Understand linear transform (2)

Yinan Huang

Oct 18, 2020

Contents

1 Introduction

This note will cover topics about linear transform (especially operators) in a more formal way. Specifically, we will first talk about linear transform, matrices, invertibility and isomorphism. Then we will go over eigenvalues and eigenvectors, and finally talk about spectral theorem and several decomposition methods.

2 Linear transform

Definition 2.1. *A linear transform $T : V \to W$, denoted as $T \in L(V, W)$, is a function that meets the following:*

- *For all $v, w \in V$, $T(\boldsymbol{v} + \boldsymbol{w}) = T\boldsymbol{v} + T\boldsymbol{w}$.*
- *For all $v \in V$ and $s \in F$, $T(s\boldsymbol{v}) = sT(\boldsymbol{v})$.*

Linear transform is the most significant element when we study linear algebra. It has a nice property that it is determined by its transform on basis.

Theorem 2.1. *Let $T \in L(V, W)$ and $\{\boldsymbol{v}_i\}$ is basis of V. Then $\{\boldsymbol{w_i}\}$ in*

$$T\boldsymbol{v_i} = \boldsymbol{w_i} \tag{1}$$

specifically determines T.

Proof 2.1. *For any $\boldsymbol{v} = \sum_i c_i \boldsymbol{v_i}$, we have*

$$T\boldsymbol{v} = \sum_i c_i T\boldsymbol{v_i} = \sum_i c_i \boldsymbol{w_i}. \tag{2}$$

So the effect of T acting on any v is well-defined by vectors $\boldsymbol{w_i}$. Based on this we can see that $T\boldsymbol{0} = \boldsymbol{0}$.

2.1 Null space, range space and invertibility

Definition 2.2. *Let $T \in L(V, W)$. Null space of T is defined as $N(T) \equiv \{v \in V | Tv = 0\}$. Range space of T is defined as $R(T) \equiv \{w \in W | \exists v \in V, Tv = w\}$.*

It is easily to prove that $R(T)$ and $N(T)$ are subspace. The next thorem tells us that $N(T)$ and $R(T)$ are closely related with invertiblity.

Theorem 2.2. *Let $T \in L(V, W)$. T is* ***injective*** *if and only if $N(T) = \{\boldsymbol{0}\}$. T is* ***surjective*** *if and only if $R(T) = W$.*

Proof 2.2. *We begin with the first statement. Suppose T is injective, then for any $\boldsymbol{v}, \boldsymbol{w} \in V$, $T\boldsymbol{v} = T\boldsymbol{w}$ implies $\boldsymbol{v} = \boldsymbol{w}$. Consider $\boldsymbol{u} \in N(T)$, then since $T(\boldsymbol{u}) = T(\boldsymbol{u} - \boldsymbol{0}) = 0$, we have $\boldsymbol{u} = \boldsymbol{0}$. So $N(T) = \{\boldsymbol{0}\}$. Suppose $N(T) = \{\boldsymbol{0}\}$, then for all $\boldsymbol{v}, \boldsymbol{w}$ such that $T(\boldsymbol{v} - \boldsymbol{w}) = 0$, we have $\boldsymbol{v} = \boldsymbol{w}$, which says T is injective.*

Then we try proving the second statement. Suppose T is surjective, then for all $w \in W$, we can find a $v \in V$ such that $Tv = w$. This is essentially equivalent to $R(T) = W$.

From the knowledge of function, we can conclude the following thorem.

Theorem 2.3. *Let $T \in L(V, W)$. T is invertible if and only if $N(T) = \{\boldsymbol{0}\}$ and $R(T) = W$, or equivalently,* $\dim\{R(T)\} = \dim\{W\} = \dim\{V\}$. *The latter one says only when* $\dim\{V\} = \dim\{W\}$ *can we find invertible transform.*

Proof 2.3. *Since T is invertible when T is surjective and injective, so the former one is proved. Then by applying the fundamental thorem of linear transform,* $\dim\{N(T)\} + \dim\{R(T)\} = \dim\{V\}$*, we can show that* $\dim\{R(T)\} = \dim\{W\} = \dim\{V\}$ *if and only if T is invertible.*

So to discuss invertibility, we restrict ourself to $T \in L(V, V)$, or $T \in L(V)$ for short. A linear transform mapping from V to V itself has a name.

Definition 2.3. *Let $T \in L(V, W)$. T is called an **operator** if $V = W$, denoted as $T \in L(V)$.*

Theorem 2.4. *Let $T \in L(V)$. The following statements are equivalent:*

- *T is invertible.*
- *T is surjective.*
- *T is injective.*

Proof 2.4. *Suppose T is surjective, then* $\dim\{R(T)\} = \dim\{V\}$*, which is equivalent to* $\dim\{N(T)\} = 0$ *and thus T is injective and invertible.*

2.2 Isomorphic vector spaces

If we say two spaces are isomorphic, we means they have the same structure though their elements may be totally distinct.

Definition 2.4. *Let V, W be vector spaces. V and W are said to be **isomorphic** if there is a **invertible** linear transform $T \in L(V, W)$. T is called **isomorphism** of V and W.*

Theorem 2.5. *V and W are isomorphic if and only if* $\dim\{V\} = \dim\{W\}$.

Proof 2.5. *Invertible linear transform exists if and only if* $\dim\{V\} = \dim\{W\}$*. So if isomorphism exists, then* $\dim\{V\} = \dim\{W\}$.

Basically, if V and W are isomorphic, then every property of space V can be directly applied to W. The power of isomorphism will be seen in the next subsection where we introduce matrix.

2.3 Matrix representation

Definition 2.5. *For vector space V and W, given the basis $\{\boldsymbol{v_1}, \boldsymbol{v_2}, ..., \boldsymbol{v_n}\} \subseteq V$ and $\{\boldsymbol{w}_1, \boldsymbol{w_2}, ..., \boldsymbol{w_m}\} \subseteq W$, **matrix representation** of linear transform is a function $M : L(V, W) \to F^{n,m}$, defined by*

$$T\boldsymbol{v}_i = \sum_j [M(T)]_{j,i}\, \boldsymbol{w}_i. \tag{3}$$

Theorem 2.6. *Matrix representation M is an **isomorphism** from $L(V, W)$ to $F^{n,m}$.*

Proof 2.6. *To prove M is an isomorphism, we need to prove it is a invertible linear transform. linearity is obvious, and since the coefficient $[M(T)]$ specifically determines $T\boldsymbol{v_i}$, and thus determines T, so M is a bijective mapping. Therefore M is an isomorphism.*

Thanks to this isomorphism, every properties applied to matrix can be found in linear transform. For examples,

- For $T \in L(X, Y)$ and $U \in L(Z, X)$, the matrix of TU is equal to the matrix of T multilpies matrix of U: $M(TU) = M(T)M(U)$.
- $\dim\{L(V, W)\} = \dim\{F^{n,m}\} = nm = \dim\{V\} \dim\{W\}$.

For vectors, we can also define a matrix representation of V given basis $\{\boldsymbol{v}_1, \boldsymbol{v_2}, ..., \boldsymbol{v_n}\}$.

Definition 2.6. *For vector space V given basis $\{\boldsymbol{v_1}, \boldsymbol{v_2}, ...\boldsymbol{v_n}\}$, **matrix representation** of vectors is a function $M : V \to F^{1,n}$ defined by*

$$\boldsymbol{v} = \sum_i M(\boldsymbol{v})_i \boldsymbol{v}_i. \tag{4}$$

Theorem 2.7. *Let $T \in L(V, W)$ and $\boldsymbol{v} \in V$. Then $M(T\boldsymbol{v}) = M(T)M(\boldsymbol{v})$ given basis $\{\boldsymbol{v_i}\}$.*

Proof 2.7. *Direct proof:*

$$T\boldsymbol{v} = \sum_i M(\boldsymbol{v})_i T\boldsymbol{v}_i = \sum_{i,j} M(\boldsymbol{v})_i \left[M(T)\right]_{j,i} \boldsymbol{w_j} = \sum_j \left[M(T\boldsymbol{v})\right]_j \boldsymbol{w}_j. \tag{5}$$

Thus $M(T)M(\boldsymbol{v}) = M(T\boldsymbol{v})$.

So when we discuss about linear transform, we can always think of matrix, prove things in matrix form and directly apply it to linear transform.

2.4 Duality

2.4.1 Dual space

Definition 2.7. *Let V be vector space on F, and linear transform $T \in L(V, F)$ is called **linear functional**.*

Definition 2.8. *Let V be vector space on F. Then linear functional space $L(V, F)$ is called **dual space** of V, denoted as V'.*

Theorem 2.8. $\dim\{V\} = \dim\{V'\}$.

Proof 2.8. *Note that* $\dim\{L(V, F)\} = \dim\{V\} \dim\{F\} = \dim\{V\}$.

Though dual space sounds pretty abstract, but they have simple relation. The matrix representation of V is column matrix, and the matrix representation of $V' = L(V, F)$ is simply row matrix.

Definition 2.9. *For basis $\{\boldsymbol{v}_i\} \subseteq V$, **dual basis** $\{\boldsymbol{v_i'}\} \subseteq V'$ is defined by*

$$\boldsymbol{v}_i'(\boldsymbol{v}_j) = \delta_{i,j}. \tag{6}$$

*And we can define **dual vector** $\boldsymbol{v}'$ of $\boldsymbol{v} = \sum_i c_i \boldsymbol{v_i}$ by*

$$\boldsymbol{v}' = \sum_i c_i^* \boldsymbol{v}_i'. \tag{7}$$

In quantum mechanics, the ket vectors $|\psi\rangle$ and bra vectors $\langle\psi|$ are simply dual vectors.

Theorem 2.9. *Dual basis are basis of dual space.*

Proof 2.9. *Since the number of dual basis is equal to* $\dim\{V'\}$*, we are left to prove its linear indepndence. We first suppose*

$$\sum_i c_i \boldsymbol{v_i'} = 0 \tag{8}$$

with at least one $c_i \neq 0$. Then let both sides act on $\boldsymbol{v_j}$:

$$\sum_i c_i \boldsymbol{v_i'}(\boldsymbol{v_j}) = c_j = 0, \tag{9}$$

which is a contradiction. So $c_i = 0$ for all i, and $\{\boldsymbol{v_i'}\}$ is linearly independent and thus a basis of V'.

2.4.2 Dual linear transform

We can also define the dual linear transform T' of $T \in L(V, W)$.

Definition 2.10. *Let $T \in L(V, W)$. $T' \in L(W', V')$ is called the **dual linear transform** of T defined by*

$$T'(\boldsymbol{w}') \equiv \boldsymbol{w}' \circ T. \tag{10}$$

Theorem 2.10. *The matrix representation of $T' \in L(W', V')$ is the transpose of $M(T)$.*

Proof 2.10. *Let matrix of T and T' be A and B. Then by definition,*

$$T'(\boldsymbol{w}_i') = \sum_j B_{j,i} \boldsymbol{v}_j'. \tag{11}$$

Acting on $\boldsymbol{v}_k$ on both sides, we have

$$\begin{aligned} \boldsymbol{w}_{\boldsymbol{i}}' T(\boldsymbol{v}_k) &= \sum_j B_{j,i} \boldsymbol{v}_{\boldsymbol{j}}'(\boldsymbol{v}_{\boldsymbol{k}}) = B_{k,i} \\ &= \boldsymbol{w}_{\boldsymbol{i}}' \left(\sum_j A_{j,k} \boldsymbol{w}_{\boldsymbol{j}} \right) = A_{i,k}. \end{aligned} \tag{12}$$

Therefore $A^T = B$, finishing the proof.

Theorem 2.11. $\dim\{R(T)\} = \dim\{R(T')\}$.

Definition 2.11. *See "Linear algebra done right".*

The duality gives us alternative way to think about transpose: transpose of vectors and matrices are the corresponding dual vectors and dual transform.

2.5 Rank

Definition 2.12. ***Column rank** of matrix M is the dimension of space its column vectors span. Similarly, **row rank** of matrix M is the dimension of space its row vectors span.*

Theorem 2.12. *The column rank of $M(T)$ is equal to* $\dim\{R(T)\}$.

Proof 2.11. *Range space $R(T) = span\{T\boldsymbol{v}_i\}$, and* $\dim\{span\{T\boldsymbol{v}_{\boldsymbol{i}}\} = \dim\{span\{M(T\boldsymbol{v}_{\boldsymbol{i}})\}$, *for which the latter is the column rank.*

Theorem 2.13. *Row rank is equal to column rank.*

Proof 2.12. *Note that Column rank of M(T)* $= \dim\{R(T)\} = \dim\{R(T')\} =$ *Column rank of $M(T') =$ Row rank of$M(T)$.*

Definition 2.13. ***Rank** of a matrix $A \in F^{n,m}$ is the column rank (or row rank) of A.*

3 Eigenvectors and eigenvalues

Definition 3.1. *Let $T \in L(V)$ be an operator. Then $\boldsymbol{v} \in V$ and $\lambda \in F$ is called the **eigenvector** and correpsonding **eigenvalue**, if*

$$Tv = \lambda v. \tag{13}$$

We are frequently encountered with eigenvectors and eigenvalues, which characterizes an operator's property.

3.1 Invariant subspaces

We are going to see that eigenvectors with the same eigenvalues span an invariant subspace of T.

Definition 3.2. *Let V be vector space and $U \subseteq V$ be subspace. If for each $\boldsymbol{u} \in U$,*

$$T\boldsymbol{u} \in U, \tag{14}$$

*then U is called the **invariant subspace** of T.*

Theorem 3.1. *$U \equiv \{\sum_i c_i \boldsymbol{v_i} | T\boldsymbol{v_i} = \lambda \boldsymbol{v_i}\}$ is an invariant subspace of T. In other word, the spance eigenvectors with the same eigenvalue span is a invariant subspace.*

Proof 3.1. *Note that $T\sum_i c_i \boldsymbol{v}_i = \lambda \sum_i c_i \boldsymbol{v}_i$, which finishes the proof.*

We call invariant space spanned by eigenvectors **eigenspace**. Eigenvectors with different eigenvalue turn out to be linearly independent.

Theorem 3.2. *Let $\{\boldsymbol{v_i}\}$ be eigenvectors with different eigenvalues λ_i, namely $T\boldsymbol{v_i} = \lambda_i \boldsymbol{v_i}$ with $\lambda_i \neq \lambda_j$ for all $i \neq j$. Then $\{\boldsymbol{v_i}\}$ are linearly independent.*

Proof 3.2. *Suppose $\boldsymbol{v_k} \in$ span$\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v_{k-1}}\}$, and $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v_{k-1}}\}$ are linearly independent, then we can say $\boldsymbol{v}_k = \sum_{i \neq k} c_i \boldsymbol{v}_i$. Apply T to both sides, getting*

$$\lambda_k \boldsymbol{v_k} = \sum_{i \neq k} c_i \lambda_i \boldsymbol{v}_i. \tag{15}$$

Multiply λ_k to both sides of $\boldsymbol{v}_k = \sum_{i \neq k} c_i \boldsymbol{v}_i$, getting

$$\lambda_k \boldsymbol{v}_k = \sum_{i \neq k} c_i \lambda_k \boldsymbol{v}_i. \tag{16}$$

Thus

$$0 = \sum_{i \neq k} c_i (\lambda_i - \lambda_k) \boldsymbol{v}_i. \tag{17}$$

Note that $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v_{k-1}}\}$ are linearly independent, which implies $c_i = 0$ for all i, given that $\lambda_i \neq \lambda_k$ for all $i < k$. So $\boldsymbol{v}_k$ turns out to be zero vector, which is a contradiction to $\boldsymbol{v}_k$ is an eigenvector.

Linear independence of eigenvectors is a nice property because if we have number of linearly independent eigenvectors equal to dimension of space, then we can use eigenvectors as basis.

Theorem 3.3. *Let V be vector space with $\dim\{V\} = n$. Then $T \in L(V)$ can have at most n distinct eigenvalues.*

Proof 3.3. *Suppose T has $m > n$ distinct eigenvalues, then the corresponding m eigenvectors are linearly indepdent, which is a contradiction to $\dim\{V\} = n$.*

3.2 Eigenvectors and diagonal matrices

Let us first look at when eigenvectors exists.

Theorem 3.4. *The following statements are equivalent:*

- *T has a eigenvalue λ.*

- $T - \lambda I$ *is not invertible/surjective/injective.*
- $\dim\{R(T - \lambda I)\} < \dim\{V\}$ *or* $\dim\{N(T - \lambda I)\} > 0$.

Proof 3.4. *Suppose* $Tv = \lambda v$*, then* $(T - \lambda I)v = 0$*, which suggests* $\dim\{N(T - \lambda I)\} > 0$*. From fundamental theorem of null space of range space,* $\dim\{R(T - \lambda I)\} = \dim\{V\} - \dim\{N(T - \lambda I)\} < \dim\{V\}$*. This also suggests* $T - \lambda I$ *is not invertible/surjective/injective.*

It will be helpful to look at the matrix representation and get insights.

Theorem 3.5. $T \in L(V)$ *with* $F = \mathbb{C}$ *always has a upper-triangular matrix in some basis.*

Proof 3.5. *We skip this proof. See "Linear algebra done right".*

It turns out that eigenvalues of T can be read from its upper-triangular matrix.

Theorem 3.6. *If* $M(T)$ *is an upper-triangular matrix, then the diagonal element* $M(T)_{i,i}$ *is eigenvalue* λ_i*, namely there exists* $\boldsymbol{v_i} \in V$ *such that* $T\boldsymbol{v_i} = M(T)_{i,i}\boldsymbol{v}_i$*.*

Proof 3.6. *Note that* $\det\{M(T) - M(T)_{i,i}I\} = 0$*, which implies* $M(T) - M(T)_{i,i}I$ *is not invertible, and thus* $M(T)_{i,i}$ *is indeed an eigenvalue.*

Theorem 3.7. $T \in L(V)$ *is invertible if and only if its upper-triangular matrix has no zero diagonal element.*

Proof 3.7. *Invertibility of* T *means* T *does not have eigenvectors with* $\lambda = 0$*. So from theorems we have already known we can see that* $M(T)_{i,i} \neq 0$ *for all* i *if* $M(T)$ *is the upper-triangular matrix of* T*.*

So from the thorem above, we can directly read eigenvalues if we have a upper-triangular matrix. If all the diagonal elements are distinct, then T has n different eigenvectors that span the whole space V.

Finally we talk about diagonal matrix.

Theorem 3.8. *The following statements are equivalent:*

- T *is diagonalizable.*
- *Eigenvectors of* T *can form a basis of* V*.*
- *There exists disjoint 1-dimensional invariant subspaces (thus eigenspaces)* $\{U_i\}$ *such that*

$$V = \oplus_i U_i. \tag{18}$$

Proof 3.8. *The proof is kind of trivial. Note that if* T *is diagonalizable, then that basis is eigenvectors by definition. Every 1-d invariant subspace is simply eigenspace (but multiple-dimensional invariant subspace is not neccessarily an eigenspace). So* $V = \oplus_i \mathit{span}\{\boldsymbol{v_i}\} = \oplus_i U_i$*.*

4 Spectral theorem and SVD

In this section we will cover spectral theorem and Sigular values decomposition of operators. Spectral theorem tells us the condition that ensure the completeness of eigenvectors. Sigular values decomposition gives an alternative way to decompose operators when there is no enough eigenvectors.

4.1 Hermitian operator and normal operators

4.1.1 Adjoint linear transform

Before introducing spectral theorem, we have to go through some concepts.

Definition 4.1. *Let $T \in L(V, W)$, then the **adjoint** transform of $T^\dagger : W \to V$ is defined by*

$$\langle Tv|w\rangle = \left\langle v\middle|T^\dagger w\right\rangle \tag{19}$$

for all $v \in V$ and $w \in W$.

Before moving on, we have to ask why $T^\dagger$ can exist? For given w and T, we can see $\langle Tv|w\rangle$ as an linear functional mapping v to $\langle Tv|w\rangle$. By the Riesz Representation theorem, we know that there exists w' such that $\langle Tv|w\rangle = \langle v|w'\rangle$, and we define $T^\dagger$ by $T^\dagger w = w'$. So the adjoint transform $T^\dagger$ is well-defined.

Theorem 4.1. *Let $T \in L(V, W)$, then $T^\dagger$ is a linear transform. Namely $T^\dagger \in L(W, V)$.*

Proof 4.1. *Easy to prove from the linear properties of inner product.*

Theorem 4.2. *The matrix representation $M(T^\dagger)$ is the conjugate transpose of $M(T)$, namely $M(T^\dagger) = [M(T)]^\dagger$.*

Proof 4.2. *Let A be matrix of T, B be matrix of $T^\dagger$. Then by definition*

$$\begin{aligned}\langle T\boldsymbol{e}_i|\boldsymbol{e}_j\rangle &= A_{j,i}\\ &= \left\langle \boldsymbol{e_i}\middle|T^\dagger \boldsymbol{e}_j\right\rangle = \sum_k B^*_{k,j}\left\langle \boldsymbol{e}_i|\boldsymbol{e_j}\right\rangle = B^\dagger_{j,i}.\end{aligned} \tag{20}$$

Thus $B = A^\dagger$, finishing the proof.

4.1.2 Hermitian operator

Definition 4.2. *Let $T \in L(V)$. T is called **Hermitian** (or self-adjoint) if and only if $T^\dagger = T$.*

Theorem 4.3. *Let $T \in L(V)$ be Hermitian operator. Then eigenvalues of T are real.*

Proof 4.3. *Suppose $T\boldsymbol{v} = \lambda\boldsymbol{v}$. Then consider inner product*

$$\begin{aligned}\langle T\boldsymbol{v}|\boldsymbol{v}\rangle &= \lambda\left\|\boldsymbol{v}\right\|^2\\ &= \langle \boldsymbol{v}|T\boldsymbol{v}\rangle = \lambda^*\left\|v\right\|^2,\end{aligned} \tag{21}$$

which imples $\lambda = \lambda^*$. Therefore eigenvalues λ must be real.

As a anology to numbers, Hermitian operators are just like real number.

4.1.3 Normal operator

Definition 4.3. *Let $T \in L(V)$. T is called a **normal** operator, if and only if $[T, T^\dagger] = 0$.*

We can see normal operator as a generalized version of Hermitian operator. For normal operators, though T is not neccessarily equal to $T^\dagger$, but they have the same normal applying on vectors :

$$\|T\boldsymbol{v}\| = \left\|T^\dagger\boldsymbol{v}\right\|. \tag{22}$$

Theorem 4.4. *If T is a normal operator, then $T^\dagger$ and T share the same eigenvectors, but the corresponding eigenvalues are complex conjugate to each other:*

$$T\boldsymbol{v} = \lambda \boldsymbol{v}, \quad T^\dagger \boldsymbol{v} = \lambda^* \boldsymbol{v}. \tag{23}$$

Proof 4.4. *Suppose $T\boldsymbol{v} = \lambda\boldsymbol{v}$. Then $(T - \lambda I)\boldsymbol{v} = 0$. Note that $T - \lambda I$ is also a norm operator, let us consider the inner product*

$$0 = \langle (T - \lambda I)\boldsymbol{v} | (T - \lambda I)\boldsymbol{v} \rangle = \left\langle (T - \lambda I)^\dagger \boldsymbol{v} \middle| (T - \lambda I)^\dagger \boldsymbol{v} \right\rangle. \tag{24}$$

So $T^\dagger \boldsymbol{v} = \lambda^* \boldsymbol{v}$.

Theorem 4.5. *If $T \in L(V)$ is a norm operator, then eigenvectors corresponding to different eigenvalues are orthogonal. Namely, eigenspaces are orthogonal.*

Proof 4.5. *Suppose $T\boldsymbol{v} = \lambda\boldsymbol{v}$ and $T\boldsymbol{w} = \beta\boldsymbol{w}$ with $\lambda \neq \beta$. Consider inner product*

$$(\lambda - \beta) \langle \boldsymbol{v} | \boldsymbol{w} \rangle = \langle \lambda \boldsymbol{v} | \boldsymbol{w} \rangle - \langle \boldsymbol{v} | \beta^* \boldsymbol{w} \rangle = \langle T\boldsymbol{v} | \boldsymbol{w} \rangle - \left\langle v \middle| T^\dagger \boldsymbol{w} \right\rangle = 0, \tag{25}$$

where we use the theorem that $T^\dagger \boldsymbol{w} = \beta^* \boldsymbol{w}$. Since $\lambda \neq \beta$, so $\langle \boldsymbol{v} | \boldsymbol{w} \rangle = 0$, finishing the proof.

4.2 Spectral theorem

The spectral theorem basically tells the equivalent condition to have an eigenbasis (eigenvectors that can span the whole space). If so, it is very nice working in such eigenbasis, since the corresponding operators are diagonal in eigenbasis.

Theorem 4.6 (Complex Spectral Theorem)**.** *Suppose $F = \mathbb{C}$ and $T \in L(V)$. Then the following are equivalent:*

(a) T is normal.

(b) T has eigenvectors that span V.

(c) $M(T)$ can be diagonalized in some basis of V.

Proof 4.6. *We first suppose (c) holds, and obviously in this basis $[T, T^\dagger] = 0$, because $M(T)$ and $M(T^\dagger)$ are both diagonal. So (c) implies (a).*

Then we suppose (a) holds. According to Schur's theorem, which says every operator has an upper-triangular matrix in some basis. Suppose A is the upper-triangular matrix of T with $A_{i,j} = 0$ when $i > j$. Then we have

$$\|T\boldsymbol{e}_i\|^2 = \|A_{1,1}\|^2,$$

and

$$\left\|T^\dagger \boldsymbol{e}_1\right\|^2 = \sum_i \|A_{1,i}\|^2.$$

Since T is normal and thus $\|T\boldsymbol{e}_1\| = \left\|T^\dagger \boldsymbol{e}_1\right\|$. Therefore we find $A_{1,i} = 0$ when $i \neq 1$. Applying the similar argument to other rows, we conclude that A is indeed a diagonal matrix in this basis. According to the defintion of eigenvectors, this basis is just the eigenbasis. So (a) implies (b).

Finally we suppose (b) holds. Then we can simply use eigenbasis to diagonalize $M(T)$. So (b) implies (c). We finish the proof.

It is worth noting that: If T is not normal, then it has no eigenbasis; if the upper-triangular matrix of T is not diagonal, then T has no eigenbasis.

Theorem 4.7 (Real Spectral Theorem)**.** *Suppose $F = \mathbb{R}$ and $T \in L(V)$. The following are equivalent:*

(a) T is Hermitian.

(b) T has eigenvectors that span V.

(c) $M(T)$ can be diagonalized in some basis of V.

Proof 4.7. *We skip the proof.*

4.3 Positive operators and isometry

4.3.1 Positive-semidefinite operators

Definition 4.4. *$T \in L(V)$ is called **positive-semidefinite**, if and only if T is Hermitian and $\langle T\boldsymbol{v}|\boldsymbol{v}\rangle \geq 0$ for all $\boldsymbol{v} \in V - \{\mathbf{0}\}$. If strict larger condition holds, then it is called **positive-definite**.*

Theorem 4.8. *Let $T \in L(V)$. Then the following are equivalent:*

(a) T is positive-semidefinite.

(b) T is Hermitian and all eigenvalues of T are nonnegative.

(c) T has a square root operator that is positive-semidefinite.

(d) T has a square root operator that is Hermitian.

(e) There exists an operator $R \in L(V)$ such that $T = R^\dagger R$.

Proof 4.8. *Suppose (a) holds. Then by definition of positive-semidefinite operators, T is Hermitian. By spectral theorem, T has a eigenbasis with real eigenvalues. In this eigenbasis, we compute $\langle T\boldsymbol{e_j}|\boldsymbol{e}_j\rangle = \lambda_j \|\boldsymbol{e_j}\|^2$. Since T is positve-semidefinite, so $\lambda_j \|\boldsymbol{e}_j\|^2 \geq 0$, and therefore $\lambda_j \geq 0$ for all j. So (a) implies (b).*

Suppose (b) holds. Since T is Hermitian, by spectral theorem, we use eigenbasis of T to diagonalize T. Then obviously in this basis, square root of T is simply diagonal matrix with diagonal matrix being square root of λ_j. Since $\lambda_j \geq 0$, so diagonal elements of $\sqrt{T}$ are also nonnegative. Thus $\sqrt{T}$ that is Hermitian and has no negative eigenvalue is thus positive-semidefinite. So (b) implies (c).

Suppose (c) holds. Then (c) implies (d) simply by definition of positive-semidefinite operators.

Suppose (d) holds. Then let R be this square root of T and $R^\dagger = R$. So $T = RR = R^\dagger R$, and thus (d) implies (e).

Suppose (e) holds. Then for all $\boldsymbol{v} \in V$, we find

$$\langle T\boldsymbol{v}|\boldsymbol{v}\rangle = \left\langle R^\dagger R\boldsymbol{v}\middle|\boldsymbol{v}\right\rangle = \langle R\boldsymbol{v}|R\boldsymbol{v}\rangle = \|R\boldsymbol{v}\|^2 \geq 0.$$

Note that $T^\dagger = R^\dagger R = T$. So T is positive-semidefinite. Thus (e) implies (a), finishing the proof.

4.3.2 Isometry

Definition 4.5. *Let $S \in L(V)$. S is called an* ***isometry*** *if and only if*

$$\|S\boldsymbol{v}\| = \|\boldsymbol{v}\|$$

for all $\boldsymbol{v} \in V$. Namely an isometry perserves the norm (and thus the induced inner product).

Theorem 4.9. *Let $S \in L(V)$. Then the following are equivalent:*

(a) S is an isometry.

(b) $\langle S\boldsymbol{v}|S\boldsymbol{w}\rangle = \langle \boldsymbol{v}|\boldsymbol{w}\rangle$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$.

(c) $\{S\boldsymbol{e_i}\}$ is an orthonormal basis if $\{\boldsymbol{e_i}\}$ is an orthonormal basis.

(d) $S^\dagger S = I$.

(e) $SS^\dagger = I$.

(f) $S^\dagger$ is an isometry.

(g) S is invertible and $S^{-1} = S^\dagger$.

Proof 4.9. *Suppose (a) holds. We know that inner product can be induced from the norm, so (a) implies (b).*

Suppose (b) holds. Then $\langle S\boldsymbol{e_i}|S\boldsymbol{e_j}\rangle = \langle \boldsymbol{e_i}|\boldsymbol{e_j}\rangle = \delta_{i,j}$. So (b) implies (c).

Suppose (c) holds. Since $\langle S\boldsymbol{v}|S\boldsymbol{w}\rangle = \langle S\boldsymbol{v}|S\boldsymbol{w}\rangle$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$, so $S^\dagger S = I$.

Suppose (d) holds. Note that if right and left inverse of S exists, then they are equal. So basically $S^\dagger S = SS^\dagger = I$.

Suppose (e) holds. Then $\left\langle S^\dagger\boldsymbol{v}\middle|S^\dagger\boldsymbol{w}\right\rangle = \left\langle \boldsymbol{v}\middle|SS^\dagger\boldsymbol{w}\right\rangle = \langle \boldsymbol{v}|\boldsymbol{w}\rangle$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$. So (e) implies (f).

Suppose (f) holds. Since we know (a) imples (d) and (e), so $S^\dagger$ is an isometry implies $S^\dagger S = I$ and $SS^\dagger = I$. So $S^{-1} = S^\dagger$.

Suppose (g) holds. Then $\langle S\boldsymbol{v}|S\boldsymbol{w}\rangle = \left\langle \boldsymbol{v}\middle|S^\dagger S\boldsymbol{v}\right\rangle = \langle \boldsymbol{v}|\boldsymbol{v}\rangle$, which suggests S is an isometry, finishing the proof.

Theorem 4.10. *If $S \in L(V)$ is an isometry with $F = \mathbb{C}$, then S has an eigenbasis with eigenvalues whose absolute values are 1.*

Proof 4.10. *Note that $[S, S^\dagger] = SS^\dagger - S^\dagger S = 0$, so an isometry is normal. By complex spectral theorem, S can be diagonalized in its eigenbasis $\{\boldsymbol{e_i}\}$. Then*

$$\langle S\boldsymbol{e_i}|S\boldsymbol{e_i}\rangle = |\lambda_i|^2 = \langle \boldsymbol{e_i}|\boldsymbol{e_i}\rangle = 1. \tag{26}$$

4.4 Polar decomposition and SVD

Finally we come to our main topics, SVD. We will use the conclusion we have proved to derive SVD theorem.

4.4.1 Polar decomposition

In some senses matrix can be treated just like numbers. For example, if we see an operator as a number, then a Hermitian operator is just like real number. Polar decomposition tells us that we can decompose an operator (complex number) into an isometry (direction of the complex number) times an positive-semidefinite operator (norm of the complex number).

Theorem 4.11 (Polar decomposition)**.** *Let $T \in L(V)$. T can be decomposed into*

$$T = S\sqrt{T^\dagger T} \tag{27}$$

with S an isometry.

Proof 4.11. *We skip this proof. We just make a few comments. This decomposition is reasonable, since*

$$\|T\boldsymbol{v}\| = \left\|\sqrt{T^\dagger T}\boldsymbol{v}\right\|.$$

And we additionaly multilply an isometry S to further make sure $T = S\sqrt{T^\dagger T}$.

4.4.2 Singular values decomposition

Definition 4.6. *Let $T \in L(V)$. **Singular values** of T is the eigenvalues of $\sqrt{T^\dagger T}$.*

Note that since $\sqrt{T^\dagger T}$ is positive-semidefinite, so by spectral theorem, T must have dim$\{V\}$ sigular values (can be repeated) corresponding to eigenbasis $\sqrt{T^\dagger T}\boldsymbol{e_i} = s_i\boldsymbol{e}_i$, and they are not less than zero.

Theorem 4.12 (SVD Theorem)**.** *Let $T \in L(V)$ with singular values $\{s_i\}$ and corresponding basis $\{\boldsymbol{e}_i\}$. Then $T\boldsymbol{v}$ for any $\boldsymbol{v} \in V$ can be decomposed into (called Sigular Values Decomposition):*

$$T\boldsymbol{v} = \sum_i \langle \boldsymbol{v}|\boldsymbol{e}_i\rangle\, s_i\boldsymbol{f_i}, \tag{28}$$

where $\{\boldsymbol{f}_i\}$ is a basis defined by $\boldsymbol{f}_i = S\boldsymbol{e}_i$ with $T = S\sqrt{T^\dagger T}$.

Proof 4.12. *For any $\boldsymbol{v}$, we have*

$$\boldsymbol{v} = \sum_i \langle \boldsymbol{v}|\boldsymbol{e_i}\rangle\, \boldsymbol{e_i},$$

Then by polar decomposition of T we find

$$T\boldsymbol{v} = S\sqrt{T^\dagger T}\sum_i \langle \boldsymbol{v}|\boldsymbol{e_i}\rangle\, \boldsymbol{e_i} = \sum_i \langle \boldsymbol{v}|\boldsymbol{e_i}\rangle\, s_iS\boldsymbol{e_i} = \sum_i \langle \boldsymbol{v}|\boldsymbol{e}_i\rangle\, s_i\boldsymbol{f_i}.$$

Since S is an isometry, so $\{f\boldsymbol{f}_i\}$ is still an orthonormal basis. If we choose $\{\boldsymbol{e_i}\}$ and $\{\boldsymbol{f_i}\}$ as basis to express T, then $M(T)$ is diagonal.