①

① Here, we are proving that the eigenvalues of AA^T and A^TA are real and non-negative.

First, we will prove that it is real.
Notice that $A \cdot A^T$ and A^TA are square symmetric matrices.
Now, we need to prove that symmetric matrices all have real eigenvalues.

Denote matrix Y as symmetric matrix, that is real.
To find eigenvalues, we have to set

$Yx = \lambda x$, where λ can be a complex eigenvalue of Y.

Now, we take the complex conjugate

$\hookrightarrow$ ex) $3+i \rightarrow 3-i$

Since Y is real, its complex conjugate is itself, giving

$$Y\bar{x} = \bar{\lambda}\bar{x}$$

Now, if we transpose both sides, we get

$\bar{x}^T Y = \bar{\lambda}\bar{x}^T$, since Y and $\bar{\lambda}$ are both symmetrix matrices.
$(\bar{\lambda}^T = \bar{\lambda})$

If we multiply x to both sides on the right, we get

$$\bar{x}^T Y x = \bar{\lambda}\bar{x}^T x$$

Notice that $Yx = \lambda x$ from the beginning. Therefore,

$\bar{x}^T Y x = \bar{x}^T \lambda x$, which equals $\lambda \bar{x}^T x$.

The complex conjugate multiplied by the original complex number always provides a positive number.
Therefore, $\bar{x}^T x$ is always positive.

Going back to the equation,

$$\bar{\lambda} x^T x = \bar{x}^T Y x = \bar{x}^T \lambda x = \lambda \bar{x}^T x.$$

Since $x^T x$ is positive, if we divide both sides by it at
$\bar{\lambda} x^T x = \lambda x^T x$,

we get $\bar{\lambda} = \lambda$.

This means that the complex conjugate of λ is equal to λ, meaning that λ is consisted of only real numbers.

Therefore, the symmetrix matrix Y has real eigenvalues.

$\boxed{\text{Since } A^TA \text{ and } AA^T \text{ are symmetric, they have real eigenvalues.}}$

Now, we have to prove that their eigenvalues are non-negative.

We can write $AA^T x = \lambda x$.

Multiplying both sides by x^T on the left, we get

$x^T A A^T x = x^T \lambda x$
$= \lambda x^T x \rightarrow$ since λ is matrix with only diagonal entries.

Then, we get $x^T A A^T x = (A^T x)^T \cdot A^T x$ by applying the rule of transpose.

Then, we get $(A^T x)^T A^T x = \lambda x^T x$
Therefore, we get $\|A^T x\|_2^2 = \lambda \|x\|_2^2$ since $A^T x$ is a vector
The values $\|A^T x\|_2^2$ and $\|x\|_2^2$ are never non-negative since they are squared length of projections and $\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$

Since they are non-negative,

$$\lambda = \frac{\|A^T x\|_2^2}{\|x\|_2^2}$$

is also non-negative.

Similarly,

$$A^T A x = \lambda x.$$

Multiplying x^T to both sides on the left, we get

$$x^T A^T A x = x^T \lambda x$$

Then, it becomes $(Ax)^T Ax = \lambda x^T x$

$$\|Ax\|^2 = \lambda \|x\|_2^2$$

$$\lambda = \frac{\|Ax\|_2^2}{\|x\|_2^2}$$

Again, $\|Ax\|^2$ and $\|x\|^2$ are non-negative values, giving λ also non-negative.

①②

Here, we have to prove that the eigenvalues of AA^T and A^TA are equal to each other.

From part 1, we proved that the eigenvalues of AA^T and A^TA are both real and non-negative. Therefore, AA^T can be written as

$$(AA^T)x = \lambda x,$$

where λ are the eigenvalues of AA^T.

If we multiply both sides by A^T at the left, we get

$$A^T(AA^Tx) = A^T\lambda x$$
$$= \lambda A^T x$$

By using associative property, we can write the equation above as

$$A^TA(A^Tx) = \lambda(A^Tx) \rightarrow A^Tx \text{ is a vector}$$

This means that the eigenvalues of A^TA is also λ, proving that the eigenvalues for A^TA and AA^T are the same.

①

③

According to the lecture, any matrix $A \in \mathbb{R}^{m \times n}$ has an orthonormal matrices. $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$, and a diagonal matrix Σ

Therefore, we can denote any matrix $A \in \mathbb{R}^{m \times n}$ as
$A = U \Sigma V^T$

The transpose of A equals to
$A^T = V \Sigma U^T$

Therefore, $A \cdot A^T = U \Sigma V^T V \Sigma U^T$
Since V is orthogonal according to the definition (and on the lecture notes), $V^T V = I$.

This means that $A \cdot A^T = U \Sigma I \Sigma U^T$
$= U \Sigma^2 U^T$

Similarly, $A^T A = V \Sigma U^T U \Sigma V^T$
$= V \Sigma^2 V^T$

From part 1.1, we noticed that the matrix $A^T A$ and AA^T are symmetric. This means that we can write

$AA^T = PDP^{-1}$, where P is the eigenvector and D is the eigenvalue matrix.

Therefore, we can write $PDP^{-1} = U \Sigma^2 U^T$.

This shows that the set of singular values, Σ, is the square root of the eigenvalues of A^TA or AA^T, and the left singular vectors are eigenvectors of AA^T. The transpose of eigenvectors of AA^T are eigenvectors of A^TA, meaning that the right singular vectors are eigenvectors of A^TA.

In summary,

$$AA^T = U \quad \Sigma \quad V^T$$

U: eigenvectors of AA^T

Σ: square root of eigenvalues of AA^T or A^TA

V^T: eigenvectors of A^TA

② ①

This question asks to prove that orthonormal rows of a square matrix implies orthonormal columns.

If a square matrix has orthonormal rows, it means that the rows of the matrix, let's say $a_1 \cdots a_n$, has the property $a_i \cdot a_i^T = 1$ for all $i \leq n$, and $a_i \cdot a_j = 0$ for all $i \neq j$, and $i, j \leq n$

This means that $A \cdot A^T = I$, implying $A = (A^T)^{-1}$

In order to prove that it has orthonormal columns, we need to prove that $A^T \cdot A = I$. By definition of inverse, $A^T \cdot (A^T)^{-1} = I$. From the previous statements, we showed that $(A^T)^{-1} = A$. By plugging in, we get $A^T \cdot A = I$, which shows that if rows of a square matrix is orthonormal, the columns are also orthonormal.

However, this statement is false when A is not a square matrix. More specifically, if the rectangular matrix has less rows that are orthonormal, it doesn't guarantee orthonormal columns.

Ex) $M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ $M^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

$M \cdot M^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ → orthonormal rows

$M^T \cdot M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ → not orthonormal columns.

②②

In this question, we are proving that a linear system $Ax=b$ is consistent if and only if $\text{rank}(A) = \text{rank}([A|b])$

If $Ax=b$ is consistent, it means that b is a linear combination of the column vectors of A.

The proof says "if and only if" meaning that we need to prove in both ways.

- Assuming $Ax=b$ is consistent, $\text{rank}(A) = \text{rank}([A|b])$
- Assuming $\text{rank}(A) = \text{rank}([A|b])$, $Ax=b$ is consistent

Let's first prove the first statement. If $Ax=b$ is consistent, it means that there exist x such that $Ax=b$ and that b is a linear combination of columns in A. In other words, b is dependent of the columns in matrix A.

By definition $[A|b]$ is the addition of column b to A. Since you are augmenting b that is dependent, it doesn't add new information regarding the column space of A. And also, the rank never decreases with the addition of a column. It will always stay the same if that column is dependent or increase if the column is independent.

Because the rank never decreases when adding column b and never increases when adding a dependent column, the rank will stay the same.

↳ $\text{rank}(A) = \text{rank}([A|b])$

Now, let's prove the second statement.

We assume $rank(A) = rank([A|b])$. Suppose the rank of matrix A is r. This means that there are r linearly independent columns of A, $a_1, \dots, a_r$. Therefore, this shows that its columns span a subspace of dimension r.

Since $rank([A|b]) = rank(A)$, it also has a rank of r. This means that there are r linearly independent columns of $[A|b]$ as well, $a_1, \dots, a_r, b^*$. Now, consider the columns $a_1, \dots, a_r$. These columns span the subspace of dimension r, identical to the subspace spanned by columns of matrix A.

The column b^* lies in the span of these columns by definition. This means that there exists a combination of vectors in the subspace spanned by columns of matrix A such that makes $Ax^* = b^*$. Plugging this vector x^* to x with all other vectors uninvolved as 0, we get $Ax = b$, showing that $rank(A) = rank([A|b])$ guarantees that $Ax = b$ is consistent.

$Rank(A) = rank([A|b])$, geometrically, means that b spans on the column space of vectors in matrix A.

(2) (3)

(a) $M = [0, 1, 2)$

$M^T = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$

$M \cdot M^T = [5]$

$M^T \cdot M = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} [0\,1\,2] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 4 \end{bmatrix}$

Now, we need to find eigenvalue of MM^T.

$M \cdot M^T \cdot x = \lambda x$

$det(|MM^T - \lambda I|) = 0$

$5 - \lambda = 0$

$\lambda = 5$

The eigenvector of MM^T is $[1]$

The eigen vector of $M^T M$ is...

knowing that $\lambda = 5, 0, 0$

$\begin{bmatrix} -\lambda & 0 & 0 \\ 0 & 1-\lambda & 2 \\ 0 & 2 & 4-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$\begin{bmatrix} -5 & 0 & 0 \\ 0 & -4 & 2 \\ 0 & 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ $\left.\begin{matrix} 0x = 0 \\ y + 2z = 0 \\ 2y + 4z = 0 \end{matrix}\right\} \rightarrow \begin{matrix} x = 0 \\ y = -2 \\ z = 1, \end{matrix} \quad \begin{matrix} x = 1 \\ y = 0 \\ z = 0 \end{matrix}$

$\left.\begin{matrix} -5x = 0 \\ -4y + 2z = 0 \\ 2y - z = 0 \end{matrix}\right\} \rightarrow \begin{matrix} x = 0, \\ y = 1 \\ z = 2 \end{matrix}$

Now, plugging in, we get

$$M = U \Sigma V^T$$
$$= [1] \begin{bmatrix} \sqrt{5} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -2 \\ 2 & 0 & 1 \end{bmatrix}^T$$

have to normalize and make orthonormal

$$M = [1] \begin{bmatrix} \sqrt{5} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ \sqrt{5}/5 & 0 & -2\sqrt{5}/5 \\ 2\sqrt{5}/5 & 0 & \sqrt{5}/5 \end{bmatrix}^T$$

$$= [1] \begin{bmatrix} \sqrt{5} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \sqrt{5}/5 & 2\sqrt{5}/5 \\ 1 & 0 & 0 \\ 0 & \frac{-2\sqrt{5}}{5} & \sqrt{5}/5 \end{bmatrix}$$

(b) By "eyeballing" M, we have to figure out the SVD of $[0,1,2]$

We know that SVD has the equation

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

According to the lecture notes, V is orthogonal, giving

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

Since Σ is simply the singular values matrix, U determines the column space of M. More specifically, since the matrix M has rank of 1, only the first u_1 span the column space of M.

Therefore, U is the column space of M, which is $[1]$.

Similarly, the lecture notes states that

$M^T U = \Sigma V$ by transposing M and since U is orthogonal, we can cancel out $U U^T$

Again, V is the column space of M^T. However, because we did transpose, V^T determines the row space of M. The matrix, once again, has rank of 1, which gives the first v_1^T as the row space of M.

There, V^T is the row space of M, which is $[0,1,2]$

Now, this matrix M has more columns than rows, which means that more entries should be added on V^T. Here comes the nullspace for the remaining two columns $(3-1=2)$. In other words, the last two columns of V^T is the null space of A.

To prove this statement, we need to prove that $M\vec{x}=0$

↳ where $\vec{x} = \sum_{i=m+1}^{n} \vec{v}_i$, where m = number of rows in M and n = number of columns in M.

$$M\vec{x} = \sum_{i=m+1}^{n} M\vec{v}_i$$

Now,

$$M\vec{x} = U \Sigma I$$

↳ since $v_i^T \cdot v_i = I$ for all i

Notice that Σ, when $m<n$, we have

$$\left(\begin{array}{c|c} \begin{matrix} a_1 & & \\ & \ddots & \\ & & a_r \end{matrix} & 0 \cdots 0 \\ \hline & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \; 0 \cdots 0 \end{array}\right)$$

for $i=m+1$ to $i=n$, there are only 0s, giving $\Sigma \cdot I = 0$

$$\begin{aligned} \text{Therefore, } M\vec{x} &= U\Sigma I \\ &= U 0 \\ &= 0 \end{aligned}$$

↳ which is the null space of $M\vec{x}$.

The nullspace of M is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}$.

The eigenvalue is simply $\sqrt{1^2+2^2} = \sqrt{5}$, with rest of two being 0

Therefore, we can write $M = U\Sigma V^T$ as

$$= [1]\,[\sqrt{5}\ \ 0\ \ 0] \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & -2 & 1 \end{bmatrix}.$$

By making them orthonormal, we get

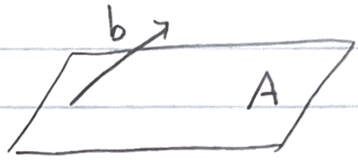
$$M = [1]\,[\sqrt{5}\ \ 0\ \ 0]\begin{bmatrix} 0 & \sqrt{5}/5 & 2\sqrt{5}/5 \\ 1 & 0 & 0 \\ 0 & -2\sqrt{5}/5 & \sqrt{5}/5 \end{bmatrix},$$

also matching as my solution in part a.

③ ①

In this question, we have to prove that the unique minimizer $x^* = (A^TA)^{-1}A^Tb$ when trying to minimize the error $\|Ax-b\|_2$.

By definition, A consists of vectors $a_1, \dots, a_m$ that span a subspace U and we are trying to project vector b on U. In picture,

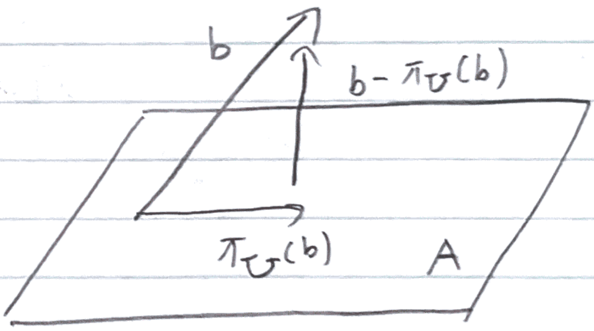


Therefore, we want to find the orthogonal projection of b onto the column space of A.

Now, there exists $x_1, \dots, x_m$ such that $\pi_U(b) = x_1 a_1 + \dots + x_m a_m$

Then, since $b - \pi_U(b)$ is orthogonal to a_i $(i = 1, \dots, m)$, we can write it as

$$A^T b = (A^TA)x^*$$



Now, according to the question, A has a full rank of n, and therefore has independent columns, which means that we can invert the square matrix A^TA because it has n pivot columns.

$$A^T b = (A^TA)x^*$$
$$x^* = (A^TA)^{-1}A^Tb$$

(3)

(2)

Here, we are solving for the same x^* when knowing that the matrix is not full rank. This means that we can't invert A^TA in the previous section.

$$\min_x \|Ax-b\|_2^2$$

We can write A as AVV^T since $VV^T = I$ because they are orthonormal matrices.

$$\|Ax-b\|_2^2 = \|AVV^Tx-b\|_2^2$$

Now, we can multiply matrix U^T to $\|Ax-b\|_2^2$, because U^T, be definition is a unitary matrix, and unitary matrix multiplied by a vector (Ax and b) only changes the shape of the vector and doesn't change the length.
We are finding the $\|Ax-b\|_2^2$, which is a vector. Therefore,

$$U^T\|Ax-b\|_2^2 = \|Ax-b\|_2^2$$

$$U^T\|AVV^Tx-b\|_2^2 = \|U^TAVV^Tx-U^Tb\|_2^2$$

Now, we write A as $U\Sigma V^T$ by applying SVD decomposition.

$$\|U^TU\Sigma V^TVV^Tx-U^Tb\|_2^2 = \|\Sigma V^Tx-U^Tb\|_2^2$$

$$= \sum_{i=1}^{r}(\sigma_i v_i^T x_i - u_i^T b)^2 + \sum_{i=r+1}^{m}(u_i^T b)^2$$

↳ since $m > n$

Now, if we set x_i = a value to make the set of equation to make $\sum_{i=1}^{r} (\sigma_i v_i^T x_i - u_i^T b)^2$ equal to 0, we can find

$\min_x \|Ax-b\|_2^2$, which will be equal to $\sum_{i=r+1}^{m} (u_i^T b)^2$.

Since there exists a square, all of $\sigma_i v_i^T x_i - u_i^T b$ should equal 0 since $\sum_{i=1}^{r} 0 = 0$.

$$\sigma_i v_i^T x_i - u_i^T b = 0 \quad \text{for } i=1,\dots,r$$

Since σ_i is a value, we can divide it to both sides.

$$v_i^T x_i = \frac{u_i^T b}{\sigma_i} \quad \text{for all } i=1,\dots,r$$

Since v_i^T is orthonormal, if we multiply v_i to both sides on the left, we can remove it

$$v_i v_i^T x_i = \frac{v_i u_i^T b}{\sigma_i} \quad \text{for all } i=1,\dots,r$$

$$x_i = \frac{v_i u_i^T b}{\sigma_i}$$

Therefore, the minimizer x^* equals $\sum_{i=1}^{r} \left(\frac{v_i u_i^T b}{\sigma_i} \right)$