y^2 + (1 \pm \sqrt{2})^2 y^2 = 1, \\
\iff &(4 \pm 2\sqrt{2})y^2 = 1, \\
&\iff y = \pm \frac{1}{\sqrt{4 \pm 2\sqrt{2}}} \\
&\iff x = \mp \frac{1 \pm \sqrt{2}}{\sqrt{4 \pm 2\sqrt{2}}}.
\end{aligned}$$

Then

$$\begin{aligned}
f(x, y) &= x^2 + 4xy + 5y^2 \\
&= (u^2 + 4u + 5)y^2 \\
&= \frac{3 \pm 2\sqrt{2} - 4 \mp 4\sqrt{2} + 5}{4 \pm 2\sqrt{2}} \\
&= \frac{4 \mp 2\sqrt{2}}{4 \pm 2\sqrt{2}} \\
&= \frac{2 \mp \sqrt{2}}{2 \pm \sqrt{2}} \\
&= \frac{1}{2}(6 \mp 4\sqrt{2}) \\
&= 3 \mp 2\sqrt{2}.
\end{aligned}$$

Thus the maximum is $3 + 2\sqrt{2}$, and so the operator norm is

$$\sqrt{3 + 2\sqrt{2}} = \sqrt{1 + 2\sqrt{2} + \sqrt{2}^2} = 1 + \sqrt{2}.$$

Lecture 31. Friday October 25

Example. Let $a > b > c > 0$. Find the minimum and maximum $f(x, y, z) = x^2 + y^2 + z^2$ sucject to $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$. Obviously the minimum is c and maximum is a.

Let g be the constraint.

$$\begin{aligned}
f'(x, y, z) &= \begin{pmatrix} 2x & 2y & 2z \end{pmatrix}, \\
g'(x, y, z) &= \begin{pmatrix} \frac{2x}{a^2} & \frac{2y}{b^2} & \frac{2z}{c^2} \end{pmatrix}. \\
\left(1 - \frac{\lambda}{a^2}\right)x &= 0, \\
\left(1 - \frac{\lambda}{b^2}\right)y &= 0, \\
\left(1 - \frac{\lambda}{c^2}\right)z &= 0.
\end{aligned}$$

At least one of x, y, z is non-zero on the ellipsoid. Thus $\lambda = a^2, b^2$ or c^2, with corresponding solutions $(\pm a, 0, 0)$, $(0, \pm b, 0)$ and $(0, 0, \pm c)$. Inspection gives the result.

Proof of proposition 2.66 (2 variables). Let $f, g\colon \mathbb{R}^2 \to \mathbb{R}$ be C^1 and (x_0, y_0) be a local extremum of f subject to $g(x, y) = 0$. Further assume that $\nabla g(x_0, y_0) \neq 0$. WLOG, we can assume $\frac{\partial g}{\partial y}(x_0, y_0) \neq 0$.

By the implicit function theorem, there exists a C^1 function $y\colon \mathbb{R} \to \mathbb{R}$ such that $g(x, y(x)) = 0$ near (x_0, y_0). Furthermore,

$$y'(x_0) = -\left(\frac{\partial g}{\partial y}(x_0, y_0)\right)^{-1}\left(\frac{\partial g}{\partial x}(x_0, y_0)\right).$$

At (x_0, y_0), f has a local extremum subject to $g = 0$. Thus $\varphi\colon x \in \mathbb{R} \mapsto f(x, y(x)) \in \mathbb{R}$ has a local extremum at x_0. Differentiating using the chain rule gives

$$f_x(x_0, y_0) + f_y(x_0, y_0)y'(x_0) = 0.$$

Substituting $y'(x_0)$ gives that ∇f and ∇g are parallel at (x_0, y_0). ■

Proof of theorem 2.64 (the general case). Assume all hypotheses of the theorem are satisfied. Since $g'(a)\colon \mathbb{R}^n \to \mathbb{R}^m$ has rank m, we must have $m \leq n$. If $n = m$, then $g'(a)$ is invertible. Thus by the inverse function theorem, g is a local diffeomorphism around a, and since $\nabla g_1(a)$, $\nabla g_2(a)$, …, $\nabla g_m(a)$ are linearly independent, they span $\mathbb{R}^n$. Thus $\nabla f(a) \in \mathbb{R}^n$ is trivially a linear combination of the $\nabla g_i(a)$.

Now assume $m < n$. Permute the g_i such that the last m rows of $g'(a)$ are linearly independent, and view g as a C^1 map from $\mathbb{R}^{n-m} \times \mathbb{R}^m$ to $\mathbb{R}^m$. Let $a = (a_x, a_y)$. By the implicit function theorem, there is a C^1 map $y\colon U' \to \mathbb{R}^m$ such that $g(x, y(x)) = 0$ for $x \in U'$, where U' is a neighbourhood of a_x. Define $\psi(x) = (x, y(x))$. (Keep in mind that $\psi(a_x) = a$.) Then

$$\mathbb{R}^{n-m} \supseteq U' \xrightarrow{\psi} \mathbb{R}^n \supseteq U \xrightarrow{f} \mathbb{R}$$

Since a is an extremum of f subject to $g = 0$, a_x is an extremum of $f \circ \psi$ on U'. Thus

$$\begin{aligned}
(f \circ \psi)'(a_x) &= 0 \\
\implies f'(\psi(a_x))\psi'(a_x) &= 0 \\
\implies f'(a)\psi'(a_x) &= 0 \\
\implies [f'(a)]_{1\times n}[\psi'(a_x)]_{n\times(n-m)} &= 0.
\end{aligned}$$

That is, $[f'(a)]$ is orthogonal to each column of $[\psi'(a_x)]$. Since $(g \circ \psi)(x) = 0$ for all $x \in U'$, we have $g'(a)\psi'(a_x) = 0$. Thus each row of $g'(a)$ is orthogonal to each column of $\psi'(a_x)$.

But $\psi'(a_x)(h) = (h, y'(a_x)(h))$, so the first $n - m$ rows of $[\psi'(a_x)]$ are the identity matrix. The columns of $[\psi'(a_x)]$ are linearly independent! $g'(a)$ has rank m and $\psi'(a_x)$ has rank $n - m$. Thus the row space of $g'(a)$ is the orthogonal complement of the column space of $\psi'(a_x)$. Then $\nabla f(a)$ is a linear combination of the $\nabla g_i(a)$. ■