

Homework 3

Math 5424

Numerical Linear Algebra

Alexander Novotny

Zuriah Quinton

October 23, 2023

1. Let \mathbf{A} be $n \times m$ with $n > m$. Show that $\|\mathbf{A}^\top \mathbf{A}\|_2 = \|\mathbf{A}\|_2^2$ and $\kappa_2(\mathbf{A}^\top \mathbf{A}) = \kappa_2(\mathbf{A})^2$.

Let \mathbf{M} be $n \times n$ and positive definite and \mathbf{L} be its Cholesky factor so that $\mathbf{M} = \mathbf{L}\mathbf{L}^\top$. Show that $\|\mathbf{M}\|_2 = \|\mathbf{L}\|_2^2$ and $\kappa_2(\mathbf{M}) = \kappa_2(\mathbf{L})^2$.

Proof. We have

$$\|\mathbf{A}^\top \mathbf{A}\|_2 = \max_i |\lambda_i(\mathbf{A}^\top \mathbf{A})|$$

since $\mathbf{A}^\top \mathbf{A}$ is normal, and

$$\|\mathbf{A}\|_2^2 = \sqrt{\max_i \lambda_i(\mathbf{A}^\top \mathbf{A})}^2 = \max_i |\lambda_i(\mathbf{A}^\top \mathbf{A})|.$$

So $\|\mathbf{A}^\top \mathbf{A}\|_2 = \|\mathbf{A}\|_2^2$.

Next, we have

$$\begin{aligned} \kappa_2(\mathbf{A}^\top \mathbf{A}) &= \|(\mathbf{A}^\top \mathbf{A})^{-1}\|_2 \|\mathbf{A}^\top \mathbf{A}\|_2 \\ &= \max_i |\lambda_i((\mathbf{A}^\top \mathbf{A})^{-1})| \max_i |\lambda_i(\mathbf{A}^\top \mathbf{A})| \\ &= \frac{1}{\sigma_m^2} \sigma_1^2 \\ &= \frac{\sigma_1^2}{\sigma_m^2}, \end{aligned}$$

where σ_i is the i^{th} singular value of \mathbf{A} in descending order. Further,

$$\begin{aligned} \kappa_2(\mathbf{A})^2 &= \left(\frac{\sigma_1}{\sigma_m} \right)^2 \\ &= \frac{\sigma_1^2}{\sigma_m^2}. \end{aligned}$$

Note, we may have $\sigma_m = 0$ if \mathbf{A} is not full rank. In this case, we still have $\kappa_2(\mathbf{A}^\top \mathbf{A}) = \infty = \kappa_2(\mathbf{A})^2$.

Next, for $\mathbf{M} = \mathbf{L}\mathbf{L}^\top$, we have

$$\begin{aligned} \|\mathbf{M}\|_2 &= \|\mathbf{L}\mathbf{L}^\top\|_2 = \max_i |\lambda_i(\mathbf{L}\mathbf{L}^\top)| \\ \text{and } \|\mathbf{L}\|_2^2 &= \sqrt{\max_i \lambda_i(\mathbf{L}^\top \mathbf{L})}^2 = \max_i |\lambda_i(\mathbf{L}^\top \mathbf{L})|. \end{aligned}$$


By lemma 1 these are equivalent.

Finally, for the condition number of \mathbf{M} we have

$$\begin{aligned}
 \kappa_2(\mathbf{M}) &= \kappa_2(\mathbf{L}\mathbf{L}^\top) \\
 &= \|(\mathbf{L}\mathbf{L}^\top)^{-1}\|_2 \|\mathbf{L}\mathbf{L}^\top\|_2 \\
 &= \max_i |\lambda_i((\mathbf{L}\mathbf{L}^\top)^{-1})| \max_i |\lambda_i(\mathbf{L}\mathbf{L}^\top)| \\
 &= \frac{1}{\lambda_m} \lambda_1 \\
 &= \frac{\sigma_1^2}{\sigma_m^2}.
 \end{aligned}$$

Since \mathbf{M} is positive definite, we know $\sigma_i > 0$ for all i , so this value is finite. Further,

$$\begin{aligned}
 \kappa_2(\mathbf{L})^2 &= \left(\frac{\sigma_1}{\sigma_m} \right)^2 \\
 &= \frac{\sigma_1^2}{\sigma_m^2}.
 \end{aligned}$$

Thus, $\kappa_2(\mathbf{M}) = \kappa_2(\mathbf{L})^2$. 


Lemma 1. If \vec{v} is an eigenvector of $\mathbf{L}\mathbf{L}^\top$, with λ its eigenvalue, then λ is an eigenvalue of $\mathbf{L}^\top\mathbf{L}$, with eigenvector $\mathbf{L}^\top\vec{v}$.

Proof. We have

$$\begin{aligned}
 \mathbf{L}\mathbf{L}^\top\vec{v} &= \lambda\vec{v} \\
 \implies (\mathbf{L}^\top\mathbf{L})(\mathbf{L}^\top\vec{v}) &= \lambda(\mathbf{L}^\top\vec{v}).
 \end{aligned} \tag{1}$$

For $\vec{v} \notin \ker(\mathbf{L}^\top)$. Suppose for contradiction that $\vec{v} \in \ker(\mathbf{L}^\top)$. Then eq. (1) is

$$\mathbf{L} \underbrace{\mathbf{L}^\top\vec{v}}_{=0} = \lambda\vec{v} \implies \lambda = 0,$$

which is absurd since the eigenvalues of a positive definite matrix are non-zero. Thus, we have $\vec{v} \notin \ker(\mathbf{L}^\top)$, and so λ is always an eigenvalue of $\mathbf{L}^\top\mathbf{L}$. 

2. In this question we will ask how to solve $\mathbf{B}\vec{y} = \vec{c}$ given a fast way to solve $\mathbf{A}\vec{x} = \vec{b}$, where $\mathbf{A} - \mathbf{B}$ is “small” in some sense.

(a) Prove the *Sherman-Morrison formula*: Let \mathbf{A} be nonsingular, \vec{u} and \vec{v} be column vectors, and $\mathbf{A} + \vec{u}\vec{v}^\top$ be nonsingular. Then $(\mathbf{A} + \vec{u}\vec{v}^\top)^{-1} = \mathbf{A}^{-1} - (\mathbf{A}^{-1}\vec{u}\vec{v}^\top\mathbf{A}^{-1})/(1 + \vec{v}^\top\mathbf{A}^{-1}\vec{u})$.

More generally, prove the *Sherman-Morrison-Woodbury formula*: Let \mathbf{U} and \mathbf{V} be $n \times k$ rectangular matrices, where $k < n$ and \mathbf{A} is $n \times n$. Then $\mathbf{T} = \mathbf{I} + \mathbf{V}^\top\mathbf{A}^{-1}\mathbf{U}$ is nonsingular if and only if $\mathbf{A} + \mathbf{U}\mathbf{V}^\top$ is nonsingular, in which case $(\mathbf{A} + \mathbf{U}\mathbf{V}^\top)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\mathbf{T}^{-1}\mathbf{V}^\top\mathbf{A}^{-1}$.

Proof. First, we have

$$\begin{aligned}
 (\mathbf{A} + \vec{u}\vec{v}^\top) \left(\mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\vec{u}\vec{v}^\top\mathbf{A}^{-1}}{1 + \vec{v}^\top\mathbf{A}^{-1}\vec{u}} \right) &= \cancel{\mathbf{A}\mathbf{A}^{-1}} - \frac{\mathbf{I}\mathbf{A}\mathbf{A}^{-1}\vec{u}\vec{v}^\top\mathbf{A}^{-1}}{1 + \vec{v}^\top\mathbf{A}^{-1}\vec{u}} + \vec{u}\vec{v}^\top\mathbf{A}^{-1} - \frac{\vec{u}\vec{v}^\top\mathbf{A}^{-1}\vec{u}\vec{v}^\top\mathbf{A}^{-1}}{1 + \vec{v}^\top\mathbf{A}^{-1}\vec{u}} \\
 &= \mathbf{I} - \frac{\vec{u}\vec{v}^\top\mathbf{A}^{-1} - \vec{u}\vec{v}^\top\mathbf{A}^{-1}(1 + \vec{v}^\top\mathbf{A}^{-1}\vec{u}) + \vec{u}\vec{v}^\top\mathbf{A}^{-1}\vec{u}\vec{v}^\top\mathbf{A}^{-1}}{1 + \vec{v}^\top\mathbf{A}^{-1}\vec{u}} \\
 &= \mathbf{I} - \frac{\cancel{\vec{u}\vec{v}^\top\mathbf{A}^{-1}} - \cancel{\vec{u}\vec{v}^\top\mathbf{A}^{-1}} - \vec{u}\vec{v}^\top\mathbf{A}^{-1}\vec{v}^\top\mathbf{A}^{-1}\vec{u} + \vec{u}\vec{v}^\top\mathbf{A}^{-1}\vec{u}\vec{v}^\top\mathbf{A}^{-1}}{1 + \vec{v}^\top\mathbf{A}^{-1}\vec{u}} \\
 &= \mathbf{I} - \frac{\vec{v}^\top\mathbf{A}^{-1}\vec{u}(-\cancel{\vec{u}\vec{v}^\top\mathbf{A}^{-1}} + \cancel{\vec{u}\vec{v}^\top\mathbf{A}^{-1}})}{1 + \vec{v}^\top\mathbf{A}^{-1}\vec{u}} \\
 &= \mathbf{I},
 \end{aligned}$$

so $(\mathbf{A} + \vec{u}\vec{v}^\top)^{-1} = \mathbf{A}^{-1} - (\mathbf{A}^{-1}\vec{u}\vec{v}^\top\mathbf{A}^{-1})/(1 + \vec{v}^\top\mathbf{A}^{-1}\vec{u})$. Then, assume $\mathbf{T} = \mathbf{I} + \mathbf{V}^\top\mathbf{A}^{-1}\mathbf{U}$ is nonsingular. We have

$$\begin{aligned} (\mathbf{A} + \mathbf{UV}^\top)(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UT}^{-1}\mathbf{V}^\top\mathbf{A}^{-1}) &= \cancel{\mathbf{AA}^{-1}} - \cancel{\mathbf{AA}^{-1}}\mathbf{UT}^{-1}\mathbf{V}^\top\mathbf{A}^{-1} \\ &\quad + \mathbf{UV}^\top\mathbf{A}^{-1} - \mathbf{UV}^\top\mathbf{A}^{-1}\mathbf{UT}^{-1}\mathbf{V}^\top\mathbf{A}^{-1} \\ &= \mathbf{I} - \mathbf{U}(\mathbf{T}^{-1} - \mathbf{I} + \underbrace{\mathbf{V}^\top\mathbf{A}^{-1}\mathbf{U}\mathbf{T}^{-1}}_{\mathbf{T}^{-1}})\mathbf{V}^\top\mathbf{A}^{-1} \\ &= \mathbf{I} - \mathbf{U}(\cancel{\mathbf{T}^{-1}} - \mathbf{I} + \cancel{\mathbf{T}\mathbf{T}^{-1}} - \cancel{\mathbf{T}^{-1}})\mathbf{V}^\top\mathbf{A}^{-1} \\ &= \mathbf{I}, \end{aligned}$$

so $\mathbf{A} + \mathbf{UV}^\top$ is nonsingular and $(\mathbf{A} + \mathbf{UV}^\top)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UT}^{-1}\mathbf{V}^\top\mathbf{A}^{-1}$. Finally, assume $\mathbf{A} + \mathbf{UV}^\top$ is nonsingular and $(\mathbf{A} + \mathbf{UV}^\top)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UBV}^\top\mathbf{A}^{-1}$ for some \mathbf{B} , $k \times k$. Then we have

$$\begin{aligned} \mathbf{I} &= (\mathbf{A} + \mathbf{UV}^\top)(\mathbf{A} + \mathbf{UV}^\top)^{-1} \\ &= (\mathbf{A} + \mathbf{UV}^\top)(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UBV}^\top\mathbf{A}^{-1}) \\ &= \cancel{\mathbf{AA}^{-1}} - \cancel{\mathbf{AA}^{-1}}\mathbf{UBV}^\top\mathbf{A}^{-1} + \mathbf{UV}^\top\mathbf{A}^{-1} - \mathbf{UV}^\top\mathbf{A}^{-1}\mathbf{UBV}^\top\mathbf{A}^{-1} \\ \Rightarrow 0 &= \mathbf{U}(-\mathbf{B} + \mathbf{I} - \mathbf{V}^\top\mathbf{A}^{-1}\mathbf{UB})\mathbf{V}^\top\mathbf{A}^{-1} \\ \Rightarrow 0 &= \mathbf{I} - \underbrace{(\mathbf{I} + \mathbf{V}^\top\mathbf{A}^{-1}\mathbf{U})}_{\mathbf{T}}\mathbf{B} \\ \Rightarrow \mathbf{I} &= \mathbf{TB}, \end{aligned}$$

therefore \mathbf{T} is invertible and non-singular, and $\mathbf{B} = \mathbf{T}^{-1}$. ☺

- (b) If you have a fast algorithm to solve $\mathbf{A}\vec{x} = \vec{b}$, show how to build a fast solver for $\mathbf{B}\vec{y} = \vec{c}$, where $\mathbf{B} = \mathbf{A} + \vec{u}\vec{v}^\top$.

Answer. We have

$$\begin{aligned} \vec{y} &= \mathbf{B}^{-1}\vec{c} \\ &= (\mathbf{A} + \vec{u}\vec{v}^\top)^{-1}\vec{c} \\ &\stackrel{2a}{=} (\mathbf{A}^{-1} - (\mathbf{A}^{-1}\vec{u}\vec{v}^\top\mathbf{A}^{-1})/(1 + \vec{v}^\top\mathbf{A}^{-1}\vec{u}))\vec{c} \\ &= \mathbf{A}^{-1}\vec{c} - (\mathbf{A}^{-1}\vec{u})\vec{v}^\top(\mathbf{A}^{-1}\vec{c})/(1 + \vec{v}^\top(\mathbf{A}^{-1}\vec{u})). \end{aligned}$$

Then, an algorithm to quickly solve this problem can be found in algorithm 1.

Algorithm 1: An algorithm to quickly solve $(\mathbf{A} + \vec{u}\vec{v}^\top)\vec{y} = \vec{c}$ given an algorithm to quickly solve $\mathbf{A}\vec{x} = \vec{b}$.

Data: \mathbf{A} , \vec{c} , \vec{u} , \vec{v}

Result: $\vec{y} = (\mathbf{A} + \vec{u}\vec{v}^\top)^{-1}\vec{c}$

- 1 Solve $\mathbf{A}\vec{w} = \vec{c}$;
 - 2 Solve $\mathbf{A}\vec{x} = \vec{u}$;
 - 3 $\vec{y} \leftarrow \vec{w} - \vec{x}\vec{v}^\top\vec{w}/(1 + \vec{v}^\top\vec{x})$;
-

3. Consider the linear system $\mathbf{A}\vec{x} = \vec{b}$ where

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 3 & 3 & 1 \\ 8 & 7 & 9 & 5 \\ 6 & 7 & 9 & 8 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 4 \\ 11 \\ 29 \\ 30 \end{bmatrix}.$$

- (a) Calculate the appropriate determinants to show that \mathbf{A} can be written as $\mathbf{A} = \mathbf{LU}$ where \mathbf{L} is a unit lower triangular matrix and \mathbf{U} is a non-singular upper triangular matrix. You may use MATLAB to compute the determinants. But the remaining parts of this problem should be done by paper-and-pencil.

Answer.

$$\det \mathbf{A}_1 = |1| = 1$$

$$\det \mathbf{A}_2 = \begin{vmatrix} 2 & 1 \\ 4 & 3 \end{vmatrix} = 2$$

$$\det \mathbf{A}_3 = \begin{vmatrix} 2 & 1 & 1 \\ 4 & 3 & 3 \\ 8 & 7 & 9 \end{vmatrix} = (2)(6) - (1)(12) + (1)(4) = 4$$

$$\det \mathbf{A}_4 = \begin{vmatrix} 2 & 1 & 1 & 0 \\ 4 & 3 & 3 & 1 \\ 8 & 7 & 9 & 5 \\ 6 & 7 & 9 & 8 \end{vmatrix} = 2 \begin{vmatrix} 3 & 3 & 1 \\ 7 & 9 & 5 \\ 7 & 9 & 8 \end{vmatrix} - 1 \begin{vmatrix} 4 & 3 & 1 \\ 8 & 9 & 5 \\ 6 & 9 & 8 \end{vmatrix} + 1 \begin{vmatrix} 4 & 3 & 1 \\ 8 & 7 & 5 \\ 6 & 7 & 8 \end{vmatrix} + 0 = 8$$

Since these determinants are all positive the LU decomposition without pivoting is possible.

- (b) Compute the LU decomposition $\mathbf{A} = \mathbf{L}\mathbf{U}$ and convert $\mathbf{A}\mathbf{x} = \mathbf{b}$ into $\mathbf{U}\mathbf{x} = \mathbf{y}$ where \mathbf{U} is upper triangular and solve for \mathbf{x} .

Answer. We have

$$\begin{bmatrix} \alpha_{11} = 2 & \vec{c}^\top = [1 & 1 & 0] \\ \vec{a} = \begin{bmatrix} 4 \\ 8 \\ 6 \end{bmatrix} & \hat{\mathbf{A}} = \begin{bmatrix} 3 & 3 & 1 \\ 7 & 9 & 5 \\ 7 & 9 & 8 \end{bmatrix} \end{bmatrix}$$

$$\text{so } \hat{l}_1 = \vec{a}/\alpha_{11} = \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix}, \text{ and}$$

$$\begin{aligned} \hat{\mathbf{A}} - \hat{l}_1 \vec{c}^\top &= \begin{bmatrix} 3 & 3 & 1 \\ 7 & 9 & 5 \\ 7 & 9 & 8 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 3 & 1 \\ 7 & 9 & 5 \\ 7 & 9 & 8 \end{bmatrix} - \begin{bmatrix} 2 & 2 & 0 \\ 4 & 4 & 0 \\ 3 & 3 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 5 \\ 4 & 6 & 8 \end{bmatrix}. \end{aligned}$$

Then the first step of LU gives

$$\begin{bmatrix} 1 & 0 \\ \hat{l}_1 & \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \vec{c}^\top \\ 0 & \hat{\mathbf{A}} - \hat{l}_1 \vec{c}^\top \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 4 & 0 & 1 & 0 \\ 3 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 3 & 5 & 5 \\ 0 & 4 & 6 & 8 \end{bmatrix}. \quad (2)$$

Repeating this process on the lower right submatrix in \mathbf{U}_1 from eq. (2) we have

$$\begin{bmatrix} \alpha_{22} = 1 & \vec{c}^\top = [1 & 1] \\ \vec{a} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} & \hat{\mathbf{A}} = \begin{bmatrix} 5 & 5 \\ 6 & 8 \end{bmatrix} \end{bmatrix}$$

$$\text{so } \hat{l}_2 = \vec{a}/\alpha_{22} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \text{ and}$$

$$\begin{aligned} \hat{\mathbf{A}} - \hat{l}_2 \vec{c}^\top &= \begin{bmatrix} 5 & 5 \\ 6 & 8 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 5 \\ 6 & 8 \end{bmatrix} - \begin{bmatrix} 3 & 3 \\ 4 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}. \end{aligned}$$

Then the second step of LU gives

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 4 & 3 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 4 \end{bmatrix}. \quad (3)$$

Repeating one final time on the lower 2×2 submatrix in eq. (3) yields

$$\begin{bmatrix} \alpha_{33} = 2 & \vec{c}^\top = [2] \\ \vec{a} = [2] & \hat{\mathbf{A}} = [4] \end{bmatrix}$$

so $\hat{l}_3 = \vec{a}/\alpha_{33} = [1]$, and

$$\begin{aligned} \hat{\mathbf{A}} - \hat{l}_3 \vec{c}^\top &= [4] - [1][2] \\ &= [2]. \end{aligned}$$

So the third, and final, step of LU gives

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 4 & 3 & 1 & 0 \\ 3 & 4 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 \end{bmatrix}. \quad (4)$$

Now, if we have $\mathbf{A}\vec{x} = \vec{b}$ then we have $\mathbf{LU}\vec{x} = \vec{b}$, where \mathbf{L} is invertible. So, $\mathbf{U}\vec{x} = \mathbf{L}^{-1}\vec{b} = \vec{y}$, and we can find \vec{y} with

$$\begin{aligned} \vec{y} = \mathbf{L}^{-1}\vec{b} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 2 & -3 & 1 & 0 \\ 3 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 11 \\ 29 \\ 30 \end{bmatrix} \\ &= \begin{bmatrix} 4 \\ -8 + 11 \\ 8 - 33 + 29 \\ 12 - 11 - 29 + 30 \end{bmatrix} \\ &= \begin{bmatrix} 4 \\ 3 \\ 4 \\ 2 \end{bmatrix}. \end{aligned}$$

Then the transformed equation is

$$\mathbf{U}\vec{x} = \vec{y} \implies \begin{bmatrix} 2 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 4 \\ 2 \end{bmatrix}.$$

Starting at the bottom, clearly $x_4 = 1$. Then we have $2x_3 + 2x_4 = 4 \implies 2x_3 + 2 = 4 \implies x_3 = 1$. Then we have $x_2 + x_3 + x_4 = 3 \implies x_2 + 1 + 1 = 3 \implies x_2 = 1$. Finally, we have $2x_1 + x_2 + x_3 = 4 \implies 2x_1 + 1 + 1 = 4 \implies x_1 = 1$. So,

$$\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

(c) Using \mathbf{L} and \mathbf{U} from the earlier steps, solve the new system $\mathbf{A}\tilde{x} = \tilde{b}$ where

$$\tilde{b} = \begin{bmatrix} 5 \\ 15 \\ 41 \\ 45 \end{bmatrix}.$$

Do NOT perform the Gaussian elimination from scratch. You already have \mathbf{U} and \mathbf{L} .

Answer. So, $U\tilde{x} = L^{-1}\tilde{b} = \tilde{y}$, and we can find \tilde{y} with

$$\begin{aligned}\tilde{y} = L^{-1}\tilde{b} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 2 & -3 & 1 & 0 \\ 3 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 15 \\ 41 \\ 45 \end{bmatrix} \\ &= \begin{bmatrix} 5 \\ -10 + 15 \\ 10 - 45 + 41 \\ 15 - 15 - 41 + 45 \end{bmatrix} \\ &= \begin{bmatrix} 5 \\ 5 \\ 6 \\ 4 \end{bmatrix}.\end{aligned}$$

Then the transformed equation is

$$U\tilde{x} = \tilde{y} \implies \begin{bmatrix} 2 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \\ \tilde{x}_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 6 \\ 4 \end{bmatrix}.$$

Starting at the bottom, clearly $\tilde{x}_4 = 2$. Then we have $2\tilde{x}_3 + 2\tilde{x}_4 = 6 \implies 2\tilde{x}_3 + 4 = 6 \implies \tilde{x}_3 = 1$. Then we have $\tilde{x}_2 + \tilde{x}_3 + \tilde{x}_4 = 5 \implies \tilde{x}_2 + 1 + 2 = 5 \implies \tilde{x}_2 = 2$. Finally, we have $2\tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 = 5 \implies 2\tilde{x}_1 + 2 + 1 = 5 \implies \tilde{x}_1 = 1$. So,

$$\tilde{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}.$$

4. Let \mathbf{A} have the following economy (thin) SVD:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0.6 \\ 0 & -0.8 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0.6 & -0.8 \\ -0.8 & -0.6 \end{bmatrix}^\top.$$

Answer the following questions without forming \mathbf{A} and without using Matlab. None of these questions require using numerical software. At most, you need to be able to perform simple matrix arithmetic using paper-and-pencil.

(a) Find $\text{rank}(\mathbf{A})$, $\|\mathbf{A}\|_2$ and $\|\mathbf{A}\|_F$.

Answer. We have

$$\begin{aligned}\text{rank}(\mathbf{A}) &= 2, \\ \|\mathbf{A}\|_2 &= \max_i \sqrt{\lambda_i(\mathbf{A}^\top \mathbf{A})} \\ &= \sqrt{\sigma_1^2(\mathbf{A})} \\ &= |\sigma_1(\mathbf{A})| = 4, \\ \|\mathbf{A}\|_F &= \sqrt{\text{tr}((\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top)^\top (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top))} \\ &= \sqrt{\text{tr}(\mathbf{V}^\top \mathbf{V} \boldsymbol{\Sigma}^\top \mathbf{U}^\top \mathbf{U} \boldsymbol{\Sigma})} \\ &= \sqrt{\text{tr}^2(\boldsymbol{\Sigma})} \\ &= |\text{tr}(\boldsymbol{\Sigma})| \\ &= 4 + 3 = 7.\end{aligned}$$

(b) Find an orthonormal basis for $\text{Range}(\mathbf{A})$, $\text{Null}(\mathbf{A})$, $\text{Range}(\mathbf{A}^\top)$, and $\text{Null}(\mathbf{A}^\top)$.

Answer. We have

$$\begin{aligned}\text{Range}(\mathbf{A}) &= \text{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0.6 \\ -0.8 \end{bmatrix} \right\}, \\ \text{Null}(\mathbf{A}) &= \text{span} \emptyset, \\ \text{Range}(\mathbf{A}^\top) &= \text{span} \left\{ \begin{bmatrix} 0.6 \\ -0.8 \end{bmatrix}, \begin{bmatrix} -0.8 \\ -0.6 \end{bmatrix} \right\}, \\ \text{Null}(\mathbf{A}^\top) &= \mathbb{R}^3 \setminus \text{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0.6 \\ -0.8 \end{bmatrix} \right\} \\ &= \text{span} \left\{ \begin{bmatrix} 0 \\ 0.8 \\ 0.6 \end{bmatrix} \right\}.\end{aligned}$$

Note that the given sets consist of orthogonal unit vectors, so they are orthonormal bases for the sets which they span.

- (c) Form the optimal rank-1 approximation \mathbf{A}_1 to \mathbf{A} in the 2-norm? What is the error matrix $\mathbf{A} - \mathbf{A}_1$, and what is the norm of the error $\|\mathbf{A} - \mathbf{A}_1\|_2$? The error matrix and the norm of the error follow directly from the SVD; no computation is needed.

Answer. We have

$$\begin{aligned}\mathbf{A}_1 &= \begin{bmatrix} 1 & 0 \\ 0 & 0.6 \\ 0 & -0.8 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.6 & -0.8 \\ -0.8 & -0.6 \end{bmatrix}^\top \\ &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} [4] \begin{bmatrix} 0.6 \\ -0.8 \end{bmatrix}^\top, \\ \mathbf{A} - \mathbf{A}_1 &= \begin{bmatrix} 1 & 0 \\ 0 & 0.6 \\ 0 & -0.8 \end{bmatrix} \left(\begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} 0.6 & -0.8 \\ -0.8 & -0.6 \end{bmatrix}^\top \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 0.6 \\ 0 & -0.8 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0.6 & -0.8 \\ -0.8 & -0.6 \end{bmatrix}^\top, \\ \|\mathbf{A} - \mathbf{A}_1\|_2 &= 3.\end{aligned}$$

5. (From Golub and van Loan) Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, let σ_1 denote the largest singular value of \mathbf{A} . Prove that

$$\sigma_1(\mathbf{A}) = \max_{\substack{\vec{y} \in \mathbb{R}^m, \\ \vec{x} \in \mathbb{R}^n}} \frac{\vec{y}^\top \mathbf{A} \vec{x}}{\|\vec{y}\|_2 \|\vec{x}\|_2}.$$

Proof. For $\vec{y}, \vec{x} \neq 0$ we have

$$\begin{aligned}\max_{\substack{\vec{y} \in \mathbb{R}^m, \\ \vec{x} \in \mathbb{R}^n}} \frac{\vec{y}^\top \mathbf{A} \vec{x}}{\|\vec{y}\|_2 \|\vec{x}\|_2} &= \max_{\substack{\vec{y} \in \mathbb{R}^m, \\ \vec{x} \in \mathbb{R}^n}} \left(\frac{\vec{y}}{\|\vec{y}\|_2} \right)^\top \frac{\mathbf{A} \vec{x}}{\|\vec{x}\|_2} \\ &= \max_{\substack{\vec{y} \in \mathbb{R}^n, \\ \|\vec{y}\|=1, \\ \vec{x} \in \mathbb{R}^n}} \vec{y}^\top \frac{\mathbf{A} \vec{x}}{\|\vec{x}\|_2}\end{aligned} \tag{5}$$

which is the inner product, or the projection of a unit vector \vec{y} onto $\frac{\mathbf{A} \vec{x}}{\|\vec{x}\|_2}$. The maximum of this quantity is achieved

when \vec{y} is the unit vector in the direction of $\frac{\mathbf{A}\vec{x}}{\|\vec{x}\|_2}$. In other words,

$$\begin{aligned}\vec{y} &= \frac{\frac{\mathbf{A}\vec{x}}{\|\vec{x}\|_2}}{\left\| \frac{\mathbf{A}\vec{x}}{\|\vec{x}\|_2} \right\|_2} \\ &= \frac{\mathbf{A}\vec{x}}{\|\mathbf{A}\vec{x}\|_2}.\end{aligned}$$

Then in eq. (5) we have

$$\begin{aligned}\max_{\substack{\vec{y} \in \mathbb{R}^m, \\ \vec{x} \in \mathbb{R}^n}} \frac{\vec{y}^\top \mathbf{A}\vec{x}}{\|\vec{y}\|_2 \|\vec{x}\|_2} &= \max_{\substack{\vec{y} \in \mathbb{R}^n, \\ \|\vec{y}\|=1, \\ \vec{x} \in \mathbb{R}^n}} \vec{y}^\top \frac{\mathbf{A}\vec{x}}{\|\vec{x}\|_2} \\ &= \max_{\vec{x} \in \mathbb{R}^n} \left(\frac{\mathbf{A}\vec{x}}{\|\mathbf{A}\vec{x}\|_2} \right)^\top \frac{\mathbf{A}\vec{x}}{\|\vec{x}\|_2} \\ &= \max_{\vec{x} \in \mathbb{R}^n} \frac{\|\mathbf{A}\vec{x}\|_2^2}{\|\mathbf{A}\vec{x}\|_2 \|\vec{x}\|_2} \\ &= \max_{\vec{x} \in \mathbb{R}^n} \frac{\|\mathbf{A}\vec{x}\|_2}{\|\vec{x}\|_2} \\ &= \|\mathbf{A}\|_2 \\ &= \sigma_1(\mathbf{A}).\end{aligned}$$



6. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ have the singular value decomposition

$$\mathbf{A} = \sum_{j=1}^r \sigma_j \vec{u}_j \vec{v}_j^\top.$$

We showed in class that the matrix

$$\mathbf{A}_k = \sum_{j=1}^k \sigma_j \vec{u}_j \vec{v}_j^\top$$

is an optimal rank- k approximation to \mathbf{A} in the 2-norm and $\|\mathbf{A} - \mathbf{A}_k\| = \sigma_{k+1}$. But this minimizer (optimal approximant) is not unique. In other words, using truncated SVD we can find another rank- k matrix $\tilde{\mathbf{A}}_k \in \mathbb{R}^{m \times n}$ such that $\|\mathbf{A} - \tilde{\mathbf{A}}_k\| = \sigma_{k+1}$. In this problem, you will prove this fact.

Define

$$\tilde{\mathbf{A}}_k = \sum_{j=1}^k (\sigma_j - \eta_j) \vec{u}_j \vec{v}_j^\top \quad \text{where} \quad 0 \leq |\eta_j| < \sigma_{k+1}. \quad (6)$$


Show that $\tilde{\mathbf{A}}_k$ in (6) has rank- k and minimizes the error, i.e., $\|\mathbf{A} - \tilde{\mathbf{A}}_k\| = \sigma_{k+1}$.

Proof. Observe that $\sum_{j=1}^k (\sigma_j - \eta_j) \vec{u}_j \vec{v}_j^\top$ is the SVD of $\tilde{\mathbf{A}}_k$, with singular values of $(\sigma_j - \eta_j)$. Since $|\eta_j| < \sigma_{k+1} \leq \sigma_j$ for all $1 \leq j \leq k$, then $\sigma_j - \eta_j > 0$, and $\tilde{\mathbf{A}}_k$ has k non-zero singular values, so it has rank k . As well, we have

$$\begin{aligned}\mathbf{A} - \tilde{\mathbf{A}}_k &= \sum_{j=1}^r \sigma_j \vec{u}_j \vec{v}_j^\top - \sum_{j=1}^k (\sigma_j - \eta_j) \vec{u}_j \vec{v}_j^\top \\ &= \sum_{j=1}^k [\sigma_j \vec{u}_j \vec{v}_j^\top - (\sigma_j - \eta_j) \vec{u}_j \vec{v}_j^\top] + \sum_{j=k+1}^r \sigma_j \vec{u}_j \vec{v}_j^\top \\ &= \sum_{j=1}^k \eta_j \vec{u}_j \vec{v}_j^\top + \sum_{j=k+1}^r \sigma_j \vec{u}_j \vec{v}_j^\top.\end{aligned}$$

Note, then, that this is the SVD of $\mathbf{A} - \tilde{\mathbf{A}}_k$, with eigenvalues of $|\eta_j|$ for $1 \leq j \leq k$ and σ_j for $k+1 \leq j \leq r$. Since $|\eta_j| < \sigma_{k+1}$ for all j and $\sigma_{k+1} \geq \sigma_k$ for all j , we must have

$$\|\mathbf{A} - \tilde{\mathbf{A}}_k\|_2 = \sigma_{k+1},$$

as required. 

7. Approximation of a Fingerprint Image via SVD: You will complete the Matlab Script `Homework3Problem7.m` to answer this question. You will attach the completed script to the .pdf file.

Before start completing your code, it will help use the commands `help svd`, `help subplot`, `help semilogy` to understand how to use these functions.

Lines 7-10 load the image into Matlab and plot it. The resulting matrix \mathbf{A} in your workspace is a 1133×784 matrix with entries consisting of only ones and zeroes; ones correspond to the white spots in the image and zeroes to the black spots.

- (a) Compute the short (reduced) SVD of \mathbf{A} using Matlab. Then plot the normalized singular values under `subplot(2,1,2)`. This figure might be hard to read due to the scale of the y -axis. Then plot only the leading 700 normalized singular values under `subplot(2,1,2)`. What do you observe? What would be your decision for the (numerical) rank of \mathbf{A} based on these comments? Justify your reasoning. Does the result of the built-in rank command coincide with your decision?

Answer. Plots of the singular values of \mathbf{A} can be found in fig. 1. We notice a significant drop off in singular values after the 654th singular value. Because of this, the numerical rank of \mathbf{A} should be 654. Indeed, this is the computed rank of \mathbf{A} .

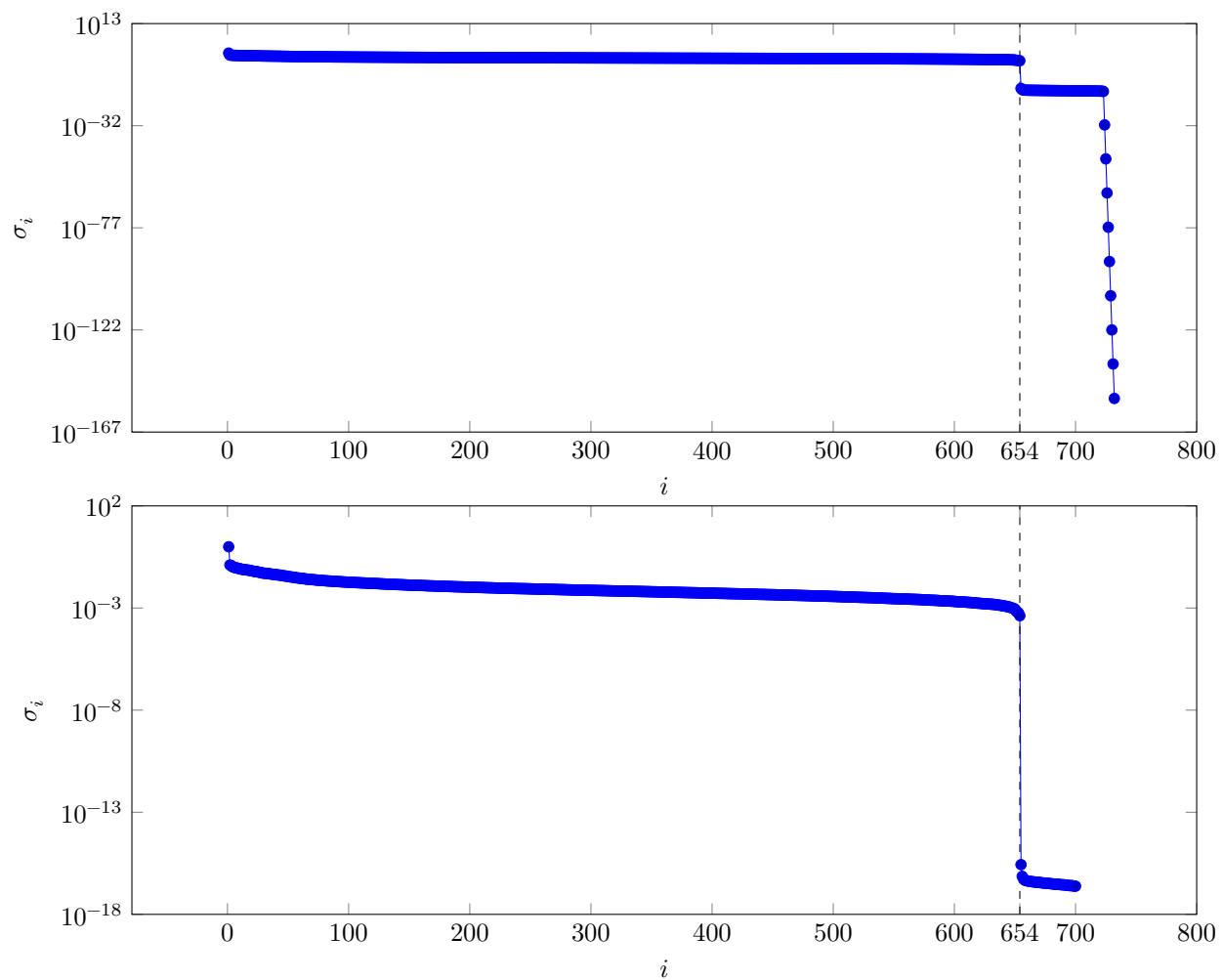


Figure 1: Plots showing all singular values of \mathbf{A} and just the first 700 singular values of \mathbf{A} . Note the large gap between values that occurs at the calculated rank of \mathbf{A} (654).

- (b) Construct the optimal rank- k approximation \mathbf{A}_k to \mathbf{A} in the 2-norm for $k = 1, k = 10$, and $k = 50$. In each case, compute the relative error $\frac{\|\mathbf{A} - \mathbf{A}_k\|_2}{\|\mathbf{A}\|_2}$. Note that you do NOT need to form $\mathbf{A} - \mathbf{A}_k$ to compute these error values; the singular values are all you need. Use `subplot(2,2,1)` to plot the original image in the top left corner. Then, use the `imshow` command on the three low-rank approximations and plot them in the `subplot(2,2,2)`, `subplot(2,2,3)`, and `subplot(2,2,4)` spots. Put appropriate titles on each plot, e.g., “original image”, “rank-1 approximation” etc., using the `title` command. Plots without appropriate labels and explanations will lose points. These plots need to be attached to the .pdf file.

Answer. The plots of the rank- k approximations, for $k = 1, 10, 50$, can be found in fig. 2, with program output displayed below:

```
Computed Rank of A = 654
Relative error for A_1 = 0.12761236428030379
Relative error for A_10 = 0.08040029616965558
Relative error for A_50 = 0.03400080504768164
```

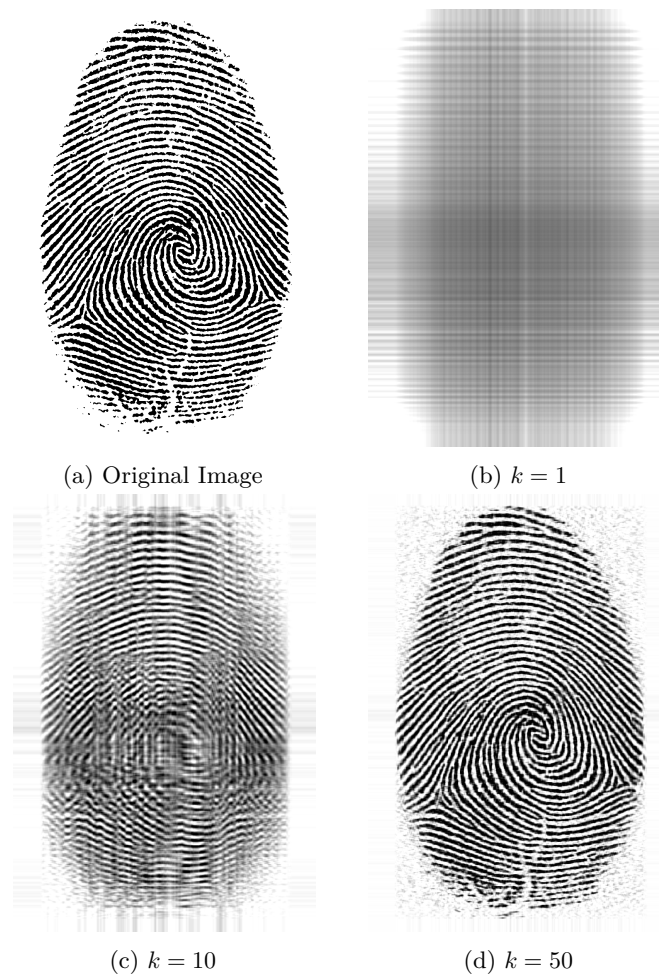


Figure 2: Original fingerprint image compared to its optimal rank- k approximations.

The code is given below:

```

1  #![feature(float_next_up_down)]
2  use image::{DynamicImage::ImageLuma8, GrayImage};
3  use nalgebra::{
4      allocator::Allocator, ComplexField, DMatrix, DefaultAllocator, Dim, DimMin, DimMinimum, Matrix,
5      RawStorage, Storage, SVD,
6  };
7  use num_traits::AsPrimitive;
8  use show_image::create_window;
9  use std::{
10     error::Error,
11     fs::{self, File},
12     io::Write,
13 };
14
15 #[show_image::main]
16 fn main() -> Result<(), Box<dyn Error>> {
17     // Open original .pbm image
18     let img = image::open("HW4_Prob6_fingerprint.pbm")?;
19     // Display image
20     let window = create_window("Original Image", Default::default())?;
21     window.set_image("Figure 1", img.clone())?;
22
23     // Assert image is 8-bit monochrome
24     let ImageLuma8(img) = img else { unreachable!() };
25     // Convert image to matrix of doubles
26     let a = DMatrix::from_row_iterator(

```

```

27     img.height() as usize,
28     img.width() as usize,
29     img.as_raw().iter().map(|x| *x as f64),
30 );
31
32 // Save size of A for later
33 let (m, n) = a.shape();
34
35 // check matrix is still the same image
36 mat_to_img_show(&a, "Original matrix as image check"?);
37
38 // Compute SVD of A
39 let svd = SVD::new(a, true, true);
40
41 // Create output file
42 fs::create_dir_all("./out")?;
43 let mut out = File::create("./out/all_singular_values7.dat")?;
44 // Print normalized singular values to output file
45 writeln!(out, "#all singular values of A"?);
46 for s in &svd.singular_values {
47     writeln!(out, "{}", s / svd.singular_values[0])?;
48 }
49 // notify Terminal of completed SVD printing
50 eprintln!("finished printing");
51
52 // The rank() function in matlab uses a default tolerance value. The nalgebra version doesn't - it needs
53 //   ↳ to be provided. So we replicate matlab's default tolerance.
54 // tol from MatLab = max(size(A))*eps(norm(A))
55 // From matlab's documentation, eps(x) is the positive distance between |x| and the next largest float.
56 // Rust's f64::next_up(x) returns that next largest float. As well, we are using the already computed
57 //   ↳ largest singular value as norm(A).
58 let tol = (m.max(n) as f64) * (f64::next_up(svd.singular_values[0]) - svd.singular_values[0]);
59
60 // Compute the Rank of A with the computed tolerance
61 println!("Computed Rank of A = {}", svd.rank(tol));
62
63 // Compute optimal rank-k approximations, display them, and save them
64 let mut approx = DMatrix::zeros(m, n);
65 let mut prev_k = 0;
66 for k in [1, 10, 50] {
67     approx = rank_k_approx(&svd, k, &approx, prev_k);
68     prev_k = k;
69     mat_to_img_show(&approx, format!("Rank_{k}_Approximation")?);
70     // Print relative errors
71     println!(
72         "Relative error for A_{k} = {}",
73         // Rust is 0 indexed, need to subtract 1
74         svd.singular_values[(k + 1) - 1] / svd.singular_values[1 - 1]
75     );
76 }
77
78 // keep images up until original window is closed
79 for _event in window.event_channel()? {}
80 Ok(())
81 }
82
83 /// Calculates the optimal rank `k` approximation of a matrix represented by its singular value decomposition.
84 /// Uses previously-computed optimal approximation `prev`, which is rank `prev_k`, which must be less than
85 ///   ↳ `k`.
86 /// For new approximation, pass zero matrix for `prev` and 0 for `prev_k`
87 fn rank_k_approx<T: ComplexField, R: DimMin<C>, C: Dim>(
88     svd: &SVD<T, R, C>,
89     k: usize,
90     prev: &Matrix<T, R, C, impl Storage<T, R, C>>,
91     prev_k: usize,
92 ) -> Matrix<T, R, C, <nalgebra::DefaultAllocator as nalgebra::allocator::Allocator<T, R, C>>::Buffer>
93 where
94     DefaultAllocator: Allocator<T, DimMinimum<R, C>, C>
95         + Allocator<T, R, DimMinimum<R, C>>
96         + Allocator<T::RealField, DimMinimum<R, C>>

```

```

94         + Allocator<T, R, C>,
95     {
96         let mut u = svd.u.as_ref().unwrap().clone();
97         // Multiply  $\sigma_i u_i$  for  $i \in [1, k]$ 
98         for i in prev_k..k {
99             let val = svd.singular_values[i].clone();
100             u.column_mut(i).scale_mut(val);
101         }
102         // Multiply  $\sum_{i=1}^k (\sigma_i u_i) v_i^\top$ 
103         prev + u.columns(prev_k, k - prev_k) * svd.v_t.as_ref().unwrap().rows(prev_k, k - prev_k)
104     }
105
106     /// Displays and saves the image stored in mat to the file "./out/<`wind_name`>.png"
107     fn mat_to_img_show<T: AsPrimitive<u8>, R: Dim, C: Dim, S: RawStorage<T, R, C>>(<
108         mat: &Matrix<T, R, C, S>,
109         wind_name: impl AsRef<str>,
110     ) -> Result<(), Box<dyn Error>> {
111         // Convert matrix to 8-bit monochrome image
112         // Annoyingly, image crate and matrix crate use different size types, so converting is required
113         let im2 = GrayImage::from_fn(mat.ncols() as u32, mat.nrows() as u32, |c, r| {
114             image::Luma([mat[(r as usize, c as usize)].as_()])
115         });
116
117         // Create window and display image
118         let window2 = create_window(wind_name.as_ref(), Default::default())?;
119         window2.set_image("f", im2.clone())?;
120
121         // Save image as png in output folder
122         im2.save_with_format(
123             format!("./out/{f}.png", wind_name.as_ref()),
124             image::ImageFormat::Png,
125         )?;
126
127         Ok(())
128     }

```

8. In the previous problem you used SVD to compress a black-and-white image. In this example, you will use SVD to compute a low-rank approximation to the color image hokiebirdwithacat.jpg. You will complete the Matlab Script Homework3Problem8.m to answer this question. You will attach the completed script to the .pdf file.

Lines 7-20 load the image into Matlab, plot the original image, and extract the images correspond to every color layer, and convert these three layer images to double precision matrices A1, A2, and A3.

- (a) Compute the SVDs of every layer A1, A2, and A3. Then compute the vector of normalized singular values for every layer. Using the subplot comment (and logarithmic y-axis), plot all three vectors in Figure 2.

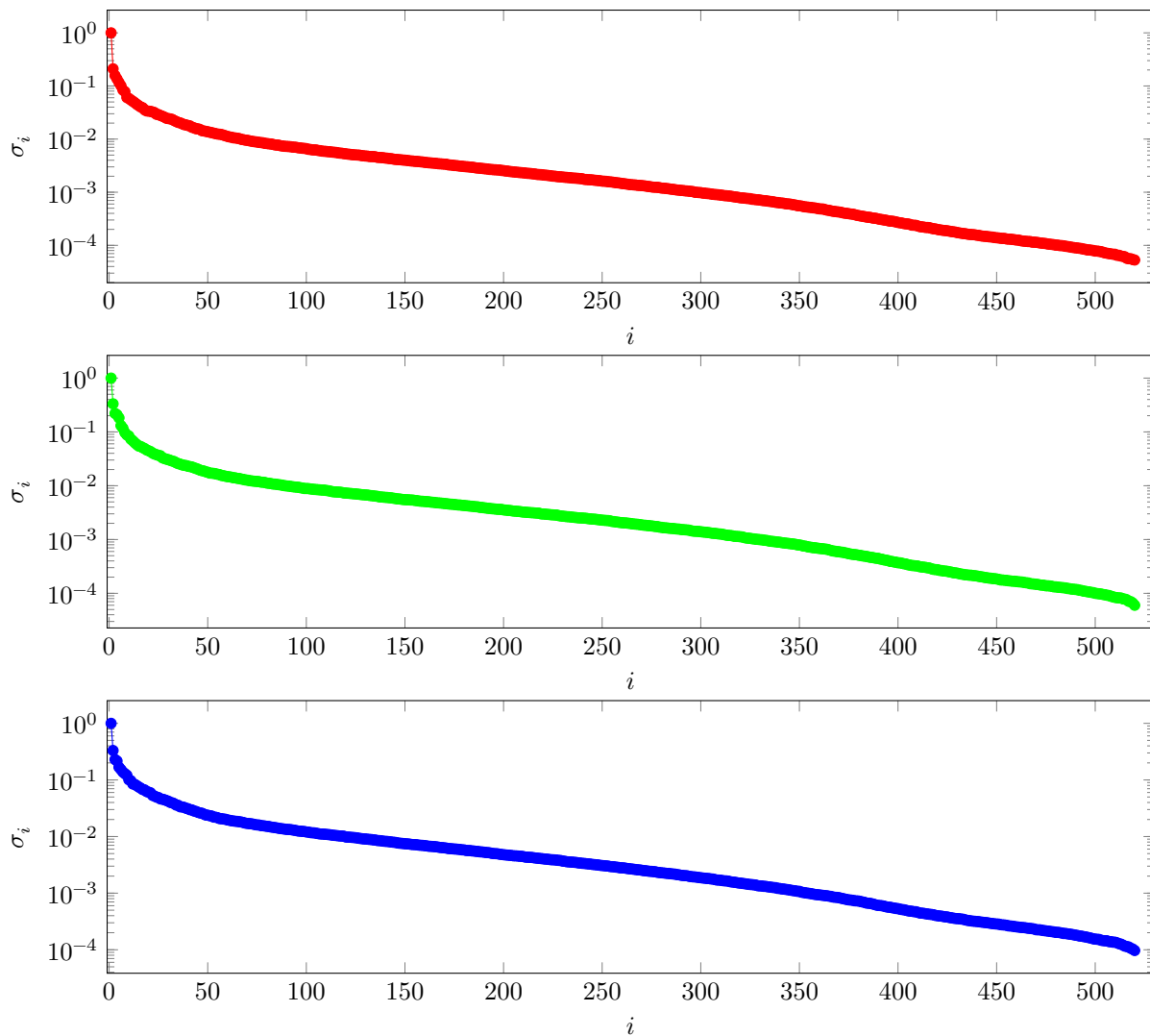


Figure 3: Plots showing all normalized singular values of hokiebirdwithacat, split into red, green, and blue components.

Answer.

- (b) Use the same error tolerance for every layer: Find the smallest rank- k for each layer such that the relative error in each layer is less than

- (i) 10% (ii) 5% (iii) 1%

So, you will have three approximations. For each case, plot the low-rank image (either as a new figure or all in one plot using subplot). These plots need to be attached to the .pdf file. Make sure to clearly indicate the rank of every layer and which plot corresponds to which error tolerance. Did you obtain the same k value for every layer? Comment on your results.

Answer.

- (c) Use different error tolerances for different layers: Now we will choose different tolerances for different values. Pick three different selections:

- 50% error for Layer 1, 1% error for Layer 2, and 1% error for Layer 3,
- 1% error for Layer 1, 50% error for Layer 2, and 1% error for Layer 3,
- 1% error for Layer 1, 1% error for Layer 2, and 50% error for Layer 3,

So, you will have three approximations. For each case, plot the low-rank image (either as a new figure or all in one plot using subplot). These plots need to be attached to the .pdf file. Make sure to clearly indicate the rank of every layer and which plot corresponds to which error tolerance. Comment on your results.

Answer.