Applied Math HW 3

Colin Williams

October 3, 2021

Question 1

Find the SVD (by hand calculation) and the pseudo-inverse of the following matrices.

$$A = \begin{bmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Answer.
Starting with matrix A. Let us first calculate AA^T and A^TA.

$$AA^T = \begin{bmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad A^TA = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$$

Since AA^T and A^TA are both diagonal and the eigenvalues of a diagonal matrix are simply its diagonal elements, we can see that the only singular value of A is $\sigma_1 = \sqrt{4} = 2$. We can now explicitly find our Σ in the SVD decomposition:

$$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Recall the eigenvectors of A^TA are the columns of V, therefore let us calculate those. As we saw above, $\lambda_1 = 4$ and $\lambda_2 = 0$ are the eigenvalues of A^TA (in this order since we require decreasing order). Thus,

$$\begin{aligned} (A^TA - \lambda_1 I)v_1 &= 0 \\ \implies (A^TA - 4I)v_1 &= 0 \\ \implies \begin{bmatrix} -4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} &= 0 \\ \implies -4x_1 &= 0 \\ \implies x_1 &= 0 \\ \implies v_1 &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned} \qquad \begin{aligned} (A^TA - \lambda_2 I)v_2 &= 0 \\ \implies A^TAv_2 &= 0 \\ \implies \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} &= 0 \\ \implies 4y_2 &= 0 \\ \implies y_2 &= 0 \\ \implies v_2 &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{aligned}$$

Thus, we can explicitly write V as

$$V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Since we only have one singular value, we only have a formula for the first column of U, namely

$$\begin{aligned} u_1 &= \frac{Av_1}{\sigma_1} \\ &= \frac{1}{2} \begin{bmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

However, it is clear that we can choose $u_2 = e_2$ and $u_3 = e_3$ to make U an orthogonal matrix as the theorem requires. Therefore, U simply the identity matrix I_3. Thus, we have our SVD given as

$$\begin{aligned} A &= U\Sigma V^T \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^T \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{aligned}$$

Therefore, our pseudo-inverse can be calculated as

$$\begin{aligned}
A^+ &= V\Sigma^+U^T \\
&= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix}
\end{aligned}$$

Next, for matrix B, let us follow the same procedure. Notice that $B = B^T$, so we have

$$BB^T = B^TB = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

Now let me compute the eigenvalues of B^TB:

$$\begin{aligned}
\det(B^TB - \lambda I) &= \det \begin{bmatrix} 2-\lambda & 2 \\ 2 & 2-\lambda \end{bmatrix} \\
&= (2-\lambda)^2 - 4 \\
&= 4 - 4\lambda + \lambda^2 - 4 \\
&= \lambda^2 - 4\lambda \\
&= \lambda(\lambda - 4)
\end{aligned}$$

Thus, it is easy to see that the roots of the characteristic equation are 0 and 4. Therefore, the only singular value of B is $\sigma_1 = \sqrt{4} = 2$. We can now explicitly find our Σ in the SVD decomposition:

$$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

Next, I will calculate the eigenvectors of B^TB. Putting the eigenvalues of B^TB in decreasing order, we have $\lambda_1 = 4$ and $\lambda_2 = 0$. Therefore,

$$\begin{aligned}
&(B^TB - \lambda_1 I)v_1 = 0 \\
\implies &(B^TB - 4I)v_1 = 0 \\
\implies &\begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = 0 \\
\implies &\begin{cases} -2x_1 + 2y_1 &= 0 \\ 2x_1 - 2y_1 &= 0 \end{cases} \\
\implies &x_1 = y_1 \\
\implies &v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
&(B^TB - \lambda_2 I)v_2 = 0 \\
\implies &B^TBv_2 = 0 \\
\implies &\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = 0 \\
\implies &\begin{cases} 2x_1 + 2y_1 &= 0 \\ 2x_1 + 2y_1 &= 0 \end{cases} \\
\implies &x_1 = -y_1 \\
\implies &v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}
\end{aligned}$$

Thus, we can explicitly write V as

$$V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Since we only have one singular value, we can only explicitly write the first column of U, namely

$$\begin{aligned}
u_1 &= \frac{Bv_1}{\sigma_1} \\
&= \frac{1}{2\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
&= \frac{1}{2\sqrt{2}} \begin{bmatrix} 2 \\ 2 \end{bmatrix} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\end{aligned}$$

Therefore, we need to choose the second column of U to be orthonormal to u_1, but notice $u_1 = v_1$, so we can choose $u_2 = v_2$ to make U an orthogonal matrix. This gives our SVD decomposition as:

$$\begin{aligned} B &= U\Sigma V^T \\ &= \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}2 & 0\\ 0 & 0\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}^T \\ &= \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}2 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} \end{aligned}$$

Therefore, the pseudo-inverse is

$$\begin{aligned} B^+ &= V\Sigma^+U^T \\ &= \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}\frac{1}{2} & 0\\ 0 & 0\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}^T \\ &= \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}\frac{1}{2} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix}\frac{1}{2} & 0\\ \frac{1}{2} & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} \\ &= \begin{bmatrix}\frac{1}{4} & 0\\ \frac{1}{4} & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} \\ &= \begin{bmatrix}\frac{1}{4} & \frac{1}{4}\\ \frac{1}{4} & \frac{1}{4}\end{bmatrix} \\ &= \frac{1}{4}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} \end{aligned}$$

Question 2

let A and B be two symmetric matrices. Show that A and B possess a common basis of eigenvectors if and only if $AB = BA$.

Proof.
First, assume that A and B possess a common basis of eigenvectors. Since A and B are both symmetric, we know that they are both orthogonally diagonalizable. In other words

$$A = PD_AP^T \qquad \text{and} \qquad B = QD_BQ^T$$

where P and Q are both orthogonal matrices and D_A and D_B are diagonal matrices with entries equal to the eigenvalues of A and B respectively. Furthermore, in the proof that symmetric matrices are orthogonally diagonalizable, we found that P has columns equal to the eigenvectors of A and Q has columns equal to the eigenvectors of B. Thus, since A and B have the same set of eigenvectors, we can say that $P = Q$. Next, it is clear that D_A and D_B are both symmetric since they are diagonal. Furthermore, their product will be symmetric as it is simply the product of corresponding diagonal entries. Therefore, we have

$$D_AD_B = D_A^TD_B^T = (D_BD_A)^T = D_BD_A$$

in other words, two diagonal matrices are commutative under multiplication. With all of this together, we have that

$$\begin{aligned} AB &= (PD_AP^T)(QD_BQ^T) \\ &= (PD_AP^T)(PD_BP^T) \\ &= PD_AD_BP^T \\ &= PD_BD_AP^T \\ &= PD_BP^TPD_AP^T \\ &= (QD_BQ^T)(PD_AP^T) \\ &= BA \end{aligned}$$

Thus, A and B are also commutative under multiplication.

Next, assume that $AB = BA$. Let v be an eigenvector for A with associated eigenvalue λ. Then, we have

$$\begin{aligned} & AB = BA \\ \implies & ABv = BAv \\ & \qquad = B(\lambda v) \\ & \qquad = \lambda Bv \\ \implies & A(Bv) = \lambda(Bv) \end{aligned}$$

If Bv is the zero vector, then $Bv = 0 = 0v$, so v is also an eigenvector for B. If Bv is not the zero vector, then we have that Bv is also an eigenvector of A with the same eigenvalue of λ.

First, assume that E_λ, the eigenspace of A associated with λ, has dimension of one. If this is the case, then since $v, Bv \in E_\lambda$ and since $\dim(E_\lambda) = 1$, then v and Bv are scalar multiples of one another. In particular, we can say $Bv = \alpha v$, so that v is an eigenvector of B with eigenvalue of α.

Next, assume that E_λ has a dimension of $p > 1$. Let $v_1, v_2, \ldots, v_p$ be a basis of E_λ consisting of orthonormal eigenvectors of A (this is always attainable with Graham-Schmidt and normalization). Just as we showed Bv must be an eigenvector of A given that v is an eigenvector, then so too must we have that Bv_k is an an eigenvector of A with eigenvalue of λ for all $1 \le k \le p$. Thus, $Bv_k \in E_\lambda$ for all $1 \le k \le p$. Since Bv_k is in the eigenspace, then we can express it as a linear combination of basis elements for that space, i.e.

$$Bv_k = c_{1k}v_1 + c_{2k}v_2 + \cdots + c_{pk}v_p$$

In particular, we can consider the matrix $C \in \mathbb{R}^{p \times p}$ and $V \in \mathbb{R}^{n \times p}$ defined by

$$B \begin{bmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_p \\ | & | & \cdots & | \end{bmatrix} = \begin{bmatrix} | & | & \cdots & | \\ Bv_1 & Bv_2 & \cdots & Bv_p \\ | & | & \cdots & | \end{bmatrix} = \begin{bmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_p \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & & \ddots & \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix} = VC$$

Let $\nu = (\nu_1, \nu_2, \ldots, \nu_p)$ be an eigenvector of C with eigenvalue α, i.e. $C\nu = \alpha\nu$ and $\nu \neq 0$. When multiplying both sides of the previous equality by ν, we get

$$\begin{aligned} BV &= VC \\ BV\nu &= VC\nu \\ BV\nu &= \alpha V\nu \end{aligned}$$

By defining y as $y = \nu_1 v_1 + \nu_2 v_2 + \cdots + \nu_p v_p = V\nu$, we see that the previous equality is equivalent to $By = \alpha y$. Therefore, y is an eigenvector of B with eigenvalue of α. On the other hand, since y is a non-trivial linear combination of $\{v_k\}_{k=1}^p$, then we know that $y \in E_\lambda$ which means that y is an eigenvector for A with eigenvalue of λ. Notice that since $C \in \mathbb{R}^{p \times p}$, we can find p eigenvectors of C. Therefore, the previous construction of y with ν could be done with any of C's p linearly independent eigenvectors, and we can denote these as $\{y_k\}_{k=1}^p$ which must also all be linearly independent. In a similar fashion as before, each y_k is an eigenvector of A and B.

Since A is a symmetric matrix, then we have that for each eigenvalue λ_k of A with eigenspace E_{λ_k}, then if A has r distinct eigenvalues, then the following equality holds

$$\sum_{k=1}^{r} \dim(E_{\lambda_k}) = n$$

However, we have shown that if $\dim(E_\lambda) = p$, we can find p linearly independent vectors which are eigenvectors of A and B. Thus, since the sum of all of these is n and each eigenspace has no overlapping elements (aside from the zero vector), we can find n linearly independent eigenvectors of A that are also eigenvectors of B, meaning the two matrices have a common basis of eigenvectors. □

Question 3

For $A \in \mathbb{R}^{m \times n}$, show that

$$\operatorname{rank}(A) = \operatorname{rank}(A^T) = \operatorname{rank}(AA^T) = \operatorname{rank}(A^T A).$$

Proof.
Let $b_i \in \mathbb{R}^n$ represent the i-th column of A^T, or equivalently the transpose of the i-th row of A. Assume that A^T has a rank of k. This means there exists some set $\{u_1, u_2, \ldots, u_k\} \subset \mathbb{R}^n$ which is a basis for the column space of A^T. Since b_i is one of the columns of A^T, we have

$$b_i = c_{i1}u_1 + c_{i2}u_2 + \cdots + c_{ik}u_k$$

Recalling that these b_i's are the columns of A^T, we have

$$
\begin{aligned}
A = (A^T)^T &= \begin{bmatrix} \rule{1cm}{0.4pt} & b_1^T & \rule{1cm}{0.4pt} \\ \rule{1cm}{0.4pt} & b_2^T & \rule{1cm}{0.4pt} \\ & \vdots & \\ \rule{1cm}{0.4pt} & b_m^T & \rule{1cm}{0.4pt} \end{bmatrix} \\
&= \begin{bmatrix} c_{11}u_1^T + c_{12}u_2^T + \cdots + c_{1k}u_k^T \\ c_{21}u_1^T + c_{22}u_2^T + \cdots + c_{2k}u_k^T \\ \vdots \\ c_{m1}u_1^T + c_{m2}u_2^T + \cdots + c_{mk}u_k^T \end{bmatrix} \\
&= \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ & \vdots & \ddots & \\ c_{m1} & c_{m2} & \cdots & c_{mk} \end{bmatrix} \begin{bmatrix} \rule{1cm}{0.4pt} & u_1^T & \rule{1cm}{0.4pt} \\ \rule{1cm}{0.4pt} & u_2^T & \rule{1cm}{0.4pt} \\ & \vdots & \\ \rule{1cm}{0.4pt} & u_k^T & \rule{1cm}{0.4pt} \end{bmatrix}
\end{aligned}
$$

Letting a_i be the i-th column of A and letting $u_j^T = [u_{1j}, u_{2j}, \ldots, u_{nj}]$, we see from this expansion that

$$
\begin{aligned}
a_i &= \begin{bmatrix} c_{11}u_{i1} + c_{12}u_{i2} + \cdots + c_{1k}u_{ik} \\ c_{21}u_{i1} + c_{22}u_{i2} + \cdots + c_{2k}u_{ik} \\ \vdots \\ c_{m1}u_{i1} + c_{m2}u_{i2} + \cdots + c_{mk}u_{ik} \end{bmatrix} \\
&= u_{i1}\begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{bmatrix} + u_{i2}\begin{bmatrix} c_{12} \\ c_{22} \\ \vdots \\ c_{m2} \end{bmatrix} + \cdots + u_{ik}\begin{bmatrix} c_{1k} \\ c_{2k} \\ \vdots \\ c_{mk} \end{bmatrix}
\end{aligned}
$$

Therefore, we see that each column of A can be expressed as the linear combination of k different vectors in $\mathbb{R}^m$. Thus, $\dim(\text{Range}(A)) = r \leq k$. On the other hand, we could start with a basis for the column space of A, say $\{v_1, v_2, \ldots, v_r\} \subset \mathbb{R}^m$ and express a column of A as a linear combination of these vectors, then realize A^T has each column of A as rows and do the same calculations as above to find a linear combination of r different vectors in $\mathbb{R}^n$ that is equal to each column of A^T. In this manner, we would show that $k \leq r$. With both inequalities in place, we have that $\text{Rank}(A) = r = k = \text{Rank}(A^T)$.

For the next equality, I will try to impose the Rank-Nullity Theorem. Therefore, let $x \in \text{Null}(A) = N(A)$. This means, the following equalities hold

$$
\begin{aligned}
Ax &= 0 && \text{by definition of Null space} \\
\implies A^TAx &= A^T0 && \text{by multiplying by } A^T \\
\implies A^TAx &= 0
\end{aligned}
$$

Thus, $x \in N(A^TA)$, so $N(A) \subset N(A^TA)$. On the other hand, let $y \in N(A^TA)$. With this in place, we have

$$
\begin{aligned}
A^TAy &= 0 && \text{by definition of Null space} \\
\implies y^TA^TAy &= y^T0 && \text{by multiplying by } y^T \\
\implies (Ay)^TAy &= 0 && \text{by property of product transpose} \\
\implies ||Ay||^2 &= 0 && \text{by definition of vector norm} \\
\implies Ay &= 0 && \text{by the property of vector norms}
\end{aligned}
$$

Thus, $y \in N(A)$ so that $N(A^TA) \subset N(A)$. With both inclusions, we can say that $N(A) = N(A^TA)$. This means we have

$$
\begin{aligned}
\text{Rank}(A) &= n - \dim(N(A)) && \text{by Rank-Nullity Theorem} \\
&= n - \dim(N(A^TA)) && \text{by above equality} \\
&= \text{Rank}(A^TA) && \text{by Rank-Nullity Theorem}
\end{aligned}
$$

Therefore $\text{Rank}(A) = \text{Rank}(A^TA)$. To show that $\text{Rank}(A^T) = \text{Rank}(AA^T)$, simply use the result that $\text{Rank}(A) = \text{Rank}(A^TA)$ applied with A set to be A^T and notice that $(A^T)^T(A^T) = AA^T$. Altogether, this gives the desired sequence of equalities. $\square$

Question 4

Let $A \in \mathbb{R}^{m\times n}$ have singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Show that

$$\operatorname{rank}(A) = r, \qquad ||A||_2 := \max_{x\neq 0} \frac{||Ax||_2}{||x||_2} = \sigma_1, \qquad ||A||_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}.$$

Proof.
If A has r singular values, then that means that $A^TA \in \mathbb{R}^{n\times n}$ has r non-zero eigenvalues, or in fact $n-r$ eigenvalues equal to zero. Notice that is v is an eigenvector corresponding to an eigenvalue of zero, then $A^TAv = 0v = 0$. Thus, $v \in N(A^TA)$. Since there are $n-r$ linearly independent eigenvectors corresponding to eigenvalues of zero, we can say that $\dim(N(A^TA)) = n-r$. Thus, by the Rank-Nullity Theorem, we have that

$$\operatorname{Rank}(A^TA) = n - \dim(N(A^TA)) = n-(n-r) = r$$

Therefore, by the result from Question 3, we have $\operatorname{Rank}(A) = \operatorname{Rank}(A^TA) = r$ which proves the first property.

Next, let $\{v_1, v_2, \ldots, v_n\}$ be an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A^TA. Since the singular values of A are the eigenvalues of A^TA, we can say that v_i is the eigenvector corresponding to σ_i^2. Thus, let $x \in \mathbb{R}^n$ be expressed as

$$x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

Next, notice that

$$||Ax||_2^2 = \langle Ax, Ax\rangle = (Ax)^T Ax = x^TA^TAx = \langle x, A^TAx\rangle$$

Notice we can explicitly write A^TAx as

$$\begin{aligned} A^TAx &= A^TA(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n) \\ &= \sigma_1^2\alpha_1 v_1 + \sigma_2^2\alpha_2 v_2 + \cdots + \sigma_n^2 \alpha_n v_n \end{aligned}$$

since each v_i is an eigenvector of A^TA. Thus, recalling that our basis of eigenvectors is orthonormal, we can compute the inner product of x and A^TAx as

$$\begin{aligned} \langle x, A^TAx\rangle &= \langle \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n, \sigma_1^2\alpha_1 v_1 + \sigma_2^2\alpha_2 v_2 + \cdots + \sigma_n^2\alpha_n v_n\rangle \\ &= \sigma_1^2\alpha_1^2 + \sigma_2^2\alpha_2^2 + \cdots + \sigma_n^2\alpha_n^2 \\ &\leq \sigma_1^2(\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2) && \text{since } \sigma_1 \text{ is the largest singular value} \\ &= \sigma_1^2 ||x||_2^2 \end{aligned}$$

Thus, by taking square roots, we get $||Ax||_2 \leq \sigma_1 ||x||_2$. Using this, we have the following inequality

$$\begin{aligned} ||A||_2 &= \max_{x\neq 0} \frac{||Ax||_2}{||x||_2} \\ &\leq \max_{x\neq 0} \frac{\sigma_1 ||x||_2}{||x||_2} \\ &= \sigma_1 \end{aligned}$$

On the other hand, taking x to be an eigenvector associated with σ_1^2, we get that

$$\begin{aligned} ||Ax||_2^2 &= \langle x, A^TAx\rangle \\ &= \langle x, \sigma_1^2 x\rangle \\ &= \sigma_1^2 ||x||_2^2 \end{aligned}$$

By taking square roots, we get $||Ax||_2 = \sigma_1 ||x||_2$. Thus, we have

$$\frac{||Ax||_2}{||x||_2} = \frac{\sigma_1 ||x||_2}{||x||_2} = \sigma_1$$

Therefore, since this expression must be no greater than the maximum for this expression, we get

$$\sigma_1 \leq \max_{x\neq 0} \frac{||Ax||_2}{||x||_2} = ||A||_2$$

Therefore, with both inequalities in place, we can conclude the second property of this question is true.

Notice this proof could have been a bit shorter if I had used the fact that the 2-norm is invariant by left or right multiplication of orthogonal matrices to get that $||A||_2 = ||U\Sigma V^T||_2 = ||\Sigma||_2$ by singular value decomposition. We proved this in the last homework, but in that proof, I used the result which I just proved about singular values. Therefore, to avoid any circular reasoning, I went for a more direct proof.

For the last property, I will use the result from the last homework that $||QBZ||_F = ||B||_F$ for Q and Z orthogonal matrices. Thus, by using the singular value decomposition of A, we have

$$||A||_F = ||U\Sigma V^T||_F = ||\Sigma||_F \qquad \text{since } U \text{ and } V^T \text{ are by definition orthogonal}$$

Thus, calculating the Frobenius norm of Σ is simple since it is simply the square root of the sum of squares of each entry of Σ and Σ only has the r non-zero entries consisting of singular values along the diagonal. Therefore,

$$||\Sigma||_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}$$

Thus, using this equality with the last relation about $||A||_F$, we get the desired property must be true. $\square$

Question 5

Show that every invertible matrix A can be written uniquely in the form $A = CU$ where C is an orthogonal matrix and U is a positive definite orthogonal matrix.

I believe the way the question is stated is currently false. If C and U are both orthogonal, then CU must also be orthogonal since

$$CU(CU)^T = CUU^TC^T = CC^T = I$$
$$(CU)^TCU = U^TC^TCU = U^TU = I$$

This would imply, however, that A was orthogonal which was not given as a hypothesis in the question. Therefore, I will assume that the restriction of U to be an orthogonal matrix was a typo and remove that restriction in my proof.

Proof.
Recall from the singular value decomposition that we can write

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

where W and V are both orthogonal and Σ consists of the singular values of A along its diagonal. Note that since A is invertible, it is full rank, so A^TA is full rank, meaning it has all non-zero eigenvalues. Thus, since the singular values are the square roots of the eigenvalues of A^TA, we know that Σ has non-zero entries along all of its diagonal components, meaning it is full rank. First, I will consider the matrix

$$C = WV^T$$

Since W and V are both orthogonal, then so too must C be orthogonal (by the same argument used in my remark preceding this proof). Next, define U as

$$U = V\Sigma V^T \implies CU = WV^TV\Sigma V^T = W\Sigma V^T = A$$

Note that this is positive definite. This can be seen by taking any arbitrary nonzero $x \in \mathbb{R}^n$. Then, if we define $y := V^Tx$, y must be nonzero since V is orthogonal (in particular orthogonal matrices are full rank). Using this, we have

$$\begin{aligned} x^TUx &= x^TV\Sigma V^Tx \\ &= (V^Tx)^T\Sigma V^Tx \\ &= y^T\Sigma y \\ &= \sigma_1 y_1^2 + \sigma_2 y_2^2 + \cdots + \sigma_n y_n^2 > 0 \end{aligned}$$

where the last inequality follows since $y \neq 0$ and each $\sigma_i > 0$. Therefore, we have found our respective C and U. Furthermore, they are unique since if $A = CU$, then we have the equality

$$A^TA = (CU)^TCU = U^TC^TCU = U^TU = (V\Sigma V^T)^T(V\Sigma V^T) = UU = U^2$$

In other words, U must be the square root of A^TA. However, the square root of a symmetric and positive definite matrix (which A^TA is both symmetric and positive definite since all of its eigenvalues are positive) is unique, so U is unique. Therefore, since U is invertible, C is uniquely determined as $C = AU^{-1}$, making the entire decomposition unique. $\square$

Question 6

Let $A \in \mathbb{R}^{m \times n}$ satisfy $\text{rank}(A) = n$. Show that the pseudo-inverse of A is given by

$$A^+ = (A^T A)^{-1} A^T.$$

Furthermore, show that $A^+ = A^{-1}$ if $m = n$.

Proof.
First, note that this matrix is well-defined. In particular, I need to show that $A^T A$ is invertible. However, since $A \in \mathbb{R}^{m \times n}$, then $A^T A \in \mathbb{R}^{n \times n}$. By Question 3, we know that $\text{rank}(A) = \text{rank}(A^T A)$, therefore $A^T A$ has rank of n, so it is full rank and square, so it is invertible.

Now, I will provide this verification of the pseudo-inverse by showing that that this matrix satisfies the Moore-Penrose Conditions. Since matrices that satisfy those conditions are uniquely determined and we know that the pseudo-inverse satisfies the Moore-Penrose conditions, then this would be sufficient to show that that this matrix is indeed the pseudo-inverse. Let $B = (A^T A)^{-1} A^T$, then the conditions to verify are:

1. $(AB) = (AB)^T$
   - $AB = A(A^T A)^{-1} A^T$
   - $(AB)^T = (A(A^T A)^{-1} A^T)^T = A[(A^T A)^{-1}]^T A^T = A[(A^T A)^T]^{-1} A^T = A(A^T A)^{-1} A^T$
2. $(BA) = (BA)^T$
   - $BA = (A^T A)^{-1} A^T A = I$
   - $(BA)^T = ((A^T A)^{-1} A^T A)^T = I^T = I$
3. $ABA = A$
   - $ABA = A(A^T A)^{-1} A^T A = AI = A$
4. $BAB = B$
   - $BAB = (A^T A)^{-1} A^T A (A^T A)^{-1} A^T = I(A^T A)^{-1} A^T = (A^T A)^{-1} A^T = B$

Thus, B satisfies the Moore-Penrose conditions, so B is indeed the pseudo inverse of A.

Next, note that if $m = n$ and $\text{rank}(A) = n$, then A and A^T are both invertible. Thus, we have

$$(A^T A)^{-1} A^T = A^{-1} (A^T)^{-1} A^T = A^{-1} I = A^{-1}$$

which proves the second identity. $\square$