Homework #8
CS 575: Numerical Linear Algebra
Spring 2023

John Tran

April 20, 2023

Important Notes

- `Python` 3.11 was used to run the notebook (i.e., to use the `match` statement)
- the partially completed notebook was used to complete the coding assignment, so many of the things asked (e.g., verification and testing) was done for us and they were modified for the mat-mat portion
- the `README` was done in `Markdown` but the raw text can still be viewed

Problem 1

If we let $A \in \mathbb{R}^{m \times m}$ have eigenvalues $\lambda_j, j = 1, 2, \ldots, m$ where $\mu \neq \lambda_j$, then we are asked to show the eigenvalues of $(A - \mu\lambda)^{-1}$ are $(\lambda_j - \mu)^{-1}$.

First, we can show that some general matrix $B \in \mathbb{R}^{m \times m}$ with eigenvalues $\sigma_j, j = 1, 2, \ldots, m$ has its inverse B^{-1} with eigenvalues $\frac{1}{\sigma_j}$ (i.e., eigenvalues are also simply inverted). We can also let $v_j, j = 1, 2, \ldots, m$ be the corresponding eigenvectors for the eigenvalue σ_j.

$$
\begin{aligned}
Bv_j &= \sigma_j v_j \\
B^{-1}Bv_j &= B^{-1}\sigma_j v_j \\
v_j &= \sigma_j(B^{-1}v_j) \\
\frac{1}{\sigma_j}v_j &= B^{-1}v_j \\
B^{-1}v_j &= \frac{1}{\sigma_j}v_j \quad \text{flip left and right sides}
\end{aligned}
$$

This will be useful later but for now, we can find the eigenvalues of $A - \mu I$ (we can use the same eigenvector notation for A as previously defined by B for simplicity, and we know for the Inverse power method, $A - \mu I$ has the same eigenvectors as A)

$$
\begin{aligned}
(A - \mu I)v_j &= Av_j - \mu I v_j \\
&= \lambda_j v_j - \mu v_j \\
&= (\lambda_j - \mu)v_j
\end{aligned}
$$

So, $A - \mu I$ has eigenvalues $\lambda_j - \mu$.

If we set $B = A - \mu I$, then the eigenvalues of $B^{-1} = (A - \mu I)^{-1}$ will be inverted using our result from above for $(\lambda_j - \mu)^{-1}$ and we are done.

Problem 2

We can use the Inverse power method to find smallest absolute magnitude eigenvalue (λ_{min}) of A by setting μ to be approximately equal to λ_{min} (not exactly equal since that make the denominator of $(\lambda_{min} - \mu)^{-1}$ be 0). Since both the Power and Inverse power method will find the largest absolute magnitude eigenvalue, we want the denominator of $(\lambda_{min} - \mu)^{-1}$ to be small as possible for λ_{min} so $(\lambda_{min} - \mu)^{-1}$ would be maximized. It does not matter if λ_{min} is positive or negative (as long as $\lambda_{min} \neq 0$) since only the magnitude is considered for the Inverse power method. To recover the sign of the smallest eigenvalue, the corresponded computed eigenvector could be used in Av_{min} to check for the sign (since the eigenvectors in A are the same as $(A - \mu I)^{-1}$).

Problem 3

Note: there was a typo in the given QR iteration with shifts (inferred correction given below)

$$\begin{aligned} A^{(k)} - \mu^{(k)} I &= Q^{(k)} R^{(k)} \\ A^{(k+1)} &= R^{(k)} Q^{(k)} + \mu^{(k)} I \end{aligned}$$

We are asked to show that the matrices $A^{(k)}$ generated in the QR iteration with shifts are similar to each other.

For this problem, we can prove the inductive case by showing that $A^{(k)}$ is similar to $A^{(k+1)}$ (for all k, which we use the general case). For similar matrices, we have the following relationship

$$A^{(k+1)} = S^{-1} A^{(k)} S$$

for some matrix S. The reversed relationship also holds (i.e., $A^{(k)}$ and $A^{(k+1)}$ are swapped above) where the resulting S we get above would consequently be swapped with its inverse S^{-1} in the reversed relationship, but the similarity would still hold.

By inspecting the forms of $A^{(k)}$ and $A^{(k+1)}$, we can see that the only major difference (once $\mu^{(k)}I$ is moved to the right hand side for $A^{(k)}$ above) is the switched ordering from $Q^{(k)}R^{(k)}$ to $R^{(k)}Q^{(k)}$. So, we can guess that $S^{-1} = R^{(k)}$ (or $S = \left(R^{(k)}\right)^{-1}$) to make the right $R^{(k)}$ vanish:

$$\begin{aligned}
A^{(k+1)} &= S^{-1}A^{(k)}S \\
&= R^{(k)}\left(Q^{(k)}R^{(k)} + \mu^{(k)}I\right)\left(R^{(k)}\right)^{-1} \\
&= R^{(k)}Q^{(k)}R^{(k)}\left(R^{(k)}\right)^{-1} + R^{(k)}\mu^{(k)}I\left(R^{(k)}\right)^{-1} \\
&= R^{(k)}Q^{(k)} + \mu^{(k)}R^{(k)}I\left(R^{(k)}\right)^{-1} \\
&= R^{(k)}Q^{(k)} + \mu^{(k)}R^{(k)}\left(R^{(k)}\right)^{-1} \\
&= R^{(k)}Q^{(k)} + \mu^{(k)}I
\end{aligned}$$

This is the exact form given for $A^{(k+1)}$ and we are done.

Problem 4

We are given that $A \in \mathbb{R}^{m\times n}$, $m \geq n$ and $A = U\Sigma V^T$ be the SVD of A. We are asked to show that AA^T has eigenvalues (another typo where the homework said singular values) that are just squares of those from the singular values of A and U are the eigenvectors of AA^T associated with those eigenvalues.

First, we can notice the SVD of A^T can be found by taking the transpose of the SVD of A

$$\begin{aligned}
A^T &= \left(U\Sigma V^T\right)^T \\
&= V\Sigma^T U^T
\end{aligned}$$

where the placement of U, V are reversed.

So,

$$\begin{aligned}
AA^T &= \left(U\Sigma V^T\right)\left(V\Sigma U^T\right) \\
&= U\Sigma\Sigma^T U^T
\end{aligned}$$

Since $\Sigma\Sigma^T \in \mathbb{R}^{m\times m}$, this is just another diagonal matrix each with diagonal entries σ_i for $i = 1, 2, \ldots, n$, then $\Sigma\Sigma^T$ has diagonal entries σ_i^2 with an extra $(m-n)$ 0s for the remaining diagonal entries.

Looking at the form we got for AA^T closely, we can see it is a symmetric matrix with our result resembling the spectral decomposition where we can let the corresponding diagonal matrix be $\Sigma\Sigma^T$, representing the diagonal matrix with eigenvalues along the diagonals and the columns of U be the corresponding eigenvectors (by definition of the spectral decomposition for symmetric matrices).

Problem 5

Note: this problem is fairly lengthy so the discussion and proofs will be shortened for brevity and any repeated logic will be mentioned. In addition, the solution is provided at the back of the textbook so the work given below will be adapted from there.

(a) First, we can make use of the fact that the singular values σ_i for $i = 1, 2, \ldots, m$ of A are the square roots of the eigenvalues of A^TA, which results in non-negative values (we have shown in class that A^TA where A is a square matrix is SPD). Now, to show that σ_{min} is strictly positive, we can use the given statement that A is invertible. We can do a simple proof by contradiction and assume A^TA is singular. This would imply $A^TAx = 0$ for some $x \in \mathbb{R}^m$, $x \neq 0$. So, multiplying both sides by x^T, $x^TA^TAx = 0, (Ax)^T(Ax) = 0, \|Ax\|_2 = 0, Ax = 0$, resulting in A being singular since $x \neq 0$. However, if A is invertible, then A^TA must also be invertible and that means none of the eigenvalues can be 0, including σ_{min}.

(b) We can start with $\|A\|_2 = \sigma_{max}$:

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\| \quad x \in \mathbb{R}^m$$
$$x = \sum_{i=1}^{m} c_i v_i$$

where c_i are some constants and v_i correspond to the eigenvectors of A^TA for eigenvalues σ_i^2. Since A^TA is symmetric, the eigenvectors are orthonormal and form a basis in $\mathbb{R}^m$, and that is why we can rewrite x in this basis.

In addition, since $\|x\|_2 = 1$, we have the constraint $\sum_{i=1}^{m} c_i^2 = 1$ making use of the fact of orthonormal eigenvectors mentioned above.

So,

$$\begin{aligned}
\|A\|_2 &= \max_{\|x\|_2=1} \|Ax\| \\
&= \max_{\|x\|_2=1} \sqrt{(Ax)^T Ax} \\
&= \max_{\|x\|_2=1} \sqrt{x^T (A^T Ax)} \\
&= \max_{\|x\|_2=1} \sqrt{\left(\sum_{i=1}^{m} c_i v_i\right)\left(\sum_{j=1}^{m} c_j \sigma_j^2 v_j\right)}
\end{aligned}$$

where the last line has two different indexing variables to prevent confusion and σ_j^2 comes from the eigenvalue equation of $A^T A$.

$$\begin{aligned}
\|A\|_2 &= \max_{\|x\|_2=1} \sqrt{\left(\sum_{i=1}^{m} c_i v_i\right)\left(\sum_{j=1}^{m} c_j \sigma_j^2 v_j\right)} \\
&= \max_{\|x\|_2=1} \sqrt{\sum_{i=1}^{m} c_i^2 \sigma_i^2}
\end{aligned}$$

where the last line uses the property of the orthonormal vectors v_i where $v_i^T v_j = 0$ if $i \neq j$ and $v_i^T v_i = 1, \forall i \in \{1, 2, \ldots, m\}$

So, the max condition can be easily satisfied if only σ_{max} is used where $\sigma_1 = \sigma_{max}$ and $c_1 = 1$ to satisfy our constraint above for $\sum_{i=1}^{m} c_i^2 = 1$. This would result in $\|A\|_2 = \sqrt{\sigma_{max}^2} = \sigma_{max}$

For $\|A^{-1}\|_2$, we can manipulate the expression of the induced 2-norm first:

$$\left\|A^{-1}\right\|_2 = \max_{\|x\|_2 \neq 0} \frac{\|A^{-1}x\|_2}{\|x\|_2}$$

If we let $y = A^{-1}x$,

$$\begin{aligned}\left\|A^{-1}\right\|_2 &= \max_{\|y\|_2 \neq 0} \frac{\|y\|_2}{\|Ay\|_2} \\ &= \frac{1}{\min_{\|y\|_2 \neq 0} \frac{\|Ay\|_2}{\|y\|_2}}\end{aligned}$$

where we flipped the max condition to a min condition by taking the reciprocal. If we look at the denominator, we can rewrite the 2-norm using the modified constraint where $\|z\|_2 = 1$ where z is the unit vector in the direction of y:

$$\min_{\|y\|_2 \neq 0} \frac{\|Ay\|_2}{\|y\|_2} = \min_{\|z\|_2 = 1} \|Az\|_2$$

From here, we can follow the exact same steps as for $\|A\|_2$ except we have a ***min*** condition in the denominator this time to optimize where

$$\left\|A^{-1}\right\|_2 = \frac{1}{\min_{\|z\|_2 = 1} \sqrt{\sum_{i=1}^{m} c_i^2 \sigma_i^2}}$$

As before, we can use a similar argument where this time, we can just use σ_{min} where $\sigma_m = \sigma_{min}$ and $c_m = 1$ to satisfy our constraint above for $\sum_{i=1}^{m} c_i^2 = 1$. This would result in $\|A^{-1}\|_2 = \frac{1}{\sqrt{\sigma_{min}^2}} = \frac{1}{\sigma_{min}}$

(c) We are asked to show $\kappa_2\left(A^T A\right) = \frac{\sigma_{max}^2}{\sigma_{min}^2}$ where $\kappa_2\left(A^T A\right) = \left\|\left(A^T A\right)^{-1}\right\|_2 \left\|A^T A\right\|_2$.

However, to make things simpler, we can notice that the only difference here is we can substitute $A^T A$ for A in part (b) and the modified calculations are rather simple to go over (looking at substituting A^TA for A in $\|A\|_2$ first):

$$\begin{aligned}
\left\|A^TA\right\|_2 &= \max_{\|x\|_2=1}\left\|A^TAx\right\| \\
&= \max_{\|x\|_2=1}\sqrt{(A^TAx)^TA^TAx} \\
&= \max_{\|x\|_2=1}\sqrt{x^T(A^TA(A^TAx))}, \quad (A^TAx)^T = A^TA \\
&= \max_{\|x\|_2=1}\sqrt{\left(\sum_{i=1}^m c_iv_i\right)\left(\sum_{j=1}^m c_j(\sigma_j^2)^2v_j\right)} \quad \text{two applications of the eigenvalue equation} \\
&= \max_{\|x\|_2=1}\sqrt{\sum_{i=1}^m c_i^2\sigma_i^4}
\end{aligned}$$

From here, we can see the only difference is the doubled factor on σ_i from 2 to 4 under the square root and this would simply square our result from earlier for $\left\|A^TA\right\|_2$ since the same logic would hold for the max condition where $\left\|A^TA\right\|_2 = \sigma_{max}^2$.

The exact same logic in substituting A^TA for A in $\|A^{-1}\|_2$ would similarly hold as we would get

$$\left\|(A^TA)^{-1}\right\|_2 = \frac{1}{\min_{\|z\|_2=1}\sqrt{\sum_{i=1}^m c_i^2\sigma_i^4}}$$

and we would use the same value $1/\sigma_{min}$ (just squared) for the min condition where $\left\|(A^TA)^{-1}\right\|_2 = 1/\sigma_{min}^2$.

Putting everything together,

$$\begin{aligned}
\kappa_2\left(A^TA\right) &= \left\|\left(A^TA\right)^{-1}\right\|_2\left\|A^TA\right\|_2 \\
&= \left(1/\sigma_{min}^2\right)\sigma_{max}^2 \\
&= \sigma_{max}^2/\sigma_{min}^2
\end{aligned}$$

Problem 6

We are given the decomposition $A = CDE$ for A with the concrete values of C, D, E.

Given the size of A is 3×2 and the familiarity of the decomposition given to the SVD of A where $A = U\Sigma V^T$, we can rearrange (or rather, permute) the matrices so it would fit the format of an SVD.

(a) First, we notice that D is similar to the required format of Σ, being similar to a diagonal matrix. However, the diagonal entries need to start at the first diagonal entry (top left) and ordered from greatest to smallest from top left to bottom right. So, to accomplish this, we can introduce a set of permutation matrices P to swap the rows of D (first and third rows) and P' to swap the columns of D (first and second columns). Looking at the dimensions of D, we can see that P has size 3×3 and P' has size 2×2 So, the permutation matrices are given below:

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad P' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

To make sure A is preserved, we can add the pair PP^{-1} in between C and D (since putting a matrix to the left would swap rows) and the pair $P'(P')^{-1}$ (on the right would swap the columns) between D and E since they reduce to just the identity I.

One thing to note for our particular matrices P and P' given above is that any permutation matrix is an orthogonal matrix (since the columns are just the canonical vectors e_i), giving us the property that the inverse is just the transpose of the matrix $P^T = P^{-1}$, and from inspection from our specific choices of P and P', they are also symmetric where $P = P^T$. So, $P^{-1} = P^T = P$ and the same applies for P'.

So, $PP^{-1} = PP$ where one of the applications would be used by D (same with P' on the opposite side) and the other application would be used by C and E, respectively.

Putting everything together, we have

$$
\begin{aligned}
A &= CDE \\
&= CPPDP'P'E \\
&= (CP)(PDP')(P'E) \\
&= U\Sigma V^T
\end{aligned}
$$

where $U = CP$, $\Sigma = PDP'$, and $V^T = P'E$ since we derived this modified decomposition anchored on getting Σ into the correct format from D and any remaining changes to C and E would become U and V^T, respectively (since the sizes also match up and both C and E are orthogonal matrices).

So, plugging in everything, we have

$$
U = \begin{bmatrix} \frac{5}{3\sqrt{5}} & 0 & -\frac{2}{3} \\ \frac{2}{3\sqrt{5}} & \frac{2}{\sqrt{5}} & \frac{1}{3} \\ \frac{4}{3\sqrt{5}} & -\frac{1}{\sqrt{5}} & \frac{2}{3} \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \qquad V^T = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix}
$$

where for U, we swap the first and third columns of C, and for V^T, we swap the first and second rows of E. Also, it is interesting to note that V^T is symmetric so $V^T = V$.

(b) We are asked to find various bases for some given subspaces. The formulations used are adapted from the textbook and the answer with a brief description will be given for each. Also, the SVD of A^T would be helpful where again, we can just take the transpose of the SVD of A above

$$A^T = V\Sigma^T U^T$$

from a previous problem and V is symmetric from earlier discussion.

(i) Range of A

For finding the range, we can use the theorem to find r where $\sigma_i > 0$ for $i \leq r$. In this particular scenario, both singular values are nonzero so $r = 2$ (indexing starting at 1). So, the range is given as the columns of the ***left*** singular vectors (columns of U for A) for the set $\{u_1, u_2\}$ where u_1, u_2 are the first and second columns of U, respectively

$$Range(A) = \left\{ \begin{bmatrix} \frac{5}{3\sqrt{5}} \\ \frac{2}{3\sqrt{5}} \\ \frac{4}{3\sqrt{5}} \end{bmatrix}, \begin{bmatrix} 0 \\ \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{bmatrix} \right\}$$

(ii) Range of A^T

We can use the same theorem as above, using the new SVD for $A^T = V\Sigma^T U^T$, looking at V this time.

Since V is symmetric, the columns of V^T found above are the same columns as V and r is the same as A above for $r = 2$, looking at the set $\{v_1, v_2\}$ where v_1, v_2 are the first and second columns of V, respectively

$$Range(A^T) = \left\{ \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}, \begin{bmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{bmatrix} \right\}$$

(iii) Nullspace of A

This time, to find the nullspace, we would find the rightmost $n - r = 2 - 2 = 0$ $(n = 2)$ ***right*** singular vectors – from V^T for A. However, since the size of V^T is 2×2, there are no relevant vectors to form the nullspace so it is just the empty set.

So,

$$Nullspace(A) = \{\}$$

(iv) Nullspace of A^T

Again, we would find the rightmost right singular vectors, but for A^T, the two matrices U, V^T are flipped so U^T is now on the right side. So, the $m - r = 3 - 2 = 1$ ($m = 3$) right singular vectors of U (always the columns vectors of the non-transposed version) is just the last column of U.

So,

$$Nullspace(A^T) = \left\{ \begin{bmatrix} -\frac{2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{bmatrix} \right\}$$