SI231B: Matrix Computations, 2025 Fall

Homework Set #1

Acknowledgements:

1) Deadline: **2025-10-26 23:59:59**
2) Please submit the PDF file to gradescope. Course entry code: N2382J.
3) You have 5 “free days” in total for all late homework submissions.
4) If your homework is handwritten, please make it clear and legible.
5) All your answers are required to be in English.
6) Write down the major steps for deriving the solution; otherwise you may loss points.

Problem 1. (Subspaces and Decompositions) (20 points)

1) Let $\mathcal{V} = \mathbb{R}^2$. Determine whether each of the following is a subspace of $\mathcal{V}$. Recall that $\mathbb{Q}$ denotes the set of rational numbers. Justify your answer.
   a) $\mathcal{S}_1 = \{(x, y) \in \mathbb{R}^2 \mid x - 2y = 0\}$. (4 points)
   b) $\mathcal{S}_2 = \{(x, y) \in \mathbb{R}^2 \mid x + y \in \mathbb{Q}\}$. (4 points)

2) Now let $\mathcal{V} = \mathbb{C}^{n\times n}$, the set of all $n \times n$ complex matrices, viewed as a vector space over the real numbers $\mathbb{R}$, where:
   a) Vector addition is the conventional matrix addition, i.e., for $A, B \in \mathcal{V}$, $(A + B)_{i,j} = A_{ij} + B_{ij}, \forall 1 \leq i, j \leq n$, and A_{ij} (or equivalently $A_{i,j}$) denotes the entry of matrix A located in the i-th row and j-th column;
   b) Scalar multiplication is defined only for real scalars, i.e., given $\alpha \in \mathbb{R}$, $(\alpha A)_{ij} = \alpha \cdot A_{ij}, \forall 1 \leq i, j \leq n$.

   Define the map $\Phi : \mathcal{V} \to \mathcal{V}$ by taking the complex conjugate of every entry of a matrix:

$$\Phi(A) = \overline{A}.$$

   It is given that Φ satisfies the following two properties:

   - For all real numbers α, β and all matrices $A, B \in \mathcal{V}$,

$$\Phi(\alpha A + \beta B) = \alpha\Phi(A) + \beta\Phi(B).$$

   - Applying Φ twice returns the original matrix:

$$\Phi(\Phi(A)) = A \quad \text{for all } A \in \mathcal{V}.$$

   a) Show that the sets

$$\mathcal{V}_+ = \{A \in \mathcal{V} \mid \Phi(A) = A\}, \quad \mathcal{V}_- = \{A \in \mathcal{V} \mid \Phi(A) = -A\}$$

   are subspaces of $\mathcal{V}$ over $\mathbb{R}$. (6 points)

   b) Prove that every matrix $A \in \mathcal{V}$ can be written **in exactly one way** as

$$A = A_+ + A_-,$$

   where $A_+ \in \mathcal{V}_+$ and $A_- \in \mathcal{V}_-$. And, show that

$$\mathcal{V} = \mathcal{V}_+ + \mathcal{V}_- \quad \text{and} \quad \mathcal{V}_+ \cap \mathcal{V}_- = \{0\}.$$

   (6 points)

Problem 2. (Span and Rank) (15 points)

1) a) A general definition of range: a function $L : \mathbb{R}^n \to \mathbb{R}^m$ is called a linear transformation (or linear map) if it satisfied the following two properties for any vector $u, v \in \mathbb{R}^n$ and any scalar $\alpha \in \mathbb{R}$:

   - **Additivity:**
$$L(u + v) = L(u) + L(v),$$
   - **Homogeneity (scalar multiplication compatibility):**
$$L(\alpha v) = \alpha L(v).$$

   Equivalently, L is linear if for all $u, v \in V$ and $\alpha, \beta \in \mathbb{R}$, $L(\alpha u + \beta v) = \alpha L(u) + \beta L(v)$. The *range* (or *image*) of a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ is the set
$$\mathcal{R}(L) := \{L(v) \in \mathbb{R}^m \mid v \in \mathbb{R}^n\}.$$
   It is a subspace of $\mathbb{R}^m$. Given vectors $w_1, \ldots, w_k \in \mathbb{R}^m$, their *span* is the set of all linear combinations:
$$\operatorname{span}\{w_1, \ldots, w_k\} := \left\{ \sum_{i=1}^{k} \alpha_i w_i \,\middle|\, \alpha_i \in \mathbb{R} \right\},$$
   which is the smallest subspace of $\mathbb{R}^m$ containing all w_i.
   Let $\{v_1, \ldots, v_n\}$ be a basis for $\mathbb{R}^n$. Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation into $\mathbb{R}^m$. Prove $\mathcal{R}(L) = \operatorname{span}\{L(v_1), \ldots, L(v_n)\}$. (5 points)

   b) Consider a simple fully-connected layer of neural network with the input $\mathbf{x} \in \mathbb{R}^n$, the weight matrix $\mathbf{W} \in \mathbb{R}^{m \times n}$, and the output $\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}$ ($b \in \mathbb{R}^m$). Prove that all the outputs through this network are $\mathcal{R}(\mathbf{W}) + \mathbf{b}$. (Hint: if the outputs given all basis vectors of $\mathbb{R}^n$ as inputs are known, then the output given any input can be predicted. A higher rank of $\mathbf{W}$ enhances the model's representational capacity. $\mathcal{R}(A) + \mathbf{b} = \{\sum_{i=1}^{n} x_i \mathbf{A}(:, i) + \mathbf{b} \mid \mathbf{x} \in \mathbb{R}^n\}$) (5 points)

2) Let $A \in \mathbb{R}^{m \times n}$ have rank p. Prove that there exist matrices $X \in \mathbb{R}^{m \times p}$ and $Y \in \mathbb{R}^{n \times p}$, both of full column rank (i.e., $\operatorname{rank}(X) = \operatorname{rank}(Y) = p$), such that
$$A = XY^\top.$$
(5 points)

Problem 3. (Flop Counting and Algorithm Complexity) (10 points)

1) Recall that the operation `y = y + a*x` for scalars $a, x, y \in \mathbb{R}$ costs **2 flops** (one multiplication and one addition). Complete the following table with the total number of floating-point operations (flops) required for each vector/matrix operation. Briefly justify your answer for the last three rows. (6 points)

Operation	Dimensions	Flops
$\alpha = \mathbf{u}^\top \mathbf{v}$	$\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$	$2p$
$\mathbf{w} = \mathbf{w} + \beta \mathbf{u}$	$\beta \in \mathbb{R},\ \mathbf{u}, \mathbf{w} \in \mathbb{R}^p$	$2p$
$\mathbf{z} = \mathbf{z} + \mathbf{M}\mathbf{u}$	$\mathbf{M} \in \mathbb{R}^{q \times p},\ \mathbf{u} \in \mathbb{R}^p,\ \mathbf{z} \in \mathbb{R}^q$	________
$\mathbf{N} = \mathbf{N} + \mathbf{v}\mathbf{u}^\top$	$\mathbf{N} \in \mathbb{R}^{q \times p},\ \mathbf{u} \in \mathbb{R}^p,\ \mathbf{v} \in \mathbb{R}^q$	________
$\mathbf{D} = \mathbf{D} + \mathbf{P}\mathbf{Q}$	$\mathbf{P} \in \mathbb{R}^{q \times r},\ \mathbf{Q} \in \mathbb{R}^{r \times p},\ \mathbf{D} \in \mathbb{R}^{q \times p}$	________

2) Let $\mathbf{H} \in \mathbb{R}^{n \times n}$ be defined element-wise by

$$h_{ij} = \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_{ik}\, b_{k\ell}\, c_{\ell j}\, d_{ij}.$$

A naive implementation that computes each h_{ij} directly requires $\mathcal{O}(n^4)$ flops. Design an algorithm to compute $\mathbf{H}$ using only $\mathcal{O}(n^3)$ flops. Express your method in terms of matrix operations. You may use:

- Matrix multiplication,
- Matrix transpose,
- Hadamard (element-wise) product: $(\mathbf{X} \circ \mathbf{Y})_{ij} = x_{ij} y_{ij}$.

(Hint: $\mathbf{H}$ can be represented as the Hadamard product of $\mathbf{ABC}$ and $\mathbf{D}$.) (4 points)

Problem 4. (Norms) (15 points)

1) Show that for any $w \in \mathbb{R}^n$, $\|w\|_1 \|w\|_\infty \leq \frac{1+\sqrt{n}}{2} \|w\|_2^2$. (5 points)
2) Suppose $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$. Show that if $E = uv^T$, then $\|E\|_F = \|E\|_2 = \|u\|_2 \|v\|_2$ (5 points)
3) Given $A \in \mathbb{R}^{m \times n}$, show that $\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}$. (5 points) Hint: Let $a_i, b_i \in \mathbb{R}$ for $i = 1, 2, \ldots, n$, then the following inequality holds:

$$\left(\sum_{i=1}^{n} a_i b_i \right)^2 \leq \left(\sum_{i=1}^{n} a_i^2 \right) \left(\sum_{i=1}^{n} b_i^2 \right)$$

Problem 5. (LU Decomposition) (25 points)

Consider $\mathbf{A} = \begin{bmatrix} 1 & 3 & 1 & -2 \\ 2 & 2 & 0 & 5 \\ -6 & 3 & 4 & 8 \\ 4 & 2 & -1 & 7 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 5 \end{bmatrix}$.

1) Please determine whether matrix $\mathbf{A}$ has an LU decomposition. If it does, provide an LU decomposition; if it does not, explain the reason. (10 points)
2) Use the conclusion from problem 1) to assist in solving $\mathbf{Ax} = \mathbf{b}$. (5 points)
3) Perform partial pivoting to derive an LU decomposition on permuted $\mathbf{A}$ (5 points)
4) Using the conclusion from problem 1) , provide an LDM decomposition of $\mathbf{A}$:

$$A = LDM^T$$

where:

- L is a unit lower triangular matrix,
- $D = \text{Diag}(d_1, d_2, \ldots, d_n)$ is a diagonal matrix,
- M is a unit lower triangular matrix (M^T is therefore a unit upper triangular matrix).

(5 points)

Problem 6. (Cholesky Decomposition) (15 points)

1) Given $A = \begin{pmatrix} 4 & 2 & 2 & 4 \\ 2 & 5 & 1 & 2 \\ 2 & 1 & 5 & 4 \\ 4 & 2 & 4 & 6 \end{pmatrix}$, provide a Cholesky decomposition of A. (7.5 points)
2) Let A be an $n \times n$ real symmetric positive definite matrix ($n \geq 2$) with Cholesky decomposition $A = GG^T$. Denote the k-th leading principal minor of A by $\Delta_k = \det(A_k)$, where A_k is the submatrix formed by the first k rows and first k columns of A, and let $\Delta_0 = 1$ by convention.
   Prove that for any $k \in \{1, 2, \ldots, n\}$, $g_{kk}^2 = \frac{\Delta_k}{\Delta_{k-1}}$; (7.5 points)
   Hint: You need to use mathematical induction. Additionally, you could partition A_k and G_k into the following form :

$$A_k = \begin{bmatrix} A_{k-1} & * \\ * & * \end{bmatrix}, \quad G_k = \begin{bmatrix} G_{k-1} & * \\ * & * \end{bmatrix}$$