CSE 203B W22 Homework 1

Due Time : 11:50pm, Wednesday Jan. 12, 2022 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 content.

1. Convex Optimization (12 pts)
1.1. Given a function $f_0(x) = x^4 - 4x^3 + 6x^2 - 4x + 1$, where $x \in \mathbb{R}$. Solve $min_x f_0(x)$ using Kuhn-Tucker conditions. Show your derivation. (2 pts)

[Solution]
We can solve $min_x f_0(x)$ using the following KT conditions:

$$\nabla^2 f_0(x) \geq 0$$

$$\nabla f_0(x^*) = 0$$

First, we have $\nabla^2 f_0(x) = 12(x-1)^2 \geq 0$ which shows that $f_0(x)$ is a convex function. This means that the local minimum of $f_0(x)$ is the global minimum. This allows us to find $min_x f_0(x)$ through the first derivative test. By solving $\nabla f_0(x^*) = ((x^*-1)^4)' = 4(x^*-1)^3 = 0$, we have $x^* = 1$, and $\min_x f_0(x) = 0$.

1.2. Given two functions $f_0(x) = x^2 - 3x + 1$, and $f_1(x) = 3x + 1$, where $x \in \mathbb{R}$. Solve $min_x f_0(x)$ subject to $f_1(x) \leq 0$ using the primal dual transform with Lagrange multipliers. Show your derivation. (10 pts)

[Solution]
The Lagrangian is $L(x, \lambda) = f_0(x) + \lambda f_1(x) = x^2 - 3x + 1 + \lambda(3x + 1)$, where λ is the Lagrange multiplier, $\lambda \in \mathbb{R}$, $\lambda \geq 0$.
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 KT conditions:

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

From the solution, we know that $x = \frac{3-3\lambda}{2}$ is the global minimum of $L(x, \lambda)$. So we can plug this into $g(\lambda)$ and get

$$g(\lambda) = \frac{-5 + 22\lambda - 9\lambda^2}{4}$$

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

$$\frac{\partial g^2(\lambda)}{\partial \lambda^2} = -\frac{9}{2} \leq 0$$
$$\frac{\partial g(\lambda)}{\partial \lambda} = \frac{11 - 9\lambda}{2} = 0$$

And we get $\lambda = \frac{11}{9}$. By plugging this back into $x(\lambda)$, we have $x^* = -\frac{1}{3}$, and $\min_x f_0(x) = \frac{19}{9}$.

2. Matrix Properties (14 pts)

2.1. Linear System (2pts)

Consider the following system of linear equations

$$\begin{aligned} x_1 + x_2 + 3x_3 &= 1 \\ 2x_1 - x_2 + 2x_3 &= -2 \\ 3x_1 + 5x_3 &= -1. \end{aligned}$$

Write the equations in a matrix form.

[Solution]

$$\begin{bmatrix} 1 & 1 & 3 \\ 2 & -1 & 2 \\ 3 & 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}$$
$$Ax = b$$

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

[Solution]

By row reducing matrix A we have:

$$\begin{bmatrix} 1 & 1 & 3 \\ 2 & -1 & 2 \\ 3 & 0 & 5 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 - 2R_1} \begin{bmatrix} 1 & 1 & 3 \\ 0 & -3 & -4 \\ 3 & 0 & 5 \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 - 3R_1} \begin{bmatrix} 1 & 1 & 3 \\ 0 & -3 & -4 \\ 0 & -3 & -4 \end{bmatrix}$$

$$\xrightarrow{R_3 \leftarrow R_3 - R_2} \begin{bmatrix} 1 & 1 & 3 \\ 0 & -3 & -4 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_2 \leftarrow -R_2/3} \begin{bmatrix} 1 & 1 & 3 \\ 0 & 1 & 4/3 \\ 0 & 0 & 0 \end{bmatrix}$$

Since the row echelon form of A contains 2 pivots, its rank is 2. The range is the span of the column vectors that contain the pivot positions, which is

$$R(A) = c_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \forall c_1, c_2 \in \mathbb{R}$$

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 $Ax = \mathbf{0}$.
By doing row reduction the same way in question 2.2 we reach the same row echelon form of A:

$$\begin{bmatrix} 1 & 1 & 3 \\ 2 & -1 & 2 \\ 3 & 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 3 \\ 0 & 1 & 4/3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Since the row echelon form of A contains 1 zero row, the dimension of nullspace is 1.

To find the nullspace, we first identify the free variable x_3 which corresponds to the free column 3. Column 3 is free because it has no pivots. Then we can represent the pivot variables x_1, x_2 with respect to the free variable x_3 by solving:

$$x_2 + 4x_3/3 = 0 \implies x_2 = -4x_3/3$$
$$x_1 + x_2 + 3x_3 = 0 \implies x_1 = -5x_3/3$$

These solutions form the nullspace of A:
$N(A) = x_3 \begin{bmatrix} -5/3 \\ -4/3 \\ 1 \end{bmatrix}, \forall x_3 \in \mathbb{R}$
Here, we have a 2-dimensional range, and a 1-dimensional nullspace, which adds up to 3, the number of columns in our matrix.
In general, for a $m \times n$ matrix A, the dimensions of $R(A)$ and $N(A)$ sums to n.

2.4. Derive the trace and determinant of the matrix in problem 2.1. Write the eigenvalues and eigenvectors. (2pts)

[Solution]

1. The trace of a matrix is the sum of the elements along the main diagonal:

$$\text{tr}(A) = \text{tr}\left(\begin{bmatrix} 1 & 1 & 3 \\ 2 & -1 & 2 \\ 3 & 0 & 5 \end{bmatrix}\right) = 1 + (-1) + 5 = 5$$

2. The determinant of matrix A can be calculated as follows:

$$\det(A) = \begin{vmatrix} 1 & 1 & 3 \\ 2 & -1 & 2 \\ 3 & 0 & 5 \end{vmatrix} = 1 \begin{vmatrix} -1 & 2 \\ 0 & 5 \end{vmatrix} - 1 \begin{vmatrix} 2 & 2 \\ 3 & 5 \end{vmatrix} + 3 \begin{vmatrix} 2 & -1 \\ 3 & 0 \end{vmatrix} = 1(-5) - 1(4) + 3(3) = 0$$

   Alternatively, since $rank(A) = 2 < 3$, the matrix is singular, thus $det(A) = 0$.

3. The eigenvectors x and the associated eigenvalues λ of a matrix A satisfy $(A - \lambda I)x = 0$ $(x \neq \mathbf{0})$.

We can find the eigenvalues by solving $det(A - \lambda I) = 0$:

$$\begin{aligned}
&\begin{vmatrix} 1-\lambda & 1 & 3 \\ 2 & -1-\lambda & 2 \\ 3 & 0 & 5-\lambda \end{vmatrix} \\
&= (1-\lambda)\begin{vmatrix} -1-\lambda & 2 \\ 0 & 5-\lambda \end{vmatrix} - 1\begin{vmatrix} 2 & 2 \\ 3 & 5-\lambda \end{vmatrix} + 3\begin{vmatrix} 2 & -1-\lambda \\ 3 & 0 \end{vmatrix} \\
&= (1-\lambda)(-1-\lambda)(5-\lambda) - (4-2\lambda) + 3(3+3\lambda) \\
&= \lambda(12 + 5\lambda - \lambda^2) = 0 \\
&\implies \lambda_1 = 0,\ \lambda_2 = \frac{5+\sqrt{73}}{2},\ \lambda_3 = \frac{5-\sqrt{73}}{2}
\end{aligned}$$

Then the eigenvectors x can be found by solving $(A - \lambda I)x = 0$. We can row reduce the matrix $A - \lambda I$ as follows:

$$\begin{bmatrix} 1-\lambda & 1 & 3 \\ 2 & -1-\lambda & 2 \\ 3 & 0 & 5-\lambda \end{bmatrix} \xrightarrow{R_1 \leftarrow \frac{1}{1-\lambda}R_1} \begin{bmatrix} 1 & \frac{1}{1-\lambda} & \frac{3}{1-\lambda} \\ 2 & -1-\lambda & 2 \\ 3 & 0 & 5-\lambda \end{bmatrix}$$

$$\xrightarrow{R_2 \leftarrow R_2 - 2R_1} \begin{bmatrix} 1 & \frac{1}{1-\lambda} & \frac{3}{1-\lambda} \\ 0 & \frac{\lambda^2-3}{1-\lambda} & \frac{-2\lambda-4}{1-\lambda} \\ 3 & 0 & 5-\lambda \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 - 3R_1} \begin{bmatrix} 1 & \frac{1}{1-\lambda} & \frac{3}{1-\lambda} \\ 0 & \frac{\lambda^2-3}{1-\lambda} & \frac{-2\lambda-4}{1-\lambda} \\ 0 & \frac{-3}{1-\lambda} & \frac{\lambda^2-6\lambda-4}{1-\lambda} \end{bmatrix}$$

$$\xrightarrow{R_2 \leftarrow \frac{1-\lambda}{\lambda^2-3}R_2} \begin{bmatrix} 1 & \frac{1}{1-\lambda} & \frac{3}{1-\lambda} \\ 0 & 1 & \frac{-2\lambda-4}{\lambda^2-3} \\ 0 & \frac{-3}{1-\lambda} & \frac{\lambda^2-6\lambda-4}{1-\lambda} \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 + \frac{3}{1-\lambda}R_2} \begin{bmatrix} 1 & \frac{1}{1-\lambda} & \frac{3}{1-\lambda} \\ 0 & 1 & \frac{-2\lambda-4}{\lambda^2-3} \\ 0 & 0 & \frac{\lambda(12+5\lambda-\lambda^2)}{\lambda^2-3} \end{bmatrix}$$

$$\xrightarrow{\lambda(12+5\lambda-\lambda^2)=0} \begin{bmatrix} 1 & \frac{1}{1-\lambda} & \frac{3}{1-\lambda} \\ 0 & 1 & \frac{-2\lambda-4}{\lambda^2-3} \\ 0 & 0 & 0 \end{bmatrix}$$

x_3 is the free variable, we can solve the following equations for x_1 and x_2 in terms of x_3:

$$x_2 - \frac{2\lambda+4}{\lambda^2-3}x_3 = 0 \implies x_2 = \frac{2\lambda+4}{\lambda^2-3}x_3$$
$$x_1 + \frac{x_2}{1-\lambda} + \frac{3x_3}{1-\lambda} = 0 \implies x_1 = \frac{3\lambda+5}{\lambda^2-3}x_3$$

Therefore, the eigenvectors are:

$$v = c\begin{bmatrix} \frac{3\lambda+5}{\lambda^2-3} \\ \frac{2\lambda+4}{\lambda^2-3} \\ 1 \end{bmatrix}, \forall c \in \mathbb{R} - \{0\}$$

By plugging in $\lambda_1, \lambda_2, \lambda_3$, we can derive the corresponding eigenvectors:

$$\lambda_1 = 0 \implies v_1 = c \begin{bmatrix} -\frac{5}{3} \\ -\frac{4}{3} \\ 1 \end{bmatrix}, \ \forall c \in \mathbb{R} - \{0\}$$

$$\lambda_2 = \frac{5+\sqrt{73}}{2} \implies v_2 = c \begin{bmatrix} \frac{-5+\sqrt{73}}{6} \\ \frac{11-\sqrt{73}}{6} \\ 1 \end{bmatrix}, \ \forall c \in \mathbb{R} - \{0\}$$

$$\lambda_3 = \frac{5-\sqrt{73}}{2} \implies v_3 = c \begin{bmatrix} \frac{-5-\sqrt{73}}{6} \\ \frac{11+\sqrt{73}}{6} \\ 1 \end{bmatrix}, \ \forall c \in \mathbb{R} - \{0\}$$

2.5. Prove the following properties. (3 pts)

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

[Solution]

1.

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

2. 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

$$\det(A) = \det(E_n E_{n-1} \dots E_1)$$

$$\det(AB) = \det(E_n E_{n-1} \dots E_1 B)$$

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

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

3. (a) **[Solution 1]**
By definition, $\det(A - \lambda I) = 0$ at $\lambda = \lambda_1, \dots, \lambda_n$, which means $\lambda_1, \dots, \lambda_n$ are the roots of the characteristic polynomial $\det(A - \lambda I)$:

$$\det(A - \lambda I) = (\lambda_1 - \lambda)\dots(\lambda_n - \lambda)$$

By setting $\lambda = 0$, we have $\det(A) = \prod_{i=1}^{n} \lambda_i$.

[Solution 2]
For any matrix $A \in \mathbb{R}^{n\times n}$, it can be transformed to Jordan canonical form J by a similarity transformation T:

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

where J is an upper triangular matrix and have A's eigenvalues $\lambda_1, \dots, \lambda_n$ on its diagonal. Using property $\det(AB) = \det(A)\det(B)$, we have

$$\det(J) = \det(T^{-1}AT) = \det(T^{-1})\det(A)\det(T) = \det(T^{-1})\det(T)\det(A)$$

$$\det(I) = \det(T^{-1}T) = \det(T^{-1})\det(T) = 1$$

Since J is a triangular matrix, $\det(J) = \prod_{i=1}^{n} \lambda_i$.
Therefore, $\det(A) = \prod_{i=1}^{n} \lambda_i$.

(b) For any matrix $A \in \mathbb{R}^{n\times n}$, it can be transformed to Jordan canonical form J by a similarity transformation T:

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

where J has A's eigenvalues $\lambda_1, \dots, \lambda_n$ on its diagonal, therefore $tr(J) = \sum_{i=1}^{n} \lambda_i$. It is easy to see that similarity transformation preserves trace by using property $tr(AB) = tr(BA)$:

$$\begin{aligned}\operatorname{tr}(J) &= \operatorname{tr}\left(T^{-1}AT\right)\\ &= \operatorname{tr}\left(T^{-1}(AT)\right)\\ &= \operatorname{tr}\left((AT)T^{-1}\right)\\ &= \operatorname{tr}\left(A\left(TT^{-1}\right)\right)\\ &= \operatorname{tr}(AI)\\ &= \operatorname{tr}(A)\end{aligned}$$

Therefore, $tr(A) = \sum_{i=1}^{n} \lambda_i$.

2.6. Suppose that you are a tutor. Devise a simple but meaningful numerical example to illustrate the three equations in problem 2.5.(3 pts)

[Solution]
We are not providing examples here, but any simple and meaningful numerical matrices illustrating the above properties are correct.

3. Matrix Operations (14 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

$$f(x) = 2b^T x,$$

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

[Solution]

$$f(x) = 2b^T x = \sum_{i=1}^{n} 2b_i x_i$$

Gradient:

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \cdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2b_1 \\ 2b_2 \\ \cdots \\ 2b_n \end{bmatrix} = 2b$$

Hessian:

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

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

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

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]

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

Gradient:

$$\begin{aligned} \nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \dots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} &= \begin{bmatrix} (\sum_{i=1}^n A_{1i} x_i + \sum_{j=1}^n x_j A_{j1}) + b_1 \\ (\sum_{i=1}^n A_{2i} x_i + \sum_{j=1}^n x_j A_{j2}) + b_2 \\ \dots \\ (\sum_{i=1}^n A_{ni} x_i + \sum_{j=1}^n x_j A_{jn}) + b_n \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^n (A_{1i} + A_{i1}) x_i + b_1 \\ \sum_{i=1}^n (A_{2i} + A_{i2}) x_i + b_2 \\ \dots \\ \sum_{i=1}^n (A_{ni} + A_{in}) x_i + b_n \end{bmatrix} \\ &= (A + A^T) x + b \end{aligned}$$

Hessian:

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

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

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

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]
Based on the results from 3.2, we have

$$\nabla_x f(x) = 2(A + A^T)x + b$$

$$\nabla_x^2 f(x) = 2(A + A^T)$$

3.4. 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 1]
Since A has full row rank and $m < n$, Ax=b has infinitely many solutions. One particular solution among those is the one with a minimal l^2-norm. Finding it can be formulated as solving the following constrained optimization problem:

$$\min_x \|x\|_2^2 = \min_x x^T x$$
$$s.t. \quad Ax = b$$

The Lagrangian is $L(x, \lambda) = x^T x + \lambda^T (Ax - b), \lambda \in \mathbb{R}^m, \lambda \geq \mathbf{0}$.
The Lagrange dual problem can be solved as follows:

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

$$\frac{\partial L}{\partial \lambda} = Ax - b = \mathbf{0} \tag{2}$$

By plugging (1) to (2), we have

$$\begin{aligned} &-\frac{1}{2} AA^T \lambda - b = \mathbf{0} \\ &\implies AA^T \lambda = -2b \\ &\implies \lambda = -2(AA^T)^{-1} b \end{aligned} \tag{3}$$

Since $rank(A) = m$, we have $rank(AA^T) = rank(A) = m$, i.e., the $m \times m$ square matrix AA^T has full rank, therefore it is invertible.
By plugging (3) back to (1), we have

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

as one particular solution.

[Solution 2]
Since $rank(A) = m$, 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 $m \times m$ full rank matrix, i.e. it is invertible. We can further rewrite $Ax = b$ as

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

where vector $x_1 \in \mathbb{R}^m$, vector $x_2 \in \mathbb{R}^{n-m}$.

Then, one particular solution can be found as $x_1 = A_1^{-1} b$ and x_2 is a zero vector.

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}. (2 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 matrix A is adj(A) is

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

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. 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}. Hint: For matrices, the multiplication does not commute, e.g. we cannot claim that $AB = BA$ (3 pts)

[Solution]
The inverse can be found by doing row reduction as follows:

$$\left[\begin{array}{cc|cc} A & B & I & 0 \\ C & D & 0 & I \end{array}\right]$$
$$\xrightarrow{R_1 \leftarrow A^{-1}R_1}$$
$$\left[\begin{array}{cc|cc} I & A^{-1}B & A^{-1} & 0 \\ C & D & 0 & I \end{array}\right]$$
$$\xrightarrow{R_2 \leftarrow R_2 - CR_1}$$
$$\left[\begin{array}{cc|cc} I & A^{-1}B & A^{-1} & 0 \\ 0 & D - CA^{-1}B & -CA^{-1} & I \end{array}\right]$$
$$\xrightarrow{R_2 \leftarrow (D-CA^{-1}B)^{-1}R_2}$$
$$\left[\begin{array}{cc|cc} I & A^{-1}B & A^{-1} & 0 \\ 0 & I & -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{array}\right]$$
$$\xrightarrow{R_1 \leftarrow R_1 - A^{-1}BR_2}$$
$$\left[\begin{array}{cc|cc} I & 0 & A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ 0 & I & -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{array}\right]$$

$$M^{-1} = \left[\begin{array}{cc} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{array}\right]$$

This holds under the assumption that $D - CA^{-1}B$ and A are invertible.