All hw should be uploaded to canvas as a *pdf*. Make sure that if you scan your handwritten notes that they are legible and appropriately oriented. If you use an online resource to solve any problem, please appropriately cite that source.

Problem 1. (Characteristic Polynomial.) One of the most celebrated linear algebra results relating to eigenvalues is the Cayley-Hamilton theorem. Recall that the characteristic polynomial is given by

$$p(\lambda) = \det(\lambda I - A) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1\lambda + \alpha_0.$$

The Cayley-Hamilton theorem states that

$$p(A) = A^n + \alpha_{n-1}A^{n-1} + \cdots + \alpha_1 A + \alpha_0 I = 0,$$

where we now view $p : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ as a mapping on the space of $\mathbb{R}^{n\times n}$ matrices.
This theorem holds in general for any matrix. In this problem you will show it holds for the following easier setting. Suppose A is diagonalizable.

a. Recall that in **Mod3-L1** we saw how to compute powers of matrices that are diagonalizable—i.e.,

$$A^k = V\Lambda^k V^{-1},$$

where V is a matrix containing the eigenvalues of A, and Λ is a diagonal matrix with the eigenvalues. Consider a polynomial $q(s) = a_m s^m + a_{m-1}s^{m-1} + \cdots + a_1 s + a_0$. Show that $q(A) = Vq(\Lambda)V^{-1}$ where $q(\Lambda) = \mathrm{diag}(q(\lambda_1), \ldots, q(\lambda_n))$.

b. Now, apply part a. to the polynomial $p(\lambda) = \det(\lambda I - A)$ to show that $p(A) = 0$.

c. **Extra Credit Challenge Problem.** Prove that $p(A) = 0$ holds in general—that is, it holds not just for diagonalizable matrices, but all matrices.
Hint: use the fact that polynomials are continuous functions, and the fact that diagonalizable matrices are dense in the space of all $n \times n$ matrices which we denote $\mathbb{R}^{n\times n}$—i.e., for any $\varepsilon > 0$ there exists a $\Delta \in \mathbb{R}^{n\times}$ with $\|\Delta\|_F \leq \varepsilon$ such that $A + \Delta$ is diagonalizable.

Problem 2. (Symmetric Matrices and Similarity Transforms .) Provide a counterexample or justification for your answer.

a. **(True/False)**. Is the statement true or false:

$$A = A^\top, \text{ and } A = VBV^{-1} \text{ with } V^\top V = I \implies B = B^\top \text{ is symmetric}$$

b. **(True/False)**. Is the statement true or false:

$$A \text{ is similar to } B \text{ and } A = A^\top \text{ is symmetric} \implies B = B^\top \text{ is symmetric}$$

c. This one is a little tricky, but not too bad.

The function $p : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ with values $p(A) = A^n + \alpha_{n-1}A^{n-1} + \cdots + \alpha_1 A + \alpha_0 I$ is continuous on $\mathbb{R}^{n\times n}$. This map is identically zero on the dense subset of $\mathbb{R}^{n\times n}$ formed by diagonalizable matrices which we showed in part b., hence by continuity it must be zero everywhere in $\mathbb{R}^{n\times n}$.

Problem 3. (Operator Norms.) In **Mod3-L2** we saw the concept of an operator norm:

$$\|A\|_p = \max_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|=1} \|Ax\|_p.$$

In lecture we saw that

$$\|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |a_{ij}| \quad \text{i.e., the max column sum}$$

where $\|x\|_1 = \sum_i |x_i|$. Show that this equality holds. That is, show that $\|A\|_1$ is in fact the max column sum.

Problem 4. (Definiteness of Stochastic Matrices.) An $n \times n$ matrix is called a Markov or Stochastic matrix if all entries are nonnegative and the sum of each column vector is equal to one, or if the sum of each row vector is equal to one. More specifically, a matrix A is column stochastic if $\mathbf{1}^\top A = \mathbf{1}^\top$ and row stochastic if $A\mathbf{1} = \mathbf{1}$. These matrices show up in a number of ML applications including reinforcement learning. For example, Markov chains have Markov matrices that are stochastic.

Show that if $A \succ 0$ and it is column (respectively, row) stochastic, then its spectral radius is $\lambda = 1$. That is, show that the maximum modulus eigenvalue is $\lambda = 1$ and all other eigenvalues have magnitude smaller than one.

Problem 5. (PCA and Low Rank Approximations.) The problem of finding a low rank approximation is as follows: given a data matrix X, we seek matrices $Y, Z^\top$ such that $X_k = YZ^\top$ is a rank k approximation of X. Recall from **Mod3-L3** we saw that we can use PCA to produce a low rank approximation via the following procedure:

step 1: Preprocess the data $(z^{(1)}, \ldots, z^{(m)})$ as before: so that the rows sum to the all-zero vector and, normalize each column

step 2: Form the covariance matrix $X^\top X$

step 3: Take the k rows of $Z^\top$ to be the top k principal components of X—the k eigenvectors $u_1, \ldots, u_k$ of $X^\top X$ with largest eigenvalues

step 4: For $i = 1, \ldots, m$, the i-th row of Y is defined as the projections $(\langle x^{(i)}, u_1\rangle, \ldots, \langle x^{(i)}, u_k\rangle)$.

step 5: produce $X_k = YZ^\top$

We also saw that SVD can produce a low rank approximation as follows: $\tilde{X}_k = \sum_{i=1}^{k} \sigma_i u_i v_i^\top$ where $X = U\Sigma V^\top$ is the SVD of X. Show that the matrix X_k and the matrix $\tilde{X}_k$ are identical.

Problem 6. (Image Reconsrtuction with SVD.) This problem is in Python, and can be found in the `hw4.ipynb` Jupyter notebook provided. In addition to submitting your python notebook, print the python notebook to a pdf and attach it at the end of document to get graded. We grade the actual pdf and collect the notebook for posterity.