Department of Mathematics, University of Utah

Analysis of Numerical Methods I

MATH 6610 – Section 001 – Fall 2025

Homework 4

Singular Values, I

Due Wednesday, September 17, 2025

Submission instructions:
Submit your assignment on gradescope.

Problem assignment:

1. (The spectral theorem for normal matrices) Prove that $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ is unitarily diagonalizable iff it's normal.
(It's helpful to recall the Schur Decomposition from the previous assignment.)

2. Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$.

(a) Assume $\boldsymbol{A}$ is Hermitian positive semi-definite. Show that the singular values and the eigenvlaues of $\boldsymbol{A}$ are the same.

(b) If $\boldsymbol{A}$ is normal, how are the eigenvalues of $\boldsymbol{A}$ and the singular values of $\boldsymbol{A}$ related?

3. Let $\boldsymbol{A} \in \mathbb{C}^{m \times n}$. Show that $\|\boldsymbol{A}\|_F^2 = \|\boldsymbol{\sigma}\|_2^2$, where $\|\cdot\|_F$ is the Frobenius norm, $\|\cdot\|_2$ is the ℓ^2 norm on vectors, and $\boldsymbol{\sigma} \in \mathbb{C}^{\min\{m,n\}}$ is the vector of singular values of $\boldsymbol{A}$.

4. (Schmidt-Eckart-Young-Mirsky in the ℓ^2 norm) With $\|\cdot\|_2$ the induced 2-norm, and for an arbitrary $k \in \mathbb{N}$ and $\boldsymbol{A} \in \mathbb{C}^{m \times n}$, prove that the matrix $\boldsymbol{A}_k$ defined as,

$$\boldsymbol{A}_k = \operatorname*{argmin}_{\text{rank}(\boldsymbol{B}) \leq k} \|\boldsymbol{A} - \boldsymbol{B}\|_2,$$

is the rank-$\min\{k, \text{rank}(\boldsymbol{A})\}$ truncated SVD.
(The typical approach: for any rank-k $\boldsymbol{B}$, construct a vector $\boldsymbol{v}$ lying both in $\ker(\boldsymbol{B})$ and in the span of the first $k+1$ right-singular vectors of $\boldsymbol{A}$, and compute a lower bound for $\|(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{v}\|_2$.)

5. (Orthogonal projectors are distance-minimizers) Let $\boldsymbol{x} \in \mathbb{C}^n$ be arbitrary, and let $\boldsymbol{P}$ be an orthogonal projector onto some subspace $V \subset \mathbb{C}^n$. Prove that,

$$\|\boldsymbol{x} - \boldsymbol{P}\boldsymbol{x}\|_2 = \min_{\boldsymbol{v} \in V} \|\boldsymbol{x} - \boldsymbol{v}\|_2.$$

(The only substantive tools you need are basic properties of orthogonal projectors and the Pythagorean Theorem.)

6. (Principal Component Analysis) Let $\boldsymbol{B} \in \mathbb{C}^{m \times n}$ be comprised of columns $\boldsymbol{b}_j$, $j \in [n]$. In this problem, we will view each column as a single piece of data in $\mathbb{C}^m$-dimensional space. In particular, we'll consider these columns as random realizations from some probability distribution on vectors. Under this model, the (empirical) mean of our set of realizations is

$$\boldsymbol{b}_0 \coloneqq \frac{1}{n} \sum_{j \in [n]} \boldsymbol{b}_j,$$

which is the "average" data point. Often we are interested in $m \gg n$, but this assumption is not needed.

For a fixed $k \leq \min\{m, n\}$, the ultimate goal of Principal Component Analysis (PCA) is to define a compression matrix $\boldsymbol{Q}_k \in \mathbb{C}^{m \times k}$, $\boldsymbol{Q}_k^* \boldsymbol{Q}_k = \boldsymbol{I}$, such that the compression of the data, $\boldsymbol{Q}_k^* \boldsymbol{b}_j$ contains as much of the "variance" of $\boldsymbol{B}$ as possible. In this problem, we'll maximize the variance of the projected data, defined as the quadratic variation of the data. Let $\boldsymbol{A} = \boldsymbol{B} - \boldsymbol{b}_0 \boldsymbol{1}^*$, where $\boldsymbol{1}^*$ is an $n \times 1$ vector of ones. Then PCA chooses the dimension-k compression/reduction as the following affine transformation:

$$\boldsymbol{c}_j = \boldsymbol{Q}_k^* \boldsymbol{a}_j, \qquad\qquad \boldsymbol{a}_j \coloneqq (\boldsymbol{b}_j - \boldsymbol{b}_0),$$

The matrix $\boldsymbol{C} \in \mathbb{C}^{k \times n}$ with columns $\boldsymbol{c}_j$ is the PCA representation of $\boldsymbol{B}$, and the lifted representation:

$$\widetilde{\boldsymbol{b}}_j = \boldsymbol{Q}_k \boldsymbol{c}_j + \boldsymbol{b}_0,$$

is the rank-k approximation of $\boldsymbol{b}_j$ in the full m-dimensional space. The particular $\boldsymbol{Q}_k$ chosen by PCA is $\boldsymbol{Q}_k = \boldsymbol{U}_k$, where $\boldsymbol{U}_k$ is the first k columns of the left singular-vector matrix of $\boldsymbol{A}$. The columns of $\boldsymbol{U}_k$ are the *principal components* of the data. We will show why this is chosen in this problem.

Our definition of "variance" will be quadratic variation from the mean $\boldsymbol{b}_0$ when orthogonally projected onto the subspace V:

$$\mathrm{var}_V(\boldsymbol{B}) \coloneqq \sum_{j \in [n]} \|\boldsymbol{P}_V \boldsymbol{a}_j\|_2^2,$$

where $\boldsymbol{P}_V$ is the orthogonal projection onto the subspace V.

(a) Show that if $P_{\boldsymbol{U}}$ is the orthogonal projector onto range($\boldsymbol{U}$), then for any $\boldsymbol{A} \in \mathbb{C}^{m \times n}$:

$$\boldsymbol{P}_{\boldsymbol{U}_k} \boldsymbol{A} = \sum_{j \in [k]} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^*,$$

where $\boldsymbol{u}_j$ and $\boldsymbol{v}_j$ are the ordered left- and right-singular vectors of $\boldsymbol{A}$, and σ_j are the corresponding singular values of $\boldsymbol{A}$. (This part really has nothing to do with PCA.)

(b) For any $\boldsymbol{A} \in \mathbb{C}^{m \times n}$, show that $\boldsymbol{U}_k$ solves the optimization problem,

$$\underset{\substack{\boldsymbol{U} \in \mathbb{C}^{m \times k} \\ \boldsymbol{U}^* \boldsymbol{U} = \boldsymbol{I}}}{\mathrm{argmin}} \|\boldsymbol{A} - \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{A}\|_F^2.$$

(Again, not explicitly related to PCA. You may use Schmidt-Eckart-Young-Mirksy here.)

(c) Prove that the PCA choice $V = \mathrm{span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ solves the optimization problem,

$$\underset{\dim V = k}{\mathrm{argmax}}\, \mathrm{var}_V(\boldsymbol{B}),$$

and hence $\boldsymbol{Q}_k = \boldsymbol{U}_k$ as in PCA maximizes variance of the projected data.