Problem Set 5

1. (i) Show that the mappings

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

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

(ii) Show that the equations

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

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

2. Consider the system of equations

$$\begin{aligned} x^2 + uy + \mathrm{e}^v \;&=\; 0 \\ 2x + u^2 - uv \;&=\; 5\,. \end{aligned}$$

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 $(x,y) = (2,5)$.

3. Let A be an $m \times n$ matrix. We define the *matrix norm* of A through

$$\|A\| \;:=\; \sup\bigl\{|Ax| : |x| \leqslant 1\bigr\}\,.$$

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

(ii) Prove that $\|\cdot\|$ is a *norm*, i.e. that it satisfies the following axioms:

- $\|aA\| = |a|\|A\|$ for all $a \in \mathbb{R}$;
- $\|A+B\| \;\leqslant\; \|A\| + \|B\|$;
- if $\|A\| = 0$ then $A = 0$.

(iii) Prove that $\|AB\| \leqslant \|A\|\|B\|$, where A is an $k \times m$ 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

$$\|A\| \leqslant |A| . \tag{1}$$

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

Hint. For (1), pick some $x \in \mathbb{R}^n$ and estimate $|Ax|^2$ using Cauchy-Schwarz.

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

(ii) Find the dimensions of the box of maximal volume, whose edges are parallel to the coordinate axes, which can be inscribed in the ellipsoid

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

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

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

A fundamental property of 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 $\overline{p}$.
Hint. Prove by induction on n that the maximum of S on $\overline{\mathcal{P}}_n \setminus \mathcal{P}_n$ is $\log(n-1)$.

(iii) In applications to statistical physics, each point of $\{1, \dots, n\}$ represents a state of a physical system with a given energy E_i. The *energy* of the probability distribution p is

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

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 $\min_i E_i < E < \max_i E$. (Why?) 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

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

for some parameter β chosen so that $H(\overline{p}) = E$. (Why does such a β exist?) 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).

6. (i) Let $p_1, \dots, p_n$ 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

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

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 $f(x) \leqslant 1$ in M.

(iii) Use (ii) to prove the inequality

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

where $a_1, \dots, a_n$ 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 than or equal to the arithmetic mean.

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, prove the identity

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

and use it 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.

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

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

(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

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

Hints. By definition, $x_{k+1} - x_k = -(f'(x_k))^{-1} f(x_k)$. Use (2) to estimate $(f'(x_k))^{-1}$. Then write x_k in $f(x_k)$ as $x_k = x_{k-1} - (f'(x_{k-1}))^{-1} f(x_{k-1})$, and do a second-order Taylor approximation of f around x_{k-1}.

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

(iv) Prove that if

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

for some $\varepsilon \in (0,1)$, then

$$|x_{k+1} - x_k| \leqslant \frac{1}{K^2 2^k}\varepsilon^{2^k} \leqslant \frac{\varepsilon}{K^2}\left(\frac{\varepsilon}{2}\right)^k.$$

Conclude, using a telescopic sum, that if $x_0 \in [-r,r]$ and

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

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.

Hints. Proceed as in the proof of Banach's fixed point theorem: Prove first that $(x_k)_{k\in\mathbb{N}}$ is a Cauchy sequence by using the telescopig identity $x_l - x_k = \sum_{i=k}^{l-1}(x_{i+1} - x_i)$, where $l > k$. To show that the limit is a fixed point of φ, take the limit in $x_{k+1} = \varphi(x_k)$ and use that φ is continuous.

(v)* Suppose that $f'(0) \neq 0$. Show that there exist $K > 0$ and $0 < r < R$ 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 $[-r,r]$).

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.

Due: Thursday, April 25, in class.