Linear Algebra Notes for Final

Date: May 5 Made by Eric

Angle of Discussion

The textbook definition of an inner product is a function that maps two vectors with order to a scaler and satisfies the following requirements.

$$\langle x+z, y\rangle = \langle x, y\rangle + \langle z, y\rangle \tag{1}$$

$$\langle cx, y\rangle = c\langle x, y\rangle \tag{2}$$

$$\overline{\langle x, y\rangle} = \langle y, x\rangle \tag{3}$$

$$\langle x, x\rangle \geq 0 \tag{4}$$

Give close attention to the third and fourth requirement. Implicitly and almost informally, they each require the inner product space to be over a field that have certain special property. In a symbolic sense, the third requirement require us to define a conjugacy relations between elements of the range of inner product function, and the fourth requirement require us to define positive and negative over $\{\langle x, x\rangle | x \in V\}$.

Without further investigation, we should not believe that only vector space over $\mathbb{R}$ or $\mathbb{C} \geq \mathbb{R}$ on which can we define an inner product, and we should not believe that the result we have in this article don't hold true on vector space over other field. However, for the future discussion of metric space and topology, we now narrow our discussion to only vector space over $\mathbb{R}$ or $\mathbb{C}$.

Basic Notion about Orthonormal basis

Definition 1. *If two vector v, w satisfy $\langle v, w\rangle = \langle w, v\rangle = 0$, then we say v and w are* ***perpendicular****, and we write $v \perp w$.*

Lemma 1. $v \perp w \iff w \perp v$

Definition 2. *We say a set of vector $\{v_1, \dots, v_n\}$ is* ***orthogonal*** *if for each two vectors v_i, v_j, we have $v_i \perp v_j$.*

Definition 3. *On an inner product space V,* ***norm*** *$\|\cdot\|$ is a function that maps a vector to a scaler defined by $\|v\| = \sqrt{\langle v, v\rangle}$.*

Definition 4. *We say a set of vector $\{v_1, \dots, v_n\}$ is* ***orthonormal*** *if $\{v_1, \dots, v_n\}$ is orthogonal and for each vector v_i, we have $\|v_i\| = 1$.*

Lemma 2. *Every finite dimensional vector space have an orthonormal basis.*

Proof. Arbitrarily pick a basis for a finite dimensional vector space and do Gran-Schmidt orthonormal process finite-sequentially. The theorem follows immediately. ■

Theorem 3. *There exists some infinite dimensional vector space that have no orthonormal hamel basis.*

Proof. Skipped for being too difficult at the moment ■

Definition 5. *Let S be a set of vector. The **orthogonal component** $\{v \in V | \forall s \in S, v \perp s\}$ is written $S^\perp$. Let v be a vector. When we write $v^\perp$, we mean $\{v\}^\perp$.*

Lemma 4. *$S^\perp$ give a subspace.*

Definition 6. *Let S, W be two set of vectors. We say S and W are **perpendicular** and write $S \perp W$ if any two vector $s \in S$ and $w \in W$ give us $s \perp w$.*

Lemma 5. $S \perp W \iff W \perp S$

Lemma 6. $S \perp S^\perp$

Proof. Arbitrarily pick v and w respectively from S and $S^\perp$. Because w satisfy $\forall s \in S, s \perp w$, so we have $v \perp w$. ■

The next Lemma shows that $S^\perp$ is the maximal orthogonal component of S.

Lemma 7. $S \perp R \implies R \subseteq S^\perp$

Proof. Arbitrarily pick a vector r from R. We have $\forall s \in S, r \perp s$, so $r \in S^\perp$. ■

Lemma 8.

$$\mathit{span}(S) \subseteq (S^\perp)^\perp$$

Proof. From Lemma 6, we know $S \perp S^\perp$. This enable us to easily deduce span$(S) \perp S^\perp$, then by Lemma 7, we know span$(S) \subseteq (S^\perp)^\perp$. ■

Goal Problem: With what condition we have $S = (S^\perp)^\perp$?

Notice that by Lemma 4, we know that **for $S = (S^\perp)^\perp$ to hold true, we need S to be a subspace**. This enable us to reduce our *goal problem* to: With what condition does $W = (W^\perp)^\perp$ hold true?

The next two theorems give us answer in finite dimensional space.

Theorem 9. *Let V be a finite dimensional vector space and W be a subspace of V*

$$V = W \oplus W^\perp$$

Proof. Let $\alpha = \{v_1, \dots, v_n\}$ be an orthonormal basis for W and extend α to an orthonormal basis $\beta = \{v_1, \dots, v_r\}$ for V. We now show $W^\perp = \text{span}(\{v_{n+1}, \dots, v_r\})$.

Arbitrarily pick a vector $\sum_{i=n+1}^r c_i v_i$ from span$(\{v_{n+1}, \dots, v_r\})$, and arbitrarily pick a vector $\sum_{j=1}^n d_j v_j$ from W. We see

$$\langle \sum_{i=n+1}^r c_i v_i, \sum_{j=1}^n d_j v_j \rangle = \sum_{i=n+1}^r c_i \sum_{j=1}^n \overline{d_j} \langle v_i, v_j \rangle = 0$$

So we know $\sum_{i=n+1}^{r} c_i v_i \in W^\perp$.

Arbitrarily pick a vector $\sum_{i=1}^{r} c_i v_i$ from $W^\perp$. For that $\forall 1 \leq j \leq n, v_j \in W$, we know that for each $1 \leq j \leq n$

$$v_j \perp (\sum_{i=1}^{r} c_i v_i)$$

In another symbol, we can write

$$c_j = \langle \sum_{i=1}^{r} c_i v_i, v_j \rangle = 0$$

By this, we can write

$$\sum_{i=1}^{r} c_i v_i = \sum_{i=n+1}^{r} c_i v_i \in \text{span}(\{v_{n+1}, \ldots, v_r\})$$

(done)

The fact that $V = W + W^\perp$ and the fact that $W \cap W^\perp = \{0\}$ follow immediately. ■

Theorem 10. ***(The answer of goal problem in finite dimensional vector space)*** *Let W be a subspace of a finite dimensional vector space.*

$$W = (W^\perp)^\perp$$

Proof. Arbitrarily pick an orthonormal basis $\alpha = \{w_1, \ldots, w_n\}$ for W and extend it to an orthonormal basis $\beta = \{w_1, \ldots, w_r\}$ for V.

In the proof of Theorem 9, we have showed $W^\perp = \text{span}(\{w_{n+1}, \ldots, w_r\})$.

Consider $\{w_1, \ldots, w_r\}$ to be a result of orthonormal extension of $\{w_{n+1}, \ldots, w_r\}$. Again in the proof of Theorem 9, we see $(W^\perp)^\perp = \text{span}(\{w_1, \ldots, w_n\}) = W$. ■

The answer of goal problem in infinite dimensional space is easy to state but difficult to prove for now, so we only state the answer with strict definition and without proof.

Theorem 11. ***(The answer of goal problem in infinite dimensional vector space)*** *Let W be a subspace of a in infinite dimensional vector space V.*

$$\overline{W} = (W^\perp)^\perp$$

The **topology** τ of V is defined by $\tau = \{S \subseteq V | \forall x \in S, \exists \varepsilon \in \mathbb{R}^+, \|y - x\| < \varepsilon \implies y \in S\}$. One can check that this form a topology. $\overline{W}$ is the **closure** of W. **(The closure of W is the intersection of all closed sets containing W)**

Basic Notion about Adjoint

Lemma 12. ***(the complex conjugation of a linear transformation from a finite dimensional space to a finite dimensional space does not depend on the choice of orthonormal bases)*** *Let $T : V \to W$ be a linear transformation where both V and W are finite dimensional. The linear transformation $T^H : W \to V$ defined by $[T^H]_\beta^\alpha = ([T]_\alpha^\beta)^*$, where α, β are all orthonormal bases, is the same linear transformation $T^{H'} : W \to V$ defined by $[T^{H'}]_{\beta'}^{\alpha'} = ([T]_{\alpha'}^{\beta'})^*$*

Proof. We only have to prove $[T^{H'}]_\beta^\alpha = [T^H]_\beta^\alpha$, then we immediately have $T^{H'} = T^H$.

We first use the identity of subtracting two coordinates-changing matrix on the side of $[T^{H'}]_\beta^\alpha$ and use the identity of double complex conjugation.

$$[T^{H'}]_\beta^\alpha = [I_V]_{\alpha'}^\alpha [T^{H'}]_{\beta'}^{\alpha'} [I_W]_\beta^{\beta'} = [I_V]_{\alpha'}^\alpha ([T]_{\alpha'}^{\beta'})^* [I_W]_\beta^{\beta'}$$
$$= ([I_V]_{\alpha'}^\alpha ([T]_{\alpha'}^{\beta'})^* [I_W]_\beta^{\beta'})^{**} = (([I_W]_\beta^{\beta'})^* [T]_{\alpha'}^{\beta'} ([I_V]_{\alpha'}^\alpha)^*)^* \tag{5}$$

We now show that the complex conjugation of a orthonormal-coordinate-changing square matrix is its inverse, by directly computing the product of a orthonormal-coordinate-changing square matrix and its complex conjugation.

Observe that

$$(([I_V]_{\alpha'}^\alpha)^* [I_V]_{\alpha'}^\alpha)_{i,j} = \langle \alpha_j, \alpha_i \rangle_{\mathbb{C}^n} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Notice that the only complex square matrix that have the property of

$$M_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

is the identity matrix.

Get back to equation (5), we have

$$[T^{H'}]_\beta^\alpha = ((([I_W]_\beta^{\beta'})^{-1} [T]_{\alpha'}^{\beta'} ([I_V]_{\alpha'}^\alpha)^{-1})^*$$
$$= ([I_W]_{\beta'}^\beta [T]_{\alpha'}^{\beta'} [I_V]_\alpha^{\alpha'})^* = ([T]_\alpha^\beta)^* = [T^H]_\beta^\alpha \text{ (notice the definition of } T^H)$$

(done) ■

Definition 7. *The **adjoint** $T^H : W \to V$ of a linear transformation $T : V \to W$ where both V and W are finite dimensional is given by $[T^H]_\beta^\alpha = ([T]_\alpha^\beta)^*$, where α and β are respectively orthonormal basis of V and W. Equivalently, one can say that T^H is the unique linear transformation that satisfy the beautiful property*

$$\forall \textbf{\textit{(orthonormal bases)}} (\alpha, \beta), [T^H]_\beta^\alpha = ([T]_\alpha^\beta)^*$$

Notice by Lemma 12, the choice of α and β doesn't matter. If we choose different orthonormal bases α', β' to construct the adjoint , we will still have the same T^H.

Lemma 13. *The slightly stronger property*

$$\forall\textbf{\textit{(orthogonal bases)}}(\alpha, \beta), [T^H]_\beta^\alpha = ([T]_\alpha^\beta)^*$$

is generally false. One must be careful when using the property given in Definition 7.

Proof. We simply raise a counterexample.

Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by lifting $T(\begin{bmatrix}1\\0\end{bmatrix}) = \begin{bmatrix}1\\2\\3\end{bmatrix}$ and $T(\begin{bmatrix}0\\1\end{bmatrix}) = \begin{bmatrix}4\\4\\4\end{bmatrix}$

Let α, β respectively be the standard ordered basis for $\mathbb{R}^2$ and $\mathbb{R}^3$. Let $\alpha' = (\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\2\end{bmatrix})$. Check that α' is orthogonal and we continue with our proof. From

$$[T]_\alpha^\beta = \begin{bmatrix}1 & 4\\2 & 4\\3 & 4\end{bmatrix}$$

We give

$$[T^H]_\beta^\alpha = ([T]_\alpha^\beta)^* = \begin{bmatrix}1 & 2 & 3\\4 & 4 & 4\end{bmatrix}$$

Now we express T and T^H with bases α' and β and see

$$[T^H]_\beta^{\alpha'} = \begin{bmatrix}1 & 2 & 3\\2 & 2 & 2\end{bmatrix} \neq \begin{bmatrix}1 & 8\\2 & 8\\3 & 8\end{bmatrix}^* = ([T]_{\alpha'}^\beta)^*$$

(done) ■

The next theorem, even thought taking place on finite dimensional vector space, enable us to generalize our notion of **adjoint** and **complex conjugation** to infinite dimensional vector space.

Theorem 14. *Let V, W be two finite dimensional vector space and $T : V \to W$ be a linear transformation. The following holds*

$$\langle T(x), y\rangle_W = \langle x, T^H(y)\rangle_V$$

for all $x \in V$ and $y \in W$

Proof. Let α and β respectively be orthonormal basis for V and W, and rewrite $\langle T(x), y\rangle_W$ and $\langle x, T^H(y)\rangle_V$ to

$$\langle T(x), y\rangle_W = \langle [T(x)]_\beta, [y]_\beta\rangle_{\mathbb{C}^m} = \langle [T]_\alpha^\beta [x]_\alpha, [y]_\beta\rangle_{\mathbb{C}^m}$$

$$\langle x, T^H(y)\rangle_V = \langle [x]_\alpha, [T^H(y)]_\beta\rangle_{\mathbb{C}^n} = \langle [x]_\alpha, [T^H]_\beta^\alpha [y]_\beta\rangle_{\mathbb{C}^n}$$

Use the symbol $[x]_\alpha = x \in \mathbb{C}^n$ and $[y]_\beta = y \in \mathbb{C}^m$ and $[T]_\alpha^\beta = A$, so we can reduce the theorem to

$$\langle Ax, y\rangle_{\mathbb{C}^m} = \langle x, A^* y\rangle_{\mathbb{C}^n} \tag{6}$$

To prove equation (6), we simply compute

$$\langle Ax, y\rangle_{\mathbb{C}^m} = \sum_{j=1}^{m}(Ax)_j\overline{y_j} = \sum_{j=1}^{m}(\sum_{i=1}^{n} A_{j,i}x_i)\overline{y_j}$$

$$\langle x, A^* y\rangle_{\mathbb{C}^n} = \sum_{i=1}^{n} x_i\overline{(A^*y)_i} = \sum_{i=1}^{n} x_i\overline{\sum_{j=1}^{m} A^*_{i,j}y_j} = \sum_{i=1,j=1}^{n,m} A_{j,i}x_i\overline{y_j} = \langle Ax, y\rangle_{\mathbb{C}^m}$$

and see the result. ■

Now we prove the "inverse" of Theorem 14, then we will have an equivalent definition of **adjoint (See Definition** 7**)**.

Theorem 15. *Let V, W be two finite dimensional vector space and $T : V \to W$ be a linear transformation. We have*

$$\forall (x,y) \in (V,W), \langle T(x), y\rangle_W = \langle x, T'(y)\rangle_V \implies T' = T^H$$

Proof. Let α and β respectively be orthonormal basis for V and W. This time, we rewrite $\langle T(x), y\rangle_W$ and $\langle x, T'(y)\rangle_V$ to **(Recall that till now, we are still not sure if $T' = T^H$ hold true)**

$$\langle T(x), y\rangle_W = \langle [T(x)]_\beta, [y]_\beta\rangle_{\mathbb{C}^m} = \langle [T]_\alpha^\beta [x]_\alpha, [y]_\beta\rangle_{\mathbb{C}^m}$$

$$\langle x, T'(y)\rangle_V = \langle [x]_\alpha, [T'(y)]_\beta\rangle_{\mathbb{C}^n} = \langle [x]_\alpha, [T']_\beta^\alpha [y]_\beta\rangle_{\mathbb{C}^n}$$

Notice that when $[x]_\alpha = e_i$ and $[y]_\beta = e_j$, we have

$$([T]_\alpha^\beta)_{j,i} = \langle T(x), y\rangle_W = \langle x, T(y)\rangle_V = \overline{([T']_\beta^\alpha)_{i,j}}$$

This tell us that $[T']_\beta^\alpha = ([T]_\alpha^\beta)^*$, which enable us to easily deduce $T' = T^H$ ■

Corollary 15.1. *Let V, W be two finite dimensional vector space and $T : V \to W$ be a linear transformation.*

$$Q \textit{ is the } \textbf{\textit{adjoint}} \textit{ of } T \iff \forall (x,y) \in (V,W), \langle T(x), y\rangle_W = \langle x, Q(y)\rangle_V$$

Say, we want to naturally generalize the notion of **adjoint** to infinite dimensional vector space. Recall the definition of adjoint in finite dimensional space given by Definition 7. Immediately we see that we can not duplicate Definition 7 to the case of infinite dimensional vector space. However, we can duplicate Corollary 15.1 to be the definition of **adjoint**, or, a weird name, **complex comjugation**, in the case of infinite dimensional vector space as below.

Definition 8. *Let V, W be two inner product space where at least one of them is infinite dimensional. Let $T : V \to W$ be a linear transformation. Q is the* ***adjoint*** *T^H of T if $\forall(x, y) \in (V, W), \langle T(x), y\rangle_W = \langle x, Q(y)\rangle_V$.*

Although at here we can not use elementary method to raise a counterexample to show that not all linear transformation on infinite dimensional vector space have an adjoint, we encourage the reader to convince himself this is true by remembering that not all infinite dimensional space have orthornormal basis, which lead to an intuitive and non-rigorous argument that some linear operator on it don't have *complex conjugation*.

However, one thing we can prove using elementary method is that if a linear operator T, in both cases of finite and infinite dimensional space, have an adjoint T^H, then that adjoint T^H is unique.

Lemma 16. ***(adjoint is always unique)*** *Let V, W be two inner product space, and $T : V \to W$ be a linear transformation.*

If Q and Q' are both adjoint of T, then $Q = Q'$

Proof. Notice that

$$\forall(x, y) \in (V, W), \langle x, Q(y) - Q'(y)\rangle_V = \langle x, Q(y)\rangle_V - \langle x, Q'(y)\rangle_V$$
$$= \langle T(x), y\rangle_W - \langle T(x), y\rangle_W = 0$$

So we can see $\forall y \in W, \|Q(y) - Q'(y)\|_V = 0$, if we let $x = Q(y) - Q'(y)$

This tell us $\forall y \in W, Q(y) = Q'(y)$

■

Chapter: SVD

For some square matrix A, it is possible for one to diagonalize the matrix A in the form of $A = PDP^{-1}$ to simplify the computation of certain problems. Yet, it is easy for any undergraduates to see that only a small amount of matrices can be diagonalized. The "harsh" requirement for a matrix to be diagonalizable force us to come up with a way of decomposing matrix similar to diagonalization with looser requirement for matrix A, if we ever want to expand the concept of diagonalization into the much complicated real world. Now, we introduce one decomposition for every complex matrix, that will give us an extremely neat way to compute in only positive real number, which is extremely surprising, consider that this decomposition is for every complex matrix. This decomposition is called "singular value decomposition", which we often just call "SVD".

Definition 9. ***(Theorem of SVD in matrix from)*** *Suppose $A \in M_{m\times n}(\mathbb{C})$. An SVD of A is a expression*

$$A = U \sum V^*$$

where U is an unitary $m \times m$ complex matrix, V is an unitary $n \times n$ complex matrix, and $\sum$ satisfy

$$\sum_{i,j} = \begin{cases} \sigma_i \in \mathbb{R}_0^+ & i = j \\ 0 & i \neq j \end{cases}$$

Here, we look deep inside the SVD and provide a proof that we can "SVD" all $A \in M_{m\times n}(\mathbb{C})$

Lemma 17. *Notice A^*A and AA^* are square matrix of the same size.*

$$A^*A \text{ and } AA^* \text{ share the same eigenvalue}$$

Proof. Suppose A^*A have eigenvalue λ, we show AA^* also have eigenvalue λ.

Suppose $A^*Av = \lambda v$. Observe

$$AA^*(Av) = A(A^*Av) = A\lambda v = \lambda(Av) \tag{7}$$

■

Lemma 18. *The eigenvalues of a self-adjoint matrix M are real.*

Proof. Suppose $Mx = \lambda x$, we wish to show $\lambda = \overline{\lambda}$.

First observe

$$\lambda\|x\| = \lambda x^*x = x^*\lambda x = x^*Mx \tag{8}$$

And then observe

$$\lambda\|x\| = x^*Mx = (\overline{x})^t(Mx) = (Mx)^t\overline{x} = x^tM^t\overline{x} \tag{9}$$

conjugate both side of the equation and we see

$$\overline{\lambda}\|x\| = \overline{x}^t\overline{M}^t x = x^*M^*x = x^*Mx = \lambda\|x\| \tag{10}$$

Because $x \neq 0$, We have shown $\overline{\lambda} = \lambda$ ■

Corollary 18.1. *A^*A and AA^* are orthonormally diagonalizable, and the eigenvalue of A^*A and AA^* are real*

Proof. Notice A^*A and AA^* are self-adjoint, so by Lemma 18, this is true. ■

Definition 10. *A self-adjoint matrix M is **positive semi-definite** if $x^*Mx \in \mathbb{R}_0^+, \forall x \in \mathbb{C}^n$*

Lemma 19. *Suppose M is self-adjoint. Then*

M is positive-definite if and only if all eigenvalues of M is a non-negative real number.

Proof. We first express M in the form

$$M = PDP^{-1} \tag{11}$$

Where P comprise an orthonormal eigenbasis

Arbitrarily pick $x \in \mathbb{C}^n$ and suppose $y = Px$

Observe

$$x^*Mx \in \mathbb{R}_0^+ \iff x^*PDP^{-1}x \in \mathbb{R}_0^+ \iff y^*Dy \in \mathbb{R}_0^+ \tag{12}$$

$(\longrightarrow)$

By specifying y to $\begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}$ and such, we see all diagonal entries of D are non-negative.

Notice that the diagonal entries of D are exactly the eigenvalues of M, and our proof is done.

$(\longleftarrow)$

Because D contain only non-negative diagonal entry we know that $\forall y \in \mathbb{C}^n, y^*Dy \in \mathbb{R}_0^+$

■

Lemma 20. *Suppose M is self-adjoint. Then*

M is positive semi-definite if and only if $M = B^*B$ for some matrix B

Proof. $(\longleftarrow)$

$$x^*Mx = (x^*B^*)(Bx) = (Bx)^*(Bx) = \|Bx\| \in \mathbb{R}_0^+ \tag{13}$$

$(\longrightarrow)$

We first express M in the form

$$M = PDP^{-1} \tag{14}$$

where D is a diagonal matrix containing only non-negative real entries, and P contain orthonormal eigenbasis.

Because P contain orthonormal basis, we know $P^{-1} = P^*$, and because D contains only non-nagative real entries, the diagonal matrix $\sqrt{D}$ defined by

$$(\sqrt{D})_{i,i} = \sqrt{D_{i,i}} \tag{15}$$

exists and it satisfy

$$(\sqrt{D})^* = \sqrt{D} \tag{16}$$

Then we can express M in the form

$$M = P\sqrt{D}(\sqrt{D})^*P^* = (P\sqrt{D})(P\sqrt{D})^* \tag{17}$$

■

Corollary 20.1. *The eigenvalues of A^*A and AA^* are non-negative*

Proof. Notice that A^*A and AA^* are positive definite, by Lemma 20. ■

Lemma 21.

$$N(A^*A) = N(A), \textit{ and "mirrorticaly"}, N(AA^*) = N(A^*) \tag{18}$$

Proof.

$$Ax = 0 \implies A^*Ax = 0 \tag{19}$$

$$A^*Ax = 0 \implies x^*A^*Ax = 0 \implies \|Ax\| = 0 \implies Ax = 0 \tag{20}$$

■

Corollary 21.1.

$$rank(A^*A) = rank(A) = rank(AA^*) \tag{21}$$

Proof. Notice that

$$rank(A) = rank(A^*) \tag{22}$$

■

Lemma 22. *Suppose $\{v_1, \ldots, v_n\}$ is the orthonormal eigenbasis corresponding to eigenvalues $\{\sigma_1^2, \ldots, \sigma_n^2\}$ for A^*A. Then we have*

$$\|Av_i\| = \sigma_i \tag{23}$$

Proof.

$$\|Av_i\| = \sqrt{v_i^*A^*Av_i} = \sqrt{v_i^*(A^*Av_i)} = \sqrt{\sigma_i^2 v_i^* v_i} = \sigma_i\|v_i\| = \sigma_i \tag{24}$$

■

Lemma 23. *Rearrange the two corresponding sets $\{v_1, \ldots, v_n\}$, $\{\sigma_1^2, \ldots, \sigma_n^2\}$ so that $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_n^2$ and suppose $\sigma_{r-1} > \sigma_r = 0$. Define $u_1, \ldots, u_r$ by*

$$u_i = \frac{Av_i}{\sigma_i} \tag{25}$$

Then

$$\{u_1, \ldots, u_r\} \textit{ are orthonormal.}$$

Proof. Observe

$$\langle u_i, u_j \rangle = u_i^* u_j = \frac{1}{\sigma_i \sigma_j} v_i^* A^* A v_j = \frac{\sigma_j}{\sigma_i} \langle v_i, v_j \rangle \tag{26}$$

Then

$$i \neq j \implies \langle u_i, u_j \rangle = 0 \text{ and } \langle u_i, u_i \rangle = \frac{\sigma_i}{\sigma_i} \langle v_i, v_i \rangle = 1 \tag{27}$$

■

With all the small Lemma above, now we are ready to prove that SVD exists for all matrix $A \in M_{m\times n}(\mathbb{C})$, with almost no effort.

Theorem 24. ***(SVD exists for all matrices)*** *Suppose $A \in M_{m\times n}(\mathbb{C})$. There exists some complex unitary $m \times m$ matrix U and complex unitary $n \times n$ matrix V such that*

$$A = U \sum V^*$$

where $\sum$ are diagonal matrix with only non-negative real entries

Proof. Let V comprise orthonormal eigenbasis $\{v_1, \dots, v_n\}$ for A^*A, $\sum$ be defined by

$$\textstyle\sum_{i,i} = \sigma_i$$

where σ_i^2 is the eigenvalue corresponding to v_i, and expand the orthonormal set $\{u_1, \dots, u_r\}$ from Lemma 23, to an orthonormal basis $\{u_1, \dots, u_m\}$ to be the columns of U.

We only have to show that U and V are unitary and $A = U \sum V^*$

To see that U and V are unitary, observe they each comprise orthonoral basis, so $U^*U = I = V^*V$. (done)

Because $\{v_1, \dots, v_n\}$ is a basis for $\mathbb{C}^n$ **(Notice that** $A : \mathbb{C}^n \to \mathbb{C}^m$ **and** $A^*A : \mathbb{C}^n \to \mathbb{C}^n$**)**, we only have to show that $Av_i = U \sum V^* v_i$

$$U \sum V^* v_i = U \sum e_i = U\sigma_i e_i = \sigma_i U e_i = \sigma_i u_i = \begin{cases} \sigma_i \frac{Av_i}{\sigma_i} = Av_i & \sigma_i > 0 \\ 0 & \sigma_i = 0 \end{cases} \tag{28}$$

Notice that by Lemma 21, we know $N(A^*A) = N(A)$ so if $\sigma_i = 0$, then $Av_i = 0$, which complete our proof. (done)

■

We now introduce **polar decomposition**.

Definition 11. *To polar decompose a square matrix A is to express A in the form*

$$A = WP \tag{29}$$

where U is a square unitary matrix and P is positive semi-definite

Theorem 25. ***(Polar decomposition exists for all square matrices)*** *Suppose $A \in M_{n\times n}(\mathbb{C})$. There exists some complex unitary $n \times n$ matrix U and complex positive semi-definite matrix P such that*

$$A = WP \tag{30}$$

Proof. We first SVD A to

$$A = U \sum V^* \tag{31}$$

Because $V^*V = I$, we can express A in the form

$$A = (UV^*)(V \sum V^*) \tag{32}$$

We only have to show that UV^* is unitary and $V \sum V^*$ is positive semi-definite.

Observe

$$(UV^*)^* UV^* = VU^*UV^* = I \text{ (done)} \tag{33}$$

and observe

$$x^* V \sum V^* x = \langle \sum V^* x, V^* x\rangle = \sum \sigma_i (V^* x)_i^2 \geq 0 \text{ (done)} \tag{34}$$

■

In most case of building of math theories, the language of abstraction often come afterward specific background of problem solving, serving as the easiest way to solve all problems of the same kind once and for all. To solve a linear equation, one can express the linear equation in the form of $y = Ax$, and if A is invertible, then we can easily solve for x with $x = A^{-1}Ax = A^{-1}y$, but what if A is not invertible? We here introduce the concept of **Pseudo Inverse**.

Definition 12. *Suppose $A \in M_{m\times n}(\mathbb{C})$. The **Pseudo Inverse** of A is a matrix A^+ that satisfy four requirement*

$$AA^+A = A \tag{35}$$

$$A^+AA^+ = A^+ \tag{36}$$

$$(AA^+)^* = AA^+ \tag{37}$$

$$(A^+A)^* = A^+A \tag{38}$$

Notice that due to the third requirement, A^+ must be of the size $n \times m$, which is compatible the the fourth requirement.

Theorem 26. *For all $A = U \sum V^* \in M_{m\times n}(\mathbb{C})$, there exists unique pseudo inverse A^+ that is given by*

$$A^+ = V \sum^+ U^*$$

Where $\sum^+ \in M_{n\times m}(\mathbb{R}_0^+)$ *is given by*

$$\sum^+ = \begin{cases} \frac{1}{\sum_{i,j}} & i = j \\ 0 & i \neq j \end{cases}$$

Proof. We first show that the pseudo inverse given above is correctly defined, and then we will show that although SVD is not unique, the pseudo inverse is unique.

$$\sum \sum^+ \sum = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \sum = \sum$$

$$\sum^+ \sum \sum^+ = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \sum^+ = \sum^+$$

Give close attention to

$$\sum \sum^+ = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \in M_{m\times m}(\mathbb{R}) \text{ and } \sum^+ \sum = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \in M_{n\times n}(\mathbb{R})$$

So

$$\sum \sum^+ = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \neq \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} = \sum^+ \sum$$

From our computation, both $\sum \sum^+$ and $\sum^+ \sum$ are self-adjoint, so $\sum^+$ is indeed an pseudo inverse of $\sum$.

$$AA^+A = (U \sum V^*)(V \sum^+ U^*)(U \sum V^*)$$

$$= U \sum \sum^+ \sum V^* = U \sum V^* = A$$

$$A^+AA^+ = (V \sum^+ U^*)(U \sum V^*)(V \sum^+ U^*)$$

$$= V \sum^+ \sum \sum^+ U^* = V \sum^+ U^* = A^+$$

$$A^+A = (V \sum^+ U^*)(U \sum V^*) = V \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} V^*$$

$$(A^+A)^* = (V \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} V^*)^* = V \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} V^* = A^+A \tag{39}$$

$$AA^+ = (U \sum V^*)(V \sum^+ U^*) = U \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} U^*$$

$$(AA^+)^* = (U \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} U^*)^* = U \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} U^* = AA^+ \text{ (done)} \tag{40}$$

Suppose X, Y are both pseudo inverse of A. Observe

$$X = XAX = X(AX)^* = XX^*A^* = XX^*(AYA)^* \tag{41}$$

$$= XX^*((AY)A)^* = XX^*A^*(AY)^* = XX^*A^*AY \tag{42}$$

$$X(AX)^*AY = XAXAY = XAY = XAYAY = XA(YA)^*Y \tag{43}$$

$$= XAA^*Y^*Y = (XA)^*A^*Y^*Y = (A(XA))^*Y^*Y \tag{44}$$

$$= (AXA)^*Y^*Y = A^*Y^*Y = (YA)^*Y = YAY = Y \text{ (done)} \tag{45}$$

■