ANALYSIS 2: HOMEWORK SOLUTIONS, SPRING 2013

STEVEN HEILMAN

Contents

1. Problem Set 1

Exercise 1.1.

(i) Prove that for any sets A, B and C we have
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \qquad A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

(ii) Let S be an index set and let A_s be a family of subsets of some set X. Let $A^c := X \smallsetminus A$. Prove De Morgan's laws
$$\left(\bigcup_{s \in S} A_s\right)^c = \bigcap_{s \in S} A_s^c, \qquad \left(\bigcap_{s \in S} A_s\right)^c = \bigcup_{s \in S} A_s^c.$$

Proof of (i). Let P, Q and R be statements. We use $\wedge$ to denote "and," and we use $\vee$ to denote "or." We begin with the following truth table

P	Q	R	$P \vee (Q \wedge R)$	$(P \vee Q) \wedge (P \vee R)$	$P \wedge (Q \vee R)$	$(P \wedge Q) \vee (P \wedge R)$
T	T	T	T	T	T	T
T	F	T	T	T	T	T
F	T	T	T	T	F	F
F	F	T	F	F	F	F
T	T	F	T	T	T	T
T	F	F	T	T	F	F
F	T	F	F	F	F	F
F	F	F	F	F	F	F

Table 1. A Truth Table

Date: May 9, 2013.

Now, let P be the statement "$x \in A$," let Q be the statement "$x \in B$," and let R be the statement "$x \in C$." By the construction of Table 1, the rows of the table exhaust the $2^3 = 8$ possibilities for the location of x. So, since the fourth and fifth columns of Table 1 are identical, we conclude that $P \vee (Q \wedge R) = (P \vee Q) \wedge (P \vee R)$. By definitions of P, Q and R, we get $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. Also, since the sixth and seventh columns of Table 1 are identical, we conclude that $P \wedge (Q \vee R) = (P \wedge Q) \vee (P \wedge R)$. By definitions of P, Q and R, we get $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. □

Proof of (ii). Let $x \in \cup_{s \in S} A_s$. Then $x \in \cup_{s \in S} A_s$ if and only if there exists $s' \in S$ such that $x \in A_{s'}$. Let P be the statement $x \in \cup_{s \in S} A_s$. The statement $x \in (\cup_{s \in S} A_s)^c$ is therefore the negation of P. The negation of the statement "there exists $s' \in S$ such that $x \in A_{s'}$" is "for all $s \in S$, $x \notin A_s$." So, $x \in (\cup_{s \in S} A_s)^c$ if and only if, for all $s \in S$, $x \notin A_s$. That is, $x \in (\cup_{s \in S} A_s)^c$ if and only if, for all $s \in S$, $x \in A_s^c$. That is, $x \in (\cup_{s \in S} A_s)^c$ if and only if, $x \in \cap_{s \in S} A_s^c$. We have therefore proved the first required law. We now proceed with the second law.

Let $x \in \cap_{s \in S} A_s$. Then $x \in \cap_{s \in S} A_s$ if and only if for all $s \in S$, $x \in A_s$. Let P be the statement $x \in \cap_{s \in S} A_s$. The statement $x \in (\cap_{s \in S} A_s)^c$ is therefore the negation of P. The negation of the statement "for all $s \in S$, $x \in A_s$" is "there exists $s' \in S$ such that $x \notin A_{s'}$." So, $x \in (\cap_{s \in S} A_s)^c$ if and only if there exists $s' \in S$ with $x \notin A_{s'}$. That is, $x \in (\cap_{s \in S} A_s)^c$ if and only if there exists $s' \in S$ with $x \in A_{s'}^c$. That is, $x \in (\cap_{s \in S} A_s)^c$ if and only if $x \in \cup_{s \in S} A_s^c$. We have therefore proved the second required law. □

Exercise 1.2. Show that the scalar product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$ satisfies the parallelogram law or polarization identity

$$|x + y|^2 + |x - y|^2 = 2\,|x|^2 + 2\,|y|^2 .$$

What does this identity mean for the parallelogram spanned by x and y, i.e. the parallelogram with vertices $0, x, y$ and $x + y$?

Proof. Let $x, y \in \mathbb{R}^n$. Then

$$\begin{aligned} |x + y|^2 + |x - y|^2 &= \langle x + y, x + y \rangle + \langle x - y, x - y \rangle \\ &= \langle x, x \rangle + 2\langle x, y \rangle + \langle y, y \rangle + \langle x, y \rangle + \langle y, y \rangle - 2\langle x, y \rangle \\ &= 2\langle x, x \rangle + 2\langle y, y \rangle \\ &= 2\,|x|^2 + 2\,|y|^2 . \end{aligned}$$

Geometrically, this identity says that the sum of the squared lengths of the diagonals of a parallelogram is equal to the sum of the squared lengths of the edges of the parallelogram. □

Exercise 1.3.

(i) Prove that if A and B are closed, then $A \cup B$ is also closed. Find an example that shows that infinite unions of closed sets are not necessarily closed.
(ii) Prove that if for each $s \in S$ the set A_s is closed, then $\cap_{s \in S} A_s$ is also closed.

Proof of (i). Let A, B be closed sets. By the definition of closedness, A^c and B^c are open. From De Morgan's law, $(A \cup B)^c = A^c \cap B^c$. We have therefore exhibited $(A \cup B)^c$ as a finite intersection of open sets. Therefore, $(A \cup B)^c$ is open, and its complement $A \cup B$ is closed, as desired.

We now claim that $(0,1) = \cup_{n\geq 3}[1/n, 1-(1/n)]$. Let $x \in (0,1)$. Then there exists $n \in \mathbb{Z}$, $n \geq 3$ such that $1/n \leq x \leq 1-(1/n)$. So, $(0,1) \subseteq \cup_{n\geq 3}[1/n, 1-(1/n)]$. Now, let $x \in \cup_{n\geq 3}[1/n, 1-(1/n)]$. Then there exists $n \in \mathbb{N}$, $n \geq 3$ such that $x \in [1/n, 1-(1/n)] \subseteq (0,1)$. Therefore, $\cup_{n\geq 3}[1/n, 1-(1/n)] \subseteq (0,1)$. Therefore $(0,1) = \cup_{n\geq 3}[1/n, 1-(1/n)]$. □

Proof of (ii). Let A_s be closed, $s \in S$. Then A_s^c is open for all $s \in S$. From De Morgan's law, $(\cap_{s\in S} A_s)^c = \cup_{s\in S} A_s^c$. Since an arbitrary union of open sets is open, we conclude that $(\cap_{s\in S} A_s)^c$ is open. Therefore, its complement $\cap_{s\in S} A_s$ is closed, as desired. □

Exercise 1.4.

(i) Show that

$$f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B), \qquad f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B).$$

(ii) Show that

$$f(A \cup B) = f(A) \cup f(B), \qquad f(A \cap B) \subseteq f(A) \cap f(B).$$

Find an example to show that, in general, $f(A \cap B)$ is not equal to $f(A) \cap f(B)$. Show that if f is injective, then we have

$$f(A \cap B) = f(A) \cap f(B).$$

Proof of (i). We first show that $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-f}(B)$.

$$\begin{aligned} x \in f^{-1}(A \cap B) &\iff f(x) \in A \cap B \\ &\iff f(x) \in A \,\text{and}\, f(x) \in B \\ &\iff x \in f^{-1}(A) \,\text{and}\, x \in f^{-1}(B) \\ &\iff x \in (f^{-1}(A)) \cap (f^{-1}(B)). \end{aligned}$$

We now show that $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.

$$\begin{aligned} x \in f^{-1}(A \cup B) &\iff f(x) \in A \cup B \\ &\iff f(x) \in A \,\text{or}\, f(x) \in B \\ &\iff x \in f^{-1}(A) \,\text{or}\, x \in f^{-1}(B) \\ &\iff x \in (f^{-1}(A)) \cup (f^{-1}(B)). \end{aligned}$$

□

Proof of (ii). We first show that $f(A \cup B) = f(A) \cup f(B)$.

$$\begin{aligned} x \in f(A \cup B) &\iff \exists\, y \in A \cup B \,\text{such that}\, f(y) = x \\ &\iff \exists\, y \in A \,\text{or}\, \exists\, y \in B \,\text{such that}\, f(y) = x \\ &\iff x \in f(A) \,\text{or}\, x \in f(B) \\ &\iff x \in f(A) \cup f(B). \end{aligned}$$

We now prove that $f(A \cap B) \subseteq f(A) \cap f(B)$. Let $x \in f(A \cap B)$. Then there exists $y \in A \cap B$ such that $f(y) = x$. So, $y \in A$ and $y \in B$, and $f(y) = x$. Then $x \in f(A)$ and $x \in f(B)$, so $x \in f(A) \cap f(B)$. Combining our implications, we conclude that $f(A \cap B) \subseteq f(A) \cap f(B)$.

We now construct a function f and two sets A, B such that $f(A\cap B)\neq f(A)\cap f(B)$. Let $f\colon \mathbb{R}\to\mathbb{R}$ with $f(x)=x^2$, let $A=[-1,0]$ and let $B=[0,1]$. Then

$$f(A\cap B)=f(0)=0\neq[0,1]=[0,1]\cap[0,1]=f(A)\cap f(B).$$

Now, assume that f is injective. We show the reverse implication $f(A\cap B)\supseteq f(A)\cap f(B)$. Let $x\in f(A)\cap f(B)$. Then $x\in f(A)$ and $x\in f(B)$. So, there exist $y\in A$ and $y'\in B$ such that $f(y)=x$ and $f(y')=x$. Since f is injective and $f(y)=f(y')$, we conclude that $y=y'$. Since $y\in A$ and $y\in B$, we conclude that $y\in A\cap B$. Since $f(y)=x$, we conclude that $x\in f(A\cap B)$. Combining our implications, we conclude that $f(A\cap B)\supseteq f(A)\cap f(B)$. □

Exercise 1.5.

(i) Let $A_1, A_2, A_3, \ldots$ be subsets of some set X, and define

$$U := \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k, \qquad V := \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k.$$

Which one of $U\cup V$ and $V\cup U$ is true? Prove your claim.

(ii) Let $(f_k)_{k\in\mathbb{N}}$ be a sequence of functions on some set X, and let f be a function on X. Show that the set of convergence C defined by

$$C := \left\{x\in X\colon \lim_{k\to\infty} f_k(x)=f(x)\right\}$$

may be written as

$$C=\bigcap_{\ell=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left\{x\in X\colon |f_k(x)-f(x)|\leq\frac{1}{\ell}\right\}.$$

Proof of (i). We show that $V\subseteq U$, but in general, $U\neq V$. Observe

$$\begin{aligned} x\in \cap_{n=1}^{\infty}\cup_{k=n}^{\infty} A_k &\Longleftrightarrow \forall\, n\geq 1,\ x\in\cup_{k=n}^{\infty}A_k \\ &\Longleftrightarrow \forall\, n\geq 1,\ \exists\, k\geq n \,\text{such that}\, x\in A_k \\ &\Longleftrightarrow \text{there exist infinitely many } k\geq 1 \text{ such that } x\in A_k. \qquad (*) \end{aligned}$$

$$\begin{aligned} x\in \cup_{n=1}^{\infty}\cap_{k=n}^{\infty} A_k &\Longleftrightarrow \exists\, n\geq 1, \text{ such that } x\in\cap_{k=n}^{\infty}A_k \\ &\Longleftrightarrow \exists\, n\geq 1, \text{ such that } \forall\, k\geq n, x\in A_k. \qquad (**) \end{aligned}$$

Now, let $x\in V$. From $(**)$, $\exists\ n\geq 1$ such that $\forall\ k\geq n$, $x\in A_k$. In particular, there exist infinitely many $k\geq n\geq 1$ such that $x\in A_k$. Therefore, from $(*)$, $x\in U$. We conclude that $V\subseteq U$.

However, in general, $V\neq U$. For example, for k even, let $A_k:=[0,1]$, and for k odd, let $A_k:=[-1,0)$. From $(**)$, $V=\emptyset$, and from $(*)$, $U=[-1,1]$. So, in this case, $U\neq V$. □

Proof of (ii). Let $U_\ell := \cup_{n=1}^{\infty}\cap_{k=n}^{\infty}\left\{x\in X\colon |f_k(x)-f(x)|\leq\frac{1}{\ell}\right\}$. From $(**)$, $x\in U_\ell$ if and only if: there exists $n\geq 1$ such that, for all $k\geq n$, $|f_k(x)-f(x)|\leq 1/\ell$. So, $x\in\cap_{\ell=1}^{\infty}$ if and only if: for all $\ell\geq 1$, there exists $n\geq 1$ such that, for all $k\geq n$, $|f_k(x)-f(x)|\leq 1/\ell$. And the latter condition is the definition of the statement $\lim_{k\to\infty} f_k(x)=f(x)$. We conclude that $C=\cap_{\ell=1}^{\infty}U_\ell$, as desired. □

Exercise 1.6. Prove that a sequence in $\mathbb{R}^n$ converges if and only if all of its components converge.

Proof. Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $\varepsilon > 0$. Let $B_2(x, \varepsilon) := \{y \in \mathbb{R}^n \colon |x - y| < \varepsilon\}$, and let $B_\infty(x, \varepsilon) := \{y \in \mathbb{R}^n \colon \max_{i=1,\ldots,n} |x_i - y_i| \leq \varepsilon\}$. Note that

$$|x| = \left(\sum_{i=1}^n x_i^2\right)^{1/2} \leq \left(n \max_{i=1,\ldots,n} x_i^2\right)^{1/2} = \sqrt{n} \max_{i=1,\ldots,n} |x_i| \leq \sqrt{n} \left(\sum_{i=1}^n x_i^2\right)^{1/2}.$$

These inequalities imply the following set containments

$$B_2(x, \varepsilon) \subseteq B_\infty(x, \varepsilon) \subseteq B_2(x, \varepsilon\sqrt{n}) \qquad (*)$$

We now prove the forward implication of the exercise. Let $x^{(j)} \to x$ in $\mathbb{R}^n$ as $j \to \infty$. Then $x^{(j)} \to x$ if and only if: $\forall\, \varepsilon > 0$, $\exists\, N \in \mathbb{N}$ such that, for all $j \geq N$, $x^{(j)} \in B_2(x, \varepsilon)$. So, using $(*)$, $\forall\, \varepsilon > 0$, $\exists\, N \in \mathbb{N}$ such that, for all $j \geq N$, $x^{(j)} \in B_\infty(x, \varepsilon)$. By the definition of $B_\infty(x, \varepsilon)$, $\forall i = 1, \ldots, n$, $\forall\, \varepsilon > 0$, $\exists\, N \in \mathbb{N}$ such that, for all $j \geq N$, $|x_i^{(j)} - x_i| < \varepsilon$. That is, each component of the sequence converges.

We now prove the reverse implication of the exercise. Suppose that, for all $i = 1, \ldots, n$, $x_i^{(j)} \to x_i$ as $j \to \infty$. Then, $\forall\, \varepsilon > 0$, $\exists\, N(i) \in \mathbb{N}$ such that, for all $j \geq N(i)$, $|x_i^{(j)} - x_i| < \varepsilon/\sqrt{n}$. Define $N' := \max_{i=1,\ldots,n} N(i)$. Then, $\forall\, \varepsilon > 0$, $\exists\, N = N' \in \mathbb{N}$ such that, for all $j \geq N$, $|x_i^{(j)} - x_i| < \varepsilon/\sqrt{n}$ $\forall\, i = 1, \ldots, n$. By the definition of $B_\infty(x, \varepsilon)$, $\forall\, \varepsilon > 0$, $\exists\, N \in \mathbb{N}$ such that, for all $j \geq N$, $x \in B_\infty(x, \varepsilon/\sqrt{n})$. Using $(*)$, $\forall\, \varepsilon > 0$, $\exists\, N \in \mathbb{N}$ such that, for all $j \geq N$, $x \in B_2(x, \varepsilon)$. That is, $x^{(j)} \to x$, as desired. □

Exercise 1.7. Given a set A, the closure $\overline{A}$ of A is defined as

$$\overline{A} := \bigcap_{\substack{B \supseteq A\colon \\ B \text{ closed}}} B.$$

(i) Prove that $\overline{A}$ is closed. Hence, $\overline{A}$ is the smallest closed set containing A.
(ii) Show that
$$\overline{A} = A \cup \{\text{limit points of } A\}.$$
(iii) Show that $\overline{A \cup B} = \overline{A} \cup \overline{B}$ and $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$.
(iv) Find an example that shows that $\overline{A \cap B} \neq \overline{A} \cap \overline{B}$ in general.

Proof of (i). From De Morgan's law, $(\overline{A})^c = \cup_{B \supseteq A\colon B \text{ closed}} B^c$. Since each set B is closed, $(\overline{A})^c$ is a union of open sets. That is, $(\overline{A})^c$ is open, and therefore $\overline{A}$ is closed. □

Proof of (ii). From the definition of $\overline{A}$, we know that $A \subseteq \overline{A}$. So, to prove our desired equality, it suffices to show that $\overline{A} \smallsetminus A = \{\text{limit points of } A\} \smallsetminus A$.

We first show that $\overline{A} \smallsetminus A \subseteq \{\text{limit points of } A\} \smallsetminus A$. Let $x \in \overline{A} \smallsetminus A$. Suppose for the sake of contradiction that $x \in (\{\text{limit points of } A\} \smallsetminus A)^c = \{\text{limit points of } A\}^c \cup A$. Since $x \notin A$, $x \in \{\text{limit points of } A\}^c$. Then, by negating the definition of the limit points, there exists $r > 0$ such that $B_2(x, r) \cap (A \smallsetminus \{x\}) = \emptyset$. Since $x \in \overline{A} \smallsetminus A$, $x \notin A$. So, there exists $r > 0$ such that $B_2(x, r) \cap A = \emptyset$. Since $B_2(x, r)$ is open, $A \subseteq B_2(x, r)^c$, and $B_2(x, r)^c$ is a closed set. So, $\overline{A} \subseteq B_2(x, r)^c$. In particular, $x \notin \overline{A}$, a contradiction. We therefore conclude that $\overline{A} \smallsetminus A \subseteq \{\text{limit points of } A\} \smallsetminus A$.

We now show that $\overline{A} \setminus A \supseteq \{\text{limit points of } A\} \setminus A$. Let x be a limit point of A such that $x \notin A$. Since x is a limit point of A, for $n \in \mathbb{N}$, there exists $x_n \in A \cap B(x, 2^{-n})$ with $x_n \neq x$. We need to show that $x \in B$. We argue by contradiction. Assume that $x \in (\overline{A} \setminus A)^c = (\overline{A})^c \cup A$. Since $x \notin A$, $x \in (\overline{A})^c$. From De Morgan's law, $x \in \cup_{B \supseteq A, B \text{ closed}} B^c$. That is, there exists a closed B such that $B \supseteq A$ and such that $x \notin B$. Since B is closed, B^c is open, so there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \cap B = \emptyset$. But if we choose N such that $2^{-N} < \varepsilon$, then $x_N \in A \cap B(x, 2^{-N}) \subseteq A \cap B(x, \varepsilon)$. But since $A \subseteq B$, $x_N \in B(x, \varepsilon) \subseteq B^c \subseteq A^c$. The latter statement contradicts $x_N \in A$. We therefore conclude that $x \in \overline{A} \setminus A$, i.e. $\overline{A} \setminus A \supseteq \{\text{limit points of } A\} \setminus A$. □

Proof of (iii). We first show that $\overline{A \cup B} \subseteq \overline{A} \cup \overline{B}$. From (ii), we know that $\overline{A \cup B} = A \cup B \cup \{\text{limit points of } A \cup B\}$. If x is a limit point of $A \cup B$, then $\forall\, r > 0$, $B(x, r) \cap ((A \cup B) \setminus \{x\}) \neq \emptyset$. So, $\forall\, n \geq 1$, there exists $x_n \in (B(x, 2^{-n}) \setminus \{x\}) \cap (A \cup B)$. By taking a subsequence of the sequence $(x_n)_{n \in \mathbb{N}}$, we may assume that either $(x_{n_j})_{j \in \mathbb{N}} \subseteq A$ or $(x_{n_j})_{j \in \mathbb{N}} \subseteq B$. So, either $B(x, r) \cap (A \setminus \{x\}) \neq \emptyset$ for all $r > 0$, or $B(x, r) \cap (B \setminus \{x\}) \neq \emptyset$ for all $r > 0$. That is, x is a limit point of A, or x is a limit point of B. So, from (ii), $x \in \overline{A} \cup \overline{B}$. That is, $\{\text{limit points of } A \cup B\} \subseteq \overline{A} \cup \overline{B}$. Therefore, $\overline{A \cup B} = A \cup B \cup \{\text{limit points of } A \cup B\} \subseteq A \cup B \cup (\overline{A} \cup \overline{B}) = \overline{A} \cup \overline{B}$, proving our first containment.

We now show that $\overline{A \cup B} \supseteq \overline{A} \cup \overline{B}$. Let $x \in \overline{A} \cup \overline{B}$. From (ii), $\overline{A} \cup \overline{B} = A \cup B \cup \{\text{limit points of } A\} \cup \{\text{limit points of } B\}$. Let $x \in \{\text{limit points of } A\} \cup \{\text{limit points of } B\}$. That is, x is a limit point of A, or x is a limit point of B. So, either $B(x, r) \cap (A \setminus \{x\}) \neq \emptyset$ for all $r > 0$, or $B(x, r) \cap (B \setminus \{x\}) \neq \emptyset$ for all $r > 0$. In either case, $B(x, r) \cap ((A \cup B) \setminus \{x\}) \neq \emptyset$ for all $r > 0$. So, x is a limit point of $A \cup B$. And from (ii), $x \in \overline{A \cup B}$. That is, $\{\text{limit points of } A\} \cup \{\text{limit points of } B\} \subseteq \overline{A \cup B}$. Therefore, $\overline{A} \cup \overline{B} = A \cup B \cup \{\text{limit points of } A\} \cup \{\text{limit points of } A\} \subseteq A \cup B \cup \overline{A \cup B} = \overline{A \cup B}$, proving our second containment.

We now show that $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$. From (ii), $\overline{A \cap B} = A \cap B \cup \{\text{limit points of } A \cap B\}$. Since $A \subseteq \overline{A}$ and $B \subseteq \overline{B}$, $A \cap B \subseteq \overline{A} \cap \overline{B}$. So, it suffices to show that $\{\text{limit points of } A \cap B\} \subseteq \overline{A} \cap \overline{B}$. Let x be a limit point of $A \cap B$. Then $\forall\, r > 0$, $B(x, r) \cap ((A \cap B) \setminus \{x\}) \neq \emptyset$. In particular, for all $r > 0$, and $B(x, r) \cap (B \setminus \{x\}) \neq \emptyset$, and for all $r > 0$, $B(x, r) \cap (A \setminus \{x\}) \neq \emptyset$. That is, x is a limit point of A and x is a limit point of B. So, from (ii), $x \in \overline{A}$ and $x \in \overline{B}$. That is, $x \in \overline{A} \cap \overline{B}$. In conclusion, $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$. □

Proof of (iv). Let $A := [0, 1] \cap \mathbb{Q}$, and let $B := [0, 1] \setminus A$. By the density of the rationals, $\forall\, x \in [0, 1]$ and for all $r > 0$, $\exists\, y \in A$ with $0 < |y - x| < r$. Also, by the density of the irrationals, $\forall\, x \in [0, 1]$ and for all $r > 0$, $\exists\, y \in A$ with $0 < |y - x| < r$. So, from (ii), $[0, 1] \subseteq \overline{B}$ and $[0, 1] \subseteq A$. Since $[0, 1]$ is closed and $[0, 1] \supseteq A$, $[0, 1] \supseteq B$, the definition of the closure shows that $[0, 1] \supseteq \overline{A}$ and $[0, 1] \supseteq \overline{B}$. Combining our containments shows that $\overline{A} = \overline{B} = [0, 1]$, so that $\overline{A} \cap \overline{B} = [0, 1]$. However, by construction, $A \cap B = \emptyset$, so by the definition of $\overline{A \cap B}$, $\overline{A \cap B} = \emptyset$. In particular, $\overline{A} \cap \overline{B} = [0, 1] \neq \emptyset = \overline{A \cap B}$. □

Exercise 1.8.

(i) Prove the following characterization of continuity. A function f is continuous at a if and only if every sequence $(x_k)_{k \in \mathbb{N}}$ that converges to a has a subsequence $(x_{k_j})_{j \in \mathbb{N}}$ such that $\lim_{j \to \infty} f(x_{k_j}) = f(a)$.

(ii) Suppose that f is a continuous, bijective function defined on a compact set A. Show that f^{-1} is also continuous.

(iii) Can you find an example of a continuous, bijective function f such that f^{-1} is not continuous?

Proof of (i). The forward implication follows directly from the limit point definition of continuity. We therefore prove the reverse implication. We first consider the case that $f\colon \mathbb{R}^n \to \mathbb{R}$ is our function. Assume that every sequence $(x_k)_{k\in\mathbb{N}}$ that converges to a has a subsequence $(x_{k_j})_{j\in\mathbb{N}}$ such that $\lim_{j\to\infty} f(x_{k_j}) = f(a)$. We prove by contradiction that f is continuous. Assume for the sake of contradiction that f is not continuous. Then there exists $a \in \mathbb{R}^n$ and there exists a sequence $(x_k)_{k\in\mathbb{N}}$ such that $x_k \to a$ as $k \to \infty$, and either (1) $\lim_{k\to\infty} f(x_k) \neq f(a)$, or (2) the limit $\lim_{k\to\infty} f(x_k)$ does not exist. In either case, let $(x_{k_j})_{j\in\mathbb{N}}$ be the subsequence guaranteed by our assumption such that $\lim_{j\to\infty} f(x_{k_j}) = f(a)$. In case (1), we have $\lim_{k\to\infty} f(x_k) = \lim_{j\to\infty} f(x_{k_j}) = f(a)$, a contradiction. In case (2), let $(a_k)_{k\in\mathbb{N}}, (b_k)_{k\in\mathbb{N}}$ be subsequences of $(x_k)_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} f(a_k) = \limsup_{k\to\infty} f(x_k)$ and such that $\lim_{k\to\infty} f(b_k) = \liminf_{k\to\infty} f(x_k)$. By assumption, we can take subsequences of a_{k_j} and b_{k_j} such that $\lim_{j\to\infty} f(a_{k_j}) = f(a)$ and such that $\lim_{j\to\infty} f(b_{k_j}) = f(a)$. But then

$$\limsup_{k\to\infty} f(x_k) = \lim_{k\to\infty} f(a_k) = \lim_{j\to\infty} f(a_{k_j}) = f(a) = \lim_{j\to\infty} f(b_{k_j}) = \lim_{k\to\infty} f(b_k) = \liminf_{k\to\infty} f(x_k).$$

That is, $\lim_{k\to\infty} f(x_k)$ exists, a contradiction. So, in any case, we achieve a contradiction. We conclude that f is in fact continuous, as desired.

We now note how to prove the general case, where our function is $f\colon \mathbb{R}^n \to \mathbb{R}^m$. As before, the forward implication follows directly from the limit point definition of continuity. We therefore prove the reverse implication. Assume that every sequence $(x_k)_{k\in\mathbb{N}}$ that converges to a has a subsequence $(x_{k_j})_{j\in\mathbb{N}}$ such that $\lim_{j\to\infty} f(x_{k_j}) = f(a)$. We prove by contradiction that f is continuous. Assume for the sake of contradiction that f is not continuous. Then there exists $a \in \mathbb{R}^n$ and there exists a sequence $(x_k)_{k\in\mathbb{N}}$ such that $x_k \to a$ as $k \to \infty$, and either (1) $\lim_{k\to\infty} f(x_k) \neq f(a)$, or (2) the limit $\lim_{k\to\infty} f(x_k)$ does not exist. Suppose Case (1) occurs. From the contrapositive of Exercise 1.6, there exists some component $i \in \{1,\ldots,m\}$ of the vectors $f(x_k)$ such that $\lim_{k\to\infty} f(x_k)_i \neq f(a)_i$. But then we repeat the argument above for the function $f_i\colon \mathbb{R}^n \to \mathbb{R}$ and Case (1), and we achieve a contradiction, as above. Now, suppose Case (2) occurs. From the contrapositive of Exercise 1.6, there exists some component $i \in \{1,\ldots,m\}$ of the vectors $f(x_k)$ such that $\lim_{k\to\infty} f(x_k)_i$ does not exist. But then we repeat the argument above for the function $f_i\colon \mathbb{R}^n \to \mathbb{R}$ and Case (2), and we achieve a contradiction, as above. The exercise is therefore concluded. □

Proof of (ii). Let $f\colon A \to Y$ be a continuous, bijective function defined on a compact set A. Since f is bijective, $f^{-1}\colon Y \to A$ is a well defined function. Let $(y_k)_{k\in\mathbb{N}}$ be a sequence in Y such that $y_k \to y$. From part (i), it suffices to show that there exists a subsequence $(y_{k_j})_{j\in\mathbb{N}}$ such that $\lim_{j\to\infty} f^{-1}(y_{k_j}) = f^{-1}(y)$. Let a_k be the unique elements of A such that $f(a_k) = y_k$, and let a be the unique element of A such that $f(a) = y$. Then $(a_k)_{k\in\mathbb{N}}$ is a sequence in the compact set A, so there exists a subsequence $(a_{k_j})_{j\in\mathbb{N}}$ and some $a' \in A$ such that $a_{k_j} \to a'$ as $j \to \infty$. Since f is continuous, $\lim_{j\to\infty} f(a_{k_j}) = f(a')$. That is, $\lim_{j\to\infty} y_{k_j} = f(a')$. However, since $y_k \to y$, we conclude that $y = f(a')$. And since $y = f(a)$, we conclude by bijectivity of f that $a = a'$. So, $a_{k_j} \to a$ as $j \to \infty$, i.e. $f^{-1}(y_{k_j}) \to f^{-1}(y)$ as $j \to \infty$, as desired. □

Proof of (iii). Define $f\colon [0,1)\cup[2,3] \to [0,2]$ so that $f(x) = x$ for $x \in [0,1)$, and $f(x) = x-1$ for $x \in [2,3]$. We first show that f is injective. Let $x, y \in [0,1)\cup[2,3]$ such that $f(x) = f(y)$. We break into three cases. In the first case $x, y \in [0,1)$, so by the definition of f, we must have $x = y$. In the second case, $x, y \in [2,3]$, so by the definition of f, $x-1 = y-1$, i.e. $x = y$. In the third and final case, $x \in [0,1)$ and $y \in [2,3]$, so $x = y-1$. But $y \in [2,3]$ implies $y-1 \in [1,2]$, so it cannot happen that $x = y-1$, i.e. this third case cannot occur. In conclusion, if $f(x) = f(y)$, we must have $x = y$, so f is injective.

We now prove surjectivity of f. Let $y \in [0,2]$. If $y \in [0,1)$, then let $x = y$, and note that $f(x) = y$ and $x \in [0,1)$. If $y \in [1,2]$, then let $x = y+1$, and note that $f(x) = y$ and $x \in [2,3]$. In conclusion, every $y \in [0,2]$ has an $x \in [0,1)\cup[2,3]$ such that $f(x) = y$, so f is surjective.

We now prove continuity of f. Let $x_n \to x$, $x_n, x \in [0,1)\cup[2,3]$. In particular, for $\varepsilon = 1/2$, there exists N such that $k \geq N$ implies $|x_k - x| < 1/2$. So, by discarding finitely many terms of the sequence $(x_n)_{n\in\mathbb{N}}$, we may assume that either (1) $x_n, x \in [0,1)\ \forall\ n \in \mathbb{N}$ or (2) $x_n, x \in [2,3]\ \forall\ n \in \mathbb{N}$. In case (1), f is continuous since $f(x_n) = x_n \to x = f(x)$. In case (2), f is continuous since $f(x_n) = x_n - 1 \to x - 1 = f(x)$. Since all cases have been exhausted, we conclude that f is continuous.

We now show that f^{-1} is not continuous. Let $y_n = 1 + (-1)^n/n$. For n even, $y_n \in [1,2]$, and for n odd, $y_n \in [0,1)$. From surjectivity of f, let x_n such that $f(x_n) = y_n$. Then for n even, $x_n = 1 - 1/n$, and for n odd, $x_n = 2 + 1/n$. In particular, the sequence $(x_n)_{n\in\mathbb{N}}$ does not converge. So, there exists a sequence y_n such that $y_n \to y$, but $\lim_{n\to\infty} f^{-1}(y_n)$ does not exist. That is, f^{-1} is not continuous, as desired. □

Another proof of (iii). For a complex number $z \in \mathbb{C}$, let $\mathrm{Re}(z)$ denote the real part of z, and let $\mathrm{Im}(z)$ denote the imaginary part of z. The set of complex numbers is identified with the set $\mathbb{R}\times\mathbb{R}$ by the bijection $z \mapsto (\mathrm{Re}(z), \mathrm{Im}(z))$, and $\mathbb{C}$ is given the topology of $\mathbb{R}\times\mathbb{R}$. Also, define $|z| := \sqrt{(\mathrm{Re}(z))^2 + (\mathrm{Im}(z))^2}$. Let $S^1 := \{z \in \mathbb{C}\colon |z|^2 = 1\}$. Let $f\colon [0,2\pi) \to S^1$ be defined by $f(t) = e^{it}$. Since $|e^{it}| = 1$ for all $t \in \mathbb{R}$, we know that the range of f is $S^1 \subseteq \mathbb{C}$. Since each component of $f(t) = (\cos(t), \sin(t))$ is continuous, f is a continuous map. We now prove bijectivity of f. Suppose $f(t) = f(t')$, $t, t' \in [0,2\pi)$. Then $e^{i(t-t')} = 1$, and $t - t' \in (-2\pi, 2\pi)$. Since $e^{iz} = 1$ only for $z = 2k\pi$, $k \in \mathbb{Z}$, we conclude that $t - t' = 0$, i.e. $t = t'$. Therefore f is injective. To see surjectivity of f, let $z \in S^1$. Then $(\mathrm{Re}(z))^2 + (\mathrm{Im}(z))^2 = 1$. So, there exists $t \in [0,2\pi)$ such that $\cos(t) = \mathrm{Re}(z)$ and $\sin(t) = \mathrm{Im}(z)$. Therefore, $f(t) = e^{it} = (\cos(t) + i\sin(t)) = z$, as desired.

We have shown that f is a continuous bijective function. We now show that the inverse map is not continuous. Let $n \geq 1, n \in \mathbb{Z}$. For n even, let $x_n = 1/n$, and for n odd, let $x_n = 2\pi - 1/n$. For $n \geq 1$, let $y_n := f(x_n)$. Then

$$|y_n - 1|^2 = |(\cos(\pm 1/n), \sin(\pm 1/n)) - (1,0)|^2 = (\cos(1/n) - 1)^2 + \sin^2(1/n) \leq 2/n^2.$$

Therefore, $y_n \to (1,0)$ as $n \to \infty$. However, $f^{-1}(y_n) = x_n$ does not converge as $n \to \infty$. If we restrict attention to even n, then $x_n \to 0$ as $n \to \infty$. But if we restrict attention to odd n, then $x_n \to 2\pi$. In particular, x_n does not converge to any number. So, we have a sequence y_n such that $y_n \to y$, but $\lim_{n\to\infty} f^{-1}(y_n)$ does not exist. We conclude that f^{-1} is not continuous, as desired. □

2. Problem Set 2

Exercise 2.1. Prove that a uniformly continuous function is continuous.

Proof. Let $f\colon X \to \mathbb{R}^n$ be uniformly continuous. Then, $\forall\, \varepsilon > 0$, $\exists\, \delta = \delta(\varepsilon) > 0$ such that, $\forall\, x, y \in X$, if $0 < |x - y| < \delta$, then $|f(x) - f(y)| < \varepsilon$. In particular, if we fix $x \in X$, then $\forall$ $\varepsilon > 0$, $\exists\, \delta = \delta(\varepsilon) > 0$ such that, $\forall\, y \in X$, if $0 < |x - y| < \delta$, then $|f(x) - f(y)| < \varepsilon$. That is, f is continuous. □

Exercise 2.2. Does the function

$$f(x,y) = \frac{x^3 - y^3}{|x-y| + y^2}$$

have a limit as $(x, y) \to 0$? If yes, give a limit. Answer the same question for the function

$$g(x,y) = \left(x + y^3, x + \frac{x}{x^2 + y^2}\right)$$

Proof. We show that $\lim_{(x,y)\to(0,0)} f(x, y) = 0$. Let $0 < \varepsilon < 1$. We need to find $\delta > 0$ such that $0 < |(x, y)| < \delta$ implies that $|f(x, y)| < \varepsilon$. We claim that $\delta := \varepsilon/10$ suffices. First, suppose (x, y) satisfies $0 < |(x, y)| < \delta$. Since $|x| = (|x|^2)^{1/2} \leq |(x, y)| < \delta$, and $|y| = (|y|^2)^{1/2} \leq |(x, y)| < \delta$, we know that $|x| < \delta$ and $|y| < \delta$.

We will consider two separate sets. Define

$$A := \{(x, y) \in \mathbb{R} \times \mathbb{R} \colon\ |y| \geq |x|/2, 0 < |(x, y)| < \delta\}$$
$$B := \{(x, y) \in \mathbb{R} \times \mathbb{R} \colon\ |y| < |x|/2, 0 < |(x, y)| < \delta\}$$

Then $A \cup B = \{(x, y) \in \mathbb{R} \times \mathbb{R} \colon 0 < |(x, y)| < \delta\}$. Let $(x, y) \in A$. Then, by the definition of A, $|y|^3 \geq |x|^3/8$, so $|x^3 - y^3| \leq |x|^3 + |y|^3 \leq 9\,|y|^3$. Also, $||x - y| + y^2| \geq y^2$. Moreover, since $(x, y) \in A$, $|(x, y)| < \delta$, so $|y| < \delta < \varepsilon/10$. Combining these observations,

$$|f(x,y)| = \frac{|x^3 - y^3|}{|x-y| + y^2} \ \leq \frac{9\,|y|^3}{y^2} = 9\,|y| \leq \frac{9}{10}\varepsilon < \varepsilon.$$

So, $(x, y) \in A$ implies that $|f(x, y)| < \varepsilon$.

Now, let $(x, y) \in B$. Then, by the reverse triangle inequality and the definition of B, $|x - y| \geq |x| - |y| > |x|/2$. Therefore, $||x - y| + y^2| \geq |x - y| > |x|/2$. Also, by the definition of B, $|x^3 - y^3| \leq |x|^3 + |y|^3 < (9/8)\,|x|^3$. Moreover, since $(x, y) \in B$, $|(x, y)| < \delta$, so $|x| < \delta < \varepsilon/10$. Combining these observations,

$$|g(x,y)| = \frac{|x^3 - y^3|}{|x-y| + y^2} \leq \frac{9}{4}\frac{|x|^3}{|x|} \leq 3\,|x|^2 \leq \frac{3}{100}\varepsilon^2 < \varepsilon$$

In the last line, we used that $\varepsilon < 1$, so $\varepsilon^2 < \varepsilon$. So, $(x, y) \in B$ implies that $|f(x, y)| < \varepsilon$.

By combining the results for A and B, we conclude that, if $(x, y) \in A \cup B$, i.e. if $|(x, y)| < \delta$, then $|f(x, y)| < \varepsilon$, as desired. In conclusion, f is continuous at 0.

We now show that $\lim_{(x,y)\to(0,0)} g(x, y)$ does not exist. We argue by contradiction. Suppose $\lim_{(x,y)\to(0,0)} g(x, y) = (a, b)$, $(a, b) \in \mathbb{R} \times \mathbb{R}$. For $n \in \mathbb{N}$, let (x_n, y_n) such that $x_n = 2^{-n}$ and $y_n = x_n$. Then $(x_n, y_n) \to (0, 0)$ as $n \to \infty$. So, since $\lim_{(x,y)\to(0,0)} g(x, y)$ exists, we must have

$$\lim_{n\to\infty} g(x_n, y_n) = (a, b). \qquad (*)$$

However,

$$g(x_n, y_n) = g(x_n, x_n) = \left(2^{-n} + 2^{-3n}, 2^{-n} + \frac{2^{-n}}{2^{-2n} + 2^{-2n}}\right) = \left(2^{-n} + 2^{-3n}, 2^{-n} + 2^{n-1}\right)$$

Now, let N such that $n \geq N$ implies $2^{n-1} > b + 1$. Then $2^{-n} + 2^{n-1} > 2^{n-1} > b + 1 > b$. So, it is impossible that $\lim_{n\to\infty} g(x_n, y_n) = (a, b)$, since $n \geq N$ implies $|g(x_n, y_n) - (a, b)| \geq |2^{-n} + 2^{n-1} - b| > 1$. Since we have achieved a contradiction, we conclude that the limit $\lim_{(x,y)\to(0,0)} g(x, y)$ does not exist. □

Exercise 2.3. Find a continuous function f and an open set U such that $f(U)$ is not open.

Proof. Let $f(x) := \sin(x)$, and let $U := (-\pi, \pi)$. We claim that $f(U) = [-1, 1]$. Indeed, $f(\pi/2) = 1$, $f(-\pi/2) = -1$, f is continuous on $[-\pi/2, \pi/2]$, so by the intermediate value theorem, $f[-\pi/2, \pi/2] \supseteq [-1, 1]$. Also, $|f(x)| \leq 1$ for $x \in \mathbb{R}$, so $f(U) \subseteq [-1, 1]$. Therefore,

$$[-1, 1] \subseteq f[-\pi/2, \pi/2] \subseteq f(U) \subseteq [-1, 1].$$

We conclude that $f(U) = [-1, 1]$, as desired. □

Exercise 2.4. A function $f : D \to \mathbb{R}^m$ for $D \subseteq \mathbb{R}^n$ is called *Lipschitz continuous* if there exists a constant L such that

$$|f(x) - f(y)| \leq L\,|x - y|\,, \quad \forall\, x, y \in D.$$

The constant L is called the *Lipschitz constant of f*.

(i) Prove that a Lipschitz continuous function is uniformly continuous.
(ii) Find an example of a uniformly continuous function that is not Lipschitz continuous.
(iii) Prove that the function $x \mapsto |x|$ is a Lipschitz continuous function from $\mathbb{R}^n \to \mathbb{R}$.
(iv) Prove that the function $(x, y) \mapsto x+y$ is Lipschitz continuous function from $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$.
(v) Prove that the inner product $(x, y) \mapsto \langle x, y\rangle$ is a Lipschitz continuous function from $D \times D$ to $\mathbb{R}$, for any bounded domain D.

Proof of (i). Suppose f is Lipschitz continuous with constant L. We now show that f is uniformly continuous. Let $\varepsilon > 0$. Then, define $\delta = \delta(\varepsilon) := \varepsilon/L$. And let $x, y \in D$. If $|x - y| \leq \delta$, then the definition of Lipschitz continuity implies that $|f(x) - f(y)| \leq L\,|x - y| \leq L\delta < \varepsilon$, where in the last line we used our definition of δ. In conclusion, f is uniformly continuous. □

Proof of (ii). Let $D := [0, 1]$, and for $x \in D$ let $f(x) := \sqrt{x}$. We first show that f is uniformly continuous. Let $\varepsilon > 0$. Then let $\delta := \varepsilon^2/2$. If $|x - y| < \delta$ with $x, y \in [0, 1]$, then by concavity of f, and the definition of δ,

$$|f(x) - f(y)| = \left|\sqrt{x} - \sqrt{y}\right| = 2\left|\frac{\sqrt{x} - \sqrt{y}}{2}\right| \leq 2\sqrt{|x - y|\,/2} < \sqrt{2\delta} = \varepsilon.$$

So, f is uniformly continuous on D. However, f is not Lipschitz, since for $x > 0$,

$$|f(0) - f(x)|\,/\,|0 - x| = \sqrt{x}/\,|x| = x^{-1/2} \to \infty \text{ as } x \to 0.$$

□

Proof of (iii). We show that $f(x) := |x|$ is Lipschitz with constant $L = 1$, i.e. $\left| |x| - |y| \right| \leq |x - y|$, $x, y \in \mathbb{R}^n$. However, this is just the reverse triangle inequality. For completeness, we prove this inequality from the usual triangle inequality. From the usual triangle inequality, $|y| = |x + (y - x)| \leq |x| + |y - x|$, so $|y| - |x| \leq |y - x|$. Similarly, $|x| = |y + (x - y)| \leq |y| + |x - y|$, so $|x| - |y| \leq |x - y|$. Therefore,

$$\left| |y| - |x| \right| = \max(|y| - |x|\,, |x| - |y|) \leq |y - x|\,.$$

□

Proof of (iv). Let $x, y, x_1, y_1, x_2, y_2 \in \mathbb{R}^n$. Let $f(x, y) := x + y$. We show that f is Lipschitz with constant $\sqrt{2}$, i.e.

$$|x_1 + y_1 - (x_2 + y_2)| \leq \sqrt{2}\,|(x_1, y_1) - (x_2, y_2)|\,.$$

We first prove an inequality for real numbers a, b.

$$|a| + |b| \leq \sqrt{2}(a^2 + b^2)^{1/2}. \qquad (*)$$

Observe

$$\begin{aligned} |a| + |b| &= 2\left(\frac{|a| + |b|}{2}\right)^{2 \cdot \frac{1}{2}} \\ &\leq 2\left(\frac{|a|^2 + |b|^2}{2}\right)^{\frac{1}{2}} \qquad \text{, by convexity of } t \mapsto t^2,\ t \in \mathbb{R} \\ &= \sqrt{2}(a^2 + b^2)^{1/2}. \end{aligned}$$

Now, we prove that f is Lipschitz continuous.

$$\begin{aligned} |x_1 + y_1 - (x_2 + y_2)| &= |(x_1 - x_2) + (y_1 - y_2)| \\ &\leq |x_1 - x_2| + |y_1 - y_2| \qquad \text{, by the triangle inequality on } \mathbb{R}^n \\ &\leq \sqrt{2}\left(|x_1 - x_2|^2 + |y_1 - y_2|^2\right)^{\frac{1}{2}} \qquad \text{, from } (*) \\ &= \sqrt{2}\,|(x_1, y_1) - (x_2, y_2)| \end{aligned}$$

□

Proof of (v). Let $D \subseteq \mathbb{R}^n$ be a bounded set. So, there exists $R > 0$ such that $D \subseteq B_R(0)$. We prove that $f(x, y) := \langle x, y \rangle$ is Lipschitz with constant $\sqrt{2}R$. Observe

$$\begin{aligned} |f(x_1, y_1) - f(x_2, y_2)| &= |\langle x_1, y_1 \rangle - \langle x_2, y_2 \rangle| \\ &= |\langle x_1, y_1 \rangle - \langle x_1, y_2 \rangle + \langle x_1, y_2 \rangle - \langle x_2, y_2 \rangle| \\ &= |\langle x_1, y_1 - y_2 \rangle - \langle x_2 - x_1, y_2 \rangle| \\ &\leq |\langle x_1, y_1 - y_2 \rangle| + |\langle x_2 - x_1, y_2 \rangle| \qquad \text{, by the triangle inequality on } \mathbb{R} \\ &\leq |x_1|\,|y_1 - y_2| + |y_2|\,|x_1 - x_2| \qquad \text{, by Cauchy-Schwarz} \\ &\leq R(|y_1 - y_2| + |x_1 - x_2|) \qquad \text{, by the definition of } R \\ &\leq \sqrt{2}R(|y_1 - y_2|^2 + |x_1 - x_2|^2)^{1/2} \qquad \text{, from } (*) \\ &= \sqrt{2}R\,|(x_1, y_1) - (x_2, y_2)|\,. \end{aligned}$$

□

Exercise 2.5. We say that $f\colon D \to \mathbb{R}^m$ is α-*Hölder continuous* if there is a constant L such that

$$|f(x) - f(y)| \leq L\,|x-y|^{\alpha}, \quad \forall\, x, y \in D.$$

(i) What does this condition mean for $\alpha = 1$? What about $\alpha = 0$?
(ii) If $\alpha > 0$, prove that an α-Hölder continuous function is uniformly continuous.
(iii) Prove that if D is bounded, $\alpha \leq \beta$, and if f is β-Hölder continuous, then f is α-Hölder continuous.
(iv) Prove that if f is α-Hölder continuous for $\alpha > 1$, then f is constant.

Proof of (i). If $\alpha = 1$, then f is Lipschitz continuous with constant L. If $\alpha = 0$, then f is a bounded function, since for any $x \in D$, $|f(x)| \leq L + |f(d)|$, $d \in D$ fixed. □

Proof of (ii). Suppose f is α-Hölder continuous of order α. We now show that f is uniformly continuous. Let $\varepsilon > 0$. Then, define $\delta = \delta(\varepsilon) := (\varepsilon/L)^{1/\alpha}$. And let $x, y \in D$. If $|x - y| \leq \delta$, then the definition of Hölder continuity implies that $|f(x) - f(y)| \leq L\,|x-y|^{\alpha} \leq L\delta^{\alpha} < \varepsilon$, where in the last line we used our definition of δ. In conclusion, f is uniformly continuous. □

Proof of (iii). Suppose f is β-Hölder continuous of order β, and D is bounded. So, there exists $R > 0$ such that $D \subseteq B_R(0)$. We show that f is α-Hölder continuous. Let $x, y \in D$. By assumption, $\beta - \alpha \geq 0$, so by the definition of R,

$$|x-y|^{\beta-\alpha} \leq (|x| + |y|)^{\beta-\alpha} \leq (2R)^{\beta-\alpha}. \qquad (*)$$

Now, the definition of β-Hölder continuity implies that

$$\begin{aligned} |f(x) - f(y)| &\leq L\,|x-y|^{\beta} = L\,|x-y|^{\beta-\alpha+\alpha} = (L\,|x-y|^{\beta-\alpha})\,|x-y|^{\alpha} \\ &\leq L(2R)^{\beta-\alpha}\,|x-y|^{\alpha} \quad \text{, from } (*) \end{aligned}$$

In conclusion, f is α-Hölder continuous. □

Proof of (iv). Let $\alpha > 1$. Let $x, y \in D$. For $i = 0, \ldots, n$, define $x_i := y + (i/n)(x - y)$. Then $x_0 = y$, and $x_n = x$, so by the triangle inequality,

$$|f(x) - f(y)| = \left|\sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))\right| \leq \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|\,.$$

We now apply the Hölder continuity of f. Note that $x_i - x_{i-1} = (1/n)(x - y)$, so

$$|f(x) - f(y)| \leq L\sum_{i=1}^{n} |x_i - x_{i-1}|^{\alpha} \leq L\sum_{i=1}^{n} n^{-\alpha}\,|x-y| = Ln^{1-\alpha}\,|x-y|$$

Since $\alpha > 1$, $1 - \alpha < 0$, and letting $n \to \infty$ shows that $|f(x) - f(y)| = 0$, i.e. f is constant. □

Exercise 2.6. A subset $A \subseteq \mathbb{R}^n$ is *dense* if $\forall\, x \in \mathbb{R}^n$ and $\forall\, \varepsilon > 0$, $A \cap B_{\varepsilon}(x) \neq \emptyset$.

(i) Prove that $\mathbb{Q}$ is dense in $\mathbb{R}$.
(ii) Using (i), prove that $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$
(iii) Let f, g be continuous functions and let A be a dense set in $\mathbb{R}^n$. Prove that if $f(x) = g(x)\ \forall\, x \in A$, then $f(x) = g(x)\ \forall\, x \in \mathbb{R}^n$.

Proof of (i). Let $x \in \mathbb{R}$ and let $\varepsilon > 0$. Then $B_\varepsilon(x) = \{y \in \mathbb{R} \colon |y - x| < \varepsilon\}$. From the Archimedean property of the real numbers, there exists $n \in \mathbb{N}$ such that $10^{-n} < \varepsilon$. Write x uniquely as an infinite decimal number without an infinite sequence of repeating 9's. That is, write $x = a_m a_{m-1} \cdots a_0.b_1 b_2 b_3 \cdots$, where $a_i, b_i \in [0,9] \cap \mathbb{Z}$ for all i. Then the number $y := a_m a_{m-1} \cdots a_0.b_1 b_2 \cdots b_n$ is rational, and $|x - y| < 10^{-n} < \varepsilon$. So, $y \in B_\varepsilon(x)$. Therefore, $\mathbb{Q}$ is dense in $\mathbb{R}$. $\square$

Proof of (ii). Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and let $\varepsilon > 0$. From equation $(*)$ in Exercise 1.6,

$$C := \{y \in \mathbb{R}^n \colon \max_{i=1,\ldots,n} |x_i - y_i| < \varepsilon/\sqrt{n}\} \subseteq B_\varepsilon(x) \qquad (\ddagger)$$

From part (i), for each $i = 1, \ldots, n$, let $y_i \in \mathbb{Q}$ such that $|x_i - y_i| < \varepsilon/\sqrt{n}$. Then, define $y := (y_1, \ldots, y_n)$. Since $y_i \in \mathbb{Q}$ for $i = 1, \ldots, n$, $y \in \mathbb{Q}^n$. Also, $\max_{i=1,\ldots,n} |x_i - y_i| < \varepsilon/\sqrt{n}$. So, $y \in C$. And from $(\ddagger)$, $y \in C \subseteq B_\varepsilon(x)$. Therefore, $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$. $\square$

Proof of (iii). Let $x \in \mathbb{R}^n$. Since A is dense, if $m \in \mathbb{N}$, then there exists $y_m \in A$ such that $|y_m - x| < 2^{-m}$. By construction, $y_m \to x$ as $m \to \infty$. Then, using our assumption on f, g, $f(y_m) = g(y_m)$ for all $m \in \mathbb{N}$, so

$$f(x) = \lim_{m\to\infty} f(x_m) = \lim_{m\to\infty} g(x_m) = g(x).$$

$\square$

Exercise 2.7. Define $f \colon \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1 & \text{, if } x \in \mathbb{Q} \\ 0 & \text{, if } x \notin \mathbb{Q}. \end{cases}$$

Show that f does not have a limit at any point in $\mathbb{R}$.

Proof. We first show that the set $\mathbb{R} \smallsetminus \mathbb{Q}$ is dense in $\mathbb{R}$. Let $q_1, q_2, \ldots$ be a countable enumeration of the rationals. Let $x \in \mathbb{R}$ and $\varepsilon > 0$. We want to find a $y \in \mathbb{R} \smallsetminus \mathbb{Q}$ such that $y \in B_\varepsilon(x)$. Consider the union of open intervals

$$V := \bigcup_{n\in\mathbb{N}} (q_n - \varepsilon 2^{-n}/4, q_n + \varepsilon 2^{-n}/4).$$

The total length of the set V is bounded by $\sum_{n=1}^{\infty} \varepsilon 2^{-n}/2 = \varepsilon/2$. But the length of $B_\varepsilon(x)$ is 2ε. So, there exists a $y \in B_\varepsilon(x)$ such that $y \in V^c$. But $\mathbb{Q} \subseteq V$ by construction of V, so $V^c \subseteq \mathbb{Q}^c = \mathbb{R} \smallsetminus \mathbb{Q}$, so $y \in V^c \subseteq \mathbb{R} \smallsetminus \mathbb{Q}$, as desired.

We now show that f does not have a limit in $\mathbb{R}$. We argue by contradiction. Suppose f has a limit at some point $x \in \mathbb{R}$. Then for $\varepsilon = 1/3$, there exists $\delta > 0$ such that, if $|x - y| < \delta$, then $|f(x) - f(y)| < 1/3$. From the density of the rationals in $\mathbb{R}$ (Exercise 2.6(i)), there exists $y_1 \in \mathbb{Q}$ with $|x - y_1| < \delta$. Also, from the density of the irrationals in $\mathbb{R}$ (proven above), there exists $y_2 \in \mathbb{R} \smallsetminus \mathbb{Q}$ with $|x - y_2| < \delta$. However, $f(y_1) = 1$ and $f(y_2) = 0$ by the definition of f. So, $f(x)$ must satisfy $|f(x) - 1| < 1/3$, and $|f(x) - 0| < 1/3$. No real number satisfies these two inequalities, i.e. we have achieved a contradiction. We conclude that f does not have a limit in $\mathbb{R}$. $\square$

Exercise 2.8. Using the one-dimensional Bolzano-Weierstrass theorem, prove that any bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Proof. Let $(x^{(m)})_{m\in\mathbb{N}}$ be a bounded sequence in $\mathbb{R}^n$. That is, there exists $R>0$ such that $|x^{(m)}|<R$ for all $m\in\mathbb{N}$. Since $x^{(m)}\in\mathbb{R}^n$, write $x^{(m)}=(x_1^{(m)},\ldots,x_n^{(m)})$, $x_i^{(m)}\in\mathbb{R}$ for $i=1,\ldots,n$. Now, for $x=(x_1,\ldots,x_n)\in\mathbb{R}^n$,

$$\max_{i=1,\ldots,n}|x_i|=(\max_{i=1,\ldots,n}|x_i|^2)^{1/2}\leq\left(\sum_{i=1}^{n}|x_i|^2\right)^{1/2}=|x|$$

So, for all $i=1,\ldots,n$, $|x_i^{(m)}|\leq\left|x^{(m)}\right|<R$. So, by applying Bolzano-Weierstrass for $i=1$, there exists a subsequence $(x_1^{(m_j)})_{j\in\mathbb{N}}\subseteq\mathbb{R}$ of the sequence $(x_1^{(m)})_{m\in\mathbb{N}}\subseteq\mathbb{R}$, and there exists $x_1\in\mathbb{R}$ such that $x_1^{(m_j)}\to x_1$ as $m\to\infty$. We label this subsequence so that $x^{(m_j)}=:x^{(j,1)}$. Since we have taken a subsequence of the original sequence, we still have $|x_i^{(j,1)}|<R$ for all $i=1,\ldots,n$. Now, apply Bolzano-Weierstrass to the sequence $(x_2^{(j,1)})_{j\in\mathbb{N}}\subseteq\mathbb{R}$. As before, we have a subsequence $(x^{(j,2)})_{j\in\mathbb{N}}$ of the sequence $(x^{(j,1)})_{j\in\mathbb{N}}$ and some $x_2\in\mathbb{R}$ such that $x_2^{(j,2)}\to x_2$ as $j\to\infty$. Since we have taken a subsequence, we still have the property $x_1^{(j,2)}\to x_1$ as $j\to\infty$. We continue this process n times. We get n nested subsequences

$$(x^{(j)})_{j\in\mathbb{N}}\supseteq(x^{(j,1)})_{j\in\mathbb{N}}\supseteq\cdots\supseteq(x^{(j,n)})_{j\in\mathbb{N}}.$$

Let $x:=(x_1,\ldots,x_n)$, where x_i is produced in the i^{th} step of this process of taking subsequences. The final subsequence $(x^{(j,n)})_{j\in\mathbb{N}}$ satisfies $x_i^{(j,n)}\to x_i$ as $j\to\infty$, for all $i=1,\ldots,n$. The proof is therefore complete. □

Exercise 2.9. Recall that a compact set A satisfies: for any open cover of A, there exists a finite subcover of A.

(i) Find an example of a bounded set together with an open cover which has no finite subcover.

(ii) Find an example of a closed set together with an open cover which has no finite subcover

Proof of (i). Let $A:=(-1,1)$, and for $n\geq 2, n\in\mathbb{N}$ let $A_n:=(-1+1/n,1-1/n)$. We first show that $\cup_{n=2}^{\infty}A_n=A$. Since each A_n satisfies $A_n\subseteq A$, it follows that $\cup_{n=2}^{\infty}A_n\subseteq A$. For the reverse inclusion, let $x\in A$. Then there exists $\varepsilon>0$ so that $|x|<1-\varepsilon$. Let $n\in\mathbb{N}$ such that $1/n<\varepsilon$. Then $|x|<1-1/n$, and $A_n=\{y\in\mathbb{R}\colon\ |y|<1-1/n\}$, so $x\in A_n$. In particular, $x\in\cup_{n=2}^{\infty}A_n$. Therefore, $A\subseteq\cup_{n=2}^{\infty}A_n$. We conclude that $A=\cup_{n=2}^{\infty}A_n$.

We now that that this cover has no finite subcover. We argue by contradiction. Suppose a finite subcover exists. That is, there exist natural numbers $n_1,\ldots,n_j$ with $n_1<n_2<\cdots<n_j$ such that $A\subseteq A_{n_1}\cup\cdots\cup A_{n_j}$. Since $N<M$ implies $A_N\subseteq A_M$, we conclude that $A\subseteq A_{n_j}$. But the point $x:=1-1/(2n_j)$ satisfies $x\notin A_{n_j}$, but $x\in A$. This statement contradicts the inclusion $A\subseteq A_{n_j}$. We conclude that this cover has no finite subcover. □

Proof of (ii). Let $A:=\mathbb{R}$, and for $n\in\mathbb{N}$, let $A_n:=(n-1,n+1)$. Recall that A is closed. We first show that $\cup_{n\in\mathbb{Z}}A_n=A$. Since each A_n satisfies $A_n\subseteq A$, it follows that $\cup_{n\in\mathbb{Z}}A_n\subseteq A$. For the reverse inclusion, let $x\in A$. Then there exists $n\in\mathbb{Z}$ such that $|n-x|<1$. So, $x\in(n-1,n+1)=A_n$. In particular, $x\in\cup_{n\in\mathbb{Z}}A_n$. Therefore, $A\subseteq\cup_{n\in\mathbb{Z}}A_n$. We conclude that $A=\cup_{n\in\mathbb{Z}}A_n$.

We now that that this cover has no finite subcover. We argue by contradiction. Suppose a finite subcover exists. That is, there exist natural numbers $n_1,\ldots,n_j$ such that $A\subseteq$

$A_{n_1} \cup \cdots \cup A_{n_j}$. Let $N := \max_{i=1,\ldots,j} |n_j|$. By the definition of A_n, we therefore have $A \subseteq A_{n_1} \cup \cdots \cup A_{n_j} \subseteq (-N-2, N+2)$. But the point $x := N+3$ satisfies $x \notin (-N-2, N+2)$, but $x \in A$. This statement contradicts the inclusion $A \subseteq (-N-2, N+2)$. We conclude that this cover has no finite subcover. □

Exercise 2.10. Let f be defined by

$$f(x,y) = \begin{cases} |y/x^2|\, e^{-|y/x^2|} & \text{, if } x \neq 0 \\ 0 & \text{, if } x = 0 \end{cases}$$

Prove that f is discontinuous at $(0,0)$. Prove that f is continuous along any line passing through the origin.

Proof. We first show that f is discontinuous at $(0,0)$. For $n \in \mathbb{N}$, let $x_n := 2^{-n}$, and let $y_n := x_n^2$. Then $(x_n, y_n) \to (0,0)$ as $n \to \infty$. Then

$$\lim_{n\to\infty} f(x_n, y_n) = e^{-1} \neq 0 = f(0,0).$$

Therefore, $f(x,y)$ is discontinuous at $(0,0)$.

We now show that f is continuous along any line passing through the origin. Since $f(x,y) = 0$ along the line $x = 0$, f is continuous on this line. Now, consider any other line, i.e. let $\lambda \in \mathbb{R}$ and consider the set $A_\lambda := \{(x,y) \in \mathbb{R} \times \mathbb{R} \colon y = \lambda x\}$. Let $(x_n, y_n) \in A_\lambda$ such that $(x_n, y_n) \to (0,0)$ as $n \to \infty$ with $(x_n, y_n) \neq (0,0)$. Then $y_n = \lambda x_n$ for all $n \in \mathbb{N}$, so $y_n/x_n^2 = \lambda x_n^{-1}$. Since $(x_n, y_n) \to (0,0)$ as $n \to \infty$, $x_n \to 0$ as $n \to \infty$, by Exercise 1.6. So,

$$\lim_{n\to\infty} f(x_n, y_n) = \lim_{n\to\infty} \lambda x_n^{-1} e^{-\lambda x_n^{-1}} = \lim_{t\to 0} \lambda t^{-1} e^{-\lambda t^{-1}} = \lim_{r\to\infty} r e^{-r} = 0 = f(0,0).$$

We conclude that f is continuous along any line through the origin. □

Exercise 2.11. Define $f\colon \mathbb{R} \to \mathbb{R}$ such that

$$f(x) = \begin{cases} 0 & \text{, if } x \text{ is irrational} \\ 1/q & \text{, if } x = p/q \text{ with } p \in \mathbb{Z} \text{ and } q \in \mathbb{Z}_{\geq 1} \text{ having no common divisor.} \end{cases}$$

Prove that f is continuous at every irrational point and discontinuous at every rational point.

Proof. We first show that f is discontinuous at every rational point. In Exercise 2.7, we showed that $\mathbb{R} \smallsetminus \mathbb{Q}$ is dense in $\mathbb{R}$. So, given any rational number x, there exists a sequence $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R} \smallsetminus \mathbb{Q}$ such that $x_n \to x$. However, by the definition of f, $f(x_n) = 0$, but $f(x) \neq 0$. Therefore, $\lim_{n\to\infty} f(x_n) = 0 \neq f(x)$, so f is not continuous at x.

We now show that f is continuous at every irrational point. Let $x \in \mathbb{R} \smallsetminus \mathbb{Q}$. Let $\varepsilon > 0$. We need to find $\delta > 0$ such that, if $|y - x| < \delta$, then $|f(x) - f(y)| < \varepsilon$. Since $x \in \mathbb{R} \smallsetminus \mathbb{Q}$, $f(x) = 0$. So, if $y \in \mathbb{R} \smallsetminus \mathbb{Q}$, $f(y) = 0$, and $|f(x) - f(y)| = 0 < \varepsilon$. So, to prove continuity at x, it suffices to only consider $y \in \mathbb{Q}$. Let $q \in \mathbb{Z}_{\geq 1}$. Consider the set

$$A_q := \{x \in \mathbb{Q} \colon \exists\, p \in \mathbb{Z} \text{ such that } x = p/q\}.$$

Then $A_q = \{\ldots, -3/q, -2/q, -1/q, 0, 1/q, 2/q, 3/q, \ldots\}$. Define $f\colon \mathbb{R} \to \mathbb{R}$ so that $f(x) = x/q$. Then $A_q = f(\mathbb{Z})$. Let $R > 2$ so that $|x| < R$, and let $T \in \mathbb{N}$ such that $1/T < \varepsilon$. Then the following set is finite

$$D := \left(\bigcup_{q=1}^{T} A_q\right) \cap B_{2R}(0).$$

Since D is a finite set and $D \subseteq \mathbb{Q}$, we know that $x \notin D$, so there exists a $\delta > 0$ such that $D \cap B_\delta(x) = \emptyset$ and such that $B_\delta(x) \subseteq B_{2R}(0)$. For example, we could take $\delta := \min_{d\in D} |x-d|/2$. In summary, $B_\delta(x) \subseteq (B_{2R}(0) \smallsetminus D)$.

Now, combining the definitions of T, D and f, we have $f(y) \leq \varepsilon$ for all $y \in B_{2R}(0) \smallsetminus D$. Since $B_\delta(x) \subseteq (B_{2R}(0) \smallsetminus D)$, we have $f(y) \leq \varepsilon$ for $y \in B_\delta(x)$. Since $f(x) = 0$, we have shown that, if $|y-x| < \delta$, then $|f(x) - f(y)| = |f(y)| = f(y) < \varepsilon$. That is, we have shown that f is continuous at every irrational point, as desired. □

3. Problem Set 3

Exercise 3.1. Show that the following functions are differentiable, and compute their differentials.

$$(i)\quad f(x,y,z) = \begin{pmatrix} xy^3 \\ z\sin y \\ x^2 - y^2 z \end{pmatrix}, \qquad (ii)\quad f(x,y) = \begin{cases} x^3/\sqrt{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

Proof of (i). To prove that f is differentiable, it suffices to show that each component of f is continuously differentiable. That is, it suffices to show that each partial derivative of each component exists and is continuous. The first the third component of f are polynomials, so these components have continuous partial derivatives. And the second component of f is a product of a polynomial and the infinitely differentiable sin function. So the second component also has continuous partial derivatives. Since each component has continuous partial derivatives, we conclude that f is differentiable. □

Proof of (ii). To prove that f is differentiable on $\mathbb{R}^2 \smallsetminus \{(0,0)\}$, it suffices to show that each partial derivative exists and is continuous on $\mathbb{R}^2 \smallsetminus \{(0,0)\}$. For $(x,y) \neq (0,0)$, f is a composition of continuously differentiable functions, so the partial derivatives of f exist and are continuously differentiable. It therefore remains to show that f is differentiable at $(0,0)$. Let $h = (h_1, h_2) \in \mathbb{R}^2$, $h \neq (0,0)$, let $L(h) := 0$ and observe

$$\frac{|f(0+h) - f(0) - L(h)|}{|h|} = \frac{\left|\frac{h_1^3}{\sqrt{h_1^2+h_2^2}} - L(h)\right|}{|h|} = \frac{|h_1|^3}{h_1^2+h_2^2} \leq \frac{(h_1^2+h_2^2)^{3/2}}{h_1^2+h_2^2} = |h|$$

So, $\lim_{h\to 0} |f(0+h) - f(0) - L(h)| / |h| = 0$, and f is differentiable at $(0,0)$. We have therefore shown that f is differentiable everywhere. □

Exercise 3.2. Suppose $f\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable and homogeneous of degree $\alpha \in \mathbb{R}$, i.e. $f(tx) = t^\alpha f(x)$ for all $x \in \mathbb{R}^n$ and $t > 0$. Prove that $f'(x)x = \alpha f(x)$.

Proof. Let $x \in \mathbb{R}^n$ and let $g\colon \mathbb{R} \to \mathbb{R}^n$ be defined by $g(t) := tx$. Write $g = (g_1, \ldots, g_n)$, so that $g_i(x)\colon \mathbb{R} \to \mathbb{R}$, $i \in \{1,\ldots,n\}$. Specifically, for $x = (x_1,\ldots,x_n) \in \mathbb{R}^n$, we have $g_i(t) = tx_i$. For $i \in \{1,\ldots,n\}$, we have $\partial g_i/\partial t = x_i$, i.e. $g'(t) = x$. So, each partial derivative of each component of g is continuously differentiable. Therefore, g is differentiable. We now apply the chain rule to the composition $f(g(t))$. Observe, $(d/dt)(f(g(t))) = f'(g(t))g'(t) = f'(tx)x$. Then, by differentiating the identity $f(tx) = t^\alpha f(x)$ with respect to t, we get

$$f'(tx)x = \alpha t^{\alpha-1} f(x).$$

Substituting $t = 1$ completes the exercise. □

Exercise 3.3. Prove that a continuously differentiable function on $\mathbb{R}^n$ is Lipschitz continuous on any compact subset of $\mathbb{R}^n$.

Proof. Let A be a compact subset of $\mathbb{R}^n$. Since A is compact, there exists $R > 0$ such that $A \subseteq B_R(0)$. Since f is continuously differentiable, the function $x \mapsto |\nabla f(x)|$ is a composition of continuous functions. So, since $\overline{B_R(0)}$ is compact, there exists $T > 0$ such that $|\nabla f(x)| \leq T$ for all $x \in \overline{B_R(0)}$. Let $x, y \in A$ with $x \neq y$, and let $v := y - x$. Write $f\colon \mathbb{R}^n \to \mathbb{R}$, and define $g\colon [0,1] \to \mathbb{R}$ by $g(t) := f((1-t)x + ty) = f(x + tv)$. Then g is continuously differentiable, since for $\lambda \in [0,1]$,

$$D_v f(x + \lambda v) = \lim_{t\to 0} \frac{f(x + tv + \lambda v) - f(x + \lambda v)}{t} = \lim_{t \to 0} \frac{g(t+\lambda) - g(\lambda)}{t} = g'(\lambda). \qquad (*)$$

By the Mean Value Theorem, there exists $\lambda \in (0,1)$ such that $g(1) - g(0) = g'(\lambda)$. So, from $(*)$, there exists $\lambda \in (0,1)$ such that

$$f(y) - f(x) = D_v f(x + \lambda v).$$

By convexity of $\overline{B_R(0)}$, $x, y \in A \subseteq \overline{B_R(0)}$ implies that $x + \lambda v = \lambda y + (1-\lambda)x \in \overline{B_R(0)}$. So, from Cauchy-Schwarz, the bound $\forall\, z \in \overline{B_R(0)}$, $|\nabla f(z)| \leq T$, and the definition of v,

$$|f(y) - f(x)| = |D_v f(x + \lambda v)| = |\langle \nabla f(x + \lambda v), v\rangle| \leq |\nabla f(x + \lambda v)|\, |v| \leq T\, |v| = T\, |y - x|\,.$$

□

Remark 3.4. It is possible to have a differentiable function on a compact set that is not Lipschitz continuous. Consider $f(x) = x^{3/2}\cos(1/x)$ on $(0,1]$ with $f(0) = 0$. Then f is defined on $[0,1]$, and f is differentiable, but the derivative of f becomes arbitrarily large as $x \to 0$, so one can show that the quantity $|f(x) - f(y)| / |x - y|$ also becomes arbitrarily large, by choosing suitable x, y that converge to 0. Therefore, f is not Lipschitz.

Exercise 3.5.

(i) Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be differentiable, and let γ be a differentiable path such that $f(\gamma(t))$ is constant. Prove that $\nabla f(\gamma(t))$ is orthogonal to $\gamma'(t)$ for all t.

(ii) Suppose $f\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}^n$. The rate of growth of f in the direction $v \in \mathbb{R}^n$ is given by the directional derivative $D_v f(a)$. Show that the direction of maximal growth, i.e. the unit vector v for which $D_v f(a)$ is maximal, is $\nabla f(a) / |\nabla f(a)|$.

(iii) Interpret (i) and (ii) in terms of the following scenario: you are hiking in mountainous terrain, and $f(x, y)$ represents the height of the terrain. If you are in possession of a map that includes contour lines, how should you walk if you want to reach a nearby summit as quickly as possible? Illustrate your argument using a sketch.

Proof of (i). Let $c \in \mathbb{R}$ such that $f(\gamma(t)) = c$. Since f, γ are differentiable, we differentiate the identity $f(\gamma(t)) = c$ with respect to t, and use the chain rule to get

$$\langle \nabla f(\gamma(t)), \gamma'(t)\rangle = 0.$$

□

Proof of (ii). Recall that $D_v f(a) = \langle \nabla f(a), v\rangle$. Let $v \in \mathbb{R}^n$ with $|v| = 1$. From Cauchy-Schwarz,

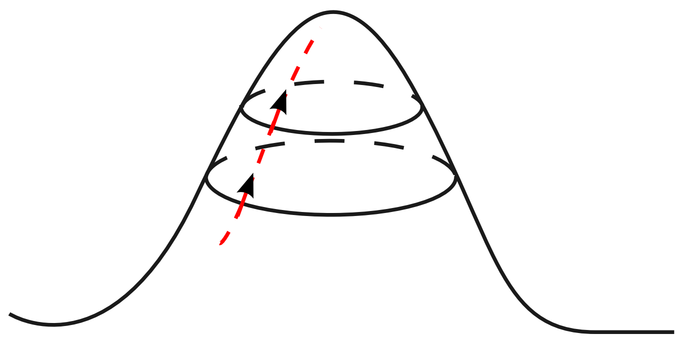
$$|D_v f(a)| = |\langle \nabla f(a), v\rangle| \leq |\nabla f(a)|\,|v| = |\nabla f(a)|\,.$$

It now suffices to find v with $|v| = 1$ such that $D_v f(a) = |\nabla f(a)|$. If $\nabla f(a) = 0$, let v be any unit vector. If $\nabla f(a) \neq 0$, let $v := \nabla f(a)/\,|\nabla f(a)|$. Then

$$\langle \nabla f(a), v\rangle = \langle \nabla f(a), \nabla f(a)\rangle^{1/2} = |\nabla f(a)|\,.$$

□

Proof of (iii). If we want to walk to a nearby summit as quickly as possible we should walk uphill, perpendicular to the level sets of f, and hopefully we will not get stuck at a saddle point. □



Exercise 3.6. Recall that a *path* is a continuous, piecewise continuously differentiable map $\gamma\colon [a,b] \to \mathbb{R}^n$. That is, there is a partition $a = a_0 < a_1 < \cdots < a_k = b$ such that γ is continuous differentiable on (a_i, a_{i+1}) for each $i = 0, \ldots, k-1$. Recall also that a 1-form λ is a continuous map from $\mathbb{R}^n$ to the space of $1 \times n$ matrices (i.e. the dual vectors). The path integral of λ along γ was defined as

$$\int_\gamma \lambda := \int_a^b \lambda(\gamma(t))\gamma'(t)dt.$$

Finally, recall that the reversal of $\gamma\colon [0,1] \to \mathbb{R}^n$ was defined by $(-\gamma)(t) := \gamma(1-t)$, and the join of two paths $\gamma_1, \gamma_2\colon [0,1] \to \mathbb{R}^n$ satisfying $\gamma_1(1) = \gamma_2(0)$ was defined as the path $(\gamma_1 \oplus \gamma_2)\colon [0,2] \to \mathbb{R}^n$ given by

$$(\gamma_1 \oplus \gamma_2)(t) := \begin{cases} \gamma_1(t) & \text{if } t \in [0,1] \\ \gamma_2(t-1) & \text{if } t \in (1,2] \end{cases}.$$

Prove that

$$\int_{-\gamma} \lambda = -\int_\gamma \lambda, \quad \text{and} \quad \int_{\gamma_1 \oplus \gamma_2} \lambda = \int_{\gamma_1} \lambda + \int_{\gamma_2} \lambda.$$

Proof. We first prove that $\int_{-\gamma} \lambda = -\int_\gamma \lambda$. Let $0 = a_0 < a_1 < \cdots < a_k = 1$ be a partition of $[0,1]$ such that γ is continuously differentiable on (a_i, a_{i+1}) for $i = 0, \ldots, k-1$. Since

$(-\gamma)(t) = \gamma(1-t)$, we know that $(-\gamma)$ is continuously differentiable on $(1-a_{i+1}, 1-a_i)$ for $i = 0, \ldots, k-1$. Now,

$$\begin{aligned}\int_{1-a_{i+1}}^{1-a_i} \lambda(\gamma(1-t))(\gamma(1-t))' dt &= \int_{1-a_{i+1}}^{1-a_i} \lambda(\gamma(1-t))(-\gamma'(1-t)) dt \quad \text{, by the chain rule} \\ &= \int_{a_{i+1}}^{a_i} \lambda(\gamma(t))\gamma'(t) dt \quad \text{, changing variables} \\ &= -\int_{a_i}^{a_{i+1}} \lambda(\gamma(t))\gamma'(t) dt.\end{aligned}$$

So,

$$\int_{-\gamma} \lambda = \sum_{i=0}^{k-1} \int_{1-a_{i+1}}^{1-a_i} \lambda(\gamma(1-t))(\gamma(1-t))' dt = -\sum_{i=0}^{k-1} \int_{a_i}^{a_{i+1}} \lambda(\gamma(t))\gamma'(t) dt = -\int_{\gamma} \lambda.$$

We now prove that $\int_{\gamma_1 \oplus \gamma_2} \lambda = \int_{\gamma_1} \lambda + \int_{\gamma_2} \lambda$. Let $0 = a_0 < a_1 < \cdots < a_k = 1$ be a partition of $[0,1]$ such that γ_1 is continuously differentiable on (a_i, a_{i+1}) for $i = 0, \ldots, k-1$. Let $0 = b_0 < b_1 < \cdots < b_j = 1$ be a partition of $[0,1]$ such that γ_2 is continuously differentiable on (b_j, b_{j+1}) for $j = 0, \ldots, \ell-1$. By the definition of $\gamma_1 \oplus \gamma_2$, the curve $\gamma_1 \oplus \gamma_2$ is continuously differentiable on (a_i, a_{i+1}) for $i = 0, \ldots, k-1$, and on $(1+b_j, 1+b_{j+1})$ for $j = 0, \ldots, \ell-1$. Also, by the definition of $\gamma_1 \oplus \gamma_2$

$$\int_{a_i}^{a_{i+1}} \lambda(\gamma_1 \oplus \gamma_2(t))(\gamma_1 \oplus \gamma_2(t))' dt = \int_{a_i}^{a_{i+1}} \lambda(\gamma_1(t))(\gamma_1(t))' dt,$$

$$\int_{1+b_j}^{1+b_{j+1}} \lambda(\gamma_1 \oplus \gamma_2(t))(\gamma_1 \oplus \gamma_2(t))' dt = \int_{b_j}^{b_{j+1}} \lambda(\gamma_2(t))(\gamma_2(t))' dt.$$

Therefore,

$$\begin{aligned}\int_{\gamma_1 \oplus \gamma_2} \lambda &= \sum_{i=0}^{k-1} \int_{a_i}^{a_{i+1}} \lambda(\gamma_1 \oplus \gamma_2(t))(\gamma_1 \oplus \gamma_2(t))' dt + \sum_{j=0}^{\ell-1} \int_{1+b_j}^{1+b_{j+1}} \lambda(\gamma_1 \oplus \gamma_2(t))(\gamma_1 \oplus \gamma_2(t))' dt \\ &= \sum_{i=0}^{k-1} \int_{a_i}^{a_{i+1}} \lambda(\gamma_1(t))(\gamma_1(t))' dt + \sum_{j=0}^{\ell-1} \int_{b_j}^{b_{j+1}} \lambda(\gamma_2(t))(\gamma_2(t))' dt \\ &= \int_{\gamma_1} \lambda + \int_{\gamma_2} \lambda.\end{aligned}$$

□

Exercise 3.7. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be differentiable with $f(0) = 0$. Prove that there exist continuous $g_i\colon \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, n$ such that

$$f(x) = \sum_{i=1}^{n} x_i g_i(x).$$

Proof. Let $x \in \mathbb{R}^n$. For $t \in [0,1]$, define $g(t) := f(tx)$. In Exercise 3.2, we showed that g is differentiable, and the chain rule applies, i.e. $g'(t) = f'(tx)x$. Since f' is continuous, we conclude that g' is continuous. So, g is continuously differentiable, and the Fundamental Theorem of Calculus may be applied to g. Applying this theorem together with $f(0) = 0$,

$$f(x) = f(x) - f(0) = g(1) - g(0) = \int_0^1 g'(t)dt = \int_0^1 \langle f'(tx), x\rangle dt = \sum_{i=1}^n x_i \int_0^1 \frac{\partial f}{\partial x_i}(tx)dt.$$

It remains to show that for each $i = 1, \ldots, n$, the function $h(x) := \int_0^1 \frac{\partial f}{\partial x_i}(tx)dt$ is continuous. For t fixed, each function $\partial f(tx)/\partial x_i$ is continuous as a function of x, so the Riemann sum

$$h_k(x) := \sum_{j=1}^k \frac{1}{k}\frac{\partial f}{\partial x_i}(xj/k) = \sum_{j=1}^k \int_{(j-1)/k}^{j/k} \frac{\partial f}{\partial x_i}(xj/k)dt$$

is a continuous function of x.

Let $A \subseteq \mathbb{R}^n$ be any compact set. To prove that h is continuous, it suffices to show that h_k converges uniformly to h on A. We therefore now show the required uniform convergence. Let $\varepsilon > 0$. The function $H(x,t) := \frac{\partial f}{\partial x_i}(tx)$ is uniformly continuous on $A \times [0,1]$, since it is continuous on the compact set $A \times [0,1]$. In particular, there exists $\delta > 0$ such that, if $|t_1 - t_2| < \delta$, then $\sup_{x\in A} |H(x,t_1) - H(x,t_2)| < \varepsilon$. Now, if $k > 1/\delta$ and $x \in A$, we get

$$\begin{aligned}
|h_k(x) - h(x)| &= \left|\sum_{j=1}^k \int_{(j-1)/k}^{j/k} \left(\frac{\partial f}{\partial x_i}(xj/k) - \frac{\partial f}{\partial x_i}(tx)\right) dt\right| \\
&\leq \sum_{j=1}^k \int_{(j-1)/k}^{j/k} \left|\frac{\partial f}{\partial x_i}(xj/k) - \frac{\partial f}{\partial x_i}(tx)\right| dt \\
&= \sum_{j=1}^k \int_{(j-1)/k}^{j/k} |H(x,j/k) - H(x,t)|\, dt \\
&\leq \sum_{j=1}^k \varepsilon/k \quad \text{, using } 1/k < \delta \\
&= \varepsilon.
\end{aligned}$$

We conclude that h_k converges uniformly to h on A, as desired. □

Exercise 3.8. In this problem we work in $\mathbb{R}^2$. Define the 1-forms

$$\mu(x,y) := (0,x), \quad \nu(x,y) := (-y,0), \quad \lambda(x,y) := (1/2)(-y,x).$$

- (i) Let γ be the path that traces the boundary of a disk of radius r once in the counterclockwise direction. Compute $\int_\gamma \mu$, $\int_\gamma \nu$ and $\int_\gamma \lambda$. Do the same when γ traces the boundary of a rectangle with side lengths a and b. What do you observe?
- (ii) Let γ be an arbitrary closed path. Prove that $\int_\gamma \mu = \int_\gamma \nu = \int_\gamma \lambda$.
- (iii) In general, if γ is a closed path that traces the boundary of an arbitrary region A in the counterclockwise direction, then any of the above integrals gives the area of A.

Proof of (i). Let $\gamma(t) := r(\cos(2\pi t), \sin(2\pi t))$. Then $\gamma'(t) = 2\pi r(-\sin(2\pi t), \cos(2\pi t))$. Observe

$$\int_\gamma \mu = \int_0^1 \langle (0, r\cos(2\pi t)), 2\pi r(-\sin(2\pi t), \cos(2\pi t))\rangle dt = 2\pi r^2 \int_0^1 \cos^2(2\pi t)dt = \pi r^2.$$

$$\int_\gamma \nu = \int_0^1 \langle(-r\sin(2\pi t),0), 2\pi r(-\sin(2\pi t),\cos(2\pi t))\rangle dt = 2\pi r^2 \int_0^1 \sin^2(2\pi t)dt = \pi r^2.$$

$$\begin{aligned}\int_\gamma \lambda &= \int_0^1 \langle(1/2)r(-\sin(2\pi t),\cos(2\pi t)), 2\pi r(-\sin(2\pi t),\cos(2\pi t))\rangle dt \\ &= \pi r^2 \int_0^1 (\sin^2(2\pi t)+\sin^2(2\pi t))dt = \pi r^2.\end{aligned}$$

Now, define

$$\sigma(t) := \begin{cases} (a/2,(b/2)(8t-1)) & \text{, if } t\in[0,1/4] \\ ((a/2)(3-8t),b/2) & \text{, if } t\in[1/4,1/2] \\ (-a/2,(b/2)(5-8t)) & \text{, if } t\in[1/2,3/4] \\ ((a/2)(8t-7),-b/2) & \text{, if } t\in[3/4,1] \end{cases}.$$

Then

$$\begin{aligned}\int_\sigma \mu &= \int_0^{1/4}\langle(0,a/2),(0,4b)\rangle dt + \int_{1/4}^{1/2}\langle(0,(a/2)(3-8t)),(-4a,0)\rangle dt \\ &\quad + \int_{1/2}^{3/4}\langle(0,-a/2),(0,-4b)\rangle dt + \int_{3/4}^{1}\langle(0,(a/2)(8t-7)),(4a,0)\rangle dt \\ &= (1/4)(2ab)+0+(1/4)(2ab)+0 = ab.\end{aligned}$$

$$\begin{aligned}\int_\sigma \nu &= \int_0^{1/4}\langle(-(b/2)(8t-1),0),(0,4b)\rangle dt + \int_{1/4}^{1/2}\langle(-b/2,0),(-4a,0)\rangle dt \\ &\quad + \int_{1/2}^{3/4}\langle(-(b/2)(5-8t),0),(0,-4b)\rangle dt + \int_{3/4}^{1}\langle(b/2,0),(4a,0)\rangle dt \\ &= 0+(1/4)2ab+0+(1/4)2ab = ab.\end{aligned}$$

$$\begin{aligned}\int_\sigma \lambda &= \frac{1}{2}\int_0^{1/4}\langle(-(b/2)(8t-1),a/2),(0,4b)\rangle dt + \frac{1}{2}\int_{1/4}^{1/2}\langle(-b/2,(a/2)(3-8t)),(-4a,0)\rangle dt \\ &\quad + \frac{1}{2}\int_{1/2}^{3/4}\langle(-b/2(5-8t),-a/2),(0,-4b)\rangle dt + \frac{1}{2}\int_{3/4}^{1}\langle(b/2,(a/2)(8t-7)),(4a,0)\rangle dt \\ &= (1/8)(2ab)+(1/8)(2ab)+(1/8)(2ab)+(1/8)(2ab) = ab.\end{aligned}$$

So, it seems that the integral over the boundary of μ, ν, or λ always gives the area of the enclosed figure. □

Proof of (ii). Define $f(x,y) := xy$. Then f is continuously differentiable, and $df = (y,x)$, so $\mu = \nu + df$. Let γ be an arbitrary closed path. From the Fundamental Theorem of Calculus, $\int_\gamma df = 0$. So, using the equality $\mu = \nu + df$, we have

$$\int_\gamma \mu = \int_\gamma(\nu+df) = \int_\gamma \nu + \int_\gamma df = \int_\gamma \nu. \qquad (*)$$

For the second equality, note that $\lambda = (1/2)(\mu + \nu)$, so using this equality and $(*)$

$$\int_\gamma \lambda = (1/2)\int_\gamma \mu + (1/2)\int_\gamma \nu = \int_\gamma \nu.$$

$\square$

Proof of (iii). Suppose we have a triangle T in $\mathbb{R}^2$ with vertices $(0,0)$, (x,y) and $(x+\Delta x, y+\Delta y)$. Also, assume that the vectors (x,y) and $(x+\Delta x, y+\Delta y)$ have a cross product with positive z-component. That is, assume that $(x,y,0)\times(x+\Delta x, y+\Delta y, 0)$ has a positive z component. Then from the cross product formula for the area of a triangle, we have

$$\begin{aligned}\mathrm{Area}(T) &= \frac{1}{2}\left|(x,y,0)\times(x+\Delta x, y+\Delta y, 0)\right| = \frac{1}{2}\left[x(y+\Delta y) - y(x+\Delta x)\right]\\ &= \frac{1}{2}(x\Delta y - y\Delta x). \qquad (*)\end{aligned}$$

Let $\gamma\colon [0,1]\to\mathbb{R}^2$ be a parametrization of the boundary ∂A of A. Since γ is piecewise continuously differentiable, the function $t\mapsto|\gamma'(t)|$ is a piecewise composition of continuous functions, so it is piecewise continuously differentiable. Since $[0,1]$ is compact, we conclude that there exists $C<\infty$ such that $|\gamma'(t)|\leq C$ for $t\in[0,1]$. Also, since γ is continuous, there exists $D<\infty$ such that $|\gamma(t)|\leq D$ for $t\in[0,1]$.

Let $a_0 = 0 < a_1 < \cdots < a_k = 1$ so that γ is continuously differentiable on (a_i, a_{i+1}) for $i = 0,\ldots,k-1$. Fix $i\in\{0,\ldots,k-1\}$, and fix $t\in[a_i,a_{i+1})$. Let $j\in\{0,\ldots,N\}$, and define $t_j := a_i + (a_{i+1}-a_i)(j/N)$. Then

$$\int_{\gamma(a_i,a_{i+1})}\lambda = \int_{a_i}^{a_{i+1}}\lambda(\gamma(t))\gamma'(t)dt = \sum_{j=0}^{N}\int_{t_j}^{t_{j+1}}\frac{1}{2}(-\gamma_2(t)\gamma_1'(t)+\gamma_1(t)\gamma_2'(t))dt. \qquad (**)$$

Let T_j be the triangle formed by 0, $\gamma(t_j)$ and $\gamma(t_{j+1})$. Since γ travels in the counterclockwise direction, the z component of the cross product $(\gamma(t_j),0)\times(\gamma(t_{j+1}),0)$ is positive. Then

$$\begin{aligned}&\int_{t_j}^{t_{j+1}}\frac{1}{2}(-\gamma_2(t)\gamma_1'(t)+\gamma_1(t)\gamma_2'(t))dt\\ &= \int_{t_j}^{t_{j+1}}\frac{1}{2}(-\gamma_2(t_j)\gamma_1'(t)+\gamma_1(t_j)\gamma_2'(t))dt + O(C(t_{j+1}-t_j)\left|\gamma(t_{j+1})-\gamma(t_j)\right|)\\ &= \frac{1}{2}\left[-\gamma_2(t_j)(\gamma_1(t_{j+1})-\gamma_1(t_j)) + \gamma_1(t_j)(\gamma_2(t_{j+1})-\gamma_2(t_j))\right]\\ &\qquad + O(C(t_{j+1}-t_j)\left|\gamma(t_{j+1})-\gamma(t_j)\right|) \quad \text{, by the Fundamental Theorem of Calculus}\\ &= \mathrm{Area}(T_j) + O(C(t_{j+1}-t_j)\left|\gamma(t_{j+1})-\gamma(t_j)\right|) \quad \text{, from } (*)\\ &= \mathrm{Area}(T_j) + O(C^2(t_{j+1}-t_j)^2) \quad \text{, by the Mean Value Theorem}\\ &= \mathrm{Area}(T_j) + O(C^2(a_{i+1}-a_i)^2/N^2) \quad \text{, by the definition of } t_j\end{aligned}$$

Then, from $(**)$,

$$\int_{\gamma(a_i,a_{i+1})}\lambda = \left(\sum_{j=0}^{N}\mathrm{Area}(T_j)\right) + O(C^2(a_{i+1}-a_i)^2/N) \qquad (***).$$

Using $(***)$, we will be done if we can show that $\sum_{j=0}^{N} \text{Area}(T_j)$ converges as $N \to \infty$ to the area of A between $\gamma(a_i)$ and $\gamma(a_{i+1})$. Let A_i denote the sectorial region of A between $\gamma(a_i)$ and $\gamma(a_{i+1})$, and let A_{ij} denote the sectorial region of A between $\gamma(t_j)$ and $\gamma(t_{j+1})$. Since $|\gamma'| \le C$, if $s, t \in (t_j, t_{j+1})$, then $|\gamma(s) - \gamma(t)| \le |s - t| C$. So, since γ travels in the counter-clockwise direction, A_{ij} is contained in the triangle with vertices

$$0,\ (1 + C(t_{j+1} - t_j)/\left|\gamma(t_j)\right|)\gamma(t_j),\ \text{and}\ (1 + C(t_{j+1} - t_j)/\left|\gamma(t_{j+1})\right|)\gamma(t_{j+1}).$$

Also, A_{ij} contains the triangle with vertices

$$0,\ \min(1 - C(t_{j+1} - t_j)/\left|\gamma(t_j)\right|, 0)\gamma(t_j),\ \text{and}\ \min(1 - C(t_{j+1} - t_j)/\left|\gamma(t_{j+1})\right|, 0)\gamma(t_{j+1}).$$

Therefore, the error between the area of A_{ij} and the area of T_j is given by the area of a sector of arc length $|\gamma(t_{j+1}) - \gamma(t_j)|$ and radius varying between $|\gamma(t_j)| + C(t_{j+1} - t_j)$ and $|\gamma(t_j)| - C(t_{j+1} - t_j)$. That is,

$$\begin{aligned}\text{Area}(T_j) - O(DC(t_{j+1} - t_j)\left|\gamma(t_{j+1}) - \gamma(t_j)\right|) &\le \text{Area}(A_{ij}) \\ &\le \text{Area}(T_j) + O(DC(t_{j+1} - t_j)\left|\gamma(t_{j+1}) - \gamma(t_j)\right|).\end{aligned}$$

Applying the Mean Value Theorem and the definition of the t_j,

$$\text{Area}(T_j) - O(C^2 D/N^2) \le \text{Area}(A_{ij}) \le \text{Area}(T_j) + O(C^2 D/N^2).$$

Summing over $j = 1, \ldots, N$, we have

$$\sum_{j=0}^{N} \text{Area}(T_j) = \text{Area}(A_i) + O(CD/N).$$

Combining this equality with $(***)$ and letting $N \to \infty$ finishes the proof. $\square$

Proof of (iii) that we may learn later. From Stokes' Theorem and the arithmetic of differential forms,

$$\begin{aligned}\int_\gamma \lambda = \int_{\partial A} \lambda = \int_A d\lambda = \int_A \frac{1}{2} d(-ydx + xdy) &= \frac{1}{2}\int_A -dy \wedge dx + dx \wedge dy \\ = \frac{1}{2}\int_A dx \wedge dy + dx \wedge dy &= \int_A dxdy = \text{Area}(A)\end{aligned}$$

$\square$

Exercise 3.9. In $\mathbb{R}^3$ it is convenient to use spherical coordinates $(r, \theta, \phi) \in [0, \infty) \times [0, \pi] \times [0, 2\pi)$. The coordinate map is $(x, y, z)^T = T(r, \theta, \phi)$ where

$$T(r, \theta, \phi) := \begin{pmatrix} r \sin\theta \cos\phi \\ r \sin\theta \sin\phi \\ r \cos\theta \end{pmatrix}.$$

(i) Give a geometric interpretation of the parameters r, θ and ϕ.
(ii) Compute T'. Show that the columns of T' are orthogonal. Interpret this result geometrically using a sketch.
(iii) Let f be differentiable on $\mathbb{R}^3$ and define $g := f \circ T$. The function g represents the function f expressed in spherical coordinates. Compute all partial derivatives of g in terms of the partial derivatives of f. Find $\partial g/\partial r$ and $\partial g/\partial \theta$ for the functions $f(x, y, z) = x^2 + y^2 + z^2$ and $f(x, y, z) = x - y$.

Proof of (i). The parameter r is the distance of (x, y, z) from the origin, since $x^2+y^2+z^2 = r^2$. The parameter θ is the angle between (x, y, z) and $(0, 0, 1)$, since $x^2 + y^2 = r^2 \sin^2\theta$ and $z^2 = r^2 \cos^2\theta$. The parameter ϕ is the angle between $(x, y, 0)$ and $(1, 0, 0)$, since z does not depend on ϕ and $(x, y, 0)/|(x, y, 0)| = (\cos\phi, \sin\phi, 0)$, for $r \neq 0$. □

Proof of (ii). Write $T(r, \theta, \phi) = (T_1, T_2, T_3)$ with $T_i \colon \mathbb{R}^3 \to \mathbb{R}$ for $i = 1, 2, 3$. Then

$$T' = \begin{pmatrix} \frac{\partial T_1}{\partial r} & \frac{\partial T_1}{\partial \theta} & \frac{\partial T_1}{\partial \phi} \\ \frac{\partial T_2}{\partial r} & \frac{\partial T_2}{\partial \theta} & \frac{\partial T_2}{\partial \phi} \\ \frac{\partial T_3}{\partial r} & \frac{\partial T_3}{\partial \theta} & \frac{\partial T_3}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\phi & r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ \sin\theta\sin\phi & r\cos\theta\sin\phi & r\sin\theta\cos\phi \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix}.$$

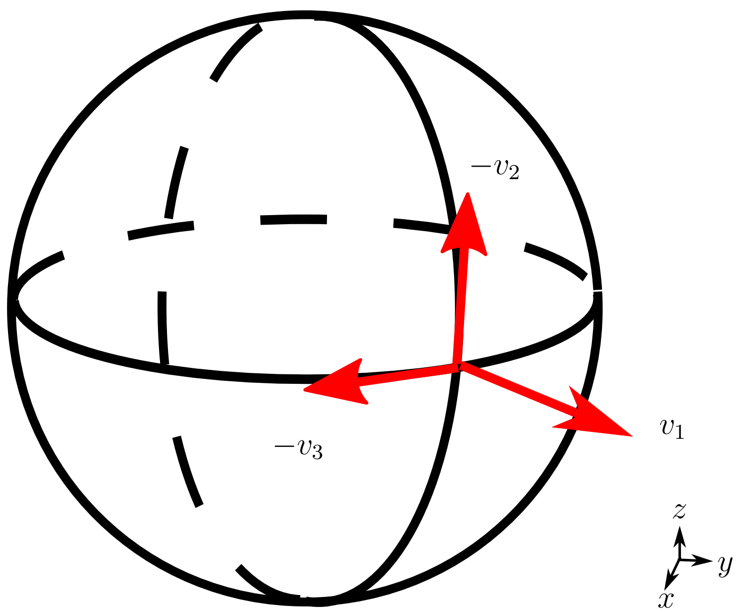
Write $T' = (v_1, v_2, v_3)$, so that v_i is the i^{th} column of T' for $i = 1, 2, 3$. Then

$$\langle v_1, v_2 \rangle = r\cos\theta\sin\theta(\cos^2\phi + \sin^2\phi) - r\cos\theta\sin\theta = 0.$$

$$\langle v_1, v_3 \rangle = -r\sin^2\theta\sin\phi\cos\phi + r\sin^2\theta\sin\phi\cos\phi = 0.$$

$$\langle v_2, v_3 \rangle = -r^2\sin\theta\cos\theta\sin\phi\cos\phi + r^2\sin\theta\cos\theta\sin\phi\cos\phi = 0.$$

So, the columns of T' are orthogonal.



□

Proof of (iii). Write $f \colon \mathbb{R}^3 \to \mathbb{R}$, $T \colon [0, \infty) \times [0, \pi] \times [0, 2\pi) \to \mathbb{R}^3$, so that $f \circ T \colon [0, \infty) \times [0, \pi] \times [0, 2\pi) \to \mathbb{R}$. From the chain rule,

$$g'(r, \theta, \phi) = f'(T(r, \theta, \phi))T'(r, \theta, \phi) = \left(\frac{\partial f}{\partial x_1}(T(r, \theta, \phi)), \frac{\partial f}{\partial x_2}(T(r, \theta, \phi)), \frac{\partial f}{\partial x_3}(T(r, \theta, \phi))\right) T'.$$

So,

$$\frac{\partial g}{\partial r} = \langle f', v_1 \rangle = \frac{\partial f}{\partial x_1}(T(r, \theta, \phi))\sin\theta\cos\phi + \frac{\partial f}{\partial x_2}(T(r, \theta, \phi))\sin\theta\sin\phi + \frac{\partial f}{\partial x_3}(T(r, \theta, \phi))\cos\theta.$$

$$\frac{\partial g}{\partial \theta} = \langle f', v_2 \rangle = \frac{\partial f}{\partial x_1}(T(r,\theta,\phi))r\cos\theta\cos\phi + \frac{\partial f}{\partial x_2}(T(r,\theta,\phi))r\cos\theta\sin\phi$$
$$+ \frac{\partial f}{\partial x_3}(T(r,\theta,\phi))(-r\sin\theta).$$
$$\frac{\partial g}{\partial \phi} = \langle f', v_3 \rangle = \frac{\partial f}{\partial x_1}(T(r,\theta,\phi))(-r\sin\theta\sin\phi) + \frac{\partial f}{\partial x_2}(T(r,\theta,\phi))r\sin\theta\cos\phi.$$

Let $f(x,y,z) = x^2 + y^2 + z^2$. From our above formulas,

$$\frac{\partial g}{\partial r} = 2(r\sin\theta\cos\phi)\sin\theta\cos\phi + 2(r\sin\theta\sin\phi)\sin\theta\sin\phi + 2(r\cos\theta)\cos\theta$$
$$= 2r\sin^2\theta + 2r\cos^2\theta = 2r.$$

$$\frac{\partial g}{\partial \theta} = 2(r\sin\theta\cos\phi)r\cos\theta\cos\phi + 2(r\sin\theta\sin\phi)r\cos\theta\sin\phi + 2(r\cos\theta)(-r\sin\theta)$$
$$= 2r^2\sin\theta\cos\theta - 2r^2\sin\theta\cos\theta = 0.$$

Now, let $f(x,y,z) = x - y$. From our above formulas,

$$\frac{\partial g}{\partial r} = \sin\theta\cos\phi + (-1)\sin\theta\sin\phi + 0 = \sin\theta(\cos\phi - \sin\phi).$$
$$\frac{\partial g}{\partial \theta} = r\cos\theta\cos\phi + (-1)r\cos\theta\sin\phi + 0 = r\cos\theta(\cos\phi - \sin\phi).$$

□

Exercise 3.10. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a polynomial in n variables, i.e. there exist $a_{k_1,\ldots,k_n} \in \mathbb{R}$ such that

$$f(x_1,\ldots,x_n) = \sum_{k_1,\ldots,k_n=0}^{m} a_{k_1,\ldots,k_n} x_1^{k_1} \cdots x_n^{k_n}.$$

Prove that f is differentiable.

Proof. It suffices to prove that each partial derivative of f exists and is continuous. Since f is a finite sum of monomials, it then suffices to show that a monomial has continuous partial derivatives. However, the partial derivative of a monomial is another monomial. Since a monomial is continuous, we have therefore shown that any monomial is continuously differentiable, as desired. □

Exercise 3.11. Let $U := \{A \in \mathbb{R}^{n\times n} \colon A \text{ is an invertible matrix}\}$.

(i) Show that U is an open subset of $\mathbb{R}^{n\times n}$.
(ii) Prove that the map $f\colon U \to U$ defined by $f(A) := A^{-1}$ is differentiable with

$$Df_A(B) = -A^{-1}BA^{-1}.$$

Proof of (i). Write the matrix A as $A = (a_{ij})_{1\le i,j\le n}$. Let S_n be the set of permutations on n elements, and let $\mathrm{sign}(\sigma)$ denote the sign of a permutation $\sigma \in S_n$. That is, if we write σ as a composition of m transpositions, then $\mathrm{sign}(\sigma) = (-1)^m$. By the definition of the determinant,

$$\det(A) = \sum_{\sigma\in S_n} \mathrm{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

That is, det is a polynomial in the entries of A. So, det is a continuous function on $\mathbb{R}^{n\times n}$ to $\mathbb{R}$. And if V is open, then $\det^{-1}(V)$ is open. Let $V := (-\infty,0)\cup(0,\infty)$. Then V is open, and $\det^{-1}(V)=U$ is open, as desired. $\square$

Proof of (ii). Let $A \in U$. Let G_{ij} denote the $(n-1)\times(n-1)$ matrix formed by removing the i^{th} row and j^{th} column from A. Let $C=(c_{ij})_{1\leq i,j\leq n}$ be defined by $c_{ij} := (\det(A))^{-1}(-1)^{i+j}\det(G_{ij})$. Cramer's rule says that $C=f(A)=A^{-1}$. That is, each component of C is a polynomial in A, divided by a polynomial in A, i.e. C is a rational function of A. Since f is defined where $\det(A)\neq 0$, we conclude that f is differentiable on A.

Let $A\in U$. From (i), there exists $\varepsilon>0$ such that $B_\varepsilon(A)\subseteq U$. Let $B'\in B_\varepsilon(A)=\{B\in\mathbb{R}^{n\times n}\colon \sum_{1\leq i,j\leq n}|a_{ij}-b_{ij}|^2<\varepsilon^2\}$. Let B be any $n\times n$ matrix. Then $|A-(A+tB)|=t\,|B|$, so there exists $\delta>0$ such that if $t\in[0,\delta]$, then $A+tB\in B_\varepsilon(A)\subseteq U$, i.e. $A+tB$ is invertible and in the domain of f. Now, $(A+tB)f(A+tB)=1$ for all $t\in[0,\delta]$. So, applying the product rule to this identity,

$$0=\frac{d}{dt}|_{t=0}[(A+tB)f(A+tB)] = A\frac{d}{dt}|_{t=0}f(A+tB)+Bf(A).$$

That is, $A(d/dt)|_{t=0}f(A+tB)=-Bf(A)=-BA^{-1}$, so $(d/dt)|_{t=0}f(A+tB)=-A^{-1}BA^{-1}$. $\square$

Exercise 3.12.

(i) Prove that $\det\colon\mathbb{R}^{n\times n}$ is differentiable. If A is invertible, prove that the differential of det at A is given by

$$D\det{}_A(B)=\mathrm{Tr}(A^{-1}B)\det(A).$$

(ii) For any $n\times n$ matrix A we define the exponential by the formula

$$\exp(A):=\sum_{k=0}^{\infty}\frac{1}{k!}A^k.$$

Prove that this series converges absolutely, componentwise.

(iii) Show that $\exp(tA)\exp(sA)=\exp((t+s)A)$.

(iv) Prove that $\exp(At)$ is differentiable in t with derivative

$$\frac{d}{dt}\exp(At)=A\exp(At).$$

(v) By (iii), $\exp(tA)$ is invertible for all t, since $\exp(-tA)$ is then the inverse of $\exp(tA)$. Use (i) to show that

$$\frac{d}{dt}\det(\exp(tA))=\mathrm{Tr}(A)\det(\exp(tA)).$$

Solve the differential equation to conclude that

$$\det(\exp(A))=\exp(\mathrm{Tr}A).$$

This problem is an example of how analysis can be used to derive identities in linear algebra.

(vi) It is not hard to see that $\exp\colon\mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n}$ is differentiable. Find its differential.

Proof of (i). As shown in Exercise 3.11(i), $\det(A)$ is a polynomial in the entries of A. So, from Exercise 3.10, det is differentiable. We now compute the derivative of $\det(A)$ at $A = 1$. We begin with a claim. Let $\sigma \in S_n$ be a permutation such that $\sigma \neq id$. Then there exist $k, \ell \in \{1, \ldots, n\}$ with $k \neq \ell$ such that $\sigma(k) \neq k$ and $\sigma(\ell) \neq \ell$.

We prove this claim by contradiction. Assume that the claim does not hold. Then there is a $\sigma \in S_n$, $\sigma \neq id$ and at most one $k \in \{1, \ldots, n\}$ such that $\sigma(k) \neq k$. But then $\sigma(i) = i$ for $i \in \{1, \ldots, n\} \smallsetminus \{k\}$, and since σ is a bijection, $\sigma(k) \notin \{1, \ldots, n\} \smallsetminus \{k\}$, i.e. $\sigma(k) = k$. Since we have achieved a contradiction, we conclude that our claim holds.

We now begin to prove the exercise. Let $B = (b_{ij})_{1 \leq i,j \leq n}$ be an $n \times n$ matrix. Then the matrix $1 + tb$ has diagonal entries $1 + tb_{ii}$, $i = 1, \ldots, n$, and all other entries are first degree monomials in t. So, when we take the determinant of $1 + tb$, the term given by the diagonal is $\prod_{i=1}^{n}(1 + tb_{ii})$, whereas all other terms in the sum definition of the determinant have a t term of degree at least 2. Specifically, we have

$$\begin{aligned}\det(1 + tB) &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} (1 + tB)_{i\sigma(i)} \\ &= \prod_{i=1}^{n} (1 + tb_{ii}) + \sum_{\sigma \in S_n, \sigma \neq id} \operatorname{sign}(\sigma) \prod_{i=1}^{n} (1 + tB)_{i\sigma(i)}\end{aligned}$$

Now, if $\sigma \neq id$, we use the Claim to write

$$\begin{aligned}\prod_{i=1}^{n} (1 + tB)_{i\sigma(i)} &= (1 + tB)_{k\sigma(k)} (1 + tB)_{\ell\sigma(\ell)} \prod_{i \neq k,\ell} (1 + tB)_{i\sigma(i)} \\ &= tB_{k\sigma(k)} tB_{\ell\sigma(\ell)} \prod_{i \neq k,\ell} (1 + tB)_{i\sigma(i)} = O(t^2).\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{d}{dt}|_{t=0} \det(1 + tB) &= \frac{d}{dt}|_{t=0} \left(\prod_{i=1}^{n} (1 + tb_{ii}) \right) = \sum_{i=1}^{n} b_{ii} \left(\prod_{j \neq i} (1 + tb_{jj})|_{t=0} \right) \\ &= \sum_{i=1}^{n} b_{ii} = \operatorname{Tr}(B). \qquad (*)\end{aligned}$$

Now, let A be any matrix. Using $(*)$, we get

$$\frac{d}{dt} \det(A + tB) = \frac{d}{dt} \det((AA^{-1})(A + tB)) = \det(A) \frac{d}{dt} \det(1 + tA^{-1}B) = \det(A)\operatorname{Tr}(A^{-1}B).$$

□

Proof of (ii). Write $A = (a_{ij})_{1 \leq i,j \leq n}$, and write $A^k = ((a^k)_{ij})_{1 \leq i,j \leq n}$ for $k \in \mathbb{N}$. Let $M := \max_{1 \leq i,j \leq n} |a_{ij}|$. Fix $k \in \mathbb{N}$, $1 \leq i, j \leq n$. We prove by induction on k that $\left|(a^k)_{ij}\right| \leq M^k n^{k-1}$. The case $k = 1$ follows by the definition of M, so we now prove the inductive step. Assume that $\left|(a^k)_{ij}\right| \leq M^k n^{k-1}$. Then $A^{k+1} = A^k A$, and by the definition of the matrix product,

$$(a^{k+1})_{ij} = \sum_{r=1}^{n} (a^k)_{ir} a_{rj}.$$

So, from the triangle inequality, the inductive hypothesis, and the definition of M,

$$\left|(a^{k+1})_{ij}\right| \leq \sum_{r=1}^{n} \left|(a^k)_{ir}\right| |a_{rj}| \leq nM^k n^{k-1} M = M^{k+1} n^k.$$

Since the inductive hypothesis is complete, we conclude that, for all $k \in N$, and for all $1 \leq i, j \leq n$,

$$\left|(a^k)_{ij}\right| \leq M^k n^{k-1}.$$

Let $\exp(A)_{ij}$ denote the i, j component of $\exp(A)$. Using the definition of $\exp(A)$, we then have by the triangle inequality and our inductive bound

$$|\exp(A)_{ij}| = \left|\sum_{k=0}^{\infty} \frac{1}{k!}(a^k)_{ij}\right| \leq \sum_{k=0}^{\infty} \frac{1}{k!}\left|(a^k)_{ij}\right| \leq \sum_{k=0}^{\infty} \frac{1}{k!} M^k n^{k-1} = e^{Mn}/n < \infty.$$

We conclude that $\exp(A)$ converges absolutely. □

Proof of (iii). Since the series defining $\exp(tA)$ converges absolutely, we can rearrange terms in its infinite summation.

$$\begin{aligned}
\exp(tA)\exp(sA) &= \sum_{r=0}^{\infty} \frac{1}{r!} t^r A^r \sum_{j=0}^{\infty} \frac{1}{j!} s^j A^j = \sum_{\ell=0}^{\infty} \sum_{r+j=\ell} \frac{1}{r!}\frac{1}{j!} t^r s^j A^\ell \\
&= \sum_{\ell=0}^{\infty} \sum_{k=0}^{\ell} \frac{1}{k!} \frac{1}{(\ell-k)!} t^k s^{\ell-k} A^\ell = \sum_{\ell=0}^{\infty} \frac{1}{\ell!} \sum_{k=0}^{\ell} t^k s^{\ell-k} \frac{\ell!}{k!(\ell-k)!} A^\ell \\
&= \sum_{\ell=0}^{\infty} \frac{1}{\ell!} \sum_{k=0}^{\ell} t^k s^{\ell-k} \binom{\ell}{k} A^\ell = \sum_{\ell=0}^{\infty} \frac{1}{\ell!}(t+s)^\ell A^\ell \\
&= \exp((t+s)A)
\end{aligned}$$

□

Proof of (iv). From part (ii), we showed that the componentwise radius of convergence of $\exp(At)$ is infinite. Therefore, the differential of the map $t \mapsto \exp(tA)$ can be found by differentiating the infinite series definition of $\exp(tA)$ term by term. That is,

$$\frac{d}{dt}\exp(tA) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d}{dt}(t^k A^k) = \sum_{k=0}^{\infty} \frac{1}{k!} k t^{k-1} A^k = A \sum_{k=1}^{\infty} \frac{1}{(k-1)!} (tA)^{k-1} = A \exp(tA).$$

□

Proof of (v). Since det and exp are differentiable, the composition $t \mapsto \det(\exp(tA))$ is differentiable. Then, applying the chain rule, (iii), and then (i),

$$\begin{aligned}
&\frac{d}{dt}\det(\exp(tA)) = \det{}'(\exp(tA))\left(\frac{d}{dt}\exp(tA)\right) = \det{}'(\exp(tA))(A\exp(tA)) \\
&\quad = D_{\exp(tA)}\det(A\exp(tA)) = \operatorname{Tr}(\exp(-tA)A\exp(tA))\det(\exp(tA)) = \operatorname{Tr}(A)\det(\exp(tA)).
\end{aligned}$$

Re-writing this expression,

$$\frac{d}{dt}\log(\det(\exp(tA))) = \operatorname{Tr}(A).$$

Integrating t from 0 to 1 and applying the Fundamental Theorem of Calculus, we get $\log(\det(\exp(A))) = \mathrm{Tr}(A)$. That is,

$$\det(\exp(A)) = \exp(\mathrm{Tr}(A)).$$

$\square$

Proof of (vi). Let $k \in \mathbb{N}$ and let $A, B \in \mathbb{R}^{n\times n}$. Consider $(A+tB)^k$ as a polynomial in t. The first degree term is of the form $\sum_{j=0}^{k-1} A^j B A^{(k-1)-j}$. Therefore,

$$\frac{d}{dt}|_{t=0}(A+tB)^k = \frac{d}{dt}|_{t=0}\left(t\sum_{j=0}^{k-1} A^j B A^{(k-1)-j}\right) = \sum_{j=0}^{k-1} A^j B A^{(k-1)-j}.$$

So, justifying the term by term differentiation as in (iv),

$$\begin{aligned}\frac{d}{dt}|_{t=0}\exp(A+tB) &= \frac{d}{dt}|_{t=0}\sum_{k=0}^{\infty}\frac{1}{k!}(A+tB)^k = \sum_{k=0}^{\infty}\frac{1}{k!}\frac{d}{dt}|_{t=0}(A+tB)^k \\ &= \sum_{k=1}^{\infty}\frac{1}{k!}\sum_{j=0}^{k-1} A^j B A^{(k-1)-j} = \sum_{k=0}^{\infty}\frac{1}{(k+1)!}\sum_{j=0}^{k} A^j B A^{k-j}. \qquad (*)\end{aligned}$$

We now wish to present the differential in a more simplified form. We begin with the following claim. For $j, \ell \in \mathbb{N}$ we have

$$\int_0^1 t^j(1-t)^\ell dt = \frac{j!\ell!}{(j+\ell+1)!}. \qquad (**)$$

To prove $(**)$, integrate by parts ℓ times to get

$$\int_0^1 t^j(1-t)^\ell dt = \frac{\ell}{j+1}\int_0^1 t^{j+1}(1-t)^{\ell-1}dt = \cdots = \frac{\ell!j!}{(j+\ell)!}\int_0^1 t^{j+\ell+1}dt = \frac{\ell!j!}{(j+\ell+1)!}.$$

From $(*)$, and $(**)$, and using absolute convergence to justify the rearrangement of terms,

$$\begin{aligned}\frac{d}{dt}|_{t=0}\exp(A+tB) &= \sum_{k=0}^{\infty}\sum_{j+\ell=k}\frac{1}{(j+\ell+1)!}A^j B A^\ell \\ &= \sum_{k=0}^{\infty}\sum_{j+\ell=k}\int_0^1\frac{1}{j!\ell!}t^j A^j B(1-t)^\ell A^\ell dt \\ &= \int_0^1\left(\sum_{j=0}^{\infty}\frac{1}{j!}t^j A^j\right)B\left(\sum_{\ell=0}^{\infty}\frac{1}{\ell!}(1-t)^\ell A^\ell\right)dt \\ &= \int_0^1 e^{At}Be^{A(1-t)}dt.\end{aligned}$$

$\square$

4. Problem Set 4

Exercise 4.1. Let f be C^{k+1} in a neighborhood of $a \in \mathbb{R}^n$. In class, we saw that f can be expressed using its Taylor series as

$$f(a+h) = P_k(h) + R_k(h),$$

where

$$P_k(h) := \sum_{r=0}^{k} \frac{1}{r!} D_h^r f(a), \qquad R_k(h) := \frac{1}{(k+1)!} D_h^{k+1} f(\xi).$$

Here ξ is in the segment between a and $a+h$.

(i) Find the Taylor polynomial of degree 2 (i.e. $P_2(h)$) for the function $f(x,y) = \sin(xy)e^{x^2}$ at $a = (0,0)$.

(ii) Find the Taylor polynomial of degree 3 (i.e. $P_3(h)$) for the function $f(x,y) = x^y$ at $a = (1,1)$.

(iii) Find the Taylor series (i.e. $\lim_{k\to\infty} P_k(h)$) for the function $f(x,y) = \sqrt{1+x^2+y^2}$ at $a = (0,0)$.

Proof of (i). Let $x, y \in \mathbb{R}$. Then

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

$$\frac{\partial^2 f}{\partial x^2} = y\cos(xy)(2x)e^{x^2} + \sin(xy)2e^{x^2} + \sin(xy)4x^2e^{x^2} - y^2\sin(xy)e^{x^2} + y\cos(xy)(2x)e^{x^2},$$

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

Note that $f \in C^2$. Let $h = (h_1, h_2) \in \mathbb{R}^2$. Then, at $(x,y) = (0,0)$,

$$\begin{aligned} P_2(h) &= f(0,0) + h_1\frac{\partial f}{\partial x}(0,0) + h_2\frac{\partial f}{\partial y}(0,0) \\ &\quad + \frac{1}{2}h_1^2\frac{\partial^2 f}{\partial x^2}(0,0) + \frac{1}{2}h_2^2\frac{\partial^2 f}{\partial y^2}(0,0) + h_1h_2\frac{\partial^2 f}{\partial x\partial y}(0,0) \\ &= h_1h_2. \end{aligned}$$

□

Proof of (ii). Let $x, y > 0$. Then

$$\frac{\partial f}{\partial x} = yx^{y-1}, \qquad \frac{\partial f}{\partial y} = x^y\log(x),$$

$$\frac{\partial^2 f}{\partial x^2} = y(y-1)x^{y-2}, \qquad \frac{\partial^2 f}{\partial y^2} = x^y(\log(x))^2, \qquad \frac{\partial^2 f}{\partial x\partial y} = yx^{y-1}\log(x) + x^{y-1},$$

$$\frac{\partial^3 f}{\partial x^3} = y(y-1)(y-2)x^{y-3}, \qquad \frac{\partial^3 f}{\partial y \partial x^2} = (y-1)x^{y-2} + yx^{y-2} + y(y-1)x^{y-2}\log(x),$$

$$\frac{\partial^3 f}{\partial x\partial y^2} = yx^{y-1}(\log(x))^2 + 2x^{y-1}\log(x), \qquad \frac{\partial^3 f}{\partial y^3} = x^y(\log(x))^3.$$

Note that $f \in C^3$. Now,

$$\begin{aligned}P_3(h) &= f(1,1) + h_1\frac{\partial f}{\partial x}(1,1) + h_2\frac{\partial f}{\partial y}(1,1)\\ &\quad + \frac{1}{2}h_1^2\frac{\partial^2 f}{\partial x^2}(1,1) + \frac{1}{2}h_2^2\frac{\partial^2 f}{\partial y^2}(1,1) + h_1h_2\frac{\partial^2 f}{\partial x\partial y}(1,1)\\ &\quad + \frac{1}{6}h_1^3\frac{\partial^3 f}{\partial x^3}(1,1) + \frac{1}{2}h_1^2h_2\frac{\partial^3 f}{\partial x^2\partial y}(1,1) + \frac{1}{2}h_1h_2^2\frac{\partial^3 f}{\partial x\partial y^2}(1,1) + \frac{1}{6}h_2^3\frac{\partial^3 f}{\partial y^3}(1,1)\\ &= 1 + h_1 + h_1h_2 + (1/2)h_1^2h_2.\end{aligned}$$

□

Proof of (iii). Let $g(r) := \sqrt{1+r} = (1+r)^{1/2}$. Note that $g(0) = 1$, $g'(0) = 1/2$, and $g''(0) = (1/2)(-1/2)$. We claim that $g^{(k)}(r) = (1/2)(-1/2)(-3/2)\cdots(-(2k-3)/2)(1+r)^{-(2k-1)/2}$ for $k \geq 1$. This claim follows by induction. So, the Taylor polynomial of g is given by

$$\sum_{k=0}^{\infty}\frac{r^k}{k!}\prod_{i=1}^{k}\left(\frac{3-2i}{2}\right) =: f(r).$$

Since the Taylor coefficients are uniformly bounded, the radius of convergence of this series is at least 1. We now show that $g(r) = f(r)$ for $|r| < 1$. Since f converges absolutely for $|r| < 1$, we can differentiate the series term by term to get

$$f'(r) = \sum_{k=1}^{\infty}\frac{r^{k-1}}{(k-1)!}\prod_{i=1}^{k}\left(\frac{3-2i}{2}\right) = \sum_{k=0}^{\infty}\frac{r^k}{k!}\prod_{i=1}^{k+1}\left(\frac{3-2i}{2}\right).$$

Therefore,

$$\begin{aligned}(1+r)f'(r) = f'(r) + rf'(r) &= \sum_{k=0}^{\infty}\frac{r^k}{k!}\prod_{i=1}^{k+1}\left(\frac{3-2i}{2}\right) + r\sum_{k=1}^{\infty}\frac{r^{k-1}}{(k-1)!}\prod_{i=1}^{k}\left(\frac{3-2i}{2}\right)\\ &= \frac{1}{2} + \sum_{k=1}^{\infty}\frac{r^k}{k!}[k + ((1/2)-k)]\prod_{i=1}^{k}\left(\frac{3-2i}{2}\right)\\ &= \frac{1}{2}\sum_{k=0}^{\infty}\frac{r^k}{k!}\prod_{i=1}^{k}\left(\frac{3-2i}{2}\right) = \frac{1}{2}f(r).\end{aligned}$$

Now,

$$\begin{aligned}\frac{d}{dx}[f(r)/\sqrt{1+r}] &= -\frac{1}{2}(1+r)^{-3/2}f(r) + (1+r)^{-1/2}f'(r)\\ &= (1+r)^{-3/2}[-(1/2)f(r) + (1+r)f'(r)] = 0.\end{aligned}$$

So, $f(r)/\sqrt{1+r}$ is constant. Since $f(0) = 1$, we conclude that $f(r) = \sqrt{1+r}$.

Since $f(r) = \sqrt{1+r}$ for $|r| < 1$, we can substitute $r = x^2 + y^2$ into the definition of f to get the Taylor series for the function $\sqrt{1+x^2+y^2}$ near $(x,y) = (0,0)$.

$$\sum_{k=0}^{\infty}\frac{(x^2+y^2)^k}{k!}\prod_{i=1}^{k}\left(\frac{3-2i}{2}\right).$$

□

Exercise 4.2. Locate the critical points and extrema of the functions

$$(i) \quad f(x,y) = x^3 + y^3 + 3xy, \qquad (ii) \quad f(x,y,z) = x^2 + y^2 + z^2 - 2xyz.$$

Proof of (i). f is a polynomial, so $f \in C^\infty$, and our theorems for locating extrema apply. Note that

$$\nabla f(x,y) = \begin{pmatrix} 3x^2 + 3y \\ 3y^2 + 3x \end{pmatrix}.$$

Assume that $\nabla f(x,y) = (0,0)$. Then $x^2 = -y$ and $y^2 = -x$. Squaring the first term, $x^4 = y^2 = -x$. If $x \neq 0$, then $x^3 = -1$, i.e. $x = -1$, and then since $y^2 = -x = 1$, we have $y = \pm 1$. If $y = 1$, then $x^2 \neq -y$, so we conclude that $(x,y) = (-1,-1)$, for $x \neq 0$. If $x = 0$, then $y = 0$. So, we have exactly two critical points: $(x,y) = (-1,-1)$ and $(x,y) = (0,0)$. Now, observe that, for $(x,y) = (0,0)$, we have

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 6x & 3 \\ 3 & 6y \end{pmatrix} = \begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix}.$$

The eigenvalues λ of this matrix satisfy $\lambda^2 - 9 = 0$, so that $\lambda = \pm 3$. That is, the critical point is indeterminate. And indeed, along the line $y = 0$, we have $f(x,y) = x^3$, so that f is neither a local max or min.

Now, observe that, for $(x,y) = (-1,-1)$, we have

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 6x & 3 \\ 3 & 6y \end{pmatrix} = \begin{pmatrix} -6 & 3 \\ 3 & -6 \end{pmatrix}.$$

The eigenvalues λ of this matrix satisfy $\lambda^2 + 12\lambda + 27 = 0$, so that $\lambda = -3, -9$. That is, the critical point is a local maximum. □

Proof of (ii). f is a polynomial, so $f \in C^\infty$, and our theorems for locating extrema apply. Then

$$\nabla f(x,y,z) = \begin{pmatrix} 2x - 2yz \\ 2y - 2xz \\ 2z - 2xy \end{pmatrix}.$$

Assume that $\nabla f(x,y) = (0,0)$. Then $x = yz$, $y = xz$ and $z = xy$. Note that if any of x, y or z is zero, then $x = y = z = 0$. So, note that $(x,y,z) = (0,0,0)$ is a critical point, and now assume that $x \neq 0, y \neq 0, z \neq 0$. Substituting the first equation into the second, we have $y = yz^2$, so $z^2 = 1$. Similarly, $x^2 = 1$ and $y^2 = 1$. Also, multiplying the first equation by x, the second by y and the third by z, we have $1 = x^2 = y^2 = z^2 = xyz$. That is, if one of x, y, z is -1, then exactly two of x, y, z are -1. We therefore have the critical points $(1,1,1)$, $(1,-1,-1)$, $(-1,1,-1)$ and $(-1,-1,1)$. We now check the matrix of second derivatives.

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial z \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial z \partial y} \\ \frac{\partial^2 f}{\partial x \partial z} & \frac{\partial^2 f}{\partial y \partial z} & \frac{\partial^2 f}{\partial z^2} \end{pmatrix}. = \begin{pmatrix} 2 & -2z & -2y \\ -2z & 2 & -2x \\ -2y & -2x & 2 \end{pmatrix}. \qquad (*)$$

For the critical point $(0,0,0)$, this matrix is 2 times the identity matrix, so f has a local minimum at $(0,0,0)$. For the critical point $(1,1,1)$, the matrix $(*)$ is

$$\begin{pmatrix} 2 & -2 & -2 \\ -2 & 2 & -2 \\ -2 & -2 & 2 \end{pmatrix}.$$

This matrix has eigenvalues $4, 4$ and -2. So, this critical point is indeterminate.

For the critical point $(1,-1,-1)$, the matrix $(*)$ is

$$\begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & -2 \\ 2 & -2 & 2 \end{pmatrix}.$$

This matrix has eigenvalues $4, 4$ and -2. So, this critical point is indeterminate. Similarly, by permutation symmetry of the variables x, y, z, the critical points $(-1,1,-1)$ and $(-1,-1,1)$ are also indeterminate. □

Exercise 4.3. (i) In class we proved that $R_k(h) = O(|h|^{k+1})$, so that for $f \in C^{k+1}$ we have

$$f(a+h) = P_k(h) + O(|h|^{k+1}).$$

In some applications, we want to make the weaker assumption $f \in C^k$. Prove using Taylor's theorem, that if $f \in C^k$, then

$$f(a+h) = P_k(h) + o(|h|^k).$$

(ii) Use (i) to prove the following stronger version of the criterion for locating extrema that we proved in class. Suppose f is C^2 in a neighborhood of $a \in \mathbb{R}^n$, and that a is a critical point of f. Then

$f''(a)$ positive definite	a is a local minimum of f
$f''(a)$ negative definite	a is a local maximum of f
$f''(a)$ indefinite	a is neither a local maximum or a local minimum.

Proof of (i). Let $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$. We first claim that

$$h_1 \cdots h_n = O(|h|^n). \qquad (\ddagger)$$

To see this, let $i \in \{1, \ldots, n\}$, and note that $|h_i| = \sqrt{h_i^2} \leq \sqrt{\sum_{j=1}^n h_j^2} = |h|$. So, $|h_1 \cdots h_n| \leq |h|^n$, proving $(\ddagger)$.

Since $f \in C^k$, we know from class, or from Question 4.1 that

$$f(a+h) = P_{k-1}(h) + R_{k-1}(h), \qquad (*)$$

where

$$P_{k-1}(h) = \sum_{r=0}^{k-1} \frac{1}{r!} D_h^r f(a), \qquad R_{k-1}(h) = \frac{1}{k!} D_h^k f(\xi)$$

for some ξ in the segment between a and $a+h$. Note that

$$D_h^k f(\xi) = (h_1 D_1 + \cdots + h_n D_n)^k f(\xi) = \sum_{j_1 + \cdots + j_n = k} \frac{k!}{j_1! \cdots j_n!} h_1^{j_1} \cdots h_n^{j_n} D_{j_1} \cdots D_{j_n} f(\xi). \qquad (**)$$

Let $j_1, \ldots, j_n \in \mathbb{Z}_{\geq 0}$ such that $j_1 + \cdots + j_n = k$. Since $f \in C^k$, $D_{j_1} \cdots D_{j_n} f(\xi)$ is continuous in ξ. Since ξ is in the segment between a and $a+h$, we therefore have

$$D_{j_1} \cdots D_{j_n} f(\xi) = D_{j_1} \cdots D_{j_n} f(a) + o(1).$$

Substituting this equation into $(**)$ and using $(\ddagger)$ gives

$$D_h^k f(\xi) = \sum_{j_1+\cdots+j_n=k} \frac{k!}{j_1! \cdots j_n!} h_1^{j_1} \cdots h_n^{j_n} [D_{j_1} \cdots D_{j_n} f(a) + o(1)] = D_h^k f(a) + o(|h|^k).$$

So, combining this equation with $(*)$ gives

$$f(a+h) = P_{k-1}(h) + \frac{1}{k!} D_h^k f(a) + o(|h|^k) = P_k(h) + o(|h|^k).$$

□

Proof of (ii). Let $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$, and define

$$q(h) := \frac{1}{2} \sum_{i,j=1}^{n} h_i h_j D_i D_j f(a).$$

Since a is a critical point of f, the first degree part of the Taylor polynomial is zero. That is, $f(a+h) = f(a) + q(h) + R_2(h)$. Since $f \in C^2$, part (i) shows that

$$f(a+h) - f(a) = q(h) + o(|h|^2). \qquad (*)$$

We now consider the three cases for the behavior of q. Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be an orthonormal set of eigenvectors of the symmetric matrix $f''(a) = (D_i D_j f(a))_{1 \leq i,j \leq n}$, and let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ be the corresponding eigenvalues. Since $v_1, \ldots, v_n$ is an orthonormal basis of $\mathbb{R}^n$, there exists $c_1, \ldots, c_n \in \mathbb{R}$ such that

$$h = c_1 v_1 + \cdots + c_n v_n.$$

By the orthonormality of the basis $v_1, \ldots, v_n$,

$$q(h) = c_1^2 \lambda_1 + \cdots + c_n^2 \lambda_n. \qquad (**)$$

Now, suppose $f''(a)$ is positive definite. Then $\lambda_i > 0$ for each $i \in \{1, \ldots, n\}$, so

$$q(h) \geq (c_1^2 + \cdots + c_n^2) \min_{i=1,\ldots,n} \lambda_i = |h|^2 \min_{i=1,\ldots,n} \lambda_i.$$

Combining this inequality with $(*)$,

$$f(a+h) - f(a) \geq |h|^2 \min_{i=1,\ldots,n} \lambda_i + o(|h|^2).$$

Let $\varepsilon > 0$. Then, there exists $\delta > 0$ such that, for all $0 < |h| < \delta$, the term $o(|h|^2)$ satisfies $o(|h|^2) < |h|^2 \min_{i=1,\ldots,n} \lambda_i / 2$. So, for $0 < |h| < \delta$,

$$f(a+h) - f(a) \geq \frac{1}{2} |h|^2 \min_{i=1,\ldots,n} \lambda_i > 0.$$

That is, $f(a)$ is a local minimum.

Now, suppose $f''(a)$ is negative definite. Then $\lambda_i < 0$ for each $i \in \{1, \ldots, n\}$, so

$$q(h) \leq (c_1^2 + \cdots + c_n^2) \max_{i=1,\ldots,n} \lambda_i = |h|^2 \max_{i=1,\ldots,n} \lambda_i.$$

Combining this inequality with $(*)$,

$$f(a+h) - f(a) \leq |h|^2 \max_{i=1,\ldots,n} \lambda_i + o(|h|^2).$$

Let $\varepsilon > 0$. Then, there exists $\delta > 0$ such that, for all $0 < |h| < \delta$, the term $o(|h|^2)$ satisfies $o(|h|^2) < |h|^2 |\max_{i=1,\ldots,n} \lambda_i| /2$. So, for $0 < |h| < \delta$,

$$f(a+h) - f(a) \leq \frac{1}{2} |h|^2 \max_{i=1,\ldots,n} \lambda_i < 0.$$

That is, $f(a)$ is a local maximum.

Now, suppose $f''(a)$ is indefinite. Then there exist $i, j \in \{1, \ldots, n\}$ such that $\lambda_i > 0$ and $\lambda_j < 0$. So,

$$q(c_i v_i) = c_i^2 \lambda_i, \qquad q(c_j h_j) = c_j^2 \lambda_j.$$

Combining this with $(*)$,

$$f(a + c_i v_i) - f(a) = c_i^2 \lambda_i + o(|c_i|^2), \qquad f(a + c_j v_j) - f(a) = c_j^2 \lambda_j + o(|c_j|^2).$$

Let $\varepsilon > 0$. Then, there exists $\delta > 0$ such that, for all $0 < |h| < \delta$, the terms $o(|c_i|^2)$ and $o(|c_j|^2)$ in both equalities are bounded by $|h|^2 \min(|\lambda_i|, |\lambda_j|)/2$. So, for $0 < |h| < \delta$,

$$f(a + c_i v_i) - f(a) \geq \frac{1}{2} c_i^2 \lambda_i > 0, \qquad f(a + c_j v_j) - f(a) \leq \frac{1}{2} c_j^2 \lambda_j < 0.$$

That is, $f(a)$ is neither a local maximum or minimum at a. □

Exercise 4.4. Let $A, B \subseteq \mathbb{R}^n$ be open, and suppose that $f \colon A \to B$ exists such that both f and f^{-1} are C^1. Prove that $f'(a)$ is an invertible matrix for all $a \in A$, and that for all $a \in A$ with $b := f(a)$ we have

$$(f^{-1})'(b) = (f'(a))^{-1}.$$

Proof. Applying the chain rule to the equality $(f \circ f^{-1})(b) = b$, we have

$$f'(f^{-1}(b))(f^{-1})'(b) = id.$$

That is, $f'(a)(f^{-1})'(b) = id$. Since the matrix $(f^{-1})'(b)$ inverts $f'(a)$, we have shown that $f'(a)$ is invertible. (Recall that, for square matrices A, B if $AB = id$, then $BAB = B$, so $BA = id$ also. That is, to find the inverse of a square matrix, it suffices to find a one-sided inverse of that matrix.) Since $f'(a)(f^{-1})'(b) = id$, we have shown $(f^{-1})'(b) = (f'(a))^{-1}$. □

Exercise 4.5. Prove that the function $f(x) := |x|$ is differentiable on $\mathbb{R}^n \setminus \{0\}$ and find $\nabla f(x)$.

Proof. Let $x \in \mathbb{R}^n \setminus \{0\}$ and let $r \in (0, \infty)$. Let $g \colon (0, \infty) \to (0, \infty)$ so that $g(r) := \sqrt{r}$, and let $h \colon \mathbb{R}^n \setminus \{0\} \to (0, \infty)$ so that $h(x) = h(x_1, \ldots, x_n) := x_1^2 + \cdots + x_n^2$. Then $f(x) = g(h(x))$. Note that g, h are differentiable, so $f \colon \mathbb{R}^n \setminus \{0\} \to (0, \infty)$ is differentiable by the chain rule. Also by the chain rule, we have

$$\nabla f(x) = g'(h(x))h'(x) = \frac{1}{2\sqrt{h(x)}} h'(x) = \frac{2x}{2\sqrt{h(x)}} = \frac{x}{|x|}.$$

□

Exercise 4.6. The Laplacian Δ is defined by

$$\Delta f(a) := \mathrm{Tr} f''(a) = \sum_{i=1}^{n} D_i^2 f(a).$$

This object appears in essentially every equation describing a law of nature, such as heat conduction, motion of waves, motion of quantum-mechanical particles, electric and magnetic fields, and Brownian motion.

(i) Prove that Δ is invariant under rotations in the following sense. Let $v_1, \ldots, v_n$ be an orthonormal basis of $\mathbb{R}^n$. Then, for all $f \in C^2(\mathbb{R}^n)$,

$$\Delta f = \sum_{i=1}^{n} D_{v_i}^2 f.$$

(ii) Let $f, g \in C^2$. Prove that

$$\Delta(fg) = g\Delta f + 2\langle \nabla f, \nabla g \rangle + f \Delta g.$$

(iii) Let $n = 2$ and consider the polar coordinates defined in class:

$$T\colon [0, \infty) \times [0, 2\pi) \to \mathbb{R}^2, \qquad T(r, \theta) := \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}.$$

Express Δ in polar coordinates. More precisely, let $f \in C^2(\mathbb{R}^2)$ and define $g := f \circ T$. Show that for $r > 0$ we have

$$(\Delta f)(T(r, \theta)) = \left(D_r^2 g + \frac{1}{r^2} D_\theta^2 g + \frac{1}{r} D_r g \right)(r, \theta).$$

(iv) Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}$ is invariant under rotations in the sense that $f(x) = g(|x|)$ for some $g \in C^2(0, \infty)$. Prove that

$$\Delta f(x) = g''(|x|) + \frac{n-1}{|x|} g'(|x|).$$

(v) Let

$$\psi\colon [0, \infty) \times \mathbb{R}^n \to \mathbb{R}, \qquad \psi(t, x) := \frac{1}{t^{n/2}} \exp\left(-\frac{|x^2|}{4t} \right).$$

Using (iii) show that

$$D_t \psi(t, x) = \Delta \psi(t, x) \qquad (t > 0, x \in \mathbb{R}^n).$$

Here Δ is the Laplacian in the variables $x_1, \ldots, x_n$. This equation is called the heat equation and its solutions, such as ψ, model the diffusion of heat through a conducting medium. The solution ψ given above corresponds to a single heat source at $x = 0$ when $t = 0$, which diffuses through space as time t increases.

(vi) Let $c > 0$ and fix a unit vector $v \in \mathbb{R}^n$ and $f \in C^2(\mathbb{R})$. Show that

$$\psi\colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}, \qquad \psi(t, x) := f(\langle x, v \rangle - ct)$$

satisfies the wave equation

$$D_t^2 \psi(t, x) = c^2 \Delta \psi(t, x).$$

As above, the Laplacian Δ only acts on $x_1, \ldots, x_n$. As its name implies, solutions ψ of the wave equation describe the motion of waves (e.g. sound, light) through space; the speed of the waves (speed of sound or light) is c.

Proof of (i). Let $e_1, \ldots, e_n$ denote the standard basis of $\mathbb{R}^n$. Let Q be an orthogonal matrix such that $Qe_i = v_i$. For each $i \in \{1, \ldots, n\}$, write $v_i = c_{i1}e_1 + \cdots + c_{in}e_n$. Then

$$D_{v_i}^2 f(a) = (c_{i1}D_1 + \cdots + c_{in}D_n)^2 f(a) = \sum_{1 \leq j,k \leq n} c_{ij}c_{ik}D_jD_kf(a) = v_i^T f''(a)v_i.$$

So, using the identity $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$,

$$\begin{aligned}\sum_{i=1}^n e_i^T f''(a)e_i &= \mathrm{Tr}(f''(a)) = \mathrm{Tr}(QQ^T f''(a)) = \mathrm{Tr}(Q^T f''(a)Q) = \sum_{i=1}^n e_i^T Q^T f''(a)Qe_i \\ &= \sum_{i=1}^n (Qe_i)^T f''(a)Qe_i = \sum_{i=1}^n v_i^T f''(a)v_i = \sum_{i=1}^n D_{v_i}^2 f(a).\end{aligned}$$

□

Proof of (ii). From the product rule,

$$D_i(fg) = fD_ig + gD_if.$$

Using the product rule again,

$$D_i^2(fg) = fD_i^2g + D_ifD_ig + gD_i^2f + D_ifD_ig.$$

So, summing from $i = 1, \ldots, n$ gives

$$\Delta(fg) = \sum_{i=1}^n D_i^2(fg) = f\sum_{i=1}^n D_i^2 g + 2\sum_{i=1}^n D_ifD_ig + g\sum_{i=1}^n D_i^2 f = f\Delta g + g\Delta f + 2\langle \nabla f, \nabla g\rangle.$$

□

Proof of (iii). Since $f \in C^2$ and $T \in C^\infty$, we can use the chain rule to compute the second derivatives of g. Observe

$$\frac{\partial g}{\partial r} = \frac{\partial f}{\partial x}(T(r,\theta))\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}(T(r,\theta))\frac{\partial y}{\partial r} = \frac{\partial f}{\partial x}(T(r,\theta))\cos\theta + \frac{\partial f}{\partial y}(T(r,\theta))\sin\theta.$$

$$\frac{\partial g}{\partial \theta} = \frac{\partial f}{\partial x}(T(r,\theta))\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}(T(r,\theta))\frac{\partial y}{\partial \theta} = \frac{\partial f}{\partial x}(T(r,\theta))(-r\sin\theta) + \frac{\partial f}{\partial y}(T(r,\theta))r\cos\theta.$$

$$\begin{aligned}\frac{\partial^2 g}{\partial r^2} &= \left(\frac{\partial}{\partial r}\frac{\partial f}{\partial x}(T(r,\theta))\right)\cos\theta + \left(\frac{\partial}{\partial r}\frac{\partial f}{\partial y}(T(r,\theta))\right)\sin\theta \\ &= \frac{\partial^2 f}{\partial x^2}(T(r,\theta))\frac{\partial x}{\partial r}\cos\theta + \frac{\partial^2 f}{\partial x\partial y}(T(r,\theta))\frac{\partial y}{\partial r}\cos\theta \\ &\qquad + \frac{\partial^2 f}{\partial y^2}(T(r,\theta))\frac{\partial y}{\partial r}\sin\theta + \frac{\partial^2 f}{\partial x\partial y}(T(r,\theta))\frac{\partial x}{\partial r}\cos\theta \\ &= \frac{\partial^2 f}{\partial x^2}(T(r,\theta))\cos^2\theta + 2\frac{\partial^2 f}{\partial x\partial y}(T(r,\theta))\sin\theta\cos\theta + \frac{\partial^2 f}{\partial y^2}(T(r,\theta))\sin^2\theta.\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 g}{\partial \theta^2} &= \left(\frac{\partial}{\partial \theta}\frac{\partial f}{\partial x}(T(r,\theta))\right)(-r\sin\theta) + \left(\frac{\partial}{\partial \theta}\frac{\partial f}{\partial y}(T(r,\theta))\right) r\cos\theta \\
&\qquad + \frac{\partial f}{\partial x}(T(r,\theta))(-r\cos\theta) + \frac{\partial f}{\partial y}(T(r,\theta)) r\sin\theta \\
&= \frac{\partial^2 f}{\partial x^2}(T(r,\theta))\frac{\partial x}{\partial \theta}(-r\sin\theta) + \frac{\partial^2 f}{\partial x \partial y}(T(r,\theta))\frac{\partial y}{\partial \theta}(-r\sin\theta) + \frac{\partial^2 f}{\partial y^2}(T(r,\theta))\frac{\partial y}{\partial \theta} r\cos\theta \\
&\qquad + \frac{\partial^2 f}{\partial x \partial y}(T(r,\theta))\frac{\partial x}{\partial \theta} r\cos\theta + \frac{\partial f}{\partial x}(T(r,\theta))(-r\cos\theta) + \frac{\partial f}{\partial y}(T(r,\theta)) r\sin\theta \\
&= \frac{\partial^2 f}{\partial x^2}(T(r,\theta)) r^2\sin^2\theta - 2\frac{\partial^2 f}{\partial x \partial y}(T(r,\theta)) r^2 \sin\theta\cos\theta + \frac{\partial^2 f}{\partial y^2}(T(r,\theta)) r^2\cos^2\theta \\
&\qquad + \frac{\partial f}{\partial x}(T(r,\theta))(-r\cos\theta) + \frac{\partial f}{\partial y}(T(r,\theta)) r\sin\theta.
\end{aligned}$$

So,

$$\begin{aligned}
&D_r^2 g + \frac{1}{r^2}D_\theta^2 g + \frac{1}{r}D_r g \\
&= \frac{\partial^2 f}{\partial x^2}(T(r,\theta))\cos^2\theta + 2\frac{\partial^2 f}{\partial x \partial y}(T(r,\theta))\sin\theta\cos\theta + \frac{\partial^2 f}{\partial y^2}(T(r,\theta))\sin^2\theta \\
&\quad + \frac{\partial^2 f}{\partial x^2}(T(r,\theta))\sin^2\theta - 2\frac{\partial^2 f}{\partial x \partial y}(T(r,\theta))\sin\theta\cos\theta + \frac{\partial^2 f}{\partial y^2}(T(r,\theta))\cos^2\theta \\
&\quad + \frac{1}{r}\frac{\partial f}{\partial x}(T(r,\theta))\cos\theta + \frac{1}{r}\frac{\partial f}{\partial y}(T(r,\theta))\sin\theta - \frac{1}{r}\frac{\partial f}{\partial x}(T(r,\theta))\cos\theta - \frac{1}{r}\frac{\partial f}{\partial y}(T(r,\theta))\sin\theta \\
&= \frac{\partial^2 f}{\partial x^2}(T(r,\theta))(\sin^2\theta + \cos^2\theta) + \frac{\partial^2 f}{\partial y^2}(T(r,\theta))(\sin^2\theta + \cos^2\theta) \\
&= (\Delta f)(T(r,\theta))
\end{aligned}$$

□

Proof of (iv). We use Question 4.5.

$$D_i f(x) = D_i(g(|x|)) = g'(|x|)\frac{x_i}{|x|}.$$

$$D_i^2 f(x) = g'(|x|)\left(\frac{|x| - x_i\left(\frac{x_i}{|x|}\right)}{|x|^2}\right) + \frac{x_i}{|x|}g''(|x|)\frac{x_i}{|x|} = g'(|x|)\frac{|x|^2 - x_i^2}{|x|^3} + g''(|(|x))\frac{x_i^2}{|x|^2}.$$

$$\Delta f(x) = \sum_{i=1}^n D_i^2 f(x) = g'(|x|)\frac{n|x|^2 - |x|^2}{|x|^3} + g''(|x|)\frac{|x|^2}{|x|^2} = g'(|x|)\frac{n-1}{|x|} + g''(|x|).$$

□

Proof of (v). Let $g(r) := t^{-n/2}\exp(-r^2/(4t))$. For $t > 0$, note that $g \in C^\infty$. Also, for $t > 0$, $\psi \in C^\infty$. Then

$$g'(r) = -(r/2)t^{-(n+2)/2}\exp(-r^2/(4t))$$
$$g''(r) = (r^2/4)t^{-(n+4)/2}\exp(-r^2/(4t)) - (1/2)t^{-(n+2)/2}\exp(-r^2/(4t))$$
$$D_t\psi(t,x) = t^{-n/2}t^{-2}(|x|^2/4)\exp(-|x|^2/(4t)) + (-n/2)t^{-(n+2)/2}\exp(-|x|^2/(4t))$$

So, using (iv),

$$\begin{aligned}
\Delta\psi(t,x) &= g''(|x|) + \frac{n-1}{|x|}g'(|x|) \\
&= \frac{|x|^2}{4t^{(n+4)/2}} \exp\left(-\frac{|x|^2}{4t}\right) - \frac{1}{2t^{(n+2)/2}} \exp\left(-\frac{|x|^2}{4t}\right) - \frac{(n-1)}{2t^{(n+2)/2}} \exp\left(-\frac{|x|^2}{4t}\right) \\
&= \frac{|x|^2}{4t^{(n+4)/2}} \exp\left(-\frac{|x|^2}{4t}\right) - \frac{n}{2t^{(n+2)/2}} \exp\left(-\frac{|x|^2}{4t}\right) \\
&= D_t\psi(t,x)
\end{aligned}$$

□

Proof of (vi). Since $f \in C^2$, we have $\psi \in C^2$. Let $v = (v_1, \dots, v_n) \in \mathbb{R}^n$ with $|v| = 1$. Then

$$D_t\psi(t,x) = \sum_{j=1}^{n} (-c)\frac{\partial f}{\partial x_j}(\langle x, v\rangle - ct), \qquad D_t^2\psi(t,x) = \sum_{j,k=1}^{n} c^2 \frac{\partial^2 f}{\partial x_j \partial x_k}(\langle x, v\rangle - ct).$$

$$D_i\psi(t,x) = v_i \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(\langle x, v\rangle - ct), \qquad D_i^2\psi(t,x) = v_i^2 \sum_{j,k=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_k}(\langle x, v\rangle - ct).$$

Using that $|v| = 1$,

$$c^2\Delta\psi(t,x) = c^2 \sum_{i=1}^{n} D_i^2\psi(t,x) = c^2 \sum_{j,k=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_k}(\langle x, v\rangle - ct) = D_t^2\psi(t,x).$$

□

Exercise 4.7. It is of fundamental importance for many arguments in analysis that there exists a C^∞ function which is positive inside the unit ball and zero outside. An example is

$$f(x) := \begin{cases} \exp(1/(|x|^2 - 1)) & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geq 1 \end{cases}.$$

Prove that $f \in \mathbb{C}^\infty(\mathbb{R}^n)$.

Proof. Let for $r \in (-1, 1)$, let $g(r) := \exp(1/(r-1))$. Note that $g'(r) = -(1/(r-1)^2)g(r)$. We prove by induction on $k \in \mathbb{Z}_{\geq 0}$ that there exists a rational function $h_k(r)$ of the form $h_k(r) = a_k(r)/(r-1)^{2^k}$, where $a_k(r)$ is a polynomial in r, such that

$$g^{(k)}(r) = h_k(r)g(r).$$

Let $k \geq 1$ and assume that $g^{(k)}(r) = [a_k(r)/(r-1)^{2^k}]g(r)$. Then, from the product rule,

$$\begin{aligned}
g^{(k+1)}(r) &= \frac{(r-1)^{2^k} a_k'(r) - a_k(r) 2^k (r-1)^{2^k-1}}{(r-1)^{2^{k+1}}} g(r) + \frac{a_k(r)}{(r-1)^{2^k}} g'(r) \\
&= \frac{(r-1)^{2^k} a_k'(r) - a_k(r) 2^k (r-1)^{2^k-1}}{(r-1)^{2^{k+1}}} g(r) - \frac{a_k(r)}{(r-1)^{2^k+2}} g(r) \\
&= \frac{(r-1)^{2^k} a_k'(r) - a_k(r) 2^k (r-1)^{2^k-1} - a_k(r)(r-1)^{2^k-2}}{(r-1)^{2^{k+1}}} g(r) \\
&=: \frac{a_{k+1}(r)}{(r-1)^{2^{k+1}}} g(r).
\end{aligned}$$

Note that $a_{k+1}(r)$ as defined here is a polynomial. So, with the inductive step completed, our claim is proven. Directly from the claim, $g \in C^\infty(-1/2, 1)$. Now, let $F(r) := g(r)$ for $|r| < 1$, and $F(r) := 0$ for $|r| \geq 1$. We will show that $F \in C^\infty(-1/2, \infty)$. It remains to show that $F^{(k)}(1)$ exists for all $k \in \mathbb{N}$ and $F^{(k)}(r)$ is continuous at $r = 1$ for all $k \in \mathbb{N}$.

From our claim, observe

$$\lim_{r \to 1^-} g^{(k)}(r) = a_k(1) \lim_{r \to 1^-} \frac{g(r)}{(r-1)^{2^k}} = a_k(1) \lim_{r \to 1^-} \frac{e^{1/(r-1)}}{(r-1)^{2^k}} = a_k(1) \lim_{t \to \infty} e^{-t} t^{2^k} = 0.$$

Also, given the existence of $g^{(k)}(1) = 0$, we have

$$g^{(k+1)}(1) = \lim_{r \to 1^-} \frac{g^{(k)}(r) - g^{(k)}(1)}{r-1} = \lim_{r \to 1^-} \frac{g^{(k)}(r)}{r-1} = a_k(1) \lim_{t \to \infty} e^{-t} t^{2^k+1} = 0.$$

So, we conclude that $g^{(k)}(1) = 0$ exists and $F^{(k)}(r)$ is continuous at $r = 1$ for all $k \in \mathbb{N}$. That is, $F \in C^\infty(-1/2, \infty)$. Finally, for $x \in \mathbb{R}^n$, note that $f(x) = F(|x|^2)$, i.e. f is the composition of two C^∞ functions. So, by the chain rule, $f \in C^\infty(\mathbb{R}^n)$, as desired. ☐

Exercise 4.8. A linear transformation $A\colon \mathbb{R}^n \to \mathbb{R}^m$ is called conformal if there is a number $\rho > 0$ such that $A^T A = \rho 1$.

(i) Prove that a conformal transformation preserves angles in the sense that, for any nonzero $v, w \in \mathbb{R}^n$, the vectors Av, Aw are also nonzero and the angle between v and w is the same as the angle between Av and Aw.

(ii) Prove that a linear map that preserves angles in the sense given in (i) is conformal. (Combining this result with (i), we deduce that a linear map is conformal if and only if it preserves angles.)

(iii) A mapping $f \in C^1(A; \mathbb{R}^m)$, where $A \subseteq \mathbb{R}^n$ is called conformal at $a \in A$ if $f'(a)$ is conformal. Let γ_1 and γ_2 be two C^1 curves in $\mathbb{R}^n$ that intersect, i.e. $\gamma_1(0) = \gamma_2(0)$. Suppose that f is conformal at this point of intersection. Prove that the angle between the tangents of γ_1 and γ_2 at time 0 is the same as the angle between the tangents of $f \circ \gamma_1$ and $f \circ \gamma_2$ at time 0.

(iv) The stereographic projection is a projection used to map the unit sphere in $\mathbb{R}^{n+1}$ to $\mathbb{R}^n$. It is defined as follows. Let

$$S^n := \{u \in \mathbb{R}^{n+1} \colon |u| = 1\}$$

be the unit sphere, and let $p := (0, \ldots, 0, 1)^T$ be the "north pole" of the sphere. For $x \in \mathbb{R}^n$, let $\ell(x)$ be the line in $\mathbb{R}^{n+1}$ that passes through the two points $(x, 0)$ and

p. Draw a sketch of $\mathbb{R}^{n+1}$ along with S^n and $\ell(x)$. Show that $\ell(x)$ intersects S^n at a unique point, which we denote by $S(x)$. Show also that S is a bijection between $\mathbb{R}^n$ and $S^n \setminus \{p\}$. S is called the stereographic projection. Prove that

$$S(x) = \frac{1}{1 + |x|^2} \begin{pmatrix} 2x \\ |x|^2 - 1 \end{pmatrix}.$$

Finally, prove that S is conformal. This fact is of great interest to cartographers: using S we may represent the surface of the earth on a flat piece of paper in such a way that all angles are preserved. The downside is that areas are not preserved; in fact, that is no projection that preserves both angles and areas (this is a mathematical theorem). We understand this theorem intuitively, since we cannot flatten a paper sphere without tearing the paper. The stereographic projection is accurate near the "south pole" $(0, \ldots, 0, -1)^T$, and becomes increasingly distorted as one approaches the north pole p.

Proof of (i). Let $v, w \in \mathbb{R}^n$, $v, w \neq 0$. Then

$$\langle Av, Aw \rangle = (Av)^T (Aw) = v^T A^T A w = v^T \rho w = \rho \langle v, w \rangle. \qquad (*)$$

Setting $v = w$, $(*)$ says that $|Av|^2 = \rho\, |v|^2$. So, if $v \neq 0$, then $\rho > 0$ implies that $\rho\, |v|^2 > 0$, so $|Av|^2 > 0$, i.e. $Av \neq 0$. Now, using $(*)$ and the identity $|Av|^2 = \rho\, |v|^2$,

$$\frac{\langle v, w \rangle}{|v|\, |w|} = \frac{\rho \langle v, w \rangle}{\sqrt{\rho}\, |v| \sqrt{\rho}\, |w|} = \frac{\langle Av, Aw \rangle}{|Av|\, |Aw|}.$$

That is, the angle between v, w is the same as the angle between Av, Aw. □

Proof of (ii). Suppose A preserves angles. That is, given $v, w \in \mathbb{R}^n$ $v, w \neq 0$, we have $Av, Aw \neq 0$, and

$$\frac{\langle Av, Aw \rangle}{|Av|\, |Aw|} = \frac{\langle v, w \rangle}{|v|\, |w|}. \qquad (*)$$

Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{R}^n$. Let $v = e_i, w = e_j$, $i, j \in \{1, \ldots, n\}$, $i \neq j$. Then $(*)$ says that $\langle Ae_i, Ae_j \rangle = 0$. Since $Ae_i \neq 0$ for $i \in \{1, \ldots, n\}$, we know that the set $Ae_1, \ldots, Ae_n$ is an orthogonal set of nonzero vectors. Also, note that

$$\langle e_i, e_j \rangle = \frac{\langle Ae_i, Ae_j \rangle}{|Ae_i|\, |Ae_j|} = \frac{e_i^T A^T A e_j}{|Ae_i|\, |Ae_j|}.$$

So, the matrix $A^T A$ has zero entries away from the diagonal, and nonzero entries along the diagonal. More specifically, the i^{th} diagonal entry of $A^T A$ is $|Ae_i|^2$. So, it remains to show that $|Ae_i| = |Ae_j|$ for $i, j \in \{1, \ldots, n\}$, $i \neq j$. Let $t \in [0, 1]$, let $v := e_i$ and let $w := e_i\sqrt{t} + e_j\sqrt{1-t}$. Then, $|v| = |w| = 1$, and using $(*)$, we have

$$\langle v, w \rangle = \sqrt{t} = \frac{\langle Av, Aw \rangle}{|Av|\, |Aw|} = \frac{\langle Ae_i, \sqrt{t} Ae_i \rangle}{|Av|\, |Aw|} = \sqrt{t} \frac{|Ae_i|^2}{|Ae_i|\, |Aw|}.$$

So,

$$|Ae_i|^2 = |Aw|^2 = |Ae_i\sqrt{t} + Ae_j\sqrt{1-t}|^2 = t\, |Ae_i|^2 + (1-t)\, |Ae_j|^2.$$

Plugging in $t = 0$ shows that $|Ae_i|^2 = |Ae_j|^2$, as desired. □

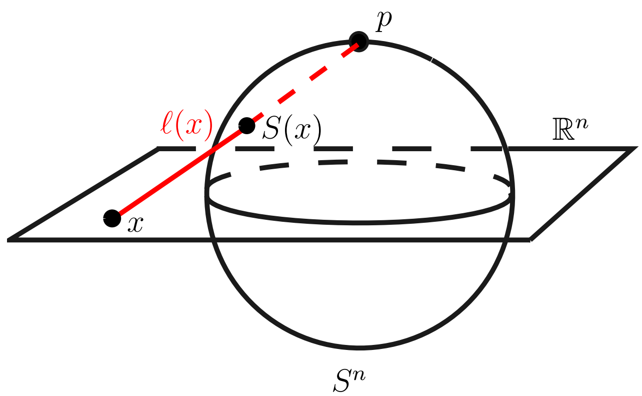
Proof of (iii). Let $i, j \in \{1, 2\}$. From the chain rule, $(f \circ \gamma_i)' = f'(\gamma_i)\gamma_i'$. So,

$$\begin{aligned}&\langle (f\circ\gamma_i)'(0), (f\circ\gamma_j)'(0)\rangle \\ &\qquad = (f'(\gamma_i)(0)\gamma_i'(0))^T(f'(\gamma_j)(0)\gamma_j'(0)) = \gamma_i'(0)^T(f'(\gamma_i(0)))^T(f'(\gamma_j)(0)\gamma_j'(0)) \\ &\qquad = \gamma_i'(0)^T(f'(\gamma_i(0)))^T(f'(\gamma_i)(0)\gamma_j'(0)) = \rho\gamma_i'(0)^T\gamma_j'(0))\end{aligned}$$

So, using the cases $i = j$ and $i \neq j$ separately,

$$\frac{\langle \gamma_1'(0), \gamma_2'(0)\rangle}{|\gamma_1'(0)|\,|\gamma_2'(0)|} = \frac{\rho\langle \gamma_1'(0), \gamma_2'(0)\rangle}{\sqrt{\rho}\,|\gamma_1'(0)|\,\sqrt{\rho}\,|\gamma_2'(0)|} = \frac{\langle (f\circ\gamma_1)'(0), (f\circ\gamma_2)'(0)\rangle}{|(f\circ\gamma_1)'(0)|\,|(f\circ\gamma_2)'(0)|}.$$

□



Proof of (iv). Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Let $w(x)$ denote the orthogonal projection of $S(x)$ onto $\mathbb{R}^n$. That is, if $S(x) = (y_1, \ldots, y_{n+1}) \in \mathbb{R}^{n+1}$, then $w(x) := (y_1, \ldots, y_n, 0)$. Note that we consider $\mathbb{R}^n$ as a subset of $\mathbb{R}^{n+1}$ via the inclusion $\mathbb{R}^n = \mathbb{R}^n \times \{0\} \subseteq \mathbb{R}^{n+1}$. We first show that $\ell(x)$ intersects S^n at a unique point. The line $\ell(x)$ is either parallel to the hyperplane $\mathbb{R}^n$, or $\ell(x)$ intersects $\mathbb{R}^n$ at precisely one point. Since $\ell(x)$ intersects $\mathbb{R}^n$ by the definition of $\ell(x)$, we know that $\ell(x)$ is not parallel to $\mathbb{R}^n$. Also, the line $\ell(x)$ is either tangent to the sphere S^n, or $\ell(x)$ intersects S^n at exactly two points. Since $\ell(x)$ is not parallel to $\mathbb{R}^n$, and $p \in \ell(x)$, we conclude that $\ell(x)$ is not tangent to the sphere S^n. So, $\ell(x)$ intersects S^n in exactly two points. Since $\ell(x)$ intersects $p \in S^n$, there exists a unique point $S(x) \neq p$ such that $\ell(x)$ intersects S^n at $S(x)$.

We now show that $S\colon \mathbb{R}^n \to S^n \setminus \{p\}$ is a bijection. We first show that S is injective. Let $x \neq y$, $x, y \in \mathbb{R}^n$. We want to show that $S(x) \neq S(y)$. The lines $\ell(x)$ and $\ell(y)$ intersect at p. Also, two distinct Euclidean lines can intersect in at most one point. So, since $\ell(x)$ intersects $S(x)$ and $\ell(y)$ intersects $S(y)$, we must have $S(x) \neq S(y)$, as desired. We now show that S is surjective. Let $s \in S^n \setminus \{p\}$. Let ℓ be the unique line that contains p and s. Since $s \neq p$, the line ℓ is not tangent to S^n at p. Therefore, the line ℓ is not parallel to the hyperplane $\mathbb{R}^n$. Therefore, there exists a unique $x \in \mathbb{R}^n$ such that ℓ intersects x. Since ℓ also intersects s and p, we conclude that $\ell(x) \cap (S^n \setminus \{p\}) = s$, so that $S(x) = s$. That is, S is surjective. In conclusion, S is bijective.

Let $h(x)$ denote the orthogonal projection of $S(x)$ onto p. So, if $S(x) = (y_1, \ldots, y_{n+1}) \in \mathbb{R}^{n+1}$, then $h(x) := (0, \ldots, 0, y_{n+1})$. Since p is orthogonal to $\mathbb{R}^n$, the Pythagorean Theorem says that $|w(x)|^2 + |h(x)|^2 = |S(x)|^2 = 1$. Note that $|S(x)| = 1$ by assumption. Note that the right triangle with edges x and p is similar to the right triangle with edges $w(x)$ and $p-h(x)$. Therefore, $|x|/1 = |w(x)|/|p-h(x)|$. In summary, using that $|h(x)| = |\langle p, h(x)\rangle|$

$$|w(x)| = |x|\,|p-h(x)|\,, \qquad |w(x)|^2 + |\langle p, h(x)\rangle|^2 = 1.$$

By substitution, $|x|^2\,|p-h(x)|^2 + |\langle p, h(x)\rangle|^2 = 1$. Expanding the inner product,

$$\begin{aligned} 1 &= |x|^2\,(1 - 2\langle p, h(x)\rangle + |h(x)|^2) + ((1 - |\langle p, h(x)\rangle|) - 1)^2 \\ &= |x|^2\,(1 - \langle p, h(x)\rangle)^2 + ((1 - |\langle p, h(x)\rangle|) - 1)^2 \end{aligned}$$

So, $(|x|^2+1)(1-\langle p, h(x)\rangle)^2 - 2(1-\langle p, h(x)\rangle) = 0$. Since $\langle p, h(x)\rangle < 1$, we can divide by $1 - \langle p, h(x)\rangle$ to get $(|x|^2+1)(1-\langle p, h(x)\rangle) = 2$, so that

$$\langle p, h(x)\rangle = \frac{-1+|x|^2}{1+|x|^2}. \qquad (*)$$

By the definition of $w(x)$, we know that $w(x)$ is a constant multiple of x. Since $|w(x)|^2 = |x|^2\,|p-h(x)|^2 = |x|^2\,(1-\langle p, h(x)\rangle)^2$, and $1 - \langle p, h(x)\rangle = 2/(|x|^2+1)$ by $(*)$, so

$$w(x) = \frac{2}{|x|^2+1}x. \qquad (**)$$

Combining $(*)$ and $(**)$,

$$S(x) = \begin{pmatrix} w(x) \\ h(x) \end{pmatrix} = \frac{1}{|x|^2+1}\begin{pmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \\ -1+|x|^2 \end{pmatrix}.$$

We now compute the $(n+1)\times n$ matrix $S'(x)$.

$$S'(x) = \begin{pmatrix} \frac{2(|x|^2+1)-4x_1^2}{(|x|^2+1)^2} & \frac{-4x_1x_2}{(|x|^2+1)^2} & \frac{-4x_1x_3}{(|x|^2+1)^2} & \cdots & \frac{-4x_1x_n}{(|x|^2+1)^2} \\ \frac{-4x_1x_2}{(|x|^2+1)^2} & \frac{2(|x|^2+1)-4x_2^2}{(|x|^2+1)^2} & \frac{-4x_2x_3}{(|x|^2+1)^2} & \cdots & \frac{-4x_2x_n}{(|x|^2+1)^2} \\ \vdots & \vdots & \ddots & & \vdots \\ \frac{-4x_1x_n}{(|x|^2+1)^2} & \frac{-4x_2x_n}{(|x|^2+1)^2} & \cdots & \cdots & \frac{2(|x|^2+1)-4x_n^2}{(|x|^2+1)^2} \\ \frac{4x_1}{(|x|^2+1)^2} & \frac{4x_2}{(|x|^2+1)^2} & \cdots & \cdots & \frac{4x_n}{(|x|^2+1)^2} \end{pmatrix}.$$

Let $v_1, \ldots, v_n$ denote the columns of $S'(x)$. Observe

$$\begin{aligned} (|x|^2+1)^4\langle v_i, v_i\rangle &= 16x_i^2\sum_{j\neq i} x_j^2 + 16x_i^2 + 16x_i^4 + 4(|x|^2+1)^2 - 16x_i^2(|x|^2+1) \\ &= 16x_i^2(|x|^2+1) + 4(|x|^2+1)^2 - 16x_i^2(|x|^2+1) \\ &= (|x|^2+1)\left(16x_i^2 + 4(|x|^2+1) - 16x_i^2\right) \\ &= 4(|x|^2+1)^2. \end{aligned}$$

Also, for $i, j \in \{1, \ldots, n\}$ with $i \neq j$,

$$\begin{aligned}
&(|x|^2+1)^4\langle v_i, v_j\rangle \\
&\quad = -8(|x|^2+1)x_ix_j + 16x_i^3x_j - 8(|x|^2+1)x_ix_j + 16x_ix_j^3 + x_ix_j\sum_{k\neq i,j} x_k^2 + 16x_ix_j \\
&\quad = -8(|x|^2+1)x_ix_j - 8(|x|^2+1)x_ix_j + 16x_ix_j\sum_{k=1}^{n} x_k^2 + 16x_ix_j \\
&\quad = -8(|x|^2+1)x_ix_j - 8(|x|^2+1)x_ix_j + 16x_ix_j(|x|^2+1) \\
&\quad = 0.
\end{aligned}$$

So, combining these observations,

$$(S'(x))^T S'(x) = \frac{4}{(|x|^2+1)^2} id.$$

□

5. Problem Set 5

Exercise 5.1.

(i) Show that the mappings

$$f(x, y) := (e^x + e^y, e^x - e^y), \qquad g(x, y) := (e^x \cos y, e^x \sin y)$$

are locally invertible around each point of $\mathbb{R}^2$.

(ii) Show that the equations

$$\sin(y + x) + \log(zx^2) = 0, \qquad e^{y+x} + xz = 0$$

implicitly define (y, z) near $(1, 1)$, as an explicit function of x near -1.

Proof of (i). Let $(x, y) \in \mathbb{R}^2$. Since f and g are the composition of C^1 functions, f and g are C^1. Now,

$$\det f'(x, y) = \det \begin{pmatrix} e^x & e^y \\ e^x & -e^y \end{pmatrix} = -2e^xe^y < 0.$$

$$\det g'(x, y) = \det \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix} = e^{2x} > 0.$$

So, f' and g' are invertible for all $(x, y) \in \mathbb{R}^2$. Since f, g are also C^1, we conclude that f, g are locally invertible around each point of $\mathbb{R}^2$, by the Inverse Function Theorem. □

Proof of (ii). Define the open set

$$D := \{(x, y, z) \in \mathbb{R}^3 \colon 9/10 < y, z < 11/10, -11/10 < x < -9/10\}.$$

Define the mapping $f \colon D \to \mathbb{R}^2$ by

$$f(x, y, z) := (\sin(y + x) + \log(zx^2),\ e^{y+x} + xz) =: (f_1(x, y, z), f_2(x, y, z)).$$

Note that f is the composition of C^1 functions, so $f \in C^1(D)$. Also,

$$\det \begin{pmatrix} \partial f_1/\partial y & \partial f_1/\partial z \\ \partial f_2/\partial y & \partial f_2/\partial z \end{pmatrix} = \det \begin{pmatrix} \cos(y + x) & 1/z \\ e^{y+x} & x \end{pmatrix} = x\cos(y + x) - e^{y+x}/z. \qquad (*)$$

Since $|y-1| < 1/10$ and $|x+1| < 1/10$, we have $|y+x| = |x-1+z+1| < 1/5$, so $\cos(y+x) > 0$. Since $9/10 < z < 11/10$, we get $10/9 > 1/z > 10/11$. So, combining these facts, the quantity in $(*)$ is negative. So, the Implicit Function Theorem says that $z = z(x)$ and $y = y(x)$ such that $f(x, y(x), z(x)) = (0,0)$ for x in a neighborhood of -1 and (y, z) in a neighborhood of $(1,1)$. □

Exercise 5.2. Consider the system of equations

$$x^2 + uy + e^v = 0$$
$$2x + u^2 - uv = 5.$$

Show that (u, v) may be solved in terms of (x, y) in a neighborhood of the point $(x, y) = (2, 5)$. Show that the mapping $(x, y) \mapsto (u, v)$ is C^1 and compute its derivative at $(x, y) = (2, 5)$.

Proof. Let $(x, y) = (2, 5)$. Then the system of equations becomes $5u + e^v + 4 = 0$ and $4 + u^2 - uv = 5$. So, $u = (1/5)(-e^v - 4)$, and substituting this into the second equation gives

$$100 + e^{2v} + 16 + 8e^v + 5ve^v + 20v = 125.$$

Simplifying this equation, we have

$$e^{2v} + 8e^v + 5ve^v + 20v = 9.$$

The function on the left is strictly increasing in v, so there exists a unique v satisfying this equation. Observe that $v = 0$ satisfies this equation, so $v = 0$ is the unique solution of this equation. Then, since $u = (1/4)(-e^v - 4)$, we must have $u = -1$. Define a function $F\colon \mathbb{R}^4 \to \mathbb{R}^2$ by

$$F(x, y, u, v) = (F_1(x, y, u, v), F_2(x, y, u, v)) := (x^2 + uy + e^v, 2x + u^2 - uv).$$
$$\det \begin{pmatrix} \partial F_1/\partial u & \partial F_1/\partial v \\ \partial F_2/\partial u & \partial F_2/\partial v \end{pmatrix} = \det \begin{pmatrix} y & e^v \\ 2u - v & -u \end{pmatrix} = -uy - (2u - v)e^v = 5 + 2 = 7 > 0.$$

That is, this determinant term is positive. Also, F is the composition of C^1 functions, so F is C^1. Therefore, the Implicit Function Theorem says that there exists an open set $W \subseteq \mathbb{R}^2$ with $(2, 5) \in W$ such that $u = u(x, y)$ and $v = v(x, y)$ for $(x, y) \in W$, and there exists an open set $V \subseteq \mathbb{R}^4$ such that $(2, 5, -1, 0) \in V$, and $F(x, y, u(x, y), v(x, y)) = (0, 5)$ for $(x, y, u, v) \in V$. Moreover, $u = u(x, y)$ and $v = v(x, y)$ are in $C^1(W)$.

We now compute the derivatives of u, v. Let $s := (u, v)$ and let $z := (x, y)$. In the set V,

$$F(z, s(z)) = (0, 5).$$

So, applying the chain rule, we have

$$D_z F(z, s) + D_s F(z, s) s'(z) = 0.$$

Since $D_s F(z, s)$ is invertible, we conclude that, when $(x, y, u, v) = (2, 5, -1, 0)$, we have

$$z'(s) = -[D_s F(z, s)]^{-1} D_z F(z, s) = \begin{pmatrix} y & e^v \\ 2u - v & -u \end{pmatrix}^{-1} \begin{pmatrix} 2x & u \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 1 \\ -2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 4 & -1 \\ 2 & 0 \end{pmatrix}.$$

□

Exercise 5.3. Let A be an $m \times n$ matrix. We define the *matrix norm* of A by

$$||A|| := \sup \{|Ax| : |x| \leq 1\}.$$

(i) Prove that for all $x \in \mathbb{R}^n$, we have $|Ax| \leq ||A||\,|x|$.

(ii) Prove that $||\cdot||$ is a *norm*, i.e. that it satisfies the following axioms:
  (1) $||aA|| = |a|\,||A||$ for all $a \in \mathbb{R}$;
  (2) $||A + B|| \leq ||A|| + ||B||$;
  (3) if $||A|| = 0$ then $A = 0$.

(iii) Prove that $||AB|| \leq ||A||\,||B||$, where A is a $k \times m$ matrix and B an $m \times n$ matrix.

(iv) Since the space of $m \times n$ matrices may be identified with $\mathbb{R}^{m\times n}$, the norm or distance defined in class reads

$$|A| = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2}\,.$$

(This is sometimes called the Hilbert-Schmidt norm.) Prove that

$$||A|| \leq |A|\,.$$

Conclude that if $A(x)$ is a continuous matrix-valued function, and if $x \to y$, then $||A(x) - A(y)|| \to 0$.

Proof of (i). Let $x \in \mathbb{R}^n$. If $x = 0$, then $|Ax| = 0 \leq 0 = ||A||\,|x|$. So, the desired inequality holds for $x = 0$. Now, let $x \neq 0$. Then, by the definition of $||A||$, and using that $|x/\,|x||= 1$,

$$|Ax| = \left|A\frac{x}{|x|}\,|x|\right| = |x|\left|A\frac{x}{|x|}\right| \leq |x|\,||A||\,.$$

□

Proof of (ii). Let A be a matrix. Let $x \in \mathbb{R}^n$ with $|x| \leq 1$, and let $a \in \mathbb{R}$. Then $|aAx| = |a|\,|Ax|$. So, taking the supremum of both sides of this equality over $\{x \in \mathbb{R}^n\colon\ |x| \leq 1\}$, we get property (1). We now prove property (2). Let $x \in \mathbb{R}^n$ with $|x| \leq 1$. From the triangle inequality,

$$|(A + B)x| \leq |Ax| + |Bx| \leq |Ax| + \sup_{|y|\leq 1} |By| = |Ax| + ||B||\,.$$

Now, taking the supremum of both sides over the set $\{x \in \mathbb{R}^n\colon\ |x| \leq 1\}$ gives (2).

We now prove property (3). We first show that

$$||A|| = \sup_{\substack{x\in\mathbb{R}^n, y\in\mathbb{R}^m\colon\\ |x|\leq 1, |y|\leq 1}} |\langle Ax, y\rangle|\,. \qquad (*)$$

If $A = 0$, then both sides are equal to zero. So, to prove $(*)$, we may assume that $A \neq 0$. Let $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ with $|x| \leq 1, |y| \leq 1$. From Cauchy-Schwarz and (i),

$$|\langle Ax, y\rangle| \leq |Ax|\,|y| \leq ||A||\,|x|\,|y| \leq ||A||\,.$$

Therefore, we get one part of $(*)$.

$$\sup_{\substack{x\in\mathbb{R}^n, y\in\mathbb{R}^m\colon\\ |x|\leq 1, |y|\leq 1}} |\langle Ax, y\rangle| \leq ||A||\,.$$

We now prove the reverse inequality. Since $A \neq 0$, there exists $x \neq 0$ with $Ax \neq 0$. Let $y := Ax/\,|Ax|$. Then

$$|\langle Ax, y\rangle| = \left|\langle Ax, \frac{Ax}{|Ax|}\rangle\right| = \frac{|Ax|^2}{|Ax|} = |Ax|\,.$$

So, taking the supremum over x with $|x| \leq 1$ shows the other part of $(*)$.

$$||A|| = \sup_{x\in\mathbb{R}^n:\ |x|\leq 1} |\langle Ax, y\rangle| \leq \sup_{\substack{x\in\mathbb{R}^n, y\in\mathbb{R}^m:\\ |x|\leq 1, |y|\leq 1}} |\langle Ax, y\rangle| .$$

We can now conclude the proof. Suppose $||A|| = 0$. By $(*)$, $|\langle Ae_j, e_i\rangle| = 0$ for all $j \in \{1, \ldots, n\}$, $i \in \{1, \ldots, m\}$. That is, each entry of A is zero. So $A = 0$, as desired. □

Proof of (iii). If $B = 0$, then $||B|| = 0$ from (ii), and $AB = 0$, so $||AB|| = ||A||\,||B|| = 0$. So, we may assume that $B \neq 0$. Then there exists $x \in \mathbb{R}^n$ with $x \neq 0$ and $Bx \neq 0$. Then

$$|ABx| = \left|A\frac{Bx}{|Bx|}\right| |Bx| \leq ||A||\,|Bx| \leq ||A||\,||B||\,|x| . \qquad (*)$$

If $Bx = 0$, then the left side of $(*)$ is zero, and $(*)$ still holds. That is, $(*)$ holds for all $x \in \mathbb{R}^n$. Finally, taking the supremum of both sides of $(*)$ over the set $\{x \in \mathbb{R}^n \colon |x| \leq 1\}$ shows that $||AB|| \leq ||A||\,||B||$. □

Proof of (iv). Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be the rows of A, and let $x \in \mathbb{R}^n$ with $|x| \leq 1$. Then, using Cauchy-Schwarz,

$$|Ax| = \sqrt{\sum_{i=1}^m |\langle x, v_i\rangle|^2} \leq \sqrt{\sum_{i=1}^m |x|^2 |v_i|^2} = |x| \sqrt{\sum_{i=1}^m |v_i|^2} = |x| \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}.$$

So, taking the supremum of both sides of this inequality over the set $\{x \in \mathbb{R}^n \colon |x| \leq 1\}$ shows that $||A|| \leq |A|$.

Suppose $x \to y$, that is $|x - y| \to 0$, and suppose that A is a continuous matrix valued function. That is, as $x \to y$, we have $|A(x) - A(y)| \to 0$. Then, using our inequality of norms, we have

$$||A(x) - A(y)|| \leq |A(x) - A(y)| \to 0 \quad \text{as} \quad x \to y$$

□

Exercise 5.4.

(i) Show that the rectangular box of maximal area that can be inscribed in the unit circle is a square.

(ii) Let $a, b, c > 0$. Find the dimensions of the box of maximal volume, whose edges are parallel to the coordinate axes, which can be inscribed in the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

Proof of (i). Let R be a rectangular box inside the unit circle, $S^1 := \{x \in \mathbb{R}^2 \colon |x| = 1\}$. Let $\rho\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation. Since $\rho(S^1) = S^1$, and since the area of R is equal to the area of ρR, we may choose a rotation ρ such that the edges of ρR are aligned with the coordinate axes, and such that ρR is still contained in S^1. Let $\lambda > 1$ and let $\delta\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the dilation defined by $\delta_\lambda(x) := \lambda x$. We now show that the rectangle must intersect S^1. The area of $\delta_\lambda R$ exceeds that of R, and also $\delta_\lambda R$ is a rectangle. If R does not intersect S^1, then there exists $\lambda > 1$ such that $\delta_\lambda R$ has larger area than R, and $\delta_\lambda R$ also does not intersect S^1. To see this, note that R and S^1 are compact, so if $R \cap S^1 = \emptyset$, then the infinimum

$\varepsilon := \inf\{|x-y| : x \in R, y \in S^1\} > 0$ is attained at some $x \in R$ and $y \in S^1$. That is, there exists $\varepsilon > 0$ such that $|x-y| \geq \varepsilon > 0$ for all $x \in R$ and for all $y \in S^1$.

So, for the purpose of finding the largest area rectangle inside of S^1, we may assume that R intersects S^1. Similarly, by applying translations and dilations separately in the x and y axes, we may assume that all four vertices of R intersect S^1.

Suppose the rectangle R has vertices at the coordinates $(x,y),(-x,y),(-x,-y),(x,-y) \in \mathbb{R}^2$, where $x \geq 0, y \geq 0$. Then the area of the rectangle is $(2x)(2y) = 4xy$. Also, since the vertices intersect S^1, we have $x^2+y^2=1$, i.e. $y=\sqrt{1-x^2}$. To find the maximum area rectangle, we therefore maximize the function $f(x) := 4x\sqrt{1-x^2}$ for $0 \leq x \leq 1$. Since $f(0)=f(1)=0$, it suffices to maximize f for $x \in (0,1)$. Since $f \in C^1(0,1)$, in order to maximize f on $(0,1)$ it suffices to check value of f at the critical points of f. Note that

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

If $f'(x)=0$, then $-2x^2+1=0$, i.e. $x=\sqrt{2}/2$. Since $f(0)=f(1)=0$, $f(\sqrt{2}/2)>0$, and $\sqrt{2}/2$ is the only critical point of f, we conclude that f has a unique maximum at $x=\sqrt{2}/2$. If $x=\sqrt{2}/2$, then $y=\sqrt{2}/2$. So, the maximum area rectangle R exists, and it is a square of side length $\sqrt{2}$. □

Proof of (ii). Let $B \subseteq \mathbb{R}^3$ be a box with edges parallel to the coordinate axes. As in part (i), by applying appropriate translations and dilations to B, we may assume that all corners of B intersect the ellipsoid.

Suppose the box B has vertices at the coordinates $(\pm x, \pm y, \pm z) \in \mathbb{R}^3$, where $x,y,z \geq 0$. Then the area of the box is $8xyz$. Also, since the vertices intersect the ellipsoid, we have $x^2/a^2+y^2/b^2+z^2/c^2=1$. To find the maximum area box, we therefore maximize the function $f(x,y,z) := xyz$ with the constraint $g(x,y,z) := x^2/a^2+y^2/b^2+z^2/c^2-1=0$, with $x,y,z, \geq 0$. Let

$$D := \{(x,y,z) \in \mathbb{R}^3 \colon x>0, y>0, z>0\}.$$

Note that $f \in C^1(D)$, $g \in C^1(D)$, and

$$\nabla f = \begin{pmatrix} yz \\ xz \\ xy \end{pmatrix}, \qquad \nabla g = \begin{pmatrix} 2x/a^2 \\ 2y/b^2 \\ 2z/c^2 \end{pmatrix}.$$

Since $\nabla g \neq 0$ on D, the Lagrange Multiplier Theorem applies (Theorem II.5.5 in Edwards).

So, suppose $(x,y,z) \in D$ is a critical point of f with respect to the constraint $g(p)=0$. Let $\lambda \in \mathbb{R}$ such that $\nabla f = \lambda \nabla g$. This system of equations says

$$yz = \lambda 2x/a^2, \qquad xz = \lambda 2y/b^2, \qquad xy = \lambda 2z/c^2.$$

Since $\nabla f \neq 0$ on D, we may assume that $\lambda \neq 0$. Substituting the first equation into the second gives $(yza^2/(2\lambda))z = 2\lambda y/b^2$, so $z^2 = 4\lambda^2/(a^2b^2)$. So, $z = 2|\lambda|/(ab)$. Substituting the second equation into the third gives $(xzb^2/(2\lambda))x = 2\lambda z/c^2$, so $x = 2|\lambda|/(bc)$. Similarly, $y = 2|\lambda|/(ac)$. Plugging these equalities for x,y,z into the condition $g(x,y,z)=0$ shows that $12\lambda^2 = a^2b^2c^2$, so $2\sqrt{3}|\lambda| = abc$. So, the only critical point we have found in D is

$$(x,y,z) = (a/\sqrt{3}, b/\sqrt{3}, c/\sqrt{3}).$$

On the boundary of D, $f = 0$. Also, $f(a/\sqrt{3}, b/\sqrt{3}, c/\sqrt{3}) = abc3^{-3/2} > 0$. So, we have found the unique maximum of f, since $f \in C^1(D)$, and by applying the Lagrange Multiplier Theorem. The box of maximal volume with edges parallel to the coordinate axes therefore has dimensions $2a/\sqrt{3}, 2b/\sqrt{3}, 2c/\sqrt{3}$. □

Exercise 5.5. Suppose that we have a probability distribution on the set $\{1, \ldots, n\}$, i.e. a sequence $p = (p_1, \ldots, p_n)$ of probabilities in the set $\overline{\mathcal{P}_n}$, where

$$\mathcal{P}_n := \left\{ p \in (0,1)^n \colon \sum_{i=1}^{n} p_i = 1 \right\}.$$

A fundamental quantity for a probability distribution p is its *entropy*

$$S(p) := -\sum_{i=1}^{n} p_i \log p_i \,.$$

(We extend the function $x \log x$ to 0 by continuity, so that $0 \log 0 := 0$.) The entropy of p measures the disorder or lack of information in p.

(i) Using Lagrange multipliers, find the critical point $\overline{p}$ of S on the set $\mathcal{P}_n$. Compute the value of S at $\overline{p}$.
(ii) Prove that S reaches its maximum on $\mathcal{P}_n$ at $\overline{p}$.
(iii) In applications to statistical physics, each point of $\{1, \ldots, n\}$ represents a state of a physical system with a given energy E_i. The *energy* of the probability distribution p is defined as

$$H(p) := \sum_{i=1}^{n} p_i E_i.$$

We now want to maximize the entropy $S(p)$ over the set $\overline{\mathcal{P}_n}$, subject to the additional constraint $H(p) = E$ for some fixed E. (The energy of the system is fixed.) We require E to satisfy $\min_{i=1,\ldots,n} E_i < E < \max_{i=1,\ldots,n} E_i$, since otherwise there may not exist a $p \in \mathcal{P}_n$ satisfying $H(p) = E$. By the method of Lagrange multipliers, prove that the unique critical point $\overline{p}$ of S in $\mathcal{P}_n \cap H^{-1}(\{E\})$ satisfies

$$\overline{p}_i \;=\; \frac{1}{Z} e^{-\beta E_i}\,, \qquad Z := \sum_{j=1}^{n} e^{-\beta E_j},$$

for some parameter β chosen so that $H(\overline{p}) = E$. This distribution is called the *canonical* or *Gibbs* distribution. The parameter β has the physical meaning of inverse temperature: $T = 1/\beta$.
(iv) Prove that S reaches its maximum on $\mathcal{P}_n \cap H^{-1}(\{E\})$ at $\overline{p}$ from (iii).

Proof of (i). We optimize S subject to the constraint $g(p) := (\sum_{i=1}^{n} p_i) - 1 = 0$. Note that $S \in C^1((0,1)^n)$, $g \in C^1((0,1)^n)$, and

$$\nabla S(p) = \begin{pmatrix} -1 - \log p_1 \\ \vdots \\ -1 - \log p_n \end{pmatrix}, \qquad \nabla g(p) = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Since $\nabla g \neq 0$ on $\mathcal{P}_n$, the Lagrange Multiplier Theorem applies (Theorem II.5.5 in Edwards)[1].

Suppose $\overline{p} \in \mathcal{P}_n$ is a critical point of S with respect to the constraint $g(\overline{p}) = 0$. Let $\lambda \in \mathbb{R}$ such that $\nabla f(\overline{p}) = \lambda \nabla g(\overline{p})$. This system of equations says

$$-1 - \log \overline{p}_1 = \lambda, \ldots, -1 - \log \overline{p}_n = \lambda.$$

So, for $i, j \in \{1, \ldots, n\}$, $\overline{p}_i = \overline{p}_j$. Since $g(\overline{p}) = 0 = (\sum_{i=1}^n \overline{p}_i) - 1 = n\overline{p}_i - 1$ for all $i \in \{1, \ldots, n\}$, we conclude that $\overline{p}_i = 1/n$ for all $i \in \{1, \ldots, n\}$. That is, we have found only one critical point $\overline{p}$ of S on $\mathcal{P}_n$. Now,

$$S(\overline{p}) = S(1/n, \ldots, 1/n) = \sum_{i=1}^{n} -(1/n)\log(1/n) = -\log(1/n) = \log(n).$$

So, for $n \geq 2$, $S(1/n, \ldots, 1/n) > 0$. In the case $n = 1$, the set $\mathcal{P}_n$ is empty, so the problem is vacuous in this case. □

Proof of (ii). We now show that $\overline{p}$ is a local maximum. Note that $S \in C^2(\mathcal{P}_n)$, $\partial^2 S/\partial p_i \partial p_j = 0$ for $i \neq j$, and $\partial^2 S/\partial p_i^2 = -1/p_i$, $i \in \{1, \ldots, n\}$. So, for $p \in \mathcal{P}_n$, the matrix of second derivatives is negative definite. Now, let $h \in \mathbb{R}^n$ such that $g(\overline{p} + h) = 0$. Note that g is linear, so $g(\overline{p} + h) = g(\overline{p}) + g(h) = g(h) = 0$. From Taylor's Theorem,

$$S(h) = S(\overline{p}) + (h - \overline{p})D_h S(\overline{p}) + \frac{1}{2}(h - \overline{p})D_h^2 S(\overline{p})(h - \overline{p}) + o(|h - \overline{p}|^2). \qquad (*)$$

Since $\overline{p}$ is a critical point of S subject to the constraint $g(\overline{p}) = 0$, and $g(\overline{p} + th) = g(\overline{p}) + tg(h) = 0$ for all $t \in [0, 1]$, we conclude that $D_h S(\overline{p}) = 0$. Also, since $D_h^2 S(\overline{p})$ is negative definite, there exists $c > 0$ such that $(h - \overline{p})D_h^2 S(\overline{p})(h - \overline{p}) \leq -c\,|h - \overline{p}|^2$, where c does not depend on h. So, from $(*)$,

$$S(h) - S(\overline{p}) \leq -c\,|h - \overline{p}|^2 + o(|h - \overline{p}|^2).$$

That is, there exists a sufficiently small neighborhood V of $\overline{p}$ such that $h \in V$ implies $S(h) - S(\overline{p}) \leq -(c/2)\,|h - \overline{p}|^2$. Since c does not depend on h, this inequality shows that $\overline{p}$ is a local maximum of S with respect to the constraint $g(\overline{p}) = 0$.

Now, the set $\{p \in [0, 1]^n : g(p) = 0\} = \overline{\mathcal{P}_n}$ is compact, so there exists a global maximum of S on this set. Note that S is the sum of strictly concave functions, so S is strictly concave. However, we have found a local maximum of S on $\mathcal{P}_n$, so the global maximum of S must occur on $\mathcal{P}_n \subseteq \overline{\mathcal{P}_n}$. To see this, we argue by contradiction. Let $p' \in \overline{\mathcal{P}_n} \setminus \mathcal{P}_n$ such that $S(p') \geq S(\overline{p})$. Let $\lambda \in (0, 1)$. Then strict concavity shows that

$$S(\lambda\overline{p} + (1 - \lambda)p') > \lambda S(\overline{p}) + (1 - \lambda)S(p') \geq S(\overline{p}).$$

Letting $\lambda \to 1$, the point $\lambda\overline{p} + (1 - \lambda)p'$ converges to $\overline{p}$. Also, since $g(p) = -1 + \sum_{i=1}^n p_i$, $g(\lambda\overline{p} + (1 - \lambda)p') = \lambda g(\overline{p}) + (1 - \lambda)g(p') = 0$. Moreover, since $\overline{p} \in \mathcal{P}_n$, $\lambda\overline{p} + (1 - \lambda)p' \in \mathcal{P}_n$. However, $S(\lambda\overline{p} + (1 - \lambda)p') > S(\overline{p})$, so $\overline{p}$ is not a local maximum of S on $\mathcal{P}_n$.

Since we have achieved a contradiction, we conclude that the global maximum of S on $\overline{\mathcal{P}_n}$, must occur on $\mathcal{P}_n$. Finally, since we have only found one critical point $\overline{p}$ of S on $\mathcal{P}_n$, and it is a local maximum, and since no other point in $\overline{\mathcal{P}_n}$ is a local maximum, we conclude that $\overline{p}$ is also a global maximum of S on $\overline{\mathcal{P}_n}$. □

[1]Strictly speaking, the theorems in Edwards begin with the assumption that there exists a local maximum or local minimum of S subject to the constraint $g = 0$. However, the theorems hold, and their proofs are identical, if we only assume that there exists a critical point of S subject to the constraint $g = 0$.

Proof of (iii). Suppose $n \geq 3$. Define $g\colon \mathbb{R}^n \to \mathbb{R}^2$, by

$$g(p) := \left(\left(\sum_{i=1}^{n} p_i \right) - 1, \left(\sum_{i=1}^{n} p_i E_i \right) - E \right) =: (g_1(p), g_2(p)).$$

We optimize S subject to the constraint $g(p) = (0,0)$. Note that $S \in C^1((0,1)^n, \mathbb{R})$, $g \in C^1((0,1)^n, \mathbb{R}^2)$, and

$$\nabla S(p) = \begin{pmatrix} -1 - \log p_1 \\ \vdots \\ -1 - \log p_n \end{pmatrix}, \qquad \nabla g_1(p) = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \qquad \nabla g_2(p) = \begin{pmatrix} E_1 \\ \vdots \\ E_n \end{pmatrix}.$$

Suppose $\overline{p} \in \mathcal{P}_n$ is a critical point of S with respect to the constraint $g(\overline{p}) = 0$.

We first discuss the constraint condition $g(p) = (0,0)$. Note that p satisfies $g(p) = (0,0)$ if and only if p lies in the intersection of the two hyperplanes $\{p\colon g_1(p) = 0\}$ and $\{p\colon g_2(p) = 0\}$. In order for our optimization problem to be nontrivial, we need the intersection of these two hyperplanes to be nontrivial, and we need the set $\{p \in (0,1)^n\colon g(p) = (0,0)\}$ to be nonempty. Since $\min_{i=1,\ldots,n} E_i < E < \max_{i=1,\ldots,n} E_i$, there exists $p \in \mathcal{P}_n$ such that $H(p) = E$, i.e. there exists $p \in (0,1)^n$ such that $g(p) = (0,0)$. To see this, reorder the numbers $E_1, \ldots, E_n$ so that

$$E_1 \leq E_2 \leq \cdots \leq E_j \leq E \leq E_{j+1} \leq \cdots \leq E_n.$$

In particular, $(1/j) \sum_{i=1}^{j} E_i \leq E \leq (1/(n-j)) \sum_{i=j+1}^{n} E_i$. So, $\exists\ t \in (0,1)$ such that

$$E = \frac{t}{j} \left(\sum_{i=1}^{j} E_i \right) + \frac{1-t}{n-j} \left(\sum_{i=j+1}^{n} E_i \right).$$

Now, define

$$r := \frac{t}{j} \left(\sum_{i=1}^{j} e_i \right) + \frac{1-t}{n-j} \left(\sum_{i=j+1}^{n} e_i \right) \in (0,1)^n.$$

Note that $\sum_{i=1}^{n} r_i = 1$ and $H(r) = E$, so $g(r) = 0$. In conclusion, the set $\{p \in (0,1)^n\colon g(p) = (0,0)\}$ is nonempty.

Also, since $\min_{i=1,\ldots,n} E_i < \max_{i=1,\ldots,n} E_i$, it cannot occur that $E_i = E_j$ for all $i, j \in \{1, \ldots, n\}$. Therefore, $\nabla g_1(\overline{p})$ and $\nabla g_2(\overline{p})$ are linearly independent. So, the Lagrange Multiplier Theorem applies (Theorem II.5.8 in Edwards). Let $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\nabla S(\overline{p}) = \lambda_1 \nabla g_1(\overline{p}) + \lambda_2 \nabla g_2(\overline{p})$. This system of equations says that, for all $i \in \{1, \ldots, n\}$,

$$-1 - \log \overline{p}_i = \lambda_1 + \lambda_2 E_i. \qquad (*)$$

Multiplying both sides by $\overline{p}_i$ shows that $-\overline{p}_i - \overline{p}_i \log \overline{p}_i = \lambda_1 \overline{p}_i + \lambda_2 \overline{p}_i E_i$. Then, by summation, and by applying the constraint $g(\overline{p}) = (0,0)$,

$$-1 + S(\overline{p}) = \sum_{i=1}^{n} (-\overline{p}_i - \overline{p}_i \log \overline{p}_i) = \sum_{i=1}^{n} (\lambda_1 \overline{p}_i + \lambda_2 \overline{p}_i E_i) = \lambda_1 + \lambda_2 E.$$

Substituting this equality into $(*)$, we get that for all $i \in \{1, \ldots, n\}$,

$$-1 - \log \overline{p}_i = -1 + S(\overline{p}) - \lambda_2 E + \lambda_2 E_i.$$

That is, $-\log \overline{p}_i = S(\overline{p}) + \lambda_2 (E_i - E)$, so

$$\overline{p}_i = e^{-\lambda_2 E_i} e^{\lambda_2 E} e^{-S(\overline{p})}. \qquad (**)$$

Summing $(**)$ over i,

$$1 = \sum_{i=1}^{n} \overline{p}_i = \sum_{i=1}^{n} e^{-\lambda_2 E_i} e^{\lambda_2 E} e^{-S(\overline{p})} = e^{\lambda_2 E} e^{-S(\overline{p})} \sum_{i=1}^{n} e^{-\lambda_2 E_i}.$$

Substituting this equality into $(**)$, we get that, for all $i \in \{1, \ldots, n\}$,

$$\overline{p}_i = \frac{e^{-\lambda_2 E_i}}{\sum_{j=1}^{n} e^{-\lambda_2 E_j}}. \qquad (\ddagger)$$

So, a critical point of S must satisfy $(\ddagger)$, for all $i \in \{1, \ldots, n\}$.

Note that, from our application of Lagrange Multipliers, we do not yet know that the critical point $\overline{p}$ satisfies $\overline{p} \in \mathcal{P}_n$. So, we now show that $\overline{p} \in \mathcal{P}_n$. From $(\ddagger)$, note that $\overline{p}_i \neq 0$. If $\overline{p}_i \geq 1$ for some $i \in \{1, \ldots, n\}$, then $e^{-\lambda_2 E_i} \geq \sum_{i=1}^{n} e^{-\lambda_2 E_i}$, which cannot occur. So, $0 < \overline{p}_i < 1$ for all $i \in \{1, \ldots, n\}$. Also from $(\ddagger)$, $\sum_{i=1}^{n} \overline{p}_i = 1$. That is, $\overline{p} \in \mathcal{P}_n$.

We now show that the constant λ_2 is unique. It suffices to show that there is a unique $\lambda_2 \in \mathbb{R}$ such that $H(\overline{p}) = E$. That is, it suffices to show there is a unique $\beta \in \mathbb{R}$ such that $\sum_{i=1}^{n} E_i e^{-\beta E_i} = E \sum_{i=1}^{n} e^{-\beta E_i}$. Define $f \colon \mathbb{R} \to \mathbb{R}$ by

$$f(\beta) := \frac{\sum_{i=1}^{n} E_i e^{-\beta E_i}}{\sum_{i=1}^{n} e^{-\beta E_i}}.$$

Note that

$$\lim_{\beta \to \infty} f(\beta) = \min_{i=1,\ldots,n} E_i < E < \max_{i=1,\ldots,n} E_i = \lim_{\beta \to -\infty} f(\beta).$$

So, to find our unique β such that $f(\beta) = E$, it suffices to show that $f'(\beta) < 0$. Observe

$$f'(\beta) = \frac{-\left(\sum_{i=1}^{n} e^{-\beta E_i}\right)\left(\sum_{i=1}^{n} E_i^2 e^{-\beta E_i}\right) + \left(\sum_{i=1}^{n} E_i e^{-\beta E_i}\right)^2}{\left(\sum_{i=1}^{n} e^{-\beta E_i}\right)^2}.$$

To show that $f'(\beta) < 0$, it therefore suffices to show that

$$\left(\frac{\sum_{i=1}^{n} E_i e^{-\beta E_i}}{\sum_{i=1}^{n} e^{-\beta E_i}}\right)^2 < \frac{\sum_{i=1}^{n} E_i^2 e^{-\beta E_i}}{\sum_{i=1}^{n} e^{-\beta E_i}}.$$

Using $(\ddagger)$, we therefore need to show that

$$\left(\sum_{i=1}^{n} E_i \overline{p}_i\right)^2 < \sum_{i=1}^{n} E_i^2 \overline{p}_i.$$

This inequality follows from the strict convexity of the map $t \mapsto t^2$, $t \in \mathbb{R}$, $0 < \overline{p}_i < 1$, and $\sum_{i=1}^{n} \overline{p}_i = 1$. Also, in order to get the strict inequality, we need to use $\min_{i=1,\ldots,n} E_i < \max_{i=1,\ldots,n} E_i$. This inequality is also known as Jensen's inequality.

In conclusion, the constant $\lambda_2 = \beta$ is unique in $(\ddagger)$. That is, there exist exactly one critical point of S on $\mathcal{P}_n \cap H^{-1}(\{E\})$.

So, the case $n \geq 3$ of the problem is concluded. It remains to check the cases $n = 1$ and $n = 2$. If $n = 1$, $\mathcal{P}_n$ is empty, so the problem is vacuous in this case. If $n = 2$, then $\overline{p}_1 + \overline{p}_2 = 1$, $0 < p\overline{p}_1, p\overline{p}_2 < 1$, and $\overline{p}_1 E_1 + \overline{p}_2 E_2 = E$. The constraints $g(\overline{p}) = (0, 0)$ give the equations of two lines in $\mathbb{R}^2$. Since ∇g_1 and ∇g_2 are linearly independent, these lines intersect in a single point. So, there is only one point p such that $g(p) = 0$. So p is trivially a critical point of S subject to $g(p) = 0$. Since $n = 2$, there exists $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $(*)$

holds, by linear algebra. The rest of the argument proceeds as above, so the conclusion of the problem holds in the case $n = 2$. $\square$

Proof of (iv). We now note that the critical point $\overline{p}$ is a local maximum of S on $\{p \in \overline{\mathcal{P}_n} \colon g(p) = 0\} = \overline{\mathcal{P}_n} \cap H^{-1}(E)$. The argument is identical to that given in part (ii). We now note that a global maximum of S on $\overline{\mathcal{P}_n} \cap H^{-1}(E)$ exists in $\mathcal{P}_n$. The argument is identical to that given in part (ii). Since we have only found one critical point $\overline{p}$ of S on $\mathcal{P}_n \cap H^{-1}(E)$ from part (iii), we conclude that $\overline{p}$ is the unique global maximum of S on $\overline{\mathcal{P}_n} \cap H^{-1}(E)$. $\square$

Exercise 5.6.

(i) Let $p_1, \ldots, p_n$ be positive real numbers satisfying $p_1 + \cdots + p_n = 1$, and define the functions

$$\varphi(x) := p_1 x_1 + \cdots + p_n x_n - 1\,, \qquad f(x) := x_1^{p_1} \cdots x_n^{p_n}.$$

Define the subset

$$M := \{x \in \mathbb{R}^n \colon \varphi(x) = 0 \text{ and } x_i > 0 \text{ for all } i\}.$$

Show that $f(x) > 0$ in M and $f(x) = 0$ in $\overline{M} \setminus M$. Conclude that f has a global maximum in M.

(ii) Find the global maximum of f in M using Lagrange multipliers. Conclude that $f(x) \leq 1$ in M.

(iii) Use (ii) to prove the inequality

$$a_1^{p_1} \cdots a_n^{p_n} \;\leq\; p_1 a_1 + \cdots + p_n a_n\,,$$

where $a_1, \ldots, a_n$ are positive real numbers. In the special case $p_i = 1/n$ for all i, this inequality reduces to the famous fact that the geometric mean is less then or equal to the arithmetic mean.

Proof of (i). Let $x \in M$. Since $x_i > 0$ and $p_i > 0$ for all $i \in \{1, \ldots, n\}$, we get $f(x) > 0$. We first show that the set $\overline{M}$ is compact. Let $x \in M$, and define $p := (p_1, \ldots, p_n) \in \mathbb{R}^n$. Since $x \in M$, $\langle x, p \rangle = 1$. Since $x_i > 0$ and $p_i > 0$ for all $i \in \{1, \ldots, n\}$, we have

$$1 = |\langle x, p \rangle| = \sum_{j=1}^{n} x_j p_j > x_i p_i.$$

So, $x_i < 1/p_i < 1/(\min_{j=1,\ldots,n} p_j) < \infty$ for all $i \in \{1, \ldots, n\}$, and therefore

$$\max_{i=1,\ldots,n} x_i < \frac{1}{\min_{j=1,\ldots,n} p_j} < \infty.$$

So, $\overline{M}$ is contained in the compact set $\{x \in \mathbb{R}^n \colon \forall\, i \in \{1, \ldots, n\}, 0 \leq x_i \leq 1/\min_{j=1,\ldots,n} p_j\}$. Since $\overline{M}$ is a closed set contained in another compact set, we conclude that $\overline{M}$ is compact.

We now show that

$$\begin{aligned}\overline{M} \setminus M = &\{x \in \mathbb{R}^n \colon \varphi(x) = 0, x_j \geq 0 \;\forall\, j \in \{1, \ldots, n\},\\ &\qquad\qquad \text{and } \exists\, i \in \{1, \ldots, n\} \text{ such that } x_i = 0\}. \qquad (*)\end{aligned}$$

Let $x \in \overline{M} \setminus M$. Then there exists $x^{(k)} \to x$ as $k \to \infty$ with $x^{(k)} \in M$. Then $x = (x_1, \ldots, x_n)$ is in the compact set $\overline{M}$. Since φ is continuous, $\varphi(x) = \lim_{k \to \infty} \varphi(x^{(k)}) = 0$. Also, since the coordinate projections are continuous, $x_i \geq \lim_{k\to\infty} x_i^{(k)} \geq 0$. If $x_i > 0$ for all $i \in \{1,\ldots,n\}$, then $x \in M$. Since $x \notin M$, We conclude that $x_i = 0$ for some $i \in \{1,\ldots,n\}$, so that x is in the right side of $(*)$. We now show the other containment. Let x in the right side of $(*)$. For $k \in \mathbb{Z}_{>0}$, define

$$y^{(k)} := (1 - 1/k)x + \frac{1}{k}\frac{1}{n}(1/p_1, \ldots, 1/p_n) \in \mathbb{R}^n.$$

Since $x_i \geq 0$ for $i \in \{1,\ldots,n\}$, we have $y_i^{(k)} > 0$ for $i \in \{1,\ldots,n\}$. Also, since $\varphi(x) = 0$ and $\varphi(1/(np_1),\ldots,1/(np_n)) = 0$, and since $\varphi(x) = \langle x, p\rangle - 1$, we conclude that

$$\varphi(y^{(k)}) = (1 - 1/k)\varphi(x) + (1/k)\varphi(1/(np_1),\ldots,1/(np_n)) = 0.$$

Finally, $y^{(k)} \to x$ as $k \to \infty$, so $x \in \overline{M} \smallsetminus M$. In conclusion, $(*)$ holds.

We can now conclude this part of the Exercise. From $(*)$, we must have $f = 0$ on $\overline{M} \smallsetminus M$. Also, since $\overline{M}$ is compact, f is continuous, and $f > 0$ on M, we conclude that there exists a global maximum of f on M. □

Proof of (ii). From part (i), it suffices to maximize f on M and ignore $\overline{M} \smallsetminus M$. Define

$$g(x) := \varphi(x) - 1.$$

We maximize f with respect to the constraint $g(x) = 0$. Note that, with this constraint, an application of Lagrange Multipliers could find a maximum of f outside of M. Note also that $f \in C^1(M)$ and $g \in C^1(M)$. Now,

$$\nabla f(x) = \begin{pmatrix} p_1 x_1^{p_1 - 1} x_2^{p_2} \cdots x_n^{p_n} \\ \vdots \\ p_n x_1^{p_1} x_2^{p_2} \cdots x_n^{p_n - 1} \end{pmatrix}, \qquad \nabla g(x) = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}.$$

Since $\nabla g \neq 0$, the Lagrange Multiplier Theorem applies (Theorem II.5.5 in Edwards). Let x be a critical point of f subject to the constraint $g(x) = 0$.

Then there exists $\lambda \in \mathbb{R}$ such that $\nabla f(x) = \lambda \nabla g(x)$. Therefore, for $i \in \{1,\ldots,n\}$,

$$x_i^{p_i - 1} \prod_{j \neq i} x_j^{p_j} = \lambda.$$

That is, $x_i = f(x)$. So, $\sum_{i=1}^n x_i p_i = \sum_{i=1}^n p_i f(x) = f(x)$. Since $\varphi(x) = 0 = \langle x, p\rangle - 1$, we have $\sum_{i=1}^n x_i p_i = \langle x, p\rangle = 1$, so $f(x) = 1$. Since $x_i = f(x) = 1$, and $\varphi(1,\ldots,1) = 0$, we conclude that $(1,\ldots,1) \in M$. Since the application of Lagrange Multipliers has only found one critical point of f, part (i) implies that $(1,\ldots,1)$ is the global maximum of f on M, so $f(x) \leq 1$ for $x \in M$. □

Proof of (iii). Let $a_1,\ldots,a_n > 0$. Define

$$x := \frac{(a_1,\ldots,a_n)}{a_1 p_1 + \cdots + a_n p_n} \in \mathbb{R}^n.$$

By the definition of x, $\langle x, p\rangle = 1$, so $\varphi(x) = 0$. Also, $x_i > 0$ for all $i \in (1,\ldots,n)$, so $x \in M$. From part (ii), $f(x) \leq 1$. Using the definitions of x and $f(x)$, and the equality $\sum_{i=1}^n p_i = 1$, we see that the inequality $f(x) \leq 1$ says that

$$\frac{a_1^{p_1} \cdots a_n^{p_n}}{a_1 p_1 + \cdots + a_n p_n} \leq 1.$$

To conclude, multiply both sides of this inequality by $\langle a, p\rangle$. □

Exercise 5.7. This problem is a digression on linear algebra and block matrices. Let A be an $n \times n$ matrix, B an $n \times m$ matrix, C an $m \times n$ matrix, and D an $m \times m$ matrix. We can put all of these matrices into a block matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

(i) Prove that

$$\det \begin{pmatrix} A & 0 \\ 0 & I_m \end{pmatrix} = \det(A).$$

(ii) Using (i) and the fact that $\det(XY) = \det(X)\det(Y)$, prove that

$$\det \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} = \det(A)\det(D).$$

(iii) Prove that

$$\det \begin{pmatrix} I_n & B \\ 0 & I_m \end{pmatrix} = 1.$$

(iv) Suppose that D is invertible. Prove that

$$\begin{pmatrix} I_n & -BD^{-1} \\ 0 & I_m \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -D^{-1}C & I_m \end{pmatrix} = \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix}.$$

(v) Suppose that D is invertible. By combining, (i) – (iv), prove that

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A - BD^{-1}C)\det(D).$$

(A special case of this formula was used in class in the proof of the implicit function theorem.)

(vi) Now that you're all warmed up with block matrices, use the identity

$$\det \left[\begin{pmatrix} I_n & B \\ -C & I_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ C & I_m \end{pmatrix} \right] = \det \left[\begin{pmatrix} I_n & 0 \\ C & I_m \end{pmatrix} \begin{pmatrix} I_n & B \\ -C & I_m \end{pmatrix} \right]$$

to prove

$$\det(I_n + BC) = \det(I_m + CB).$$

This is one of the most useful identities in linear algebra, and its proof without block matrices is much harder.

Proof of (i). Let $\sigma \in S_{n+m}$, and let $M := \begin{pmatrix} A & 0 \\ 0 & I_m \end{pmatrix}$. Let i with $n+1 \leq i \leq n+m$. Suppose $\sigma(i) \neq i$. By the definition of M, we then have $M_{i\sigma(i)} = 0$. Now, let S_n' be the set of $\sigma \in S_{n+m}$ with $\sigma(i) = i$ for $n+1 \leq i \leq n+m$. Then there exists a bijection $\phi\colon S_n' \to S_n$ defined by

$$\phi(\sigma)(1, \ldots, n) := \sigma(1, \ldots, n, n+1, \ldots, n+m).$$

Moreover, $\text{sign}(\phi(\sigma)) = \text{sign}(\sigma)$. Therefore,

$$\begin{aligned}
\det M &= \sum_{\sigma\in S_{n+m}} \text{sign}(\sigma) \prod_{j=1}^{n+m} M_{j\sigma(j)} = \sum_{\substack{\sigma\in S_{n+m}\colon \sigma(i)=i,\\ \forall\, n+1\leq i\leq n+m}} \text{sign}(\sigma) \prod_{j=1}^{n+m} M_{j\sigma(j)} \\
&= \sum_{\sigma\in S_n'} \text{sign}(\sigma) \prod_{j=1}^{n} M_{j\sigma(j)} = \sum_{\sigma\in S_n} \text{sign}(\sigma) \prod_{j=1}^{n} M_{j\sigma(j)} \\
&= \sum_{\sigma\in S_n} \text{sign}(\sigma) \prod_{j=1}^{n} A_{j\sigma(j)} = \det(A).
\end{aligned}$$

□

Proof of (ii). We first suppose that $\det(A) \neq 0$. Define

$$M := \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}, \qquad H := \begin{pmatrix} A^{-1} & 0 \\ 0 & I_m \end{pmatrix}.$$

Using part (i),

$$\det(D) = \det(MH) = \det(M)\det(H) = \det(M)\det(A^{-1}) = \det(M)(\det(A))^{-1}.$$

So, $\det(M) = \det(A)\det(D)$, as desired.

Now, assume that $\det(A) = 0$. Let R be an $n \times n$ matrix that is a composition of row operations on A such that RA has a zero row. Specifically, we choose row operation matrices $R_1, \ldots, R_j$ to be upper triangular with ones along their diagonal. Note that $\det(R_i) \neq 0$ for $i \in \{1, \ldots, j\}$. We then compose these matrices together to get $R := R_j R_{j-1} \cdots R_2 R_1$. Then $\det(R) = \prod_{i=1}^{j} \det(R_i) \neq 0$. So, define

$$K := \begin{pmatrix} R & 0 \\ 0 & I_m \end{pmatrix}.$$

Since RA has a zero row, KM has a zero row, and so $\det(KM) = 0$. Then, using part (i),

$$0 = \det \begin{pmatrix} RA & 0 \\ 0 & D \end{pmatrix} = \det(KM) = \det(K)\det(M) = \det(R)\det(M).$$

Since $\det(R) \neq 0$, we conclude that $\det(M) = 0$. Since $\det(D) = 0$, we conclude that $\det(M) = \det(A)\det(D)$, as desired. □

Proof of (iii). Let $\sigma \in S_{n+m}$. Assume that there exist $1 \leq i \leq n$ and $n+1 \leq j \leq n+m$ such that $\sigma(i) = j$. We claim that there exists $1 \leq i' \leq n$ and $n+1 \leq j' \leq n+m$ such that $\sigma(j') = i'$. We argue by contradiction. Assume that, for all $n+1 \leq k \leq n+m$ we have $\sigma(k) \in \{n+1, n+2, \ldots, n+m\}$. Since σ is injective, there must be some $n+1 \leq k \leq n+m$ such that $\sigma(k) = j$, by the pigeonhole principle. But $\sigma(i) = j$ and $i \neq k$, so we have violated injectivity of σ. Since we have achieved a contradiction, the claim is proven.

We now prove the required result. Define

$$M := \begin{pmatrix} I_n & B \\ 0 & I_m \end{pmatrix}.$$

Let $\sigma \in S_{n+m}$. Assume that there exist $1 \le i \le n$ and $n+1 \le j \le n+m$ such that $\sigma(i) = j$. From the claim, there exist $1 \le i' \le n$ and $n+1 \le j' \le n+m$ such that $\sigma(j') = i'$. That is, $M_{j'\sigma(j')} = 0$. Therefore,

$$\det(M) = \sum_{\sigma \in S_{n+m}} \operatorname{sign}(\sigma) \prod_{i=1}^{n+m} M_{i\sigma(i)} = \sum_{\substack{\sigma \in S_{n+m} \\ \sigma(j) \le n, \forall 1 \le j \le n}} \operatorname{sign}(\sigma) \prod_{i=1}^{n+m} M_{i\sigma(i)}.$$

Let $\sigma \in S_{n+m}$ such that $\sigma(j) \le n$ for all $1 \le j \le n$. Then $\prod_{i=1}^{n} M_{i\sigma(i)} = 0$, unless $\sigma(j) = j$ for all $1 \le j \le n$. Also, since $\sigma(j) \in \{1, \ldots, n\}$ for all $1 \le j \le n$, we have $\sigma\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$, and $\sigma\colon \{n+1, \ldots, n+m\} \to \{n+1, \ldots, n+m\}$. And then $\prod_{i=n+1}^{n+m} M_{i\sigma(i)} = 0$, unless $\sigma(j) = j$ for all $n+1 \le j \le n+m$. So, the only nonzero term in the determinant comes from $\sigma = id$. That is,

$$\begin{aligned}\det(M) &= \sum_{\substack{\sigma \in S_{n+m} \\ \sigma(j) \le n, \forall 1 \le j \le n}} \operatorname{sign}(\sigma) \prod_{i=1}^{n+m} M_{i\sigma(i)} = \sum_{\substack{\sigma \in S_{n+m} \\ \sigma(j) \le n, \forall 1 \le j \le n}} \operatorname{sign}(\sigma) \prod_{i=1}^{n} M_{i\sigma(i)} \prod_{i'=n+1}^{n+m} M_{i'\sigma(i')} \\ &= \prod_{i=1}^{n+m} M_{ii} = 1.\end{aligned}$$

□

Proof of (iv).

$$\begin{aligned}\begin{pmatrix} I_n & -BD^{-1} \\ 0 & I_m \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -D^{-1}C & I_m \end{pmatrix} &= \begin{pmatrix} A - BD^{-1}C & 0 \\ C & D \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -D^{-1}C & I_m \end{pmatrix} \\ &= \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix}.\end{aligned}$$

□

Proof of (v).

$$\begin{aligned}\det(A - BD^{-1}C)\det(D) &= \det \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \quad \text{, by (ii)} \\ &= \det \left[\begin{pmatrix} I_n & -BD^{-1} \\ 0 & I_m \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -D^{-1}C & I_m \end{pmatrix} \right] \quad \text{, by (iv)} \\ &= \det \begin{pmatrix} I_n & -BD^{-1} \\ 0 & I_m \end{pmatrix} \det \begin{pmatrix} A & B \\ C & D \end{pmatrix} \det \begin{pmatrix} I_n & 0 \\ -D^{-1}C & I_m \end{pmatrix} \\ &= \det \begin{pmatrix} A & B \\ C & D \end{pmatrix} \det \begin{pmatrix} I_n & 0 \\ -D^{-1}C & I_m \end{pmatrix} \quad \text{, by (iii)} \\ &= \det \begin{pmatrix} A & B \\ C & D \end{pmatrix} \det \begin{pmatrix} I_n & (-D^{-1}C)^T \\ 0 & I_m \end{pmatrix} = \det \begin{pmatrix} A & B \\ C & D \end{pmatrix} \quad \text{, by (iii).}\end{aligned}$$

□

Proof of (vi). Define

$$M := \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \qquad M' := \begin{pmatrix} B^T & D^T \\ A^T & C^T \end{pmatrix}, \qquad M'' := \begin{pmatrix} D & C \\ B & A \end{pmatrix}.$$

We begin with a claim. We claim that

$$\det M = \det M''. \qquad (\ddagger)$$

To get this equality, we first perform m cyclic permutations on the columns of the block matrix, and we then perform n cyclic permutations on the rows of the block matrix. For $\sigma, \tau \in S_{n+m}$ with τ a cyclic permutation, i.e. $\tau(i) := i+1$ for $i \in \{1, \ldots, n+m-1\}$, and $\tau(n+m) = 1$, we use the identity

$$\mathrm{sign}(\sigma(\tau)) = \mathrm{sign}(\sigma)\mathrm{sign}(\tau) = \mathrm{sign}(\sigma)(-1)^{n+m-1}.$$

This identity follows since τ can be written as a product of $n+m-1$ transpositions. Now,

$$\begin{aligned}
\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} &= \sum_{\sigma \in S_{n+m}} \mathrm{sign}(\sigma) \prod_{i=1}^{n+m} M_{i\sigma(i)} = (-1)^{m(n+m-1)} \sum_{\sigma \in S_{n+m}} \mathrm{sign}(\sigma(\tau^m)) \prod_{i=1}^{n+m} M_{i\sigma(i)} \\
&= (-1)^{m(n+m-1)} \sum_{\sigma \in S_{n+m}} \mathrm{sign}(\sigma) \prod_{i=1}^{n+m} M_{i\sigma(\tau^{-m}(i))} \\
&= (-1)^{m(n+m-1)} \det \begin{pmatrix} B & A \\ D & C \end{pmatrix} = (-1)^{m(n+m-1)} \det \begin{pmatrix} B^T & D^T \\ A^T & C^T \end{pmatrix}. \qquad (*)
\end{aligned}$$

Also,

$$\begin{aligned}
\det \begin{pmatrix} B^T & D^T \\ A^T & C^T \end{pmatrix} &= \sum_{\sigma \in S_{n+m}} \mathrm{sign}(\sigma) \prod_{i=1}^{n+m} M'_{i\sigma(i)} = (-1)^{(n+m-1)n} \sum_{\sigma \in S_{n+m}} \mathrm{sign}(\sigma(\tau^n)) \prod_{i=1}^{n+m} M'_{i\sigma(i)} \\
&= (-1)^{(n+m-1)n} \sum_{\sigma \in S_{n+m}} \mathrm{sign}(\sigma) \prod_{i=1}^{n+m} M'_{i\sigma(\tau^{-n}(i))} \\
&= (-1)^{(n+m-1)n} \det \begin{pmatrix} D^T & B^T \\ C^T & A^T \end{pmatrix} = (-1)^{(n+m-1)n} \det \begin{pmatrix} D & C \\ B & A \end{pmatrix}. \qquad (**)
\end{aligned}$$

Note that $n(n+m-1) + m(n+m-1) = (n+m)(n+m-1)$, which is an even number, since it is the product of two consecutive integers. So, $(-1)^{n(n+m-1)}(-1)^{m(n+m-1)} = 1$. And $(*)$ combined with $(**)$ therefore proves our Claim $(\ddagger)$.

With this claim completed, we can now finish the exercise.

$$\begin{aligned}
\det(I_n + BC) &= \det \begin{pmatrix} I_n + BC & B \\ 0 & I_m \end{pmatrix} \quad \text{, by } (v) \\
&= \det \left[\begin{pmatrix} I_n & B \\ -C & I_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ C & I_m \end{pmatrix} \right] = \det \left[\begin{pmatrix} I_n & 0 \\ C & I_m \end{pmatrix} \begin{pmatrix} I_n & B \\ -C & I_m \end{pmatrix} \right] \\
&= \det \begin{pmatrix} I_n & B \\ 0 & CB + I_m \end{pmatrix} = \det \begin{pmatrix} CB + I_m & 0 \\ B & I_n \end{pmatrix} \quad \text{, by } (\ddagger) \\
&= \det(I_m + CB) \quad \text{, by } (v).
\end{aligned}$$

□

Exercise 5.8. Recall that Newton's method is an algorithm for finding zeros of a function f. It consists in iterating the map

$$\varphi(x) := x - (f'(x))^{-1} f(x)\,.$$

Thus, we start with some given x_0 and define $x_1 := \varphi(x_0)$, $x_2 := \varphi(x_1)$, etc.

(i) Suppose you are trying to find $\sqrt{a}$ for some $a > 0$. This amounts to finding the positive zero of the function $f(x) = x^2 - a$. Derive an algorithm for finding $\sqrt{a}$ using Newton's method. You should recover the Babylonian method given in class.

The rest of this problem is devoted to an analysis of the convergence of Newton's method. For simplicity, we work in one dimension, i.e. we set $n = 1$. Without loss of generality, we assume that the zero of f we are interested in is at the origin: $f(0) = 0$. We shall show that, assuming $f'(0)$ is invertible and f is C^2 in a neighborhood of 0, the sequence $(x_k)_{k\in\mathbb{N}}$ converges to 0 provided x_0 is close enough to 0.

Let $R > 0$ and $K > 1$ and suppose that

$$\forall\, x \in [-R, R]: \quad K^{-1} \le |f'(x)| \le K\,, \quad |f''(x)| \;\le\; K\,. \tag{2}$$

(ii) We begin by estimating $|x_{k+1} - x_k|$ in terms of $|x_k - x_{k-1}|$. Suppose that $x_k, x_{k-1} \in [-R, R]$. Prove that

$$|x_{k+1} - x_k| \;\le\; \frac{K^2}{2}\,|x_k - x_{k-1}|^2\,.$$

(iii) Prove that $|x_1 - x_0| \le K^2\,|x_0|$.

(iv) Prove that if

$$|x_0| \le \frac{\varepsilon}{K^4}$$

for some $\varepsilon \in (0,1)$, then $|x_{k+1} - x_k| \le K^{-2} 2^{-k} \varepsilon^{2^k} \le \varepsilon(\varepsilon/2)^k K^{-2}$. Conclude that if $x_0 \in [-r, r]$ and

$$r \le \frac{\varepsilon}{K^4}\,, \qquad R \ge \varepsilon\left(\frac{1}{K^4} + \frac{2}{K^2}\right), \tag{3}$$

then the sequence $(x_k)_{k\in\mathbb{N}}$ converges in $[-R, R]$. Show that this limit is a fixed point of φ, and hence 0.

(v) Show that there exist $0 < r < R$ such that (2) and (3) are satisfied. This shows that Newton's method will find the zero of f at 0 provided one starts sufficiently close to it (in this case in the interval $[-r, r]$).

Proof of (i). Let $f(x) := x^2 - a$, $a > 0$. Note that $f(\sqrt{a}) = a - a = 0$. So, to find $\sqrt{a}$, we want to find a zero of f, via Newton's method. Let $x_0 = x \in \mathbb{R}$. For $i \in \mathbb{Z}_{>0}$, Newton's Method says that we should define

$$x_{i+1} := x_i - \frac{f(x_i)}{f'(x_i)} = x_i - \frac{x_i^2 - a}{2x_i} = \frac{1}{2}\left(x_i - \frac{a}{x_i}\right).$$

□

Proof of (ii). Since $f \in C^2[-R, R]$, Taylor's Theorem applies at $x = x_{k-1}$. Let $x \in [-R, R]$. Then there exists $\xi \in [-R, R]$ such that

$$f(x_{k-1} + y) = f(x_{k-1}) + y f'(x_{k-1}) + (1/2) y^2 f''(\xi). \qquad (*)$$

So, using the definition of φ and x_k, and using $(*)$,

$$x_{k+1} - x_k = \varphi(x_k) - x_k = -\frac{f(x_k)}{f'(x_k)} = -\frac{f\left(x_{k-1} - \frac{f(x_{k-1})}{f'(x_{k-1})}\right)}{f'(x_k)}$$

$$= -\frac{f(x_{k-1}) - \frac{f(x_{k-1})}{f'(x_{k-1})}f'(x_{k-1}) + (1/2)\left(-\frac{f(x_{k-1})}{f'(x_{k-1})}\right)^2 f''(\xi)}{f'(x_k)}$$

$$= -\frac{f''(\xi)}{2f'(x_k)}\left(\frac{f(x_{k-1})}{f'(x_{k-1})}\right)^2 = -\frac{f''(\xi)}{2f'(x_k)}\left(x_k - x_{k-1}\right)^2.$$

Note also that $f' \neq 0$ on $[-R, R]$ by our assumptions, so we can freely divide by f', since $x_k, x_{k-1} \in [-R, R]$. Now, using the above equality and our derivative bounds,

$$|x_{k+1} - x_k| \leq \frac{1}{2} \sup_{\xi \in [-R,R], y \in [-R,R]} |f''(\xi)| \, |f'(y)|^{-1} \, |x_k - x_{k-1}|^2 \leq \frac{K^2}{2} \, |x_k - x_{k-1}|^2.$$

□

Proof of (iii). As in part (i), $f \in C^2[-R, R]$ and $f' \neq 0$ on $[-R, R]$, so we apply Taylor's Theorem at $x = 0$. Let $x \in [-R, R]$. Then $\exists\, \xi \in [-R, R]$ such that $f(x) = xf'(\xi)$. Therefore,

$$|x_1 - x_0| = \left|\frac{f(x_0)}{f'(x_0)}\right| = \left|x_0 \frac{f'(\xi)}{f'(x_0)}\right| \leq K^2 \, |x_0|.$$

□

Proof of (iv). We prove by induction on k that $|x_{k+1} - x_k| \leq K^{-2}2^{-k}\varepsilon^{2^k}$ for some $0 < \varepsilon < 1$. Since $|x_0| \leq \varepsilon K^{-4}$, part (iii) and $K > 1$ show that $|x_1 - x_0| \leq \varepsilon K^{-2}$. So, we now prove the inductive step. Let $k \geq 0$ such that $|x_{k+1} - x_k| \leq K^{-2}2^{-k}\varepsilon^{2^k}$. Now, by part (ii),

$$|x_{k+2} - x_{k+1}| \leq \frac{K^2}{2} \, |x_{k+1} - x_k|^2 \leq \frac{K^2}{2} K^{-4} 2^{-2k} \varepsilon^{2^{k+1}} = K^{-2}2^{-2k-1}\varepsilon^{2^{k+1}} \leq K^{-2}2^{-(k+1)}\varepsilon^{2^{k+1}}.$$

The inductive step is complete, so the claim is complete. Also, note that $k + 1 \leq 2^k$, so

$$|x_{k+1} - x_k| \leq K^{-2}2^{-k}\varepsilon^{2^k} \leq K^{-2}\varepsilon(\varepsilon/2)^k.$$

Now, let $x_0 \in [-r, r]$, $r \leq \varepsilon K^{-4}$, $R \geq \varepsilon(K^{-4} + 2K^{-2})$. Then $|x_0| \leq \varepsilon K^{-4}$, so the claim above implies that $|x_{k+1} - x_k| \leq \varepsilon(\varepsilon/2)^k K^{-2}$. We deduce from this condition that the sequence $\{x_k\}_{k\geq 0}$ is Cauchy. Let $m, n \geq 0$ with $m > n$. Then

$$|x_m - x_n| = \left|\sum_{k=n}^{m-1}(x_{k+1} - x_k)\right| \leq \sum_{k=n}^{m-1} |x_{k+1} - x_k| \leq \sum_{k \geq n} \varepsilon(\varepsilon/2)^k K^{-2} = K^{-2}\varepsilon \frac{(\varepsilon/2)^n}{1 - \varepsilon/2}. \qquad (*)$$

So, the sequence $\{x_k\}_{k\geq 0}$ is Cauchy, since the last quantity in $(*)$ becomes arbitrarily small as $n \to \infty$. Then, there exists x such that $x_k \to x$ as $k \to \infty$. Taking $n = 0$ in $(*)$, and using that $0 < \varepsilon < 1$, we have $|x_m - x_0| < K^{-2}2\varepsilon$. Therefore,

$$|x| \leq |x - x_0| + |x_0| \leq |x_0| + \lim_{m\to\infty} |x_m - x_0| \leq \varepsilon K^{-4} + 2\varepsilon K^{-2} \leq R.$$

That is, $\{x_k\}_{k\geq 0}$ converges to a point $x\in[-R,R]$, and $x_k\in[-R,R]$ for all $k\geq 0$. Since $K^{-1}\leq|f'(y)|\leq K$ for all $y\in[-R,R]$, $f'(y)\neq 0$ for all $y\in[-R,R]$. So, $\varphi(x)$ and $\varphi(x_k)$ are defined for all $k\geq 0$. Moreover, φ is a composition of continuous functions, so

$$\varphi(x)=\lim_{k\to\infty}\varphi(x_k)=\lim_{k\to\infty}x_{k+1}=x.$$

So, $x=x-f(x)/f'(x)$, and $f(x)/f'(x)=0$, so $f(x)=0$. Since $f\in C^1[-R,R]$ and $K^{-1}\leq|f'(y)|\leq K$ for all $y\in[-R,R]$, the Intermediate Value Theorem implies that $f|_{[-R,R]}$ is equal to zero only at the point $y=0$. We therefore conclude that $x=0$. □

Proof of (v). Since $f'(0)\neq 0$ and f' is continuous in a neighborhood of the origin, there exists $R'>0$ and $k>1$ such that, for all $y\in[-R',R']$, $k^{-1}\leq|f'(y)|\leq k$. Also, since f is C^2 in a neighborhood of the origin, there exists $R>0$ and $K>1$ such that $R\leq R'$ and $K\geq k$ such that $|f''(y)|\leq K$ for all $y\in[-R,R]$. Since $R\leq R'$ and $K\geq k$, we conclude that $K^{-1}\leq|f'(y)|\leq K$ for all $y\in[-R,R]$. Now, let $\varepsilon>0$ small such that $R\geq\varepsilon(K^{-4}+2K^{-2})$, choose any r with $r\leq\varepsilon K^{-4}$, and choose any x_0 with $|x_0|\leq\varepsilon K^{-4}$. □

Exercise 5.9. Prove that if $f\in C^k$ satisfies the assumptions of the inverse function theorem, then the local inverse f^{-1} is also C^k. Formulate and prove a similar statement for the implicit function theorem.

Proof. Let U,V be open sets such that $f^{-1}\colon V\to U$. Define the set

$$GL(n,\mathbb{R}):=\{A\in\mathbb{R}^{n\times n}\colon\ \det(A)\neq 0\}.$$

Define $\iota\colon GL(n,\mathbb{R})\to GL(n,\mathbb{R})$ by $\iota(A):=A^{-1}$. Recall that $D(f^{-1})(x)=[D(f)(f^{-1}(x))]^{-1}$. That is, $D(f^{-1})(x)$ is the composition of three functions: $f^{-1}(x)$, followed by $D(f)$, followed by ι. We know from the Inverse Function Theorem that $F^{-1}\in C^1(V)$. We want to show that $f^{-1}\in C^k(V)$. We argue by induction. Assume that $f^{-1}\in C^j(V)$ for some $1\leq j<k$. Then $f^{-1}\in C^j(V)$, $D(f)\in C^j(U)$ since $j<k$ and $f\in C^k(U)$, and $\iota\in C^\infty(GL(n,\mathbb{R});GL(n,\mathbb{R}))$, since ι is a rational function whose denominator is the determinant function. The latter fact was found using Cramer's rule in Exercise 3.11(ii). So, from our composition formula $D(f^{-1})(x)=\iota[D(f)(f^{-1}(x))]$, we conclude that $D(f^{-1})\in C^j(U)$, since each of the three compositions are C^j. Since $D(f^{-1})\in C^j(U)$, we conclude that $f^{-1}\in C^{j+1}(U)$. The inductive step is therefore complete. We conclude that $f^{-1}\in C^k(U)$, as desired. □

Theorem 5.10. *Let $x,a\in\mathbb{R}^m$, $b,y\in\mathbb{R}^n$, $x=(x_1,\ldots,x_m)$, $y=(y_1,\ldots,y_n)$. Let $G\colon\mathbb{R}^{m+n}\to\mathbb{R}^n$ be C^k in a neighborhood of (a,b). Assume that $G(a,b)=0$. Assume also that the $n\times n$ matrix $(\partial G_i/\partial y_j(a,b))_{1\leq i\leq n,1\leq j\leq n}$ is not singular. Then there exists $U\subseteq\mathbb{R}^m$ and $W\subseteq\mathbb{R}^{m+n}$ open such that $a\in U$, $(a,b)\in W$, and there exists $h\colon U\to\mathbb{R}^n$ that is C^k such that $y=h(x)$ solves $G(x,y)=0$ for $(x,y)\in W$.*

Proof. Define $f(x,y):=(x,G(x,y))$, $f\colon\mathbb{R}^{m+n}\to\mathbb{R}^{m+n}$. Then $f(a,b)=(a,0)$. Also,

$$f'=\begin{pmatrix} I_m & 0\\ \left(\frac{\partial G_i}{\partial x_j}(a,b)\right)_{\substack{1\leq i\leq n\\ 1\leq j\leq m}} & \left(\frac{\partial G_i}{\partial y_j}(a,b)\right)_{\substack{1\leq i\leq n\\ 1\leq j\leq n}}\end{pmatrix}.$$

Using Claim (‡) of Exercise 5.7(vi), and using the result of Exercise 5.7(vi),

$$\det f'=\det\begin{pmatrix}\left(\frac{\partial G_i}{\partial y_j}(a,b)\right)_{\substack{1\leq i\leq n\\ 1\leq j\leq n}} & \left(\frac{\partial G_i}{\partial x_j}(a,b)\right)_{\substack{1\leq i\leq n\\ 1\leq j\leq m}}\\ 0 & I_m\end{pmatrix}=\det\left(\frac{\partial G_i}{\partial y_j}(a,b)\right)_{\substack{1\leq i\leq n\\ 1\leq j\leq n}}.$$

So, using our assumption, $\det f' \neq 0$. Also, by the definition of f, and using that G is C^k, we know that f is also C^k in a neighborhood of (a,b). So, using our modified version of the Inverse Function Theorem (modeled after Theorem III.3.3 in Edwards), $\exists\, W, V \subseteq \mathbb{R}^{m+n}$ open such that $(a,b) \in W$, $(a,0) \in V$, and $\exists\, g\colon V \to W$, $g \in C^k$ such that $g(a,0) = (a,b)$, and $g = f^{-1}$. Let $U := V \cap (\mathbb{R}^m \times \{0\})$, $U \subseteq \mathbb{R}^m$. Since $(a,0) \in V$, we have $(a,0) \in U$. Write $g(x,y) =: (A(x,y), B(x,y))$, $A(x,y) \in \mathbb{R}^m$, $B(x,y) \in \mathbb{R}^n$. Then for $(x,y) \in V$,

$$(x,y) = f(g(x,y)) = f(A(x,y), B(x,y)) = (A(x,y), G(A(x,y), B(x,y)))$$

Therefore, $A(x,y) = x$, and so

$$g(x,y) = (x, B(x,y)).$$

For $x \in U$, let $h(x) := B(x,0)$. Then

$$(x,0) = f(g(x,0)) = f(x, B(x,0)) = (x, G(x, B(x,0))).$$

Therefore,

$$0 = G(x, B(x,0)) = G(x, h(x)).$$

So, we have shown that $x \in U$ implies that $G(x,h(x)) = 0$. We now prove the converse. Suppose $G(x,y) = 0$ for $(x,y) \in W$. Then $f(x,y) = (x, G(x,y)) = (x,0)$, so

$$(x,y) = g(f(x,y)) = g(x,0) = (x, B(x,0)) = (x, h(x)).$$

Therefore, $y = h(x)$. And $g(x,0) = (x,h(x))$. □

6. Problem Set 6

Recall that a subset $M \subseteq \mathbb{R}^n$ is called a k-dimensional C^ℓ-manifold if every $p \in M$ has an open neighborhood U with $p \in U$, and there exists a diffeomorphism $\Psi \in C^\ell(U;\mathbb{R}^n)$ from U onto $V := \Psi(U)$ such that

$$\Psi(U \cap M) = V \cap (\mathbb{R}^k \times \{0\}).$$

Recall that in class we saw two ways of generating a manifold.

(a) *Using a graph (Example 15).* Let $W \subseteq \mathbb{R}^k$ be open and let $g \in C^\ell(W;\mathbb{R}^{n-k})$. Then the graph of g, defined as

$$M = G(g) := \{(x, g(x)) \colon x \in W\},$$

is a k-dimensional C^ℓ manifold.

(b) *As the preimage of a regular value (Example 16).* Let $\varphi \in C^\ell(U;\mathbb{R}^{n-k})$ for some open set $U \subseteq \mathbb{R}^n$. Then the set $M := \varphi^{-1}(0)$ is a k-dimensional C^ℓ manifold if 0 is a regular value of φ, i.e. if $\operatorname{rank}(\varphi'(p)) = n-k$ for all $p \in M$.

Exercise 6.1. In class we saw that the tangent space of a graph $M = G(g)$ as in (a) is given by $T_pM = \Phi'(x)\mathbb{R}^k$, where $\Phi(x) := (x, g(x))$ and $p = \Phi(x)$. Consider the special case $k = 1$ and $n = 2$, and verify that this expression for T_pM coincides with the expression for the tangent line, translated to pass through the origin, of the graph $y = g(x)$ that you learned in high school or in calculus.

Proof. Suppose $n = 2, k = 1$. Then $g\colon \mathbb{R} \to \mathbb{R}$ and $\Phi\colon \mathbb{R} \to \mathbb{R}^2$, with $\Phi'(x) = (1, g'(x))$. And

$$\Phi'(x)\mathbb{R}^k = \{(t, tg'(x)) \in \mathbb{R}^2 \colon t \in \mathbb{R}\}.$$

Recall that the tangent line of g at the point x is given by the formula $(y-g(x)) = g'(x)(t-x)$, where $y,t\in\mathbb{R}$ are variables. If we translate this line to pass through the origin, it then has the formula $y=g'(x)t$. So, this line consists of all points $(t,g'(x)t)$ with $t\in\mathbb{R}$, as desired. □

Exercise 6.2. Find the tangent space of the graph of the function $g(x,y):=x^2+y^2\cos(x)$.

Proof. Let $(x,y)\in\mathbb{R}^2$. Let $\Phi(x,y):=(x,y,g(x,y))$. Then $g\colon\mathbb{R}^2\to\mathbb{R}$, $\Phi\colon\mathbb{R}^2\to\mathbb{R}^3$, so $\Phi'(x,y)\colon\mathbb{R}^2\to\mathbb{R}^3$, and

$$\begin{aligned}T_{(x,y)}G(g)=\Phi'(x,y)\mathbb{R}^2&=\begin{pmatrix}1&0\\0&1\\2x-y^2\sin(x)&2y\cos(x)\end{pmatrix}\mathbb{R}^2\\&=\{(a,b,a(2x-y^2\sin(x))+b(2y\cos(x)))\in\mathbb{R}^3\colon (a,b)\in\mathbb{R}^2\}\\&=\{(a,b,c)\in\mathbb{R}^3\colon (2x-y^2\sin(x))a+(2y\cos(x))b-c=0\}.\end{aligned}$$

□

Exercise 6.3. Let $M:=\varphi^{-1}(0)$ be the preimage of a regular value of $\varphi\in C^1$, as in (b) above. Prove the following fact that was mentioned in class: the tangent space T_pM is

$$T_pM=\text{Nullspace}(\varphi'(p)).$$

Proof. It is assumed in the problem that 0 is a regular value of φ. That is, if $\varphi\colon U\to\mathbb{R}^{n-k}$ with $U\subseteq\mathbb{R}^n$, then $\text{rank}(\varphi'(p))=n-k$ for all $p\in M$. Let $p\in M$, and write $p=(a,b)$ with $a\in\mathbb{R}^k$, $b\in\mathbb{R}^{n-k}$. Note that $\varphi'(p)$ is an $(n-k)\times n$ matrix. Since $\varphi'(p)$ has rank $n-k$, we may assume, after permuting the coordinates if necessary, that the columns $k+1,k+2,\ldots,n$ of $\varphi'(p)$ have nonzero determinant. That is, we may assume that

$$\det\left(\varphi'(p)_{\substack{1\leq i\leq n-k\\k+1\leq j\leq n}}\right)\neq 0.$$

So, since $\varphi\in C^1(U)$, the Implicit Function Theorem gives $V\subseteq\mathbb{R}^k$ and $W\subseteq\mathbb{R}^n$ open such that $a\in V$, $p=(a,b)\in W$, $g\colon V\to\mathbb{R}^k$, $g\in C^1(V;\mathbb{R}^{n-k})$, and such that $y=g(x)$ solves $\varphi(x,y)=0$ for $(x,y)\in W$, $x\in\mathbb{R}^k$, $y\in\mathbb{R}^{n-k}$. That is, in a neighborhood of p, M is equal to $\{(x,g(x))\colon x\in V\}$.

Let $x\in V$ and define $\Phi(x):=(x,g(x))$, $\Phi\colon\mathbb{R}^k\to\mathbb{R}^n$. Then $\Phi'(a)\colon\mathbb{R}^k\to\mathbb{R}^n$, and

$$T_pM=\Phi'(a)\mathbb{R}^k=\begin{pmatrix}I_k\\g'(a)\end{pmatrix}\mathbb{R}^k.\qquad(*)$$

Here $p=\Phi(a)$, since $\Phi(a)=(a,g(a))=(a,b)=p$.

Let $(x,y)\in W$. Then $\varphi(\Phi(x))=0$. Applying the chain rule,

$$\varphi'(\Phi(x))\Phi'(x)=0.\qquad(**)$$

Let $v\in T_pM$. Then there exists $z\in\mathbb{R}^k$ such that $\Phi'(a)z=v$. But then, from $(**)$,

$$\varphi'(\Phi(a))\Phi'(a)z=0=\varphi'(\Phi(a))v.$$

So, $v\in\text{Nullspace}(\varphi'(\Phi(a)))=\text{Nullspace}(\varphi'(p))$, since $p=\Phi(a)$. That is, we have shown

$$T_pM\subseteq\text{Nullspace}(\varphi'(p)).\qquad(\ddagger)$$

Now, $\varphi'(p)$ is an $(n-k)\times n$ matrix of rank $n-k$. So, its nullspace has dimension at most k. On the other hand, from $(*)$, T_pM has dimension at least k. Since T_pM and $\text{Nullspace}(\varphi'(p))$ are then linear subspaces of the same dimension via $(\ddagger)$, we see that $(\ddagger)$ is an equality. □

Exercise 6.4. A *torus* is a doughnut-shaped surface in $\mathbb{R}^3$ that can be constructed as follows. Let $a > b > 0$ and consider the circle $\mathcal{C}$ of radius a in the xy-plane. By definition, the torus $T_{a,b}$ is the set of points $(x, y, z) \in \mathbb{R}^3$ that lie at a distance b from the circle $\mathcal{C}$.

(i) Draw a sketch of $T_{a,b}$ and prove that it is a 2-dimensional C^∞ manifold.
(ii) Find the tangent space of $T_{a,b}$ at a point $p = (x, y, z) \in T_{a,b}$.

Proof of (i). Let $(x, y, z) \in \mathbb{R}^3$ and define $r := \sqrt{x^2 + y^2}$. Then r is the distance from the point $(x, y, 0)$ to the origin. Let $(x, y, z) \in T_{a,b}$. Consider the closed ball

$$\overline{B_b(x, y, z)} = \{(c, d, e) \in \mathbb{R}^3 \colon\ |(x, y, z) - (c, d, e)| \le b\}.$$

Since $(x, y, z) \in T_{a,b}$, $\overline{B_b(x, y, z)}$ intersects $\mathcal{C}$ in at least one point. We now show that $\overline{B_b(x, y, z)}$ intersects $\mathcal{C}$ in exactly one point. For $(c, d) \in \mathbb{R}^2$, define the function $f(c, d) := (c - x)^2 + (d - y)^2 + z^2$. We minimize f subject to the constraint $g(c, d) := c^2 + d^2 - b^2 = 0$. That is, with the point $(x, y, z) \in T_{a,b}$ fixed, we minimize the distance of (x, y, z) from a variable point $(c, d) \in \mathcal{C}$. Note that $r \neq 0$, since $r = 0$ implies that $x = y = 0$, so the distance from (x, y, z) to $\mathcal{C}$ exceeds a. But this distance must be less than a, so $r \neq 0$.

Since $\mathcal{C}$ is compact, a minimum of f exists on $\mathcal{C}$. Let (c, d) denote this minimum. Using Lagrange Multipliers, we have $\nabla f = (2(x - c), 2(y - d))$, $\nabla g = (2c, 2d, 0)$, and there exists $\lambda > 0$ such that $\nabla f(c, d) = \lambda \nabla g(c, d)$. That is, $(x - c) = \lambda c$, $(y - d) = \lambda d$. So, $(x, y) = (\lambda + 1)(c, d)$. That is, (x, y) and (c, d) are parallel. Since $(x, y, z) \in \mathbb{R}^3$ and $\mathcal{C} \subseteq \mathbb{R}^2 \times \{0\}$, there exist only two vectors in $\mathcal{C}$ such that $(x, y, 0)$ and $(c, d, 0)$ are parallel. Namely, we have $(c, d, 0) = (a/r)(x, y, 0)$, and $(c, d, 0) = -(a/r)(x, y, 0)$. Note that

$$|(a/r)(x, y, 0) - (x, y, z)|^2 = z^2 + a^2, \qquad |-(a/r)(x, y, 0) - (x, y, z)| = z^2 + (a + r)^2.$$

Since $r > 0$, the point (x, y, z) achieves its minimum distance from $\mathcal{C}$ at the point $(a/r)(x, y, 0)$. At this point, we have

$$|(x, y, z) - (a/r)(x, y, 0)|^2 = b^2.$$

That is,

$$b^2 = (x^2 + y^2)(1 - a/r)^2 + z^2 = (r - a)^2 + z^2 = r^2 + a^2 - 2ra + z^2.$$

So, $r^2 + z^2 + a^2 - b^2 = 2ra$, and $(r^2 + z^2 + a^2 - b^2)^2 = 4r^2a^2$, so that

$$(x^2 + y^2 + z^2 + a^2 - b^2)^2 = 4a^2(x^2 + y^2).$$

Now, define $\varphi \colon \mathbb{R}^3 \to \mathbb{R}$ by

$$\varphi(x, y, z) := (x^2 + y^2 + z^2 + a^2 - b^2)^2 - 4a^2(x^2 + y^2).$$

To conclude, we want to show that $\nabla\varphi \neq 0$ for (x, y, z) with $\varphi(x, y, z) = 0$. Observe

$$\nabla\varphi(x, y, z) = \begin{pmatrix} 2(x^2 + y^2 + z^2 + a^2 - b^2)(2x) - 8a^2x \\ 2(x^2 + y^2 + z^2 + a^2 - b^2)(2y) - 8a^2y \\ 2(x^2 + y^2 + z^2 + a^2 - b^2)(2z) \end{pmatrix}.$$

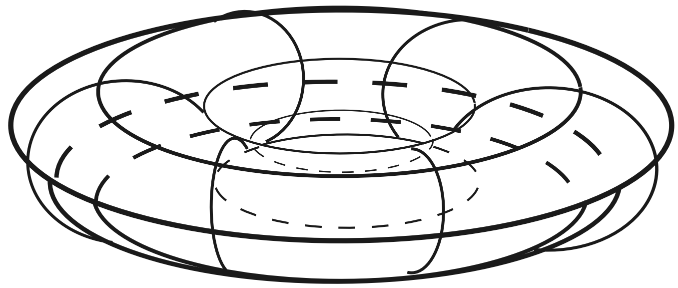
Since $a > b$, the third term of $\nabla\varphi(x, y, z)$ is zero if and only if $z = 0$. So, we now argue by contradiction, and assume that $z = 0$ and $\nabla\varphi(x, y, z) = 0$. Then the first and second terms of $\nabla\varphi(x, y, z)$ are zero, so

$$x[4(x^2 + y^2 + a^2 - b^2)]/(8a^2) = x, \qquad y[4(x^2 + y^2 + a^2 - b^2)]/(8a^2) = y.$$

Since $r > 0$, at least one of x, y is nonzero. So,

$$(x^2 + y^2 + a^2 - b^2) = 2a^2.$$

Since $\varphi(x, y, z) = 0$ and $z = 0$, we conclude that $4a^4 = 4a^2r^2$, so $r = a$. But $r = a$ implies that $(x, y, z) \in \mathcal{C}$, since $z = 0$. However, (x, y, z) has positive distance $b > 0$ from $\mathcal{C}$. We have finally arrived at a contradiction. We conclude that $\varphi'(x, y, z) \neq 0$ if $\varphi(x, y, z) = 0$. Since $\varphi \in C^\infty(\mathbb{R}^3)$, we conclude that $\varphi^{-1}(0) = T_{a,b}$ is a C^∞ manifold. □



Proof of (ii). Let $p = (x, y, z)$ and let $M := \varphi^{-1}(0) = T_{a,b}$. Using Exercise 6.3,

$$T_pM = \text{Nullspace}(\varphi'(p)) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \colon \langle (x_1, x_2, x_3), \nabla\varphi(p)\rangle = 0\}$$
$$= \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \colon \left\langle \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} 2(x^2 + y^2 + z^2 + a^2 - b^2)(2x) - 8a^2x \\ 2(x^2 + y^2 + z^2 + a^2 - b^2)(2y) - 8a^2y \\ 2(x^2 + y^2 + z^2 + a^2 - b^2)(2z) \end{pmatrix} \right\rangle = 0 \right\}.$$

□

Exercise 6.5. The *special linear group* is defined as

$$SL(n) := \{X \in \mathbb{R}^{n\times n} \colon \det X = 1\}.$$

(i) Prove that $SL(n)$ is an $(n^2 - 1)$-dimensional C^∞ manifold in the space of $n \times n$ matrices.
(ii) Show that the tangent space $T_{I_n}SL(n)$ is the space of matrices whose trace is zero.

Proof of (i). Let $X \in SL(n)$. Recall that, in Exercise 3.11(i), we found that det is polynomial in the entries of X. Therefore, $\det \in C^\infty(SL(n))$. It therefore remains to show that if $\det X = 1$, then $\det'(X) \neq 0$. Recall that, in Exercise 3.12(i), we showed that

$$\frac{d}{dt}|_{t=0} \det(X + tB) = \det(X)\text{Tr}(X^{-1}B).$$

In particular, if we take $B := X$, then $(d/dt)|_{t=0} \det(X + tX) = \det(X)\text{Tr}(I_n) = n \det(X) = n \neq 0$. Finally, note that $\det\colon \mathbb{R}^{n\times n} \to \mathbb{R}^1$, so the manifold $\det^{-1}(1)$ has dimension n^2-1. □

Proof of (ii). Recall that $SL(n) = \det^{-1}(1)$. Using Exercise 6.3,

$$T_{I_n}SL(n) = \text{Nullspace}(\nabla \det(I_n)) = \{Y \in \mathbb{R}^{n\times n} \colon \textstyle\sum_{i,j=1}^n Y_{ij}\text{Tr}(I_n^{-1}e_{ij}) = 0\}$$
$$= \{Y \in \mathbb{R}^{n\times n} \colon \textstyle\sum_{i,j=1}^n Y_{ij}\text{Tr}(e_{ij}) = 0\} = \{Y \in \mathbb{R}^{n\times n} \colon \sum_{i=1}^n Y_{ii} = 0\}$$
$$= \{Y \in \mathbb{R}^{n\times n} \colon \text{Tr}(Y) = 0\rangle\}.$$

□

Exercise 6.6. The *orthogonal group* is defined as

$$O(n) := \{X \in \mathbb{R}^{n\times n} \colon X^T X = I_n\}.$$

(i) Show that $O(n)$ is an $\frac{n(n-1)}{2}$-dimensional C^∞ manifold in the space of $n\times n$ matrices.
(ii) Show that the tangent space $T_{I_n}O(n)$ is the space of antisymmetric matrices.

Proof of (i). Let $X\in\mathbb{R}^{n\times n}$. Define $\varphi(X) := X^TX - I_n$. Note that $(\varphi(X))^T = (X^TX)^T - I_n = \varphi(X)$. So, $\varphi\colon \mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n}_{sym}$, i.e. the image of φ is contained in the space of symmetric $n\times n$ matrices. Note that φ has polynomial components, so $\varphi\in C^\infty(\mathbb{R}^{n\times n};\mathbb{R}^{n\times n}_{sym})$. Also, $\varphi^{-1}(0) = O(n)$. So, to show that $O(n)$ is a C^∞ manifold, it suffices to show that $\varphi'(X)$ has maximal rank if $\varphi(X)=0$. Note that $\mathbb{R}^{n\times n}_{sym}$ has dimension $n+(n-1)+\cdots+1$, i.e. it has dimension $\sum_{i=1}^n i = n(n+1)/2$. So, it remains to show that $\varphi'(X)$ has rank $n(n+1)/2$ if $\varphi(X)=0$. Given this fact, note the dimension of $O(n)$ would be $n^2-(n)(n+1)/2 = n(n-1)/2$, as desired. Let $Y\in\mathbb{R}^{n\times n}$, and observe

$$\begin{aligned}\frac{d}{dt}|_{t=0}\varphi(X+tY) &= \frac{d}{dt}|_{t=0}\big((X+tY)^T(X+tY)-I_n\big)\\ &= \frac{d}{dt}|_{t=0}\big(X^TX+tY^TX+tX^TY+t^2Y^TY-I_n\big)\\ &= Y^TX+X^TY. \qquad (*)\end{aligned}$$

So, if we set $Y := Xe_{ij}$, and use that $\varphi(X) = X^TX-I_n = 0$, we get

$$\frac{d}{dt}|_{t=0}\varphi(X+tY) = e_{ij}^TX^TX + X^TXe_{ij} = e_{ji}+e_{ij}. \qquad (**)$$

Now, define

$$E := \{Xe_{ij}\}_{i\geq j\colon\, i,j\in\{1,\ldots,n\}}.$$

Then E contains $n(n+1)/2$ distinct elements, and the set

$$\{(d/dt)|_{t=0}\varphi(X+tY)\colon Y\in E\}$$

consists of $n(n+1)/2$ linearly independent vectors in $\mathbb{R}^{n\times n}$, by $(**)$. That is, the rank of $\varphi'(X)$ is at least $n(n+1)/2$. Therefore, $\varphi'(X)$ has maximal rank if $\varphi(X)=0$, as desired. □

Proof of (ii). Note that $O(n)=\varphi^{-1}(0)$. From part (i), if $X=I_n$, we saw that the set E gives a spanning set for the derivative of φ. That is, if we define $E := \{e_{ij}\}_{i\geq j\colon\, i,j\in\{1,\ldots,n\}}$ then the set $\{(d/dt)|_{t=0}\varphi(I_n+tY)\colon Y\in E\}$ is a spanning set for $\varphi'(I_n)$. So, using our derivative formula $(*)$ from part (i), and Exercise 6.3,

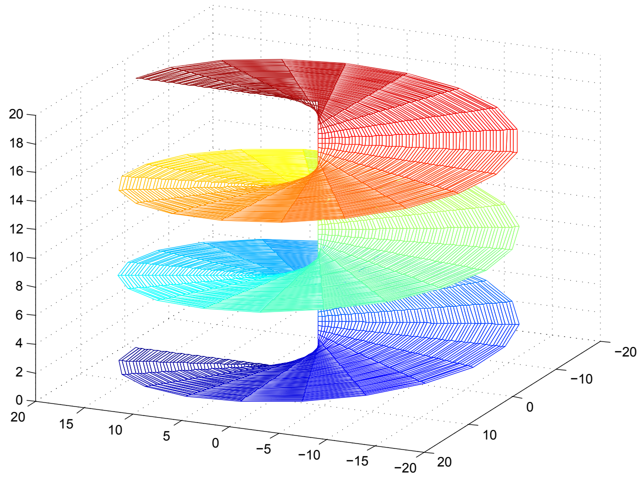
$$\begin{aligned}T_{I_n}O(n) = \mathrm{Nullspace}(\varphi'(I_n)) &= \{Y\in\mathbb{R}^{n\times n}\colon \forall\, 1\leq i,j\leq n,\ \langle Y,(e_{ij}^TI_n+I_n^Te_{ij})\rangle = 0\}\\ &= \{Y\in\mathbb{R}^{n\times n}\colon \forall\, 1\leq i,j\leq n,\ \langle Y,e_{ji}+e_{ij}\rangle = 0\}\\ &= \{Y\in\mathbb{R}^{n\times n}\colon \forall\, 1\leq i,j\leq n,\ Y_{ji}+Y_{ij} = 0\}\\ &= \{Y\in\mathbb{R}^{n\times n}\colon \forall\, 1\leq i,j\leq n,\ Y_{ij} = -Y_{ji}\}.\end{aligned}$$

□

Exercise 6.7. Fix $h>0$ and define the function $f\colon U\to\mathbb{R}^3$ where $U := (0,\infty)\times\mathbb{R}$ and

$$f(r,\theta) := (r\cos\theta, r\sin\theta, h\theta).$$

Sketch the set $M := f(U)$ and prove that it is a 2-dimensional C^∞ manifold.



Proof. Let $(x, y) \in \mathbb{R}^2$. In polar coordinates in $\mathbb{R}^2$, we write $x = r\cos\theta$, $y = r\sin\theta$, $\theta \in (-\pi/2, 3\pi/2]$. So, for $x \neq 0$, we have $y/x = \tan\theta$, so that $\theta = \tan^{-1}(y/x) + \pi 1_{\{x<0\}}(x)$. Let $D, E \subseteq \mathbb{R}^2$ be the following open sets

$$D := \{(x, y) \in \mathbb{R}^2 \colon x \neq 0\}, \qquad E := \{(x, y) \in \mathbb{R}^2 \colon y \neq 0\}.$$

We are then led to define $g \colon D \to \mathbb{R}$ and $h \colon E \to \mathbb{R}$ by

$$g(x, y) := (x, y, h[\pi 1_{\{x<0\}}(x) + \tan^{-1}(y/x)]).$$

$$h(x, y) := (x, y, h[\pi 1_{\{y<0\}}(y) + \pi/2 + \tan^{-1}(-x/y)]).$$

Note that g, h are the composition of C^∞ functions, so $g \in C^\infty(D; \mathbb{R}^3)$, $h \in C^\infty(E; \mathbb{R}^3)$. We will now create a sequence of maps from g, h that cover $M = f(U)$. For $k \in \mathbb{Z}$, define

$$g_k(x, y) := g(x, y) + (0, 0, 2\pi k h), \qquad h_k(x, y) := h(x, y) + (0, 0, 2\pi k h).$$

We claim that the set of coordinate patches $\{(g_k, D)\}_{k\in\mathbb{Z}} \cup \{(h_k, E)\}_{k\in\mathbb{Z}}$ cover M. This claim will conclude the exercise. To prove the claim, let $(x, y, z) \in f(U)$. Then there exists $r > 0$ and $\theta \in \mathbb{R}$ such that $(x, y, z) = f(r, \theta) = (r\cos\theta, r\sin\theta, h\theta)$. If $x \neq 0$, then $(x, y) \in D$ and $\tan^{-1}(y/x) + \pi 1_{\{x<0\}}(x) = \tan^{-1}(\tan(\theta)) + \pi 1_{\{x<0\}}(x) = \theta + 2k\pi$ for some $k \in \mathbb{Z}$. So, $g_k(x, y) = f(r, \theta)$. Similarly, if $y \neq 0$, then $(x, y) \in E$ and $\tan^{-1}(-x/y) + \pi/2 + \pi 1_{\{y<0\}}(y) = \tan^{-1}(-\cot(\theta)) + \pi/2 + \pi 1_{\{y<0\}}(y) = \theta + 2k\pi$ for some $k \in \mathbb{Z}$, so $h_k(x, y) = f(r, \theta)$.

We now prove the converse let $(x, y) \in D$. Then using polar coordinates, there exist $(r, \theta') \in (0, \infty) \times \mathbb{R}$ such that $r\cos\theta' = x$ and $r\sin\theta' = y$. Let $k \in \mathbb{Z}$ such that $\theta := \theta' + 2k\pi$ satisfies $\pi 1_{\{x<0\}}(x) + \tan^{-1}(y/x) = \theta$. Then $g_k(x, y) = (x, y, h\theta) = f(r, \theta)$. Similarly, let $(x, y) \in E$. Then there exist $(r, \theta') \in (0, \infty) \times \mathbb{R}$ such that $r\cos\theta' = x$ and $r\sin\theta' = y$. Let $k \in \mathbb{Z}$ such that $\theta := \theta' + 2k\pi$ satisfies $\pi 1_{\{y<0\}} + \pi/2 + \tan^{-1}(-x/y) = \theta$. Then $h_k(x, y) = (x, y, h\theta) = f(r, \theta)$. In conclusion,

$$M = f(U) = \bigcup_{k\in\mathbb{Z}} (g_k(D) \cup h_k(E)).$$

□

7. Appendix: Notation

Let n be a positive integer. Let $x=(x_1,\ldots,x_n)\in\mathbb{R}^n$, so that $x_i\in\mathbb{R}$ for $i\in\{1,\ldots,n\}$. Let $r>0$, and let $A,B\subseteq\mathbb{R}^n$ be sets. Let $v\in\mathbb{R}^n$. Let $f\colon\mathbb{R}^n\to\mathbb{R}^m$, $f_i\colon\mathbb{R}^n\to\mathbb{R}$, $i=1,\ldots,m$, $f=(f_1,\ldots,f_m)$. Assume that f is differentiable and $g\colon\mathbb{R}^n\to\mathbb{R}$ is differentiable.

$\mathbb{Z} := \{\ldots,-3,-2,-1,0,1,2,3,\ldots\}$, the integers

$\mathbb{Z}_{\geq 0} := \{0,1,2,3,4,5,\ldots\}$

$\mathbb{Q} := \{m/n\colon m,n\in\mathbb{Z}, n\neq 0\}$, the rationals

$\mathbb{R}$ denotes the set of real numbers

$\mathbb{C} := \{x+y\sqrt{-1} : x,y\in\mathbb{R}\}$, the complex numbers

$\emptyset$ denotes the empty set

$A\smallsetminus B := \{x\in A\colon x\notin B\}$

$A^c := \mathbb{R}^n\smallsetminus A$, the complement of A

$A\cap B$ denotes the intersection of A and B

$A\cup B$ denotes the union of A and B

max (or min) denotes the maximum (or minimum) of a set of numbers

sup (or inf) denotes the supremum (or infimum) of a set of numbers

$|x| = ||x||_2 := (\sum_{i=1}^n x_i^2)^{1/2}$, the 2-norm, or ℓ_2-norm

$||x||_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$, the p-norm, or ℓ_p-norm for $1\leq p<\infty$

$||x||_\infty := \sup_{i\in\{1,\ldots,n\}} |x_i|$, the sup-norm, or ∞-norm, or ℓ_∞ norm

$B_r(x) = B_2(x,r) := \{y\in\mathbb{R}^n\colon |y-x|<r\}$, the open ball of radius r

$B_p(x,r) := \{y\in\mathbb{R}^n\colon ||y-x||_p<r\}$, the open ℓ_p ball of radius r for $1\leq p\leq\infty$

Tr denotes the trace function

det denotes the determinant function

S_n denotes the set of permutations on n elements

$\mathrm{sign}(\sigma) = (-1)^N$, where $\sigma\in S_n$ is the composition of N transpositions

$id = id_n = I_n$ denotes the identify map, or the $n\times n$ identity matrix

x^T denotes the transpose of x

$e_1,\ldots,e_n$ denotes the standard basis of $\mathbb{R}^n$

$D_v f(x)$ denotes the derivative of f at x in the direction v

$\partial g/\partial x_j = D_{e_j} g = D_j g$ denotes the partial derivative of g with respect to x_j

$f'(x)$ denotes the matrix of partial derivatives $(\partial f_i/\partial x_j)_{1\leq i\leq m, 1\leq j\leq n}$

$\nabla g(x) = (D_1 g(x),\ldots,D_n g(x))$ denotes the gradient of g

Courant Institute, New York University, New York NY 10012
E-mail address: heilman@cims.nyu.edu