CSE 203B W25 Homework 1

Due Time : 11:50pm, Thursday Jan. 16, 2025 Submit to Gradescope
Gradescope: https://gradescope.com/

In this homework, we work on the basic concepts of convex optimization and linear algebra.
All the problems are graded by completion.

1. Convex Optimization (10 pts)
1.1. Given a function $f_0(x) = x^4 - 2x^3 + 3x^2 - 2x - 4$, where $x \in \mathbb{R}$. Solve $min_x f_0(x)$ using Kuhn-Tucker conditions. Show your derivation. (5 pts)

Solution. The following two KT conditions:

$$\nabla^2 f_0(x) \geq 0 \quad (1)$$
$$\nabla f_0(x^*) = 0 \quad (2)$$

First, we have that $\nabla^2 f_0(x) = 12x^2 - 12x + 6$, which implies $f_0(x)$ is not necessarily convex. By solving $\nabla f_0(x*) = 0$, we get that $\nabla f_0(x*) = 4x^3 - 6x^2 + 6x - 2 = 0$, and it's real roots are $x^* = 0.5$. $\nabla^2 f_0(x^*) \geq 0$, so this is one local min and $f(x^*) = -4.4375$
1.2. Given two functions $f_0(x) = x^2 - 6x + 9$, and $f_1(x) = 2x + 3$, where $x \in \mathbb{R}$. Solve $min_x f_0(x)$ subject to $f_1(x) \leq 0$. (5 pts)

Solution. The Lagrangian is $L(x, \lambda) = f_0(x) + \lambda f_1(x) = x^2 - 6x + 9 + \lambda(2x + 3)$, where λ is a Lagrange multiplier, $\lambda \geq 0 \in \mathbb{R}$. The primal problem is $\min_x \max_\lambda L(x, \lambda)$ and the dual problem is $\max_\lambda \min_x L(x, \lambda) = \max_\lambda g(\lambda)$. To solve the dual problem, we first solve $\min_x L(x, \Lambda)$ using the KT conditions:

$$\frac{\partial^2 L(x, \lambda)}{\partial x^2} = 2 \geq 0 \quad (3)$$
$$\frac{\partial L(x, \lambda)}{\partial x} = 2x - 6 + 2\lambda \quad (4)$$

By setting $\frac{\partial L(x,\lambda)}{\partial x} = 0$, we get that $x = (6 - 2\lambda)/2 = 3 - \lambda$ is the global minimum of $L(x, \lambda)$. Plugging this into $g(\lambda)$ yields

$$g(\lambda) = -\lambda^2 + 9\lambda$$

Then, we solve $\max_\lambda g(\lambda)$—again using the KT conditions:

$$\frac{\partial^2 g(\lambda)}{\partial \lambda^2} = -2 \leq 0 \quad (5)$$
$$\frac{\partial g(\lambda)}{\partial \lambda} = 9 - 2\lambda \quad (6)$$

$\frac{\partial g(\lambda)}{\partial \lambda} = 9 - 2\lambda = 0$ implies $\lambda = 4.5$. Plugging back into $x(\lambda)$ yields $x^* = -3/2$ and $f(x^*) = 81/4$.
2. Matrix Properties (16 pts)
2.1. Linear System:

Consider the following system of linear equations

$$\begin{aligned} x_1 \quad +2x_2 \quad +3x_3 &= 1 \\ x_1 \quad +x_2 \quad +x_3 &= -2 \\ x_2 \quad +2x_3 &= 3. \end{aligned}$$

Write the equations in a matrix form. (2 pts)

Solution.

$$Ax = b$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \tag{7}$$

2.2. For the matrix in problem 2.1, derive its range. What's the rank of this matrix? (2pts)

Solution. Perform Gaussian Elimination on A & reduce to row-echelon form. The range is the span of the associated pivots and $\text{rank}(A) = 2$.

2.3. Derive the nullspace of the matrix in problem 2.1. What's the relation between the range and nullspace of a matrix? (2pts)

Solution. The nullspace of A consists of all solutions x to the system $Ax = 0$. In general, for an $m \times n$ matrix A, the dimensions of $R(A)$ and $N(A)$ sum to n.

2.4. Derive the trace and determinant of the matrix in problem 2.1. Write the eigenvalues and eigenvectors. (3pts)

2.5. Prove the following properties. (3 pts)

- For $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times m}$, $\text{tr}AB = \text{tr}BA$.
- For $A, B \in \mathbb{R}^{n \times n}$, $\det AB = \det A \det B$.
- For $A \in \mathbb{R}^{n \times n}$, $\det A = \prod_{i=1}^{n} \lambda_i$, and $\text{tr}A = \sum_{i=1}^{n} \lambda_i$, where $\lambda_i, i = 1, \ldots, n$ are the eigenvalues of A.

Solution.

Commutativity of Trace

$$\begin{aligned} \text{trace}(AB) &= \sum_{i=1}^{m} (AB)_{ii} = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ji} \\ &= \sum_{j=1}^{n} \sum_{i=1}^{m} B_{ji} A_{ij} = \sum_{j=1}^{n} (BA)_{jj} = \text{trace}(BA) \end{aligned}$$

Distributivity of Determinant

If A is not invertible, then AB is not invertible, we have $\det(AB) = \det(A)\det(B) = 0$. If A is invertible, A can be row reduced to an identity matrix I by a finite number of elementary row operations $E_1, E_2, \ldots, E_n$, i.e.

$$A = E_n E_{n-1} \ldots E_1 I$$

Multiplying the LHS and RHS by B, we have

$$AB = E_n E_{n-1} \ldots E_1 B$$

Taking the determinant of LHS and RHS, we have

$$\begin{aligned}\det(A) &= \det(E_n E_{n-1} \ldots E_1)\\ \det(AB) &= \det(E_n E_{n-1} \ldots E_1 B)\end{aligned}$$

If E is an elementary row operation, we have $\det(EA) = \det(E)\det(A)$ (verify yourself). So,

$$\begin{aligned}\det(E_n E_{n-1} \ldots E_1 B) &= \det(E_n)\det(E_{n-1} \ldots E_1 B)\\ &= \det(E_n) \ldots \det(E_1)\det(B)\\ &= \det(E_n \ldots E_1)\det(B)\\ &= \det(A)\det(B)\end{aligned}$$

Determinants & Eigenvalues
Method 1. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be eigenvalues of A. By definition, $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the roots of the characteristic polynomial of A.

$$p_A(t) = \det(A - tI) = (\lambda_1 - t)(\lambda_2 - t) \ldots (\lambda_n - t)$$

Then, for $t = 0$ we have:

$$p_A(0) = \det(A) = \lambda_1 \lambda_2 \ldots \lambda_n = \prod_{i=1}^{n} \lambda_i$$

Method 2. Any matrix $A \in \mathbb{R}^{n \times n}$ can be transformed to Jordan canonical form J by a *similarity* transformation T:

$$J = T^{-1} A T$$

Where J is upper-triangular with diagonal corresponding to the eigenvalues of A $\lambda_1, \ldots, \lambda_n$. Correspondingly, $\mathrm{tr}(J) = \sum_{i=1}^{n} \lambda_i$. Note that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. Then by some algebra:

$$\mathrm{tr}(J) = \mathrm{tr}(T^{-1}AT) = \mathrm{tr}(T^{-1}(AT)) = \mathrm{tr}((AT)T^{-1}) = \mathrm{tr}(ATT^{-1}) = \mathrm{tr}(AI) = \mathrm{tr}(A)$$

2.6. Use the following example of matrices A and B to illustrate the equations in problem 2.5. (4 pts)

$$A = \begin{bmatrix} -2 & -4 & 2 \\ -2 & 1 & 2 \\ 4 & 2 & 5 \end{bmatrix},\ B = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 1 & 3 \end{bmatrix},$$

3. Matrix Operations (24 pts)

Gradient: consider a function $f : \mathbb{R}^n \to \mathbb{R}$ that takes a vector $x \in \mathbb{R}^n$ and returns a real value. Then the gradient of f (w.r.t. x) is the vector of partial derivatives, defined as

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}.$$

Hessian: consider a function $f : \mathbb{R}^n \to \mathbb{R}$ that takes a vector $x \in \mathbb{R}^n$ and returns a real value. Then the Hessian matrix of f (w.r.t. x) is the $n \times n$ matrix of partial derivatives, defined as

$$\nabla_x^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

3.1. Write the gradient and Hessian matrix for the linear function

$$f(x) = b^T x,$$

where $x \in \mathbb{R}^n$ and vector $b \in \mathbb{R}^n$. (2 pts)

Solution.

$$f(x) = b^\top x = \sum_{i=1}^{n} b_i x_i$$

Gradient:

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = b$$

Hessian:

$$\nabla_x^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

3.2. Write the gradient and Hessian matrix of the quadratic function

$$f(x) = x^T A x + 2b^T x + c,$$

where $x \in \mathbb{R}^n$, matrix $A \in \mathbb{R}^{n\times n}$, vector $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$. (2 pts)

Solution.
Note that A is not necessarily symmetric.

$$f(x) = x^\top Ax + b^\top x + c = \sum_{j=1}^{n}\sum_{i=1}^{n} x_j A_{ji} x_i + 2\sum_{i=1}^{n} b_i x_i + c$$

Gradient:

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} (\sum_{i=1}^n a_{1i}x_i + \sum_{j=1}^n x_j a_{j1}) + 2b_1 \\ (\sum_{i=1}^n a_{2i}x_i + \sum_{j=1}^n x_j a_{j2}) + 2b_2 \\ \vdots \\ (\sum_{i=1}^n a_{ni}x_i + \sum_{j=1}^n x_j a_{jn}) + 2b_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n (a_{1i} + a_{i1})x_i + 2b_1 \\ \sum_{i=1}^n (a_{2i} + a_{il2})x_i + 2b_2 \\ \vdots \\ \sum_{i=1}^n (a_{ni} + a_{in})x_i + 2b_n \end{bmatrix} = (A+A^\top)x+2b$$

Hessian:

$$\nabla_x^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} = \begin{bmatrix} 2a_{11} & a_{12} + a_{21} & \cdots & a_{1n} + a_{n1} \\ a_{21} + a_{12} & 2a_{22} & \cdots & a_{2n} + a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + a_{1n} & a_{n2} + a_{2n} & \cdots & 2a_{nn} \end{bmatrix} = A + A^\top$$

3.3. Given matrix $A \in \mathbb{R}^{m\times n}$ where $m < n$ and $rank(A) = m$, and vector $b \in \mathbb{R}^m$, find a solution $x \in \mathbb{R}^n$ such that $Ax = b$. (3 pts)

Solution.
Method 1. Since A has full row rank and $m < n$, $Ax = b$ has infinitely many solutions. One particularly interesting solution is the one with minimal ℓ_2-norm. Finding it can be formulated as solving the following constrained optimization problem:

$$\min_{x\in\mathbb{R}^n} \{f(x) := ||x||_2 = x^\top x\}$$
$$\text{s.t. } Ax = b$$

The Lagrangian is $L(x, \lambda) = x^\top x + \lambda^\top(Ax - b)$, $\lambda \geq 0 \in \mathbb{R}^m$. The first-order conditions can then be solved:

$$\frac{\partial L}{\partial x} = 2x + A^\top\lambda = 0 \implies x = -\frac{1}{2}A^\top\lambda \tag{8}$$

$$\frac{\partial L}{\partial \lambda} = Ax - b = 0 \tag{9}$$

Plugging (1) into (2) yields

$$-\frac{1}{2}AA^\top\lambda - b = 0 \tag{10}$$

$$\implies AA^\top\lambda = -2b \tag{11}$$

$$\implies \lambda = -2(AA^\top)^{-1}b \tag{12}$$

Since $\text{rank}(A) = m$, we have $\text{rank}(AA^\top) = \text{rank}(A) = m$, i.e., the $m \times m$ square matrix $AA^\top$ has full rank, therefore it is invertible. By plugging (12) back into (8), we have one solution corresponding to the *normal equations* from linear least squares.

$$x = A^\top (AA^\top)^{-1} b$$

There are many interpretations (see wiki on Moore Penrose psedoinverse: `https://en.wikipedia.org/wiki/Moore-Penrose_inverse`).

Method 2. Since $\text{rank}(A) = m$, we can rearrange the columns of A such that

$$A = [A_1 A_2]$$

where A_1 contains m linearly independent columns of A, and A_2 contains the rest $n - m$ columns. A_1 is therefore a full-rank $m \times m$ matrix, i.e. invertible. We can then re-write the system $Ax = b$ as

$$\begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = b$$

where $x_1 \in \mathbb{R}^m$ and $x_2 \in \mathbb{R}^{n-m}$. Then, one solution is given by $x_1 = A_1^{-1}$ and $x_2 = 0$.

3.4. Given a nonsingular matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where elements $A, B, C, D \in \mathbb{R}^{2\times 2}$, write an analytic solution of M^{-1}.

a. Assume that matrix A is not singular. (2 pts)

b. Assume that matrix D is not singular. (2 pts)

c. Assume that both matrices A and D are singular. (4 pts)

Hint: Refer to Theorem 2.1 on Page 2 of the paper: *Lu, Tzon-Tzer and Sheng-Hua Shiou. "Inverses of 2 × 2 block matrices." Computers & Mathematics With Applications 43 (2002): 119-129*

Solution.

(a.)—(b.) Via Schur Complement

(c.) If M is nonsingular, $M^\top M$ is also nonsingular. Consider the Schur Complement of $(M^\top M)^{-1} M$

3.5. Given a nonsingular matrix

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$

write the analytic solution of A^{-1}. (4 pts)

Solution. The cofactor matrix C is

$$C = \begin{bmatrix} ei - fh & fg - di & dh - eg \\ ch - bi & ai - cg & bg - ah \\ bf - ce & cd - af & ae - bd \end{bmatrix}$$

The adjoint of a matrix A; $\text{adj}(A) = C^\top$. The determinant of A is

$$\begin{aligned}\det(A) &= a(ei - fh) - b(di - fg) + c(dh - eg)\\ &= aei - afh - bdi + bfg + cdh - ceg\end{aligned}$$

And the inverse of A is

$$A^{-1} = \frac{1}{\det(A)}\text{adj}(A)$$

3.6. Consider a matrix $A = [a_{i,j}] \in R^{2\times 2}$ which is **not singular** Derive the analytical form of the derivative of f over matrix A (i.e. $[u_{i,j}] = \nabla_A f$, where $u_{i,j} = \partial f/\partial a_{i,j}$) for function $f = trA^{-1}$. (5 pts)

Solution. (eq. 36 & 40 in Matrix Cookbook) 3 useful facts: (1.) Chain rule: $\partial f(A) = \partial f(\partial A)$ (2.) Derivative of trace: $\partial\text{tr}(A) = \text{tr}(\partial A)$ (3.) Derivative of inverse $\partial A^{-1} = -A^{-1}(\partial A)A^{-1}$. Thus, we have that

$$\partial\text{tr}(A^{-1}) = \text{tr}(\partial A^{-1}) = \text{tr}(-A^{-1}(\partial A)A^{-1}) = -\text{tr}(A^{-1}(\partial A)A^{-1})$$

Let $U = \partial A$ so $tr(A^{-1}UA^{-1}) = tr(A^{-1}A^{-1}U) = \langle A^{-2}, U\rangle$, $U = 1_{u_{ij}}$ the indicator for the partial derivative, so in matrix form, the solution is $A^{-2\top}$.

Alternatively, you can directly assume a 2×2 matrix and derive and analytic solution for the inverse (as done in 3.5 for a 3×3 matrix) and then compute its trace