3. Review

3.1. Vector space.

Definition 3.1.1. A **vector space** V is a set V with two operations addition $+$ and scalar product $\cdot$ which satisfies eight axioms (which are omitted here).

Definition 3.1.2. The set of all linear combinations of a list of vectors $\mathcal{B} = \{v_1, \ldots, v_m\} \subset V$ is called the **span** of $v_1, \ldots, v_m$, denoted $\texttt{Span}(\mathcal{B}) = \texttt{Span}(v_1, \ldots, v_m)$. In other words,

$$\texttt{Span}(\mathcal{B}) = \texttt{Span}(v_1, \ldots, v_m) = \{a_1 v_1 + \ldots a_m v_m \mid a_1, \ldots, a_m \in \mathbb{F}\}.$$

Definition 3.1.3. A list $\{v_1, \ldots, v_m\} \subset V$ is called **linearly independent** if the only choice of $a_1, \ldots, a_m \in \mathbb{F}$ that makes $a_1 v_1 + \ldots + a_m v_m = 0$ is $a_1 = \ldots = a_m = 0$.

Definition 3.1.4. A **basis** $\mathcal{B}$ of a vector space V is a set of vectors $\mathcal{B} = \{v_1, v_2, \ldots, v_n\} \subset V$ such that

(1) $\texttt{Span}(\mathcal{B}) = V$.

(2) $\mathcal{B}$ is linearly independent,

3.2. Coordinates.

The primary use of bases is to set up coordinates. Choose a vector space V and choose a basis $\mathcal{B} = \{v_1, \ldots, v_m\}$. Then each vector can be written as a linear combination of these basis vectors and this expression is unique. That is, for any vector $v \in V$, we can find only one set of coefficients $a_1, \ldots, a_m \in \mathbb{F}$ such that $v = a_1 v_1 + \ldots + a_m v_m = \sum_{i=1}^m a_i v_i$. In this case we can write

$$\Big[v\Big]_{\mathcal{B}} = \Big[a_1 v_1 + \ldots + a_m v_m\Big]_{\mathcal{B}} = \Big[\textstyle\sum_{i=1}^m a_i v_i\Big]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix},$$

and call this column vector the $\mathcal{B}$-coordinates of v.

Question: How do you find the coordinates?

Answer: Solving linear equations $v = a_1 v_1 + \ldots + a_m v_m$ for variables $a_1, \ldots, a_m$.

Remark 3.2.1. If able, you **ARE REQUIRED** to use matrices to solve linear equations.

Example 3.2.2. The standard basis of $\mathtt{Mat}_{2\times 2}(\mathbb{F})$ is

$$\mathcal{E} = \left\{ E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

Any 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ can be expressed uniquely as $aE_{11}+bE_{12}+cE_{21}+dE_{22}$. Therefore we can write matrix A as a column vector with respect to the standard basis:

$$\begin{bmatrix} A \end{bmatrix}_{\mathcal{E}} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}.$$

3.3. **Linear maps.**

Definition 3.3.1. A map $L : V \to W$ between vector spaces is said to be *linear* if it preserves scalar multiplication and addition in the following way:

$$L(ax) = aL(x),$$

$$L(x+y) = L(x) + L(y),$$

where $a \in \mathbb{F}$, $x, y \in V$.

3.4. **Matrix representations.** Let $L : V \to W$ be a linear map. Choose a basis $\mathcal{E} = \{e_1, \ldots, e_n\}$ of V, and a basis $\mathcal{F} = \{f_1, \ldots, f_m\}$ of W. Then we can express L as a matrix in the following way.

For each $L(e_i)$, since it is in W, we can write it as a linear combination of $\mathcal{F}$. That is

$$\begin{bmatrix} L(e_i) \end{bmatrix}_{\mathcal{F}} = \begin{bmatrix} \sum_{j=1}^{m} a_{ji} f_j \end{bmatrix}_{\mathcal{F}} = \begin{bmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{bmatrix}.$$

We put all these column vectors together to form a matrix which is denoted by $\left[L\right]_{\mathcal{F}\leftarrow\mathcal{E}}$:

$$\left[L\right]_{\mathcal{F}\leftarrow\mathcal{E}} = \begin{bmatrix} \left[L(e_1)\right]_{\mathcal{F}} & \left[L(e_2)\right]_{\mathcal{F}} & \cdots & \left[L(e_n)\right]_{\mathcal{F}} \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}.$$

We call it the **matrix of the linear map** L with respect to the basis $\mathcal{E}$ of V and the basis $\mathcal{F}$ of W. If there is no confusion, we can simply denote it by $\left[L\right]$.

Remark 3.4.1. In the Axler's book the notation is a little different. Let $L : V \to W$ be a linear map, $\mathcal{E}$ be a basis of V and $\mathcal{F}$ be a basis of W. The relations between different notations are:

Concept	Axler's notation	My notation
Linear map	L	L
Matrix of L	$\mathcal{M}(L)$	$\left[L\right]$
Matrix of L w.r.t bases $\mathcal{E}$ and $\mathcal{F}$	$\mathcal{M}(T, \mathcal{E}, \mathcal{F})$	$\left[L\right]_{\mathcal{F}\leftarrow\mathcal{E}}$

Example 3.4.2. Compute a matrix representation for $L : \mathtt{Mat}_{2\times 2}(\mathbb{F}) \to \mathtt{Mat}_{1\times 2}(\mathbb{F})$ defined by $L(X) = \begin{bmatrix} 1 & -1 \end{bmatrix} X$ using the standard bases: $E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, $E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ and $E_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and $E_2 = \begin{bmatrix} 0 & 1 \end{bmatrix}$.

Solve. Let $V = \mathtt{Mat}_{2\times 2}(\mathbb{F})$, $W = \mathtt{Mat}_{1\times 2}(\mathbb{F})$. Follow the instructions exactly:

(1) Apply L to the basis vector $L(E_{11})$:

$$L(E_{11}) = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

(2) Since L is a map from V to W, $L(E_{11})$ should be a vector in W.

(3) Since $L(E_{11}) \in W$, and $\mathcal{F} = \{E_1, E_2\}$ be a basis of W, we can write $L(E_{11})$ as a linear combination of $\mathcal{F}$. (Think that why here I write $\mathcal{F} = \{E_1, E_2\}$ instead of $\mathcal{F} = \{E_1, E_2, \dots, E_m\}$.)

$$\begin{bmatrix} 1 & 0 \end{bmatrix} = a_{11}E_1 + a_{21}E_2 = a_{11} \begin{bmatrix} 1 & 0 \end{bmatrix} + a_{21} \begin{bmatrix} 0 & 1 \end{bmatrix}.$$

Solve this equation. The solution is $a_{11} = 1$, $a_{21} = 0$. So

$$\left[L(E_{11})\right]_{\mathcal{F}} = \left[\begin{bmatrix}1 & 0\end{bmatrix}\right]_{\mathcal{F}} = \left[1 \cdot E_1 + 0 \cdot E_2\right]_{\mathcal{F}} = \begin{bmatrix}1\\0\end{bmatrix}.$$

(4) Repeat the process to other basis vectors of V. We have

$$\left[L(E_{12})\right]_{\mathcal{F}} = \begin{bmatrix}0\\1\end{bmatrix}, \quad \left[L(E_{21})\right]_{\mathcal{F}} = \begin{bmatrix}-1\\0\end{bmatrix}, \quad \left[L(E_{22})\right]_{\mathcal{F}} = \begin{bmatrix}0\\-1\end{bmatrix}.$$

(5) Put all these column vectors together, we have:

$$\left[L\right]_{\mathcal{F}\leftarrow\mathcal{E}} = \begin{bmatrix}1 & 0 & -1 & 0\\0 & 1 & 0 & -1\end{bmatrix}.$$

□

3.5. How do we use the matrix $\left[L\right]$?

Theorem 3.5.1.

$$\left[L(v)\right]_{\mathcal{F}} = \left[L\right]_{\mathcal{F}\leftarrow\mathcal{E}}\left[v\right]_{\mathcal{E}}.$$

Example 3.5.2. Let $L : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear map defined by embedding $\mathbb{R}^2$ into $\mathbb{R}^3$ as the xy-plane. We want to compute $L\left(\begin{bmatrix}1\\2\end{bmatrix}\right)$. There are two ways.

Directly embedding: The vector $\begin{bmatrix}1\\2\end{bmatrix}$, when embedding into $\mathbb{R}^3$, is considered as directly adding a z-coordinate which should be 0. Therefore $L\left(\begin{bmatrix}1\\2\end{bmatrix}\right) = \begin{bmatrix}1\\2\\0\end{bmatrix}$.

Using matrix: First choose a basis of $\mathbb{R}^2$ and $\mathbb{R}^3$. We can use the standard bases $\mathcal{E} = \left\{e_1 = \begin{bmatrix}1\\0\end{bmatrix}, e_2 = \begin{bmatrix}0\\1\end{bmatrix}\right\} \subset \mathbb{R}^2$ and $\mathcal{F} = \left\{f_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, f_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, f_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}\right\} \subset \mathbb{R}^3$. To find the matrix of L, we need

$$L(e_1) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = f_1, \quad L(e_2) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = f_2.$$

Then we have

$$\left[L(e_1)\right]_{\mathcal{F}} = \left[f_1\right]_{\mathcal{F}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \left[L(e_2)\right]_{\mathcal{F}} = \left[f_2\right]_{\mathcal{F}} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Then

$$\left[L\right]_{\mathcal{F}\leftarrow\mathcal{E}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Since $\left[\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right]_{\mathcal{E}} = \left[e_1 + 2e_2\right]_{\mathcal{E}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, then we have

$$\left[L\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right)\right]_{\mathcal{F}} = \left[L\right]_{\mathcal{F}\leftarrow\mathcal{E}}\left[\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right]_{\mathcal{E}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

Therefore

$$L\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) = 1f_1 + 2f_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

Note that all the blue column vectors are the coordinates with respect to the bases $\mathcal{E}$ or $\mathcal{F}$. The black column vectors are the "true" vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$.

Example 3.5.3. Let $V = \texttt{Span}\{\cos x, \sin x\}$ with $\mathbb{F} = \mathbb{R}$. Let $L : V \to V$ be taking the derivative. We want to compute $L(\cos x + 2\sin x)$. There are two ways.

Directly taking derivative: $L(\cos x + 2\sin x) = (\cos x + 2\sin x)' = -\sin x + 2\cos x$.

Using matrix: First choose a basis of V. We can use $\mathcal{B} = \{b_1 = \cos x, b_2 = \sin x\}$. To find the matrix of L, we need

$$L(b_1) = L(\cos x) = (\cos x)' = -\sin x,$$

$$L(b_2) = L(\sin x) = (\sin x)' = \cos x.$$

Then we have

$$\Big[L(b_1)\Big]_{\mathcal{B}} = \Big[-\sin x\Big]_{\mathcal{B}} = \Big[-b_2\Big]_{\mathcal{B}} = \begin{bmatrix} 0 \\ -1 \end{bmatrix},$$

$$\Big[L(b_2)\Big]_{\mathcal{B}} = \Big[\cos x\Big]_{\mathcal{B}} = \Big[b_1\Big]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Then

$$\Big[L\Big]_{\mathcal{B}\leftarrow\mathcal{B}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Since $\Big[\cos x + 2\sin x\Big]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, then we have

$$\Big[L(\cos x + 2\sin x)\Big]_{\mathcal{B}} = \Big[L\Big]_{\mathcal{B}\leftarrow\mathcal{B}}\Big[\cos x + 2\sin x\Big]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

Therefore

$$L(\cos x + 2\sin x) = 2b_1 + (-1)b_2 = 2\cos x - \sin x.$$

3.6. **Exercises.**

Exercise 3.1. Find the coordinate of the vector $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^3$ with respect to the basis

$$\mathcal{U} = \left\{ u_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad u_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

Exercise 3.2. Let P_3 be the vector space of all polynomials of variable t with degree no higher than 3. Find the matrix representation for taking derivative $D : P_3 \to P_3$ with respect to the basis

$$\mathcal{F} = \{ f_1 = t^3, \quad f_2 = t^3 + t^2, \quad f_3 = t^3 + t^2 + t, \quad f_4 = t^3 + t^2 + t + 1 \}.$$

Exercise 3.3. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map defined by

$$T\left(\begin{bmatrix} a \\ b \end{bmatrix} \right) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \quad \text{for} \begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2.$$

Consider the basis

$$\mathcal{B} = \left\{ x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

(1) Compute the matrix representation of T with respect to $\mathcal{B}$.

(2) Use the above matrix to compute $T\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} \right)$.

The homework is due on Apr. 5.

4. Introduction to 132

4.1. **Changing bases.** Let V be the vector space, and $\mathcal{B} = \{v_1, \ldots, v_n\}$ and $\mathcal{E} = \{e_1, \ldots, e_n\}$ be two bases of V. For a vector $v \in V$, we can write it as

$$\begin{aligned} v &= b_1 v_1 + \ldots + b_n v_n = \sum_{i=1}^{n} b_i v_i \\ &= c_1 e_1 + \ldots + c_n e_n = \sum_{i=1}^{n} c_i e_i. \end{aligned} \tag{4.1}$$

(Think: Can we say that "$b_i = c_i$ for $i = 1, \ldots, n$ by the linearly independence of basis"?)

In other words, we have

$$\begin{bmatrix} v \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}, \quad \begin{bmatrix} v \end{bmatrix}_{\mathcal{E}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

We want to know the relation between $\begin{bmatrix} v \end{bmatrix}_{\mathcal{B}}$ and $\begin{bmatrix} v \end{bmatrix}_{\mathcal{E}}$. We start from the relations between the two bases. Since $\mathcal{E}$ form a basis of V, and vectors in $\mathcal{B}$ are also vectors in V, we can find their $\mathcal{E}$-coordinates. That is, write vectors in $\mathcal{B}$ as linear combinations of vectors in $\mathcal{E}$. For any $i = 1, \ldots, n$, we have

$$v_i = p_{1i} e_1 + p_{2i} e_2 + \ldots + p_{ni} e_n = \sum_{j=1}^{n} p_{ji} e_j. \tag{4.2}$$

So

$$\begin{bmatrix} v_i \end{bmatrix}_{\mathcal{E}} = \begin{bmatrix} p_{1i} \\ \vdots \\ p_{ni} \end{bmatrix}.$$

Then we form a **change-of-basis matrix**

$$P_{\mathcal{E} \leftarrow \mathcal{B}} = \begin{bmatrix} p_{11} & \ldots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \ldots & p_{nn} \end{bmatrix}.$$

How do we use this matrix?

Theorem 4.1.1.

$$\Big[v\Big]_{\mathcal{E}} = P_{\mathcal{E}\leftarrow\mathcal{B}}\Big[v\Big]_{\mathcal{B}}.$$

Proof. Let us start from equation (4.1) and (4.2). Since

$$v = \sum_{i=1}^{n} b_i v_i \quad \text{and} \quad v_i = \sum_{j=1}^{n} p_{ji} e_j \text{ for any } i,$$

we have

$$v = \sum_{i=1}^{n} b_i v_i = \sum_{i=1}^{n} b_i (\sum_{j=1}^{n} p_{ji} e_j) = \sum_{j=1}^{n} (\sum_{i=1}^{n} p_{ji} b_i) e_j.$$

This suggest that

$$\Big[v\Big]_{\mathcal{E}} = \begin{bmatrix} \sum_{i=1}^{n} p_{1i} b_i \\ \vdots \\ \sum_{i=1}^{n} p_{ni} b_i \end{bmatrix} = \begin{bmatrix} p_{11} & \dots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \dots & p_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = P_{\mathcal{E}\leftarrow\mathcal{B}}\Big[v\Big]_{\mathcal{B}}.$$

□

It is also easy to see that $P_{\mathcal{E}\leftarrow\mathcal{B}}$ is invertible and

$$P^{-1}_{\mathcal{E}\leftarrow\mathcal{B}} = P_{\mathcal{B}\leftarrow\mathcal{E}}.$$

Example 4.1.2. Let $V = \mathbb{R}^2$. Consider two bases: $\mathcal{E} = \left\{ e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ and $\mathcal{B} = \left\{ v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, v_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$. Let $w = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Since $w = e_1 + e_2$, we have $\Big[w\Big]_{\mathcal{E}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Now there are at least two different ways to find $\Big[w\Big]_{\mathcal{B}}$.

Directly: To find $\Big[w\Big]_{\mathcal{B}}$, we need to write w as a linear combination of vectors in $\mathcal{B}$. That is $w = b_1 v_1 + b_2 v_2$. Then we have a linear system:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = b_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + b_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Then we have

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}.$$

Therefore $b_1 = b_2 = 1/3$. So $w = \frac{1}{3}v_1 + \frac{1}{3}v_2$. Then

$$\Big[w\Big]_{\mathcal{B}} = \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}.$$

Use change-of-basis matrix: Since $v_1 = e_1 + 2e_2$, $v_2 = 2e_1 + e_2$, we have

$$\Big[v_1\Big]_{\mathcal{E}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \Big[v_2\Big]_{\mathcal{E}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Then the change-of-basis matrix is $P_{\mathcal{E}\leftarrow\mathcal{B}} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Therefore

$$\Big[w\Big]_{\mathcal{B}} = P_{\mathcal{B}\leftarrow E}\Big[w\Big]_{\mathcal{E}} = P^{-1}_{\mathcal{E}\leftarrow B}\Big[w\Big]_{\mathcal{E}} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}.$$

4.2. **Changing basis to change the matrix representation of a linear map.**

Let $L : V \to W$ be a linear map. Let $\mathcal{E}$ and $\mathcal{F}$ be two bases of V, $\mathcal{B}$ and $\mathcal{C}$ be two bases of W. Then we can have at least two matrix representations of L under different bases: $\Big[L\Big]_{\mathcal{B}\leftarrow\mathcal{E}}$ and $\Big[L\Big]_{\mathcal{C}\leftarrow\mathcal{F}}$. Let $P_{\mathcal{E}\leftarrow\mathcal{F}}$ be the change-of-matrix (from $\mathcal{F}$ to $\mathcal{E}$) on V and $Q_{\mathcal{B}\leftarrow\mathcal{C}}$ be the change-of-basis matrix (form $\mathcal{C}$ to $\mathcal{B}$) on W. I use Q here because I want to emphasize that these two change-of-basis matrices are on different vector spaces. What is the relation between them?

Theorem 4.2.1.

$$\Big[L\Big]_{\mathcal{B}\leftarrow\mathcal{E}} = Q_{\mathcal{B}\leftarrow\mathcal{C}}\Big[L\Big]_{\mathcal{C}\leftarrow\mathcal{F}} P_{\mathcal{F}\leftarrow\mathcal{E}}.$$

Proof. Recall that for any vector $v \in V$, we have

$$\Big[L(v)\Big]_{\mathcal{B}} = \Big[L\Big]_{\mathcal{B}\leftarrow\mathcal{E}}\Big[v\Big]_{\mathcal{E}}, \quad \text{and} \quad \Big[L(v)\Big]_{\mathcal{C}} = \Big[L\Big]_{\mathcal{C}\leftarrow\mathcal{F}}\Big[v\Big]_{\mathcal{F}}.$$

Since $\Big[v\Big]_{\mathcal{F}} = P_{\mathcal{F}\leftarrow\mathcal{E}}\Big[v\Big]_{\mathcal{E}}$ and $\Big[L(v)\Big]_{C} = Q_{\mathcal{C}\leftarrow\mathcal{B}}\Big[L(v)\Big]_{\mathcal{B}}$, we have

$$Q_{\mathcal{C}\leftarrow\mathcal{B}}\Big[L\Big]_{\mathcal{B}\leftarrow\mathcal{E}}\Big[v\Big]_{\mathcal{E}} = Q_{\mathcal{C}\leftarrow\mathcal{B}}\Big[L(v)\Big]_{\mathcal{B}} = \Big[L(v)\Big]_{C} = \Big[L\Big]_{\mathcal{C}\leftarrow\mathcal{F}}\Big[v\Big]_{\mathcal{F}} = \Big[L\Big]_{\mathcal{C}\leftarrow\mathcal{F}}P_{\mathcal{F}\leftarrow\mathcal{E}}\Big[v\Big]_{\mathcal{E}}$$

for any $v \in V$. Therefore it is easy to see that

$$Q_{\mathcal{C}\leftarrow\mathcal{B}}\Big[L\Big]_{\mathcal{B}\leftarrow\mathcal{E}} = \Big[L\Big]_{\mathcal{C}\leftarrow\mathcal{F}}P_{\mathcal{F}\leftarrow\mathcal{E}}.$$

In other words,

$$\Big[L\Big]_{\mathcal{B}\leftarrow\mathcal{E}} = Q_{\mathcal{B}\leftarrow\mathcal{C}}\Big[L\Big]_{\mathcal{C}\leftarrow\mathcal{F}}P_{\mathcal{F}\leftarrow\mathcal{E}}.$$

□

Example 4.2.2. Let $V = \mathbb{R}^2$. Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Let $L : V \to V$ be defined by

$$L(v) = Av \quad \text{for any } v \in V.$$

Let $\mathcal{S} = \left\{ e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ be a basis. Since

$$L(e_1) = Ae_1 = \begin{bmatrix} 0 \\ -1 \end{bmatrix} = 0e_1 + (-1)e_2, \quad L(e_2) = Ae_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1e_1 + 0e_2,$$

we have

$$\Big[L(e_1)\Big]_{\mathcal{S}} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \quad \Big[L(e_2)\Big]_{\mathcal{S}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Therefore

$$\Big[L\Big]_{\mathcal{S}\leftarrow S} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Now choose another basis $B = \left\{ b_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, b_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$. Then since

$$b_1 = 1e_1 + 1e_2, \quad b_2 = 1e_1 + (-1)e_2,$$

we have

$$\begin{bmatrix} b_1 \end{bmatrix}_{\mathcal{S}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} b_2 \end{bmatrix}_{\mathcal{S}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Therefore $P_{\mathcal{S}\leftarrow B} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

Now we have two methods to compute $\begin{bmatrix} L \end{bmatrix}_{\mathcal{B}\leftarrow B}$. Either directly use the definition of matrix representations under the basis B, or use Theorem 4.2.1.

Using the definition of matrix representations under the basis B:

(1) $L(b_1) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = x_{11}b_1 + x_{21}b_2$. Solving the vector equation, $x_{11} = 0$, $x_{21} = 1$. Therefore $\begin{bmatrix} L(b_1) \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

(2) $L(b_2) = \begin{bmatrix} -1 \\ -1 \end{bmatrix} = x_{12}b_1 + x_{22}b_2$. Solving the vector equation, $x_{12} = -1$, $x_{22} = 0$. Therefore $\begin{bmatrix} L(b_2) \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$.

Therefore the matrix representation is $\begin{bmatrix} L \end{bmatrix}_{\mathcal{B}\leftarrow\mathcal{B}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Using Theorem 4.2.1: Since $P_{\mathcal{S}\leftarrow\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, then $P_{\mathcal{B}\leftarrow\mathcal{S}} = P^{-1}_{\mathcal{S}\leftarrow\mathcal{B}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$. So

$$\begin{bmatrix} L \end{bmatrix}_{\mathcal{B}\leftarrow\mathcal{B}} = P_{\mathcal{B}\leftarrow\mathcal{S}} \begin{bmatrix} L \end{bmatrix}_{\mathcal{S}\leftarrow\mathcal{S}} P_{\mathcal{S}\leftarrow\mathcal{B}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Example 4.2.3. Let $V = \mathbb{C}$ with the field being $\mathbb{R}$. Let $L : V \to V$ be defined by

$$L(z) = -iz \quad \text{for any } z \in V.$$

Let $\mathcal{S} = \{e_1 = 1, e_2 = i\}$ be a basis. Since

$$L(e_1) = -i \cdot 1 = -i = 0e_1 + (-1)e_2, \quad L(e_2) = -i \cdot i = 1 = 1e_1 + 0e_2,$$

we have

$$\begin{bmatrix} L(e_1) \end{bmatrix}_{\mathcal{S}} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} L(e_2) \end{bmatrix}_{\mathcal{S}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Therefore

$$\begin{bmatrix} L \end{bmatrix}_{\mathcal{S} \leftarrow \mathcal{S}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Now choose another basis $\mathcal{B} = \{b_1 = 1 + i, b_2 = 1 - i\}$. Then since

$$b_1 = 1e_1 + 1e_2, \quad b_2 = 1e_1 + (-1)e_2,$$

we have $\begin{bmatrix} b_1 \end{bmatrix}_{\mathcal{S}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} b_2 \end{bmatrix}_{\mathcal{S}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Therefore $P_{\mathcal{S} \leftarrow \mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

Now we have two methods to compute $\begin{bmatrix} L \end{bmatrix}_{\mathcal{B} \leftarrow \mathcal{B}}$. Either directly use the definition of matrix representations under the basis B, or use Theorem 4.2.1.

Using the definition of matrix representations under the basis $\mathcal{B}$:

(1) $L(b_1) = -i(1 + i) = 1 - i = x_{11}b_1 + x_{21}b_2$. Solving the vector equation, $x_{11} = 0$, $x_{21} = 1$.

Therefore $\begin{bmatrix} L(b_1) \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

(2) $L(b_2) = -i(1-i) = -1-i = x_{12}b_1 + x_{22}b_2$. Solving the vector equation, $x_{12} = -1$, $x_{22} = 0$.

Therefore $\begin{bmatrix} L(b_2) \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$.

Therefore the matrix representation is $\begin{bmatrix} L \end{bmatrix}_{\mathcal{B} \leftarrow \mathcal{B}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Using Theorem 4.2.1: Since $P_{\mathcal{S} \leftarrow \mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, then $P_{\mathcal{B} \leftarrow \mathcal{S}} = P_{\mathcal{S} \leftarrow \mathcal{B}}^{-1} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$. So

$$\begin{bmatrix} L \end{bmatrix}_{\mathcal{B} \leftarrow \mathcal{B}} = P_{\mathcal{B} \leftarrow \mathcal{S}} \begin{bmatrix} L \end{bmatrix}_{\mathcal{S} \leftarrow \mathcal{S}} P_{\mathcal{S} \leftarrow \mathcal{B}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Example 4.2.4. Let $V = \texttt{Span}\{\cos(x), \sin(x)\}$ with the field being $\mathbb{R}$. Let $L : V \to V$ be defined by

$$L(f) = f' \quad \text{for any } f \in V.$$

Let $\mathcal{S} = \{e_1 = \cos(x), e_2 = \sin(x)\}$ be a basis. Since

$$\Big[L(e_1)\Big]_{\mathcal{S}} = \Big[\cos(x)'\Big]_{\mathcal{S}} = \Big[-\sin(x)\Big]_{\mathcal{S}} = \Big[0e_1 + (-1)e_2\Big]_{\mathcal{S}} = \begin{bmatrix} 0 \\ -1 \end{bmatrix},$$

$$\Big[L(e_2)\Big]_{\mathcal{S}} = \Big[\sin(x)'\Big]_{\mathcal{S}} = \Big[\cos(x)\Big]_{\mathcal{S}} = \Big[1e_1 + 0e_2\Big]_{\mathcal{S}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

we have $\Big[L\Big]_{\mathcal{S}\leftarrow\mathcal{S}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

Now choose another basis $\mathcal{B} = \{b_1 = \cos(x) + \sin(x), b_2 = \cos(x) - \sin(x)\}$. Then since

$$\Big[b_1\Big]_{\mathcal{S}} = \Big[1e_1 + 1e_2\Big]_{\mathcal{S}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \Big[b_2\Big]_{\mathcal{S}} = \Big[1e_1 + (-1)e_2\Big]_{\mathcal{S}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

we have $P_{\mathcal{S}\leftarrow\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

Now we have two methods to compute $\Big[L\Big]_{\mathcal{B}\leftarrow\mathcal{B}}$. Either directly use the definition of matrix representations under the basis $\mathcal{B}$, or use Theorem 4.2.1.

Using the definition of matrix representations under the basis $\mathcal{B}$:

(1) $L(b_1) = (\cos(x) + \sin(x))' = \cos(x) - \sin(x) = x_{11}b_1 + x_{21}b_2$. Solving the vector equation, $x_{11} = 0$, $x_{21} = 1$. Therefore $\Big[L(b_1)\Big]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

(2) $L(b_2) = (\cos(x) - \sin(x))' = -\cos(x) - \sin(x) = x_{12}b_1 + x_{22}b_2$. Solving the vector equation, $x_{12} = -1$, $x_{22} = 0$. Therefore $\Big[L(b_2)\Big]_{\mathcal{B}} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$.

Therefore the matrix representation is $\Big[L\Big]_{\mathcal{B}\leftarrow\mathcal{B}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Using Theorem 4.2.1: Since $P_{\mathcal{S}\leftarrow\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, then $P_{\mathcal{B}\leftarrow\mathcal{S}} = P_{\mathcal{S}\leftarrow\mathcal{B}}^{-1} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$. So

$$\Big[L\Big]_{\mathcal{B}\leftarrow\mathcal{B}} = P_{\mathcal{B}\leftarrow\mathcal{S}}\Big[L\Big]_{\mathcal{S}\leftarrow\mathcal{S}} P_{\mathcal{S}\leftarrow\mathcal{B}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

4.3. **Example: Choose a specific basis.**

4.3.1. *An example.* Let's consider the following example. Let $V = \mathbb{R}^4$, $W = \mathbb{R}^3$. Let $\mathcal{S}_V = \{e_1, e_2, e_3, e_4\}$ be the standard basis of V, $\mathcal{S}_W = \{f_1, f_2, f_3\}$ be the standard basis of W. Let $\mathcal{T} : V \to W$ be a linear transformation defined by

$$\Big[\mathcal{T}\Big]_{\mathcal{S}_W\leftarrow\mathcal{S}_V} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 2 \\ 1 & 3 & 2 & 4 \end{bmatrix} =: T.$$

Then we have the following information:

$$\Big[\mathcal{T}(e_1)\Big]_{\mathcal{S}_W} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \Big[\mathcal{T}(e_2)\Big]_{\mathcal{S}_W} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \quad \Big[\mathcal{T}(e_3)\Big]_{\mathcal{S}_W} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}, \quad \Big[\mathcal{T}(e_4)\Big]_{\mathcal{S}_W} = \begin{bmatrix} 4 \\ 2 \\ 4 \end{bmatrix}.$$

Then we have

$$\mathcal{T}(e_1) = f_1 + f_3, \quad \mathcal{T}(e_2) = 2f_1 + f_2 + 3f_3, \quad \mathcal{T}(e_3) = 3f_1 + f_2 + 2f_3, \quad \mathcal{T}(e_4) = 4f_1 + 2f_2 + 4f_3.$$

4.3.2. *Change basis of W.* Now we want to change basis. Let's start from W. Basic idea is that we want to use $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ 2 \\ 4 \end{bmatrix}$, or some of them to be basis vectors. To do it, we need to find a basis from these four vectors. Thus the first three column vectors in T form a basis of $\texttt{Col}(T)$. Then in $W=\mathbb{R}^3$, I choose a new basis $\mathcal{C}_W=\{w_1,w_2,w_3\}$ where

$$w_1=\begin{bmatrix}1\\0\\1\end{bmatrix},\quad w_2=\begin{bmatrix}2\\1\\3\end{bmatrix},\quad w_3=\begin{bmatrix}3\\1\\2\end{bmatrix}.$$

In addition, we know that the fourth column vector satisfies $\begin{bmatrix}4\\2\\4\end{bmatrix}=-w_1+w_2+w_3$. Therefor under the basis $\mathcal{S}_V$ of V and $\mathcal{C}_W$ of W, we have:

$$\Big[\mathcal{T}(e_1)\Big]_{\mathcal{C}_W}=\Big[w_1\Big]_{\mathcal{C}_W}=\begin{bmatrix}1\\0\\0\end{bmatrix},\quad \Big[\mathcal{T}(e_2)\Big]_{\mathcal{C}_W}=\Big[w_2\Big]_{\mathcal{C}_W}=\begin{bmatrix}0\\1\\0\end{bmatrix},$$

$$\Big[\mathcal{T}(e_3)\Big]_{\mathcal{C}_W}=\Big[w_3\Big]_{\mathcal{C}_W}=\begin{bmatrix}0\\0\\1\end{bmatrix},\quad \Big[\mathcal{T}(e_4)\Big]_{\mathcal{C}_W}=\Big[-w_1+w_2+w_3\Big]_{\mathcal{C}_W}=\begin{bmatrix}-1\\1\\1\end{bmatrix}.$$

Then the matrix of $\mathcal{T}$ under basis $\mathcal{S}_V$ and $\mathcal{C}_W$ becomes to

$$\Big[\mathcal{T}\Big]_{\mathcal{C}_W\leftarrow\mathcal{S}_V}=\begin{bmatrix}1&0&0&-1\\0&1&0&1\\0&0&1&1\end{bmatrix}.$$

4.3.3. *Some understanding about the above computation.*

(1) The above w_1,w_2,w_3 comes from $w_1=\mathcal{T}(e_1)$, $w_2=\mathcal{T}(e_2)$, $w_3=\mathcal{T}(e_3)$ directly. That is to say, we want to make choose a basis of W based on $\{\mathcal{T}(e_i)\}$. If $\{\mathcal{T}(e_i)\}$ cannot make a basis, just choose the linear independent part and extend it to be a basis of W.

(2) Since we change the basis of W, we have the following formula

$$\Big[\mathcal{T}\Big]_{\mathcal{S}_W\leftarrow\mathcal{S}_V}=P_{\mathcal{S}_W\leftarrow\mathcal{C}_W}\Big[\mathcal{T}\Big]_{\mathcal{C}_W\leftarrow\mathcal{S}_V}.$$

Here $P_{\mathcal{S}_W \leftarrow \mathcal{C}_W}$ is a 3×3 invertible matrix. Recall from MATH 031 that to multiply a invertible matrix on the left of a matrix T is equivalent to do row transformation to the matrix T. Then this means that the matrix $\left[\mathcal{T}\right]_{\mathcal{S}_W \leftarrow \mathcal{S}_V}$ is row equivalent to the matrix $\left[\mathcal{T}\right]_{\mathcal{C}_W \leftarrow \mathcal{S}_V}$. Note that $\left[\mathcal{T}\right]_{\mathcal{C}_W \leftarrow \mathcal{S}_V}$ is in reduced row echelon form. Then this process is actually the **row reduction algorithm**.

4.3.4. *Change basis of V.* Now we want to change basis of V to make this matrix even simpler. The idea is to change e_4 since the first three already looks good. Since

$$\mathcal{T}(e_4) = -w_1 + w_2 + w_3 = -\mathcal{T}(e_1) + \mathcal{T}(e_2) + \mathcal{T}(e_3),$$

we have $\mathcal{T}(e_4 + e_1 - e_2 - e_3) = 0$. We will change the basis based on this observation. Let $\mathcal{C}_V = \{v_1, v_2, v_3, v_4\}$ where

$$v_1 = e_1, \quad v_2 = e_2, \quad v_3 = e_3, \quad v_4 = e_1 - e_2 - e_3 + e_4.$$

Now using basis $\mathcal{C}_V$ of V and basis $\mathcal{C}_W$ of W, we have

$$\mathcal{T}(v_1) = w_1, \quad \mathcal{T}(v_2) = w_2, \quad \mathcal{T}(v_3) = w_3, \quad \mathcal{T}(v_4) = 0.$$

Therefore

$$\left[\mathcal{T}(v_1)\right]_{\mathcal{C}_W} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \left[\mathcal{T}(v_2)\right]_{\mathcal{C}_W} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \left[\mathcal{T}(v_3)\right]_{\mathcal{C}_W} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \left[\mathcal{T}(v_4)\right]_{\mathcal{C}_W} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Then $\mathcal{T}$ under basis $\mathcal{C}_V$ and $\mathcal{C}_W$ can be written as

$$\left[\mathcal{T}\right]_{\mathcal{C}_W \leftarrow \mathcal{C}_V} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

4.3.5. *Relations to change-of-basis matrices.* First let's write down the change-of-basis matrices.

$$P_{\mathcal{S}_V \leftarrow \mathcal{C}_V} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad P_{\mathcal{S}_W \leftarrow \mathcal{C}_W} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 3 & 2 \end{bmatrix}.$$

Here we have the formula

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = P^{-1}_{\mathcal{S}_W \leftarrow \mathcal{C}_W} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 2 \\ 1 & 3 & 2 & 4 \end{bmatrix} P_{\mathcal{S}_V \leftarrow \mathcal{C}_V}.$$

Remark 4.3.1. The above example is not an isolated example. If you have a linear transformation between two different linear spaces, you can always choose a special basis to make the matrix as an identity matrix with extra zero rows/columns. This form is called the **canonical form** of the matrix under elementary transformations.

4.4. **Similar transformation.** Let $L : V \to V$ be a linear operator. Let B and C be two bases of V. Let $P = P_{\mathcal{C} \leftarrow \mathcal{B}}$ be the change-of-basis matrix on V. Therefore the previous formula is

$$\Big[L\Big]_{\mathcal{B} \leftarrow \mathcal{B}} = P_{\mathcal{B} \leftarrow \mathcal{C}} \Big[L\Big]_{\mathcal{C} \leftarrow \mathcal{C}} P_{\mathcal{C} \leftarrow \mathcal{B}} = P^{-1} \Big[L\Big]_{\mathcal{C} \leftarrow \mathcal{C}} P.$$

Definition 4.4.1. Two matrices A_1, $A_2 \in \mathtt{Mat}_{n \times n}(\mathbb{F})$ are said to be *similar* if there is an invertible matrix $B \in \mathtt{Mat}_{n \times n}(\mathbb{F})$ such that

$$A_1 = B^{-1} A_2 B.$$

The results in the previous subsection tells us that two matrix representations of the same linear operator under different bases are similar.

Example 4.4.2. Since

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ are similar to each other.

Example 4.4.3. Let $\mathcal{T}$ be a linear operator $\mathbb{R}^2 \to \mathbb{R}^2$ defined by $\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$. This matrix can be treated as the matrix of the linear map with respect to the standard basis $\mathcal{S} = \left\{ e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, e_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ of $\mathbb{R}^2$. That is, $\begin{bmatrix} \mathcal{T} \end{bmatrix}_{\mathcal{S} \leftarrow \mathcal{S}} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$. Now we choose another basis $\mathcal{B} = \left\{ v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$. Then since

$$\mathcal{T}(v_1) = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = v_1, \quad \mathcal{T}(v_2) = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 2v_2,$$

we have

$$\begin{bmatrix} \mathcal{T} \end{bmatrix}_{\mathcal{B} \leftarrow \mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

We have the change-of-basis matrix $P_{\mathcal{S} \leftarrow \mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Then by Theorem 4.2.1 we have

$$\begin{bmatrix} \mathcal{T} \end{bmatrix}_{\mathcal{S} \leftarrow \mathcal{S}} = P_{\mathcal{S} \leftarrow \mathcal{B}} \begin{bmatrix} \mathcal{T} \end{bmatrix}_{\mathcal{B} \leftarrow \mathcal{B}} P^{-1}_{\mathcal{S} \leftarrow \mathcal{B}}.$$

Then $\begin{bmatrix} \mathcal{T} \end{bmatrix}_{\mathcal{S} \leftarrow \mathcal{S}}$ and $\begin{bmatrix} \mathcal{T} \end{bmatrix}_{\mathcal{B} \leftarrow \mathcal{B}}$ are similar to each other. In addition, since $\begin{bmatrix} \mathcal{T} \end{bmatrix}_{\mathcal{B} \leftarrow \mathcal{B}}$ is a diagonal matrix, we say that $\begin{bmatrix} \mathcal{T} \end{bmatrix}_{\mathcal{S} \leftarrow \mathcal{S}}$ is **diagonalizable**.

Remark 4.4.4. In the example, we treat a matrix as a linear map on a space. Then similar transformation is the same as changing bases. If we find a good enough basis, we can make the matrix a diagonal matrix.

Consider how we write down the matrix of a linear map. In general when we choose a basis $\{w_1, w_2\}$, we have $\mathcal{T}(w_1) = \epsilon_1 w_1 + \mu_2 w_2$, $\mathcal{T}(w_2) = \mu_1 w_1 + \epsilon_2 w_2$, and the matrix is $\begin{bmatrix} \epsilon_1 & \mu_1 \\ \mu_2 & \epsilon_2 \end{bmatrix}$.

Then the key that a matrix is diagonalizable is the existence of a basis $\{v_1, v_2\}$ such that $\mathcal{T}(v_1) = \lambda_1 v_1 + 0v_2$ and $\mathcal{T}(v_2) = 0v_1 + \lambda_2 v_2$, and in this case the matrix is $\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$.

Example 4.4.5. Consider a matrix $\begin{bmatrix} 2 \end{bmatrix}$. This is a 1×1 matrix and it defines a linear map $\mathcal{T}$ on $\mathbb{R}^1$. Let $\mathcal{B} = \left\{ v = \begin{bmatrix} a \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^1$. The matrix with respect to the basis is

$$\begin{bmatrix} \mathcal{T} \end{bmatrix}_{\mathcal{B} \leftarrow \mathcal{B}} = \begin{bmatrix} 2 \end{bmatrix}.$$

You can see that the matrix is independent of choice the basis: no matter what a is, the matrix won't be changed. Therefore you can see that in some cases, the matrix cannot be simplified by changing bases.

Example 4.4.6. Let $\mathcal{T}$ be a linear operator $\mathbb{R}^2 \to \mathbb{R}^2$ defined by $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. This matrix can be treated as the matrix of the linear map with respect to the standard basis $\mathcal{S} = \left\{ e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, e_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ of $\mathbb{R}^2$. That is, $\begin{bmatrix} \mathcal{T} \end{bmatrix}_{\mathcal{S} \leftarrow \mathcal{S}} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Now we want to diagonalize this matrix. Then we assume that we can find a basis $\mathcal{C} = \{w_1, w_2\}$ that $\mathcal{T}(w_1) = \lambda_1 w_1$ and $\mathcal{T}(w_2) = \lambda_2 w_2$. Let $w_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $w_2 = \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}$. We have

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda_1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix} = \lambda_2 \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}.$$

Then we have a system of equations:

$$x_1 + x_2 = \lambda_1 x_1, \quad x_2 = \lambda_1 x_2,$$

$$x_3 + x_4 = \lambda_2 x_3, \quad x_4 = \lambda_2 x_4.$$

If we treat λ_1 and λ_2 as (unknown) constants, we can make two linear systems:

$$\begin{cases} (1-\lambda_1)x_1 + x_2 = 0, \\ \qquad (1-\lambda_1)x_2 = 0. \end{cases}, \quad \begin{cases} (1-\lambda_2)x_3 + x_4 = 0, \\ \qquad (1-\lambda_2)x_4 = 0. \end{cases}$$

If $\lambda_1 \neq 1$, the first system has exactly one solution $x_1 = x_2 = 0$. If $\lambda_1 = 1$, the first system has infinite many solutions and the solution space is $\mathtt{Span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$. The second system has the same solutions.

Then this means it is impossible to find two linearly independent vectors w_1 and w_2 such that $\mathcal{T}(w_1) = \lambda_1 w_1$ and $\mathcal{T}(w_2) = \lambda_2 w_2$. In other words, the basis $\mathcal{C} = \{w_1, w_2\}$ such that $\mathcal{T}(w_1) = \lambda_1 w_1$ and $\mathcal{T}(w_2) = \lambda_2 w_2$ doesn't exist. Then the matrix **CANNOT** be diagonlized.

Remark 4.4.7. From this example you can see the importance of the so-called **eigenvectors**, **eigenvalues** and **eigenspaces**.

4.5. **Exercises.**

Exercise 4.1. Let $\mathcal{T} : V \to W$ be a linear transformation from $V = \mathbb{R}^3$ to $W = \mathbb{R}^4$ defined by the matrix $A = \begin{bmatrix} -1 & 2 & 3 \\ 0 & 1 & 0 \\ 1 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$. Please find a good basis for both V and W to make the matrix of $\mathcal{T}$ as simple as possible. Please also write down the change-of-basis matrices and the matrix of $\mathcal{T}$ under the new basis.

Exercise 4.2. Let $\mathcal{T} : V \to W$ be a linear transformation from $V = \mathbb{R}^3$ to $W = \mathbb{R}^2$, defined by the matrix $A = \begin{bmatrix} 1 & 3 & 2 \\ -1 & -3 & -1 \end{bmatrix}$. Please find a good basis for both V and W to make the matrix of $\mathcal{T}$ as simple as possible. Please also write down the change-of-basis matrices and the matrix of $\mathcal{T}$ under the new basis.

Exercise 4.3. Let $\mathcal{T} : V \to W$ be a linear transformation from $V = \mathbb{R}^2$ to itself, defined by the matrix $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$. Please find a good basis for V to make the matrix of $\mathcal{T}$ as simple as possible. Please also write down the change-of-basis matrices and the matrix of $\mathcal{T}$ under the new basis.

The homework is due on Apr. 12.

5. Eigenvalues, Eigenvectors, and Invariant Subspaces

5.1. **Invariant subspaces.** Let V be a vector space. Let $\mathcal{L}(V)$ be the set of operators on V. In other words, $\mathcal{L}(V)$ is the set of linear maps from V to itself.

Definition 5.1.1. Suppose $T \in \mathcal{L}(V)$. A subspace $U \subset V$ is called **invariant** under T if $\forall u \in U$, $T(u) \in U$.

Remark 5.1.2. In other words, U is invariant under T if $T|_U$ makes sense. Then $T|_U$ is an operator on U.

Proposition 5.1.3. *Suppose $T \in \mathcal{L}(V)$. The following subspaces of V is invariant under T:*

(1) $\{0\}$.

Proof. 0 is the only one vector in $\{0\}$. Since $T(0) = 0 \in \{0\}$, $\{0\}$ is invariant under T. □

(2) V.

Proof. Since V is the codomain, the image has to be in V. Then V is invariant under T. □

(3) $\texttt{Nul}(T)$.

Proof. Recall that $\texttt{Nul}(T) = \{v \in V \mid T(v) = 0\}$. Then for any $u \in \texttt{Nul}(T)$, $T(u) = 0 \in \texttt{Nul}(T)$. So $\texttt{Nul}(T)$ is invariant under T. □

(4) $\texttt{im}(T)$.

Proof. Recall that $\texttt{im}(T) = \{v \in V \mid \exists u \in V \text{ such that } T(u) = v\}$. Then for any $w \in \texttt{im}(T)$, $T(w) \in \texttt{im}(T)$. Then $\texttt{im}(T)$ is invariant under T. □

Example 5.1.4. Let P_n be the space of polynomials whose order is no more than n. Then we automatically have $P_1 \subset P_2 \subset P_3 \subset P_4 \subset \dots$. Consider P_4. The derivative operator D is an linear operator on P_4. It is easy to check that P_1, P_2, P_3 are all invariant under D.

Example 5.1.5. Consider the rotation $R(\theta)$ about z-axis in $\mathbb{R}^3$ by an angle θ. $R(\theta)$ is a linear operator on $\mathbb{R}^3$. z-axis and xy-plane are two invariant subspaces under $R(\theta)$.

Example 5.1.6. Consdier the rotation $R(\theta)$ about the origin in $\mathbb{R}^2$ by an angle θ. $R(\theta)$ is a linear operator on $\mathbb{R}^2$. It is easy to check that the only invariant subspace is $\{0\}$.

5.2. **Eigenvalues and Eigenvectors.**

Definition 5.2.1. Suppose $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbb{F}$ is called an **eigenvalue** of T if there exists $v \in V$ such that $v \neq 0$ and $T(v) = \lambda v$. Then in this case v is called an **eigenvector** of T corresponding to the eigenvalue λ.

It is straightforward that to find the eigenvalues and the eigenvectors we hope that the equation $T(v) = \lambda v$ has non-zero solutions. The equation can be rewritten as $(T - \lambda I)(v) = 0$. Then if the equation has non-zero solutions, the square matrix $T - \lambda I$ cannot be invertible. These observations are summarized below.

Proposition 5.2.2. *Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then the following are equivalent:*

(1) λ is an eigenvalue of T;

(2) $T - \lambda I$ is not injective;

(3) $T - \lambda I$ is not surjective;

(4) $T - \lambda I$ is not ivnertible.

Theorem 5.2.3. *Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and $v_1, \ldots, v_m$ are corresponding eigenvectors. Then $\{v_1, \ldots, v_m\}$ is linearly independent.*

Proof. Use contradiction. Assume that $\{v_1, \ldots, v_m\}$ is NOT linearly independent. Then there is a smallest number k such that $v_k \in \texttt{Span}\{v_1, \ldots, v_{k-1}\}$. The smallestness means that $\{v_1, \ldots, v_{k-1}\}$ is linear independent. Then there exists $a_1, \ldots, a_k \in \mathbb{F}$ such that

$$v_k = a_1 v_1 + a_2 v_2 + \ldots + a_{k-1} v_{k-1}. \tag{5.1}$$

Then we have

$$\lambda_k v_k = \lambda_k (a_1 v_1 + a_2 v_2 + \ldots + a_{k-1} v_{k-1}) = a_1 \lambda_k v_1 + a_2 \lambda_k v_2 + \ldots + a_{k-1} \lambda_k v_{k-1}. \tag{5.2}$$

Now apply T on both sides of Equation 5.1. Since $T(v_i) = \lambda_i v_i$ for any $i = 1, \ldots, n$, we have

$$\lambda_k v_k = a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2 + \ldots + a_{k-1} \lambda_{k-1} v_{k-1}. \tag{5.3}$$

Therefore using Equation 5.3 substract Equation 5.2, we have

$$a_1(\lambda_1 - \lambda_k)v_1 + a_2(\lambda_2 - \lambda_k)v_2 + \ldots + a_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} = 0.$$

Since $\{v_1, \ldots, v_{k-1}\}$ is linearly independent,

$$a_1(\lambda_1 - \lambda_k) = a_2(\lambda_2 - \lambda_k) = \ldots = a_{k-1}(\lambda_{k-1} - \lambda_k) = 0.$$

Since all λ_i's are distinct, $\lambda_i - \lambda_k \neq 0$ for all $i = 1, \ldots, k-1$. Therefore $a_1 = a_2 = \ldots = a_{k-1} = 0$. Then $v_k = 0$. This is a contradiction. Then the assumption is wrong. Therefore $\{v_1, \ldots, v_m\}$ is linearly independent. □

Corollary 5.2.4. *Suppose V is finite-dimensional. Then each operator on V has at most* $\dim V$ *distinct eigenvalues.*

Proof. Exercise. □

5.3. **Eigenspaces.**

Definition 5.3.1. Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. The **eigenspace** of T corresponding to λ, denoted V_λ, is defined by $V_\lambda = \texttt{Nul}(T - \lambda I)$. In other words, V_λ is the set of all eigenvectors of T corresponding to λ, along with the 0 vector.

An eigenspace can be treated as the span of eigenvectors corresponding to the same eigenvectors.

Theorem 5.3.2 (Sum of eigenspaces is a direct sum)**.** *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Suppose also that $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T. Then $V_{\lambda_1} + \ldots V_{\lambda_m}$ is a direct sum. Furthermore*

$$\dim V_{\lambda_1} + \ldots + \dim V_{\lambda_m} \leq \dim V.$$

Proof. By Theorem 5.2.3, $V_{\lambda_i} \cap V_{\lambda_j} = \{0\}$ for $i \neq j$. Then the sum is a direct sum and the sum of the dimension is smaller or equal to the dimension of V. □

Review: Polynomials.

Definition 5.3.3 (Polynomial)**.** A function $p : \mathbb{F} \to \mathbb{F}$ is called a **polynomial** with coefficients in $\mathbb{F}$ if there exist $a_0, \ldots, a_m \in \mathbb{F}$ such that for all $t \in \mathbb{F}$

$$p(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_m t^m.$$

Definition 5.3.4. A polynomial p is said to have **degree** m if there exists scalars $a_0, \ldots, a_m \in \mathbb{F}$ with $a_m \neq 0$ such that $p(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_m t^m$ for all $t \in \mathbb{F}$. We write $\deg p = m$. The degree of 0 is set to be $-\infty$.

Proposition 5.3.5. *(1) The set of all polynomials with addition of functions and scalar product of functions is a vector space. It is denoted by $P(\mathbb{F})$.*

(2) The subset of all polynomials with degree at most m is a subspace of $P(\mathbb{F})$. It is denoted by $P_m(\mathbb{F})$. If there are no confusions, we may just write P_m.

Definition 5.3.6 (Product of polynomials)**.** If $p, q \in P(\mathbb{F})$, then $pq \in P(\mathbb{F})$ is defined by $(pq)(t) = p(t)q(t)$ for any $t \in \mathbb{F}$.

Remark 5.3.7. The product of polynomials are different from the scalar product of polynomials.

Proposition 5.3.8. *Suppose a polynomial $p(t) = a_0 + a_1 t + \ldots + a_m t^m = 0$ for all $t \in \mathbb{F}$. Then $a_0 = a_1 = \ldots = a_m = 0$.*

Theorem 5.3.9. *$\{1, t, t^2, \ldots, t^m\}$ forms a basis of $P_m(\mathbb{F})$. Then $\dim P_m = m + 1$.*

Definition 5.3.10. A number $a \in \mathbb{F}$ such that $p(a) = 0$ is called a **zero** (or **root**) of the polynomial p.

Definition 5.3.11. A polynomial $s \in P(\mathbb{F})$ is called a **factor** (or **divisor**) of $p \in P(\mathbb{F})$ if there exists a polynomial $q \in P(\mathbb{F})$ such that $p = sq$.

Proposition 5.3.12 (Each zero of a polynomial corresponds to a degree-1 factor)**.** *Suppose $p \in P(\mathbb{F})$ and $\lambda \in \mathbb{F}$. Then $p(\lambda) = 0$ if and only if there is a polynomial $q \in P(\mathbb{F})$ such that $p(t) = (t - \lambda)q(t)$ for every $t \in \mathbb{F}$.*

Corollary 5.3.13 (A polynomial has at most as many zeros as its degree)**.** *Suppose $p \in P(\mathbb{F})$ is a polynomial with degree $m \geq 0$. Then p has at most m distinct zeros in $\mathbb{F}$.*

Theorem 5.3.14 (Fundamental Theorem of Algebra)**.** *Every non-constant polynomial with complex coefficients has a zero.*

Then every non-constant polynomial with complex coefficients has a degree-1 factor. Then every non-constant polynomial with complex coefficients can be written as a product of degree-1 polynomials. To summarize:

Corollary 5.3.15. *If $p \in P(\mathbb{C})$ is a non-constant polynomial, then p has a unique factorization (up to reordering the factors) of the form $p(t) = c(t - \lambda_1)\dots(t - \lambda_m)$ where $c, \lambda_1, \dots, \lambda_m \in \mathbb{C}$.*

Definition 5.3.16. Suppose $T \in \mathcal{L}(V)$ and $m \in \mathbb{N}$. Then T^m is defined by $T^m = T \cdots T$, and T^0 is defined by $T^0 = I$. If T is invertible, with the inverse T^{-1}, then T^{-m} is defined by $T^{-m} = (T^{-1})^m$.

Proposition 5.3.17. $T^{m+n} = T^m T^n$, $(T^m)^n = T^{mn}$.

Definition 5.3.18. Suppose $T \in \mathcal{L}(V)$ and $p \in P(\mathbb{F})$ is a polynomial given by $p(t) = a_0 + a_1 t + \dots + a_m t^m$ for $t \in \mathbb{F}$. Then $p(T)$ is the operator defined by

$$p(T) = a_0 I + a_1 T + \dots + a_m T^m \in \mathcal{L}(V).$$

Proposition 5.3.19. *Suppose $p, q \in P(\mathbb{F})$ and $T \in \mathcal{L}(V)$. Then*

(1) $(pq)(T) = p(T)q(T)$;

(2) $p(T)q(T) = q(T)p(T)$.

Remark 5.3.20. Note that complex number plays a very important role in the theory. We will review complex numbers later. At current stage you need to know that complex numbers are essential for solving equations but I won't ask you to do computations involving complex numbers before the review of complex numbers.

5.4. **Upper-triangular matrices.**

Theorem 5.4.1 (Over $\mathbb{C}$, every operator has an upper-triangular matrix)**.** *Suppose V is a finite-dimensional non-zero complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V.*

To prove this Theorem, we need the help from the following Theorem.

Theorem 5.4.2 (Operators on complex vector spaces have an eigenvalue)**.** *Every operator on a finite-dimensional, nonzero, complex vector space has an eigenvalue.*

Remark 5.4.3. Let us look at what it means by "an operator has an upper-triangular matrix". Let V be a 3-dimensional vector space. Assume that the matrix of a linear operator with respect to a basis $\mathcal{B} = \{v_1, v_2, v_3\}$ is an upper-triangular matrix of size 3×3:

$$\Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}.$$

Then we have that

$$T(v_1) = a_{11}v_1,$$

$$T(v_2) = a_{12}v_1 + a_{22}v_2,$$

$$T(v_3) = a_{13}v_1 + a_{23}v_2 + a_{33}v_3.$$

Therefore

$$T(v_1) \in \texttt{Span}\,\{v_1\},$$

$$T(v_1), T(v_2) \in \texttt{Span}\,\{v_1, v_2\},$$

$$T(v_1), T(v_2), T(v_3) \in \texttt{Span}\,\{v_1, v_2, v_3\}.$$

In other words, we have a sequence of invariant subspaces:

$$\{0\} \subset \texttt{Span}\,\{v_1\} \subset \texttt{Span}\,\{v_1, v_2\} \subset \texttt{Span}\,\{v_1, v_2, v_3\} = V.$$

To prove that a linear operator can be written as a upper-triangular matrix, we just need to prove that such a basis (or a chain of invariant subspaces) exists.

Review of quotient spaces.

Definition 5.4.4. Let V be a vector space and W a subspace. W defines an equivalence relation on V by $v \sim u$ for $v, u \in V$ if $v - u \in W$. Let $v + W$ be the equivalence class of $v \in V$. Define $(v + W) + (u + W) = (v + u) + W$ and $c(v + W) = cv + W$ for $v, u \in V$ and $c \in \mathbb{F}$. The theorem says that these two operations are well-defined and the set of equivalence classes with these two operations is a vector space. It is called the **quotient space** of V by W and is denoted by V/W. In a class $v + W$, the $v \in v + W$ is called a **representative**, and it can be any vector in the class.

Remark 5.4.5. Another notation for the class $v + W$ is $[v]$ for $v \in V$. Therefore the above formulas can also be

(1) $[v] = [u]$ if and only if $v - u \in W$. In particular, $[v] = [0]$ if and only if $v \in W$.
(2) $[v] + [u] = [v + u]$, $c[v] = [cv]$.
(3) $u \in V$ is called a representative of $[v]$ if $u \in [v] = v + W$.

Remark 5.4.6. (1) Sometimes we just use 0 to denote $[0]$. You need to read from contexts to determine whether 0 is a real 0 or a class $[0]$.
(2) Note that you need to read from contexts to determine whether $[v]$ is an equivalence class or a coordinate.

Example 5.4.7. Let $V = \mathbb{R}^2$ and $W = \texttt{Span}\left(\begin{bmatrix}1\\1\end{bmatrix}\right)$. Then the equivalence relation is that $v \sim u$ if and only if $v - u = \begin{bmatrix}1\\1\end{bmatrix}$. For example, $\begin{bmatrix}1\\2\end{bmatrix} \sim \begin{bmatrix}2\\3\end{bmatrix}$, $\begin{bmatrix}100\\99\end{bmatrix} \sim \begin{bmatrix}0\\-1\end{bmatrix}$, $\begin{bmatrix}1\\1\end{bmatrix} \not\sim \begin{bmatrix}2\\1\end{bmatrix}$. Then

$$\begin{bmatrix}1\\2\end{bmatrix} + W, \quad \begin{bmatrix}2\\3\end{bmatrix} + W, \quad \begin{bmatrix}100\\99\end{bmatrix} + W, \quad \begin{bmatrix}0\\-1\end{bmatrix} + W, \quad \begin{bmatrix}1\\1\end{bmatrix} + W, \quad \begin{bmatrix}2\\1\end{bmatrix} + W \in V/W,$$

and

$$\begin{bmatrix}1\\2\end{bmatrix} + W = \begin{bmatrix}2\\3\end{bmatrix} + W, \quad \begin{bmatrix}100\\99\end{bmatrix} + W = \begin{bmatrix}0\\-1\end{bmatrix} + W, \quad \begin{bmatrix}1\\1\end{bmatrix} + W \neq \begin{bmatrix}2\\1\end{bmatrix} + W.$$

We can also perform addition and scalar product. For example,

$$\left(\begin{bmatrix}1\\2\end{bmatrix}+W\right)+\left(\begin{bmatrix}2\\3\end{bmatrix}+W\right)=\begin{bmatrix}3\\5\end{bmatrix}+W,$$

$$\left(\begin{bmatrix}100\\99\end{bmatrix}+W\right)+\left(\begin{bmatrix}0\\-1\end{bmatrix}+W\right)=\begin{bmatrix}100\\98\end{bmatrix}+W,$$

$$3\left(\begin{bmatrix}2\\1\end{bmatrix}+W\right)=\begin{bmatrix}6\\3\end{bmatrix}+W.$$

Proposition 5.4.8. *Let $T \in \mathcal{L}(V)$ and W is an invariant subspace under T. Then the operation $\overline{T} : V/W \to V/W$ defined by*

$$\overline{T}([v]) = [T(v)]$$

is well-defined and gives a linear operator on V/W. This operator is called the ***quotient operator***.

Proof. We will mainly focus on the well-definedness part. The linear operator part is straightforward. Pick two representatives v and u from a class $[v]$. Then by definition $v - u \in W$. Then $\overline{T}([v]) = [T(v)]$ and $\overline{T}([u]) = [T(u)]$. Since W is invariant, and $v - u \in W$, then $T(v - u) \in W$. So $T(v) - T(u) \in W$. Then $[T(v)] = [T(u)]$. This means that $\overline{T}$ is independent of choice of representatives. Therefore it is well-defined. □

Example 5.4.9. Let T be the rotation about the z-axis by $90°$ counterclockwise in $V = \mathbb{R}^3$. z-axis is an invariant space. Denote it by W. The quotient space V/W consists of classes $[(x, y)] = \{(x, y, z) \mid z \in \mathbb{R}\}$. The quotient operator behaves as $\overline{T}([(x, y)]) = [(-y, x)]$.

Remark 5.4.10. The key observation of the quotient operator $\overline{T}$ is that $\overline{T}([v]) = [T(v)] = [u]$ means that $T(v) - u \in W$. In other words, $T(v)$ is a linear combination of vectors of W and u. If we choose a basis of W, $\mathcal{C} = \{w_1, \ldots, w_m\}$, then $T(v) = a_u u + a_1 w_1 + \ldots + a_m w_m$ for some constants $a_u, a_1, \ldots, a_m \in \mathbb{F}$. Then $T(v) \in \texttt{Span}\,\{u, w_1, \ldots, w_m\}$. Note that when $u \notin W$, u and $\mathcal{C}$ are linearly independent. Therefore in this case $T(v) \in \texttt{Span}(u) \oplus W$.

5.4.1. *Proof of Theorem 5.4.1.* Use induction on $\dim V$.

When $\dim V = 1$, the matrix is a 1×1 matrix which is an upper-triangular matrix.

Assume that for any k-dimensional vector space the linear operator on it admits a basis such that the matrix is an upper-triangular matrix. Let V be an arbitrary $k+1$-dimensional vector space and $\mathcal{T} \in \mathcal{L}(V)$. By Theorem 5.4.2, $\mathcal{T}$ has an eigenvalue. Then there exists $\lambda \in \mathbb{C}$, $v \in V$ such that $v \neq 0$ and $\mathcal{T}(v) = \lambda v$. Then $\mathtt{Span}(v)$ is a 1-dimensional invariant subspace of V under $\mathcal{T}$. Then by Proposition 5.4.8, we have an operator $\overline{\mathcal{T}}$ on $V/\mathtt{Span}(v)$. Since $\dim V = k+1$, $\dim V/\mathtt{Span}(v) = k$. Then by the induction assumption, there exists a basis of $V/\mathtt{Span}(v)$ such that the matrix of $\overline{\mathcal{T}}$ is upper-triangular.

By Remark 5.4.3, there exists a basis $\{[v_1], \ldots, [v_k]\}$ of $V/\mathtt{Span}(v)$ such that

$$\overline{\mathcal{T}}([v_1]) = a_{11}[v_1] = [a_{11}v_1],$$

$$\overline{\mathcal{T}}([v_2]) = a_{12}[v_1] + a_{22}[v_2] = [a_{12}v_1 + a_{22}v_2],$$

$$\cdots\cdots$$

$$\overline{\mathcal{T}}([v_k]) = a_{1k}[v_1] + a_{2k}[v_2] + \ldots + a_{kk}[v_k] = [a_{1k}v_1 + \ldots + a_{kk}v_k].$$

By Remark 5.4.10, we have

$$\mathcal{T}(v_1) \in \mathtt{Span}(v, v_1),$$

$$\mathcal{T}(v_2) \in \mathtt{Span}(v, v_1, v_2),$$

$$\cdots\cdots$$

$$\mathcal{T}(v_k) \in \mathtt{Span}(v, v_1, v_2, \ldots, v_k).$$

What's more, since v is an eigenvector of $\mathcal{T}$, $\mathcal{T}(v) \in \mathtt{Span}(v)$. If we can prove that $\mathcal{B} = \{v, v_1, \ldots, v_k\}$ can form a basis of V, this is the chain of invariant subspaces we need from Remark 5.4.3. Then under the basis $\mathcal{B}$, the matrix of $\mathcal{T}$ is upper-triangular.

Then by induction, for any finite-dimensional vector space there exists a basis such that the matrix of $\mathcal{T}$ is upper-triangular. $\square$

Proof of $\mathcal{B}$ being a basis. Consider the quotient space $V/\texttt{Span}(v)$. If $w \in [u] \in V/\texttt{Span}(v)$ for some $w \in V$, then by definition of quotient spaces, $w - u \in \texttt{Span}(v)$. In other words, $\exists r \in \mathbb{F}$ such that $w = rv + u$.

Then for any $w \in V$, $w \in [w] \in V/\texttt{Span}(v)$. Since $\{[v_1], \ldots, [v_k]\}$ is a basis of $V/\texttt{Span}(v)$, there exists $c_1, \ldots, c_k \in \mathbb{F}$ such that

$$[w] = c_1[v_1] + \ldots + c_k[v_k] = [c_1 v_1 + \ldots + c_k v_k].$$

Then there exists $r \in \mathbb{F}$ such that

$$w = rv + c_1 v_1 + \ldots + c_k v_k.$$

So $w \in \texttt{Span}(v, v_1, \ldots, v_k)$. Then $\texttt{Span}(v, v_1, \ldots, v_k) = V$. Since we already know that $\dim V = k + 1$, then $\{v, v_1, \ldots, v_k\}$ has to be linearly independent. Therefore it is a basis of V. □

Remark 5.4.11. It is possible to prove the linearly independence directly by definition without computing the dimensions. It is an exercise.

5.4.2. *Proof of Theorem 5.4.2.* Let V be a finite-dimensional nonzero complex vector space and $T \in \mathcal{L}(V)$. Let $v \in V$ be a vector. Consider the set $\{v, T(v), T^2(v), \ldots\}$. Since V is finite-dimensional, there exists a maximal $k \geq 1$ such that $\{v, T(v), T^2(v), \ldots, T^{k-1}(v)\}$ is linearly independent. By definition of k, $T(T^{k-1}(v)) \in \texttt{Span}\left(v, T(v) \ldots, T^{k-1}(v)\right)$. Then there exists some $a_0, a_1, \ldots, a_{k-1} \in \mathbb{C}$ such that

$$\begin{aligned} T^k(v) = T(T^{k-1}(v)) &= a_0 v + a_1 T(v) + \ldots + a_{k-1} T^{k-1}(v) \\ &= \left(a_0 I + a_1 T + a_2 T^2 + \ldots + a_{k-1} T^{k-1}\right)(v). \end{aligned}$$

Then we have

$$\left(T^k - a_0 I - a_1 T - a_2 T^2 - \ldots - a_{k-1} T^{k-1}\right)(v) = 0.$$

Let $p(x) = x^k - a_0 - a_1 x - a_2 x^2 - \ldots - a_{k-1} x^{k-1}$ be a polynomial in $\mathbb{C}$. Then we have $p(T)(v) = 0$. Let $p(x) = (x - c_1)(x - c_2) \ldots (x - c_k)$ be the factorization in $\mathbb{C}$ for some $c_1, \ldots, c_k \in \mathbb{C}$. Then by Proposition 5.3.19, we have

$$(T-c_1I)(T-c_2I)\dots(T-c_kI)(v)=0.$$

Now consider the sequence of vectors $v_{k+1}, v_k, v_{k-1}, \dots, v_2, v_1$ for

$$\begin{aligned}
v_{k+1} &= v,\\
v_k &= (T-c_kI)(v) = (T-c_kI)(v_{k+1}),\\
v_{k-1} &= (T-c_{k-1}I)(T-c_kI)(v) = (T-c_{k-1}I)(v_k),\\
v_{k-2} &= (T-c_{k-2}I)(T-c_{k-1}I)(T-c_kI)(v) = (T-c_{k-2}I)(v_{k-1}),\\
&\dots\dots\\
v_2 &= (T-c_2I)\dots(T-c_k)(v) = (T-c_2I)(v_3),\\
v_1 &= (T-c_1I)(T-c_2I)\dots(T-c_kI)(v) = (T-c_1I)(v_2) = 0.
\end{aligned}$$

Since $v_{k+1}\neq 0$ and $v_1 = 0$, there has to be a $1\leq r\leq k$ such that $v_r = 0$ and $v_{r+1}\neq 0$. Then $(T-c_rI)(v_{r+1}) = 0$ and $v_{r+1}\neq 0$. Therefore v_{r+1} is an eigenvector of T corresponding to the eigenvalue c_r. □

Example 5.4.12. The proof of the above two theorems actually gives us a strategy to find a basis to change a matrix into upper-triangular matrix. Let $A = \begin{bmatrix} 2 & 0 & 0\\ 0 & 2 & 1\\ 1 & 0 & 2\end{bmatrix}$. We might treat it as the matrix of linear operator T on $\mathbb{C}^3$ with respect to the standard basis. Following the steps of the proofs:

(1) Starting from an arbitrary vector. Let $v = \begin{bmatrix}1\\0\\0\end{bmatrix}$. Then since $T(v) = \begin{bmatrix}2\\0\\1\end{bmatrix} \notin \mathtt{Span}(v)$, $T^2(v) = \begin{bmatrix}4\\1\\4\end{bmatrix} \notin \mathtt{Span}(v, T(v))$, $\{v, T(v), T^2(v)\}$ forms a basis of $\mathbb{C}^3$. Then $T^3(v) = \begin{bmatrix}8\\6\\12\end{bmatrix} = 8v - 12T(v) + 6T^2(v)$. Thus $(T^3 - 6T^2 + 12T - 8I)(v) = 0$.

Since $x^3 - 6x^2 + 12x - 8 = (x-2)^3$, we have

$$(T - 2I)^3(v) = 0.$$

Consider the sequence of vectors:

$$v = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad (T-2I)(v) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad (T-2I)^2(v) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad (T-2I)^3(v) = 0.$$

Then $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ is an eigenvector, with eigenvalue 2. Let $v_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

(2) Consider $\mathbb{C}^3/\,\mathtt{Span}(v_1)$. There is an obvious map $\mathbb{C}^3/\,\mathtt{Span}(v_1) \to \mathbb{C}^2$ by

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \mathtt{Span}(v_1) \mapsto \begin{bmatrix} c_1 \\ c_3 \end{bmatrix}.$$

It is easy to check that this is an isomorphism. We can use it to simplify our computation. In this case the quotient map $\overline{T}$ is gotten by

$$\overline{T}\left(\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}\right) \simeq \overline{T}\left(\begin{bmatrix} c_1 \\ 0 \\ c_2 \end{bmatrix} + \mathtt{Span}(v_1)\right) = \left[T\left(\begin{bmatrix} c_1 \\ 0 \\ c_2 \end{bmatrix}\right) + \mathtt{Span}(v_1)\right]$$

$$= \begin{bmatrix} 2c_1 \\ 0 \\ c_1 + 2c_2 \end{bmatrix} + \mathtt{Span}(v_1) \simeq \begin{bmatrix} 2c_1 \\ c_1 + 2c_2 \end{bmatrix}.$$

Therefore the matrix of $\overline{T}$ with the basis $\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$ in $\mathbb{C}^2 \simeq \mathbb{C}^3/\,\mathtt{Span}(v_1)$ is $\begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$.

(3) Apply the first step to this new matrix $\begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$. We find an eigenvector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ with eigenvalue 2. The vector is corresponding to $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \texttt{Span}(v_1) \subset \mathbb{C}^3$. We can pick an arbitrary representative from the class. For example, we may pick $v_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$. Now we have the second invariant subspace $\texttt{Span}(v_1, v_2)$.

(4) Apply the second step to this new invariant subspace. Consider $\mathbb{C}^3/\texttt{Span}(v_1, v_2)$. There is an obvious isomorphism $\mathbb{C}^3/\texttt{Span}(v_1, v_2) \simeq \mathbb{C}$ by

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \texttt{Span}(v_1, v_2) \mapsto \begin{bmatrix} c_1 \end{bmatrix}.$$

The matrix of $\overline{T}$ is easy to be computed, and it is $\begin{bmatrix} 2 \end{bmatrix}$. Since $\mathbb{C}^3/\texttt{Span}(v_1, v_2)$ is 1-dimensional, any non-zero vector is an eigenvector. Therefore $\begin{bmatrix} 1 \end{bmatrix}$ is an eigenvector with eigenvalue 2. The vector is corresponding to $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \texttt{Span}(v_1, v_2) \subset \mathbb{C}^3$. We can pick an arbitrary representative from the class to be our basis vector. For example, we may use $v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

(5) Since $\texttt{Span}(v_1, v_2, v_3) = \mathbb{C}^3$, the search is finished. $\{v_1, v_2, v_3\}$ is a basis we need. It is easy to check that the matrix of T with respect to this basis is

$$\begin{bmatrix} 2 & 3 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}.$$

Remark 5.4.13. This example just reproduce the process to prove Theorem 5.4.1 and Theorem 5.4.2. What we get is an upper-triangular matrix. An upper-triangular matrix is not the easiest form. In the following lectures we will talk about more theorems to refine the results to make the matrix easier.

5.4.3. *Some applications.* One of the reason that we want an upper-triangular matrix is that it is good for computation. Recall that the matrix of the same linear operator under different bases are similar to each other. Let A and U be two similar matrices. Then there exists invertible P such that

$$A = PUP^{-1}.$$

Since $\det(AB) = \det(A)\det(B)$, we have $\det(A) = \det(P)\det(U)\det(P)^{-1} = \det(U)$. Then we can use the determinant of the upper-triangular matrix to compute the determinant of the original matrix. The determinant of the upper-triangular matrix is just the product of diagonal, so the computation is very easy.

Corollary 5.4.14. *Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then T is invertible if and only if all the diagonal entries of the upper-triangular matrix are non-zero.*

Corollary 5.4.15. *Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then the eigenvalues of T are precisely the entries on the diagonal of that upper-triangular matrix.*

Proof. Recall from Proposition 5.2.2 that the eigenvalues are those numbers λ which make $T - \lambda I$ is NOT invertible. Then to find eigenvalues we only need to find any basis of V to write T as a matrix A and solve the equation $\det(A - \lambda I) = 0$ for λ. Note that we can choose a basis to make A an upper-triangular matrix $\begin{bmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$. The equation $\det(A - \lambda I) = 0$ is precisely $(\lambda_1 - \lambda)(\lambda_2 - \lambda)\dots(\lambda_n - \lambda) = 0$. The solutions $\lambda_1, \dots, \lambda_n$ are eigenvalues, which are also entries on the diagonal of that upper-triangular matrix. $\square$

5.5. **Diagonalizable operators.** Any matrix is similar to an upper-triangular matrix. A special type of upper-triangular matrix is diagonal matrix.

Definition 5.5.1 (diagonalizable)**.** An operator $T \in \mathcal{L}(V)$ is called **diagonalizable** if the operator has a diagonal matrix with respect to some basis of V.

Theorem 5.5.2 (Conditions equivalent to diagonalizability)**.** *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T. Then the following are equivalent:*

(1) T is diagonalizable;

(2) V has a basis consisting of eigenvectors of T;

(3) there exist 1-dimensional subspaces $U_1, \ldots, U_n$ of V, each invariant under T, such that $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$;

(4) $V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \ldots \oplus V_{\lambda_m}$;

(5) $\dim V = \dim V_{\lambda_1} + \ldots + \dim V_{\lambda_m}$.

Proof. To prove TFAE, we will prove the following chain:

$$(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (5) \Rightarrow (2) \Rightarrow (1).$$

(1) $\Rightarrow$ (2)**:** T is diagonalizable, then by definition there is a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ such that $\Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}}$ is a diagonal matrix. Let the diagonal be $\lambda_1, \ldots, \lambda_n$. Then we have $T(v_i) = \lambda_i v_i$ for any $i = 1, \ldots, n$. Then the basis $\mathcal{B}$ consisting of eigenvectors of T.

(2) $\Rightarrow$ (3)**:** Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be the basis consisting of eigenvectors of T. Then $U_i = \texttt{Span}(v_i)$ are all 1-dimensional invariant subspace under T for any $i = 1, \ldots, n$. Then by the definition of basis, $V = U_1 \oplus \ldots \oplus U_n$.

(3) $\Rightarrow$ (4)**:** Since U_i's are all 1-dimensional invariant subspaces, they are all eigenspaces. Then each vectors can be written as a linear combination of eigenvectors. So $V = V_{\lambda_1} + V_{\lambda_2} + \ldots + V_{\lambda_m}$. By Theorem 5.3.2, the sum is a direct sum. Then $V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \ldots \oplus V_{\lambda_m}$.

(4) $\Rightarrow$ (5)**:** This is obvious.

(5) $\Rightarrow$ (2)**:** Choose a basis for each V_{λ_i} for each $i = 1, \ldots, m$, and put all these vector together to form a set $\mathcal{B}$. This is a set of eigenvectors. Since $V_{\lambda_i} \cap V_{\lambda_j} = \{0\}$ for $i \neq j$, $\mathcal{B}$ is a linearly

independent set. We can extend the set to be a basis of V. Since $\dim V = \dim V_{\lambda_1} + \ldots + \dim V_{\lambda_m}$, we don't need add any vectors to get a basis. This means that $\mathcal{B}$ is a basis. Then we get a basis of B consisting of eigenvectors of T.

(2) $\Rightarrow$ (1): Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be the basis consisting egenvectors of T. Let the corresponding eigenvalues be $\lambda_1, \ldots, \lambda_n$. Then since $T(v_i) = \lambda_i v_i$ for any $i = 1, \ldots, n$, the matrix $\Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}}$ is a diagonal matrix.

□

Corollary 5.5.3. *If a matrix is diagonalizable, the diagonal of the digonalized matrix consists of eigenvalues of the orginal matrix.*

Proof. This is obvious from the computation in the above proof (1) $\Rightarrow$ (2). □

Example 5.5.4. $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is not diagonalizable. Since

Reason 1: You can try to find a basis as in Section 4 to make a diagonal matrix, but you will see that the equation has no solutions.

Reason 2: A only has one 1-dimensional eigenspace, then eigenvectros cannot make a basis. This violates (2)-(5) in the Theorem above.

Example 5.5.5. $A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$. Compute eigenvalues:

$$\det(A - \lambda I) = \det\left(\begin{bmatrix} 2-\lambda & 0 & 0 \\ 1 & 3-\lambda & 1 \\ 0 & 0 & 3-\lambda \end{bmatrix}\right) = (2-\lambda)(3-\lambda)^2.$$

Set up the equation $\det(A - \lambda I) = 0$. The solution is $\lambda_1 = 2$ and $\lambda_2 = 3$.

$\lambda_1 = 2$: Solve the equation $(A - 2I)x = 0$:

$$\left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right].$$

The solution is that $\texttt{Span}\left(\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}\right)$.

$\lambda_2 = 3$**:** Solve the equation $(A - 3I)x = 0$:

$$\left[\begin{array}{ccc|c} -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The solution is that $\texttt{Span}\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right)$.

Since we have only two eigenspaces each of which is 1-dimensional, the matrix A is not diagonalizable.

From Theorem 5.5.2, we know that to have a diagonalizable operator, we need to have a basis consisting of eigenvectors. Here we have a speical case which is easy.

Theorem 5.5.6 (Enough eigenvalues implies diagonalizability)**.** *If* $T \in \mathcal{L}(V)$ *has* $\dim V$ *distinct eigenvalues, then* T *is diagonalizable.*

Proof. It follows Theorem 5.2.3 and Theorem 5.5.2. □

Example 5.5.7. $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Solve $\det(A - \lambda I) = 0$. The solutions are $\frac{5}{2} \pm \frac{\sqrt{33}}{2}$. Since there are two distinct eigenvalues, A is diagonalizable, and $A \sim \begin{bmatrix} \frac{5}{2} + \frac{\sqrt{33}}{2} & 0 \\ 0 & \frac{5}{2} - \frac{\sqrt{33}}{2} \end{bmatrix}$.

5.6. **Summary and introduction to Section 8.** Let $T \in \mathcal{L}(V)$ and $U \subset V$ be an invariant subspace under T. Then T restricted on U is also an operator, which is called a restricted operator. When considering a restricted operator you can totally ignore its action outside U and treat it as an operator on U only. It obeys all rules that an operator on U obeys.

Example 5.6.1. Let $V = \mathbb{C}^4$ and $\mathcal{S}$ be the standard basis. Let $T \in \mathcal{L}(V)$ is defined by

$$\left[T\right]_{\mathcal{S}\leftarrow\mathcal{S}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This means that

$$T(e_1) = e_1, \quad T(e_2) = e_2 + e_1, \quad T(e_3) = e_3 + e_2, \quad T(e_4) = e_4 + e_3.$$

Then $\texttt{Span}(e_1)$, $\texttt{Span}(e_1, e_2)$, $\texttt{Span}(e_1, e_2, e_3)$ and $\texttt{Span}(e_1, e_2, e_3, e_4) = V$ are all invariant subspaces. However all other combinations, like $\texttt{Span}(e_2)$, $\texttt{Span}(e_3, e_4)$, $\texttt{Span}(e_2, e_3, e_4)$, etc. are not invariant. Actually you can prove that, it is impossible to find two invariant subspaces $U, W \subset V$ such that $U, W \neq \{0\}$ and $V = U \oplus W$.

Proof. Assume that such a decomposition exists. Since $U, W \subset V$ are both invariant subspaces, then $T|_U \in \mathcal{L}(U)$ and $T|_W \in \mathcal{L}(W)$. Then T has an eigenvector in U and an eigenvector in W. By definition of eigenvectors, both eigenvectors are also eigenvectors of T in V. Since $V \cap W = \{0\}$, these two eigenvectors are different. However it is easy to compute that T has only one eigenvector in V. Therefore such an decomposition doesn't exist. □

Example 5.6.2. Let $V = \mathbb{C}^4$ and $\mathcal{S}$ be the standard basis. Let $T \in \mathcal{L}(V)$ is defined by

$$\left[T\right]_{\mathcal{S}\leftarrow\mathcal{S}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This means that

$$T(e_1) = e_1, \quad T(e_2) = e_2 + e_1, \quad T(e_3) = e_3, \quad T(e_4) = e_4 + e_3.$$

Then $\texttt{Span}(e_1, e_2)$ and $\texttt{Span}(e_3, e_4)$ are both invariant subspaces (as well as many others which we ignore here). In addition, $\texttt{Span}(e_1, e_2) \oplus \texttt{Span}(e_3, e_4) = V$.

Definition 5.6.3. Let $T \in \mathcal{L}(V)$. If there are NOT two invariant subspaces $U, W \subset V$ under T such that $U, W \neq \{0\}$ and $U \oplus W = V$, then V is called **indecomposable**. Otherwise it is called **decomposable**. If V is decomposable and $V = U \oplus W$, W is called a **complement** of U.

Example 5.6.4. In Example 5.6.1, V is indecomposable under T while in Example 5.6.2, V is decomposable under T.

If an linear operator is decomposable, then we use the bases of each direct summands as the basis, then the matrix of the linear operator is a block-diagonal matrix. On the other side, if a linear operator has a basis to make it a block-diagonal matrix, then there exists a invariant subspace decomposition.

Therefore here is a summary: Let V be a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$.

(1) There is always a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ such that

$$0 \subset \texttt{Span}(v_1) \subset \texttt{Span}(v_1, v_2) \subset \ldots \subset \texttt{Span}(v_1, \ldots, v_{n-1}) \subset \texttt{Span}(v_1, \ldots, v_n) = V$$

is a chain of invariant subspaces.

(2) For each invariant subspace, complements don't have to exists.

(3) If T is diagonalizable with basis $\mathcal{B}$, then all these invariant subspaces above are decomposable.

(4) If a space is decomposable, and we know how the linear operator acts on each piece, the operator actions on the whole space is very simple: it is a block-diagonal matrix.

(5) Now we know that diagonalizable operators are just a special case of linear operators. We want to study more general operators on indecomposable spaces.

Exercises.

Exercise 5.1. Suppose V is finite-dimensional. Then each operator on V has at most $\dim V$ distinct eigenvalues.

Exercise 5.2.

(1) Suppose $S, T \in \mathcal{L}(V)$ are such that $ST = TS$. Prove that $\mathtt{Nul}(S)$ is invariant under T.

(2) Suppose $S, T \in \mathcal{L}(V)$ are such that $ST = TS$. Prove that $\mathtt{im}(S)$ is invariant under T.

Exercise 5.3. See the proof of Theorem 5.4.1. Let $v \in V$. Let $\{[v_1], \ldots, [v_k]\}$ be a basis $V/\mathtt{Span}(v)$. Please show that $\{v, v_1, \ldots, v_k\}$ is linearly independent using the definition of linearly independence.

Exercise 5.4. Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Prove that T has an invariant subspace of dimension k for each $k = 1, \ldots, \dim V$.

Exercise 5.5. Suppose W is a complex vector space and $T \in \mathcal{L}(W)$ has no eigenvalues. Prove that every subspace of W invariant under T is either $\{0\}$ or infinite-dimensional.

Exercise 5.6. Let $T \in \mathcal{L}(\mathbb{C}^3)$ which is defined by the matrix

$$A = \begin{bmatrix} 2 & -2 & 0 \\ 0 & 3 & 0 \\ 1 & 6 & 2 \end{bmatrix}.$$

Find a basis $\mathbb{C}^3$ to write T as an upper-triangular matrix.

The homework is due on Apr. 19.

Exercise 5.7. Check whether the following matrices are diagonalizable. Note that you do NOT need compute the diagonalized matrix or the change-of-basis matrix.

(1) $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{bmatrix}$.

(2) $B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

(3) $C = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 3 & 0 \\ 1 & 1 & 2 \end{bmatrix}$

Exercise 5.8.

(1) Suppose $T \in \mathcal{L}(V)$ is diagonalizable. Prove that $V = \texttt{Nul}(T) \oplus \texttt{im}(T)$.

(2) State the converse of the statement above. Prove it or give a counterexample.

Exercise 5.9. Give an example that $R, T \in \mathcal{L}(\mathbb{C}^4)$ such that R and T each have $2, 6, 7$ as eigenvalues and no other eigenvalues, and there does not exist an invertible operator $S \in \mathcal{L}(\mathbb{C}^4)$ such that $R = S^{-1}TS$.

Exercise 5.10. Let V be finite-dimensional, and $T, S \in \mathcal{L}(V)$. Suppose T has $\dim V$ distinct eigenvalues, and S has the same eigenvectors as T (not necessarily with the same eigenvalues). Prove that $ST = TS$.

Exercise 5.11. The **Fibonacci sequence** $F_1, F_2, \ldots$ is defined by

$$F_1 = 1, \quad F_2 = 1, \quad \text{and} \quad F_n = F_{n-2} + F_{n-1} \text{ for } n \geq 3.$$

Define $T \in \mathcal{L}(\mathbb{R}^2)$ by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} y \\ x+y \end{bmatrix}.$$

(1) Show that $T^n\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} F_n \\ F_{n+1} \end{bmatrix}$ for each positive integer n.

(2) Find the eigenvalues of T.

(3) Find a basis of $\mathbb{R}^2$ consisting of eigenvectors of T.

(4) Use the basis from part (c) to compute $T^n\left(\begin{bmatrix}0\\1\end{bmatrix}\right)$. Conclude that

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

for each positive integer n.

The homework is due on Apr. 26.

6. Inner Product Spaces

6.1. **Inner Products and Norms.**

Remark 6.1.1. When talking about inner product spaces, we focus on $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Then we can talk about conjugation, that is $\overline{\lambda} = \lambda$ for $\lambda \in \mathbb{R}$ and $\overline{a+bi} = a - bi$ for $a, b \in \mathbb{R}$.

6.1.1. *Inner products.*

Definition 6.1.2. Let V be a vector space. An **inner product** on V is a function $V \times V \to \mathbb{F}$ which is denoted by $\langle u, v \rangle$ for $u, v \in V$, such that

positivity: $\langle v, v \rangle \geq 0$ for all $v \in V$,

definiteness: $\langle v, v \rangle = 0$ if and only if $v = 0$,

additivity in first slot: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ for all $u, v, w \in V$,

homogeneity in first slot: $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$ for all $u, v \in V$ and $\lambda \in \mathbb{F}$,

conjugate symmetry: $\langle u, v \rangle = \overline{\langle v, u \rangle}$ for all $u, v \in V$.

Definition 6.1.3. An **inner product space** is a vector space V along with an inner product on V.

Proposition 6.1.4. *Let V be an inner product space over $\mathbb{F}$.*

(1) $\langle 0, u \rangle = 0$ *for every* $u \in V$.

(2) $\langle u, 0 \rangle = 0$ *for every* $u \in V$.

(3) $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ *for all* $u, v, w \in V$.

(4) $\langle u, \lambda v \rangle = \overline{\lambda} \langle u, v \rangle$ *for all* $\lambda \in \mathbb{F}$, $u, v \in V$.

Proof. (1) $\langle 0, u \rangle = \langle u - u, u \rangle = \langle u, u \rangle - \langle u, u \rangle = 0$.

(2) $\langle u, 0 \rangle = \overline{\langle 0, u \rangle} = \overline{0} = 0$.

(3) $\langle u, v + w \rangle = \overline{\langle v + w, u \rangle} = \overline{\langle v, u \rangle + \langle w, u \rangle} = \overline{\langle v, u \rangle} + \overline{\langle w, u \rangle} = \langle u, v \rangle + \langle u, w \rangle$.

(4) $\langle u, \lambda v \rangle = \overline{\langle \lambda v, u \rangle} = \overline{\lambda \langle v, u \rangle} = \overline{\lambda}\overline{\langle v, u \rangle} = \overline{\lambda} \langle u, v \rangle$.

□

Remark 6.1.5. In $\mathbb{R}$, $\lambda = \overline{\lambda}$. Therefore sometimes in $\mathbb{R}$ we call the inner product **bilinear**. In $\mathbb{C}$ we call it **linear in the first slot, and conjugate linear in the second slot**.

Example 6.1.6. The dot product on $\mathbb{R}^n$ is an inner product. It is defined by

$$\left\langle \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \right\rangle = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n.$$

Example 6.1.7. There is a dot-product-like inner product on $\mathbb{C}^n$. It is defined by

$$\left\langle \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \right\rangle = a_1 \overline{b_1} + a_2 \overline{b_2} + \ldots + a_n \overline{b_n}.$$

We call it **the dot product** on $\mathbb{C}^n$. Actually the dot product on $\mathbb{R}^n$ can also be defined by this formula since $\overline{b} = b$ in $\mathbb{R}$.

6.1.2. *Norm.*

Definition 6.1.8. Let V be an inner product space. For $v \in V$, the **norm** of v, denoted by $\|v\|$, is defined by $\|v\| = \sqrt{\langle v, v\rangle}$.

Proposition 6.1.9. *Let V be an inner product space and $v \in V$.*

(1) $\|v\| = 0$ if and only if $v = 0$.

(2) $\|\lambda v\| = |\lambda|\|v\|$ for all $\lambda \in \mathbb{F}$.

Proof. (1) $\|v\| = 0 \Leftrightarrow \langle v, v\rangle = 0 \Leftrightarrow v = 0$.

(2) $\|\lambda v\| = \sqrt{\langle \lambda v, \lambda v\rangle} = \sqrt{\lambda\overline{\lambda}\langle v, v\rangle} = \sqrt{|\lambda|^2 \langle v, v\rangle} = |\lambda|\sqrt{\langle v, v\rangle} = |\lambda|\|v\|$.

□

6.1.3. *Orthogonal.*

Definition 6.1.10. Two vectors $u, v \in V$ are called **orthogonal** if $\langle u, v\rangle = 0$.

Remark 6.1.11. We can define the angle between two vectors by $\langle u, v\rangle = \|u\|\|v\| \cos\theta$. However we won't take this approach in the rest of this course.

Proposition 6.1.12. *Let V be an inner product space.*

(1) 0 *is orthogonal to every vector in* V*.*

(2) 0 *is the only vector in* V *that is orthogonal to itself.*

Proof. Exercise. ☐

The above Proposition is easy but very important that we use it in the following way.

Corollary 6.1.13. *If* $\langle v, w\rangle = \langle v, u\rangle$ *for any* $v \in V$*, then* $w = u$*.*

Proof. If $\langle v, w\rangle = \langle v, u\rangle$ for any $v \in V$, then $\langle v, w-u\rangle = 0$ for any $v \in V$. Then $\langle w-u, w-u\rangle = 0$. So $w - u = 0$. So $w = u$. ☐

Theorem 6.1.14 (Pythagorean Theorem)**.** *Suppose* u *and* v *are orthogonal vectors in* V*. Then* $\|u+v\|^2 = \|u\|^2 + \|v\|^2$*.*

Proof. Since u and v are orthogonal, $\langle u, v\rangle = \langle v, u\rangle = 0$. Then

$$\begin{aligned}\|u+v\|^2 = \left(\sqrt{\langle u+v, u+v\rangle}\right)^2 &= \langle u+v, u+v\rangle = \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \langle u, u\rangle + \langle v, v\rangle = \|u\|^2 + \|v\|^2.\end{aligned}$$

☐

6.2. **Orthonormal Bases.**

Definition 6.2.1. A list of vectors is called **orthonormal** is each vector in the list has norm 1 and is orthogonal to all the other vectors in the list.

In other words, a list $e_1, \dots, e_m$ of vectors in V is orthonormal if

$$\langle e_j, e_k\rangle = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Proposition 6.2.2. *If* $e_1, \dots, e_m$ *is an orthonormal list of vectors in* V*, then*

$$\|a_1e_1 + \ldots + a_me_m\|^2 = |a_1|^2 + \ldots + |a_m|^2$$

for all $a_1, \dots, a_m \in \mathbb{F}$*.*

Theorem 6.2.3. *An orthonormal list is linearly independent.*

Proof. Let $\{e_1, \ldots, e_k\}$ is an orthonormal list of vectors in V. Consider the equation $a_1e_1 + \ldots + a_ke_k = 0$. Then for any $i = 1, \ldots, k$, $a_i = \langle e_i, a_1e_1 + \ldots + a_ke_k \rangle = \langle e_i, 0 \rangle = 0$. Then $\{e_1, \ldots, e_k\}$ is linearly independent. □

Definition 6.2.4. An **orthonormal basis** of V is an orthonormal list of vectors in V that is also a basis of V.

Example 6.2.5. In $\mathbb{R}^n$ (or $\mathbb{C}^n$) with the dot product, the standard basis is an orthonormal basis.

Theorem 6.2.6. *Every orthonormal list of vectors in V with length $\dim V$ is an orthonormal basis of V.*

Remark 6.2.7. After finding an orthonormal basis, the inner product is fully understood, by the following formulas.

Proposition 6.2.8. *Suppose $\mathcal{N} = \{e_1, \ldots, e_n\}$ is an orthonormal basis of V and $v \in V$. Then*

$$v = \langle v, e_1 \rangle e_1 + \ldots + \langle v, e_n \rangle e_n$$

and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \ldots + |\langle v, e_n \rangle|^2.$$

Proof. Let $v = a_1e_1 + \ldots + a_ne_n$. Then for any $i = 1, \ldots, n$, $a_i = \langle e_i, a_1e_1 + \ldots + a_ne_n \rangle = \langle e_i, v \rangle$. Then $v = \langle v, e_1 \rangle e_1 + \ldots + \langle v, e_n \rangle e_n$. Then

$$\begin{aligned}
\|v\|^2 &= \langle a_1e_1 + \ldots + a_ne_n, a_1e_1 + \ldots + a_ne_n \rangle \\
&= \langle a_1e_1, a_1e_1 \rangle + \ldots + \langle a_1e_1, a_ne_n \rangle + \ldots + \langle a_ne_n, a_1e_1 \rangle + \ldots + \langle a_ne_n, a_ne_n \rangle \\
&= |a_1|^2 + \ldots + |a_n|^2 = |\langle v, e_1 \rangle|^2 + \ldots + |\langle v, e_n \rangle|^2.
\end{aligned}$$

□

Theorem 6.2.9 (Gram-Schmidt Procedure)**.** *Suppose $v_1, \ldots, v_m$ is a linearly independent list of vectors in V. Let $e_1 = v_1/\|v_1\|$. For $j = 2, \ldots, m$, define e_j inductively by*

$$e_j = \frac{v_j - \langle v_j, e_1 \rangle e_1 - \ldots - \langle v_j, e_{j-1} \rangle e_{j-1}}{\|\textit{the above numerator}\|}.$$

Then $\{e_1, \ldots, e_m\}$ is an orthonormal list of vectors in V such that for any $j = 1, \ldots, m$,

$$\mathtt{Span}(v_1, \ldots, v_j) = \mathtt{Span}(e_1, \ldots, e_j)$$

Remark 6.2.10. The G-S procedure is to change a list of linearly independent vectors into an orthonormal list, while keeping the "towel" structure of these vectors in the following sense:

$$\begin{array}{ccccccccc}
\mathtt{Span}(v_1) & \xrightarrow{\subset} & \mathtt{Span}(v_1, v_2) & \xrightarrow{\subset} & \mathtt{Span}(v_1, v_2, v_3) & \xrightarrow{\subset} & \ldots & \xrightarrow{\subset} & \mathtt{Span}(v_1, \ldots, v_m) \\
\downarrow{=} & & \downarrow{=} & & \downarrow{=} & & & & \downarrow{=} \\
\mathtt{Span}(e_1) & \xrightarrow{\subset} & \mathtt{Span}(e_1, e_2) & \xrightarrow{\subset} & \mathtt{Span}(e_1, e_2, e_3) & \xrightarrow{\subset} & \ldots & \xrightarrow{\subset} & \mathtt{Span}(e_1, \ldots, e_m)
\end{array}$$

This is important because of the following Theorems.

Theorem 6.2.11. *Every finite-dimensional inner product space has an orthonormal basis.*

Proof. Let the vector space be V. Pick an arbitrary basis of V. Apply the Gram-Schmidt procedure. We will get an orthonormal list of vectors in V with length $\dim V$. Then it is an orthonormal basis. □

Corollary 6.2.12. *Suppose V is finite-dimensional. Then every orthonormal list of vectors in V can be extended to an orthonormal basis of V.*

Proof. Suppose $\{e_1, \ldots, e_m\}$ is an orthonormal list of vectors in V. Then we can extend it to be a basis $\{e_1, \ldots, e_m, v_m + 1, \ldots, v_n\}$ in V. Now apply the Gram-Schmidt procedure to it and we can get an orthonormal basis $\{f_1, \ldots, f_n\}$. Note that from the formula in Gram-Schmidt procedure, if the vector and all previous vectors form an orthonormal list, then that vector won't be changed. Therefore $f_1 = e_1, \ldots, f_m = e_m$. Therefore an orthonormal list of vectors can be extended to an orthonormal basis. □

Theorem 6.2.13 (Upper-triangular matrix with respect to orthonormal basis)**.** *Suppose $T \in \mathcal{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V.*

Proof. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be the basis such that $\Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}}$ is upper-triangular. Then all $\texttt{Span}(v_1, \ldots, v_j)$ is T-invariant for any $j = 1, \ldots, n$.

Then apply the Gram-Schmidt procedure to it, we can get an orthonormal basis $\mathcal{E} = \{e_1, \ldots, e_n\}$ such that $\texttt{Span}(e_1, \ldots, e_j) = \texttt{Span}(v_1, \ldots, v_j)$ for any $j = 1, \ldots, n$. Then all $\texttt{Span}(e_1, \ldots, e_j)$ is T-invariant for any $j = 1, \ldots, n$. Then $\Big[T\Big]_{\mathcal{E}\leftarrow\mathcal{E}}$ is upper-triangular. □

Theorem 6.2.14 (Schur's Theorem)**.** *Suppose V is a finite-dimensional complex inner product vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V.*

Proof. Since V is a finite-dimensional complex inner product vector space and $T \in \mathcal{L}(V)$, then T has an upper-triangular matrix with respect to some basis of V. Then by Theorem 6.2.13, T has an upper-triangular matrix with respect to some orthonormal basis of V. □

6.3. **Exercises.**

Exercise 6.1. Let V be an inner product space.

(1) 0 is orthogonal to every vector in V.

(2) 0 is the only vector in V that is orthogonal to itself.

Exercise 6.2. Suppose V is a real inner product space.

(1) Show that $\langle u+v, u-v\rangle = \|u\|^2 - \|v\|^2$ for any $u, v \in V$.

(2) Show that if $\|u\| = \|v\|$, then $u+v$ is orthogonal to $u-v$.

Exercise 6.3. Prove or disprove: there is an inner product on R^2 such that the associated norm is given by $\left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\| = \max\{x, y\}$ for all $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$.

Exercise 6.4. Let V be an inner product space with the inner product $\langle \cdot, \cdot \rangle$. Suppose $S \in \mathcal{L}(V)$ is an injective operator on V. Define a new pairing $\langle \cdot, \cdot \rangle_S$ by $\langle u, v\rangle_S = \langle Su, Sv\rangle$ for $u, v \in V$.

(1) Please show that $\langle \cdot, \cdot \rangle_S$ is an inner product on V.

(2) Please give a counter example that $\langle \cdot, \cdot \rangle_S$ is not an inner product when S is not injective.

Exercise 6.5. Let $\mathbb{R}^3$ be the inner product space with the usual dot product. Let $T \in \mathcal{L}(\mathbb{R}^3)$ has an upper-triangular matrix with respect to the basis $\left\{ w_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, w_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, w_3 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \right\}$. Find an orthonormal basis of $\mathbb{R}^3$ with respect to which T has an upper-triangular matrix.

The homework is due on May 31.

7. Operators on Inner Product Spaces

7.1. Self-Adjoint and Normal Operators.

7.1.1. *Adjoint operators.* Recall that $\mathcal{L}(V, W)$ is the space of all linear maps from V to W.

Definition 7.1.1 (adjoint)**.** Suppose $T \in \mathcal{L}(V, W)$. The **adjoint** of T is the function $T^* : W \to V$ such that

$$\langle Tv, w \rangle = \langle v, T^* w \rangle$$

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

Remark 7.1.2. To compute adjoint operators, one of the most important tool is Corollary 6.1.13: If $\langle v, w \rangle = \langle v, u \rangle$ for any $v \in V$, then $w = u$. To compute $T^* w$, we usually evaluate its pairing with any vector $v \in V$. It also holds for the first slot.

Proposition 7.1.3. *The adjoint is a linear map. That is, if $T \in \mathcal{L}(V, W)$, then $T^* \in \mathcal{L}(W, V)$.*

Proof. We need to check that T^* preserve addition and scalar product.

Addition: For any $v \in V$, $w, u \in W$,

$$\langle v, T^*(w+u) \rangle = \langle Tv, w+u \rangle = \langle Tv, w \rangle + \langle Tv, u \rangle = \langle v, T^* w \rangle + \langle v, T^* u \rangle = \langle v, T^* w + T^* u \rangle .$$

Then $T^(w+u) = T^* w + T^* u$.

Scalar product: For any $v \in V$, $w \in W$, $\lambda \in \mathbb{F}$,

$$\langle v, T^*(\lambda w) \rangle = \langle Tv, \lambda w \rangle = \overline{\lambda} \langle Tv, w \rangle = \overline{\lambda} \langle v, T^* w \rangle = \langle v, \lambda T^* w \rangle .$$

Then $T^*(\lambda w) = \lambda T^* w$.

To sum up, $T^* : W \to V$ is linear. Then $T^* \in \mathcal{L}(W, V)$. □

Proposition 7.1.4. *Let U, V, W be three inner product spaces over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.*

(1) $(S+T)^* = S^* + T^*$ for all $S, T \in \mathcal{L}(V, W)$.

(2) $(\lambda T)^* = \overline{\lambda} T^*$ for all $\lambda \in \mathbb{F}$ and $T \in \mathcal{L}(V, W)$.

(3) $(T^*)^* = T$ for all $T \in \mathcal{L}(V, W)$.

(4) $I^* = I$ where I is the identity operator on V.

(5) $(ST)^* = T^*S^*$ for all $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$.

Proof. (1) For any $w \in W$, $v \in V$,

$$\langle v, (S+T)^*w \rangle = \langle (S+T)v, w \rangle = \langle Sv + Tv, w \rangle = \langle Sv, w \rangle + \langle Tv, w \rangle$$

$$= \langle v, S^*w \rangle + \langle v, S^*w \rangle = \langle v, (S^* + T^*)w \rangle .$$

Then $(S+T)^*w = (S^* + T^*)w$ for any $w \in W$. So $(S+T)^* = S^* + T^*$.

(2) For any $v \in V$, $w \in W$, $\lambda \in \mathbb{F}$,

$$\langle v, (\lambda T)^*w \rangle = \langle \lambda Tv, w \rangle = \lambda \langle Tv, w \rangle = \lambda \langle v, T^*w \rangle = \left\langle v, \overline{\lambda} T^*w \right\rangle .$$

Then $(\lambda T)^*w = \overline{\lambda} T^*w$ for any $w \in W$. Then $(\lambda T)^* = \overline{\lambda} T^*$.

(3) For any $v \in V$, $w \in W$,

$$\langle Tv, w \rangle = \langle v, T^*w \rangle = \overline{\langle T^*w, v \rangle} = \overline{\langle w, (T^*)^*v \rangle} = \langle (T^*)^*v, w \rangle .$$

Then $Tv = (T^*)^*v$ for any $v \in V$. Then $T = (T^*)^*$.

(4) For any $v, w \in V$,

$$\langle v, I^*w \rangle = \langle Iv, w \rangle = \langle v, w \rangle .$$

Then $I^*w = w$ for any $w \in V$. Then $I^* = I$.

(5) For any $v \in V$, $w \in W$ and $u \in U$,

$$\langle v, (ST)^*u \rangle = \langle STv, u \rangle = \langle S(Tv), u \rangle = \langle Tv, S^*u \rangle = \langle v, T^*S^*u \rangle .$$

Then $(ST)^*u = T^*S^*u$ for any $u \in U$. Then $(ST)^* = T^*S^*$.

□

Definition 7.1.5. The **conjugate transpose** of an $m \times n$ matrix is the $n \times m$ matrix obtained by interchanging the rows and columns and then taking the complex conjugate of each entry. The conjugate transpose of A is denoted by A^H. It is also denoted by some other notations: $\overline{A}^T$, $A^\dagger$, A^*, etc.. We will use A^H in this course.

Let $A = \left(a_{ji}\right)_{1\le j\le m, 1\le i\le n}$ be a $m\times n$ matrix and $B = \left(b_{pq}\right)_{1\le p\le n, 1\le q\le m}$ be a $n\times m$ matrix. Then if $B = A^H$ if and only if $a_{ji} = \overline{b_{ji}}$ for any $j = 1,\dots,m$ and $i = 1,\dots,n$.

Theorem 7.1.6 (The matrix of T^*)**.** *Let $\mathcal{E}$ (resp. $\mathcal{F}$) be an orthonormal basis of V (resp. W). Then* $\left[T^*\right]_{\mathcal{E}\leftarrow\mathcal{F}} = \left[T\right]^H_{\mathcal{F}\leftarrow\mathcal{E}}$.

Proof. Let $\mathcal{E} = \{e_1,\dots,e_n\}$ and $\mathcal{F} = \{f_1,\dots,f_m\}$. Let the matrix of T with respect to these two bases be

$$\left[T\right]_{\mathcal{F}\leftarrow\mathcal{E}} = \left(a_{ji}\right)_{1\le j\le m, 1\le i\le n},$$

and the matrix of T^* with respect to these two bases be

$$\left[T^*\right]_{\mathcal{E}\leftarrow\mathcal{F}} = \left(b_{pq}\right)_{1\le p\le n, 1\le q\le m}.$$

Then we have

$$T(e_i) = a_{1i}f_1 + \ldots + a_{mi}f_m = \sum_{j=1}^{m} a_{ji}f_j, \quad \text{for any } i = 1,\dots,n,$$

$$T^*(f_q) = b_{1q}e_1 + \ldots + b_{nq}e_n = \sum_{p=1}^{n} b_{pq}e_p, \quad \text{for any } q = 1,\dots,m.$$

Then for any $i = 1,\dots,n$ and $q = 1,\dots,m$, we have

$$\langle e_i, T^*f_q\rangle = \left\langle e_i, \sum_{p=1}^{n} b_{pq}e_p \right\rangle = \sum_{p=1}^{n} \overline{b_{pq}}\,\langle e_i, e_p\rangle = \sum_{p=1}^{n} \overline{b_{pq}}\delta_{i,p} = \overline{b_{iq}},$$

and

$$\langle e_i, T^*f_q\rangle = \langle Te_i, f_q\rangle = \left\langle \sum_{j=1}^{m} a_{ji}f_j, f_q \right\rangle = \sum_{j=1}^{m} a_{ji}\,\langle f_j, f_q\rangle = \sum_{j=1}^{m} a_{ji}\delta_{j,q} = a_{qi}.$$

Therefore for any $i = 1,\dots,n$ and $q = 1,\dots,m$, $\overline{b_{iq}} = a_{qi}$. This shows that

$$\left[T^*\right]_{\mathcal{E}\leftarrow\mathcal{F}} = \left[T\right]^H_{\mathcal{F}\leftarrow\mathcal{E}}.$$

□

7.1.2. *Sefl-adjoint operators.*

Definition 7.1.7 (self-adjoint)**.** An operator $T \in \mathcal{L}(V)$ is called **self-adjoint** if $T = T^*$. It is also called **Hermitian** operator.

In other words, T is self-adjoint if and only if $\langle Tv, w\rangle = \langle v, Tw\rangle$ for all $v, w \in V$. Also if we choose an orthonormal basis, the matrix satisfies $\left[T\right] = \left[T\right]^H$.

Proposition 7.1.8. *The sum of two self-adjoint operators is self-adjoint and the product of a real scalar and a self-adjoint operator is self-adjoint.*

Proof. It is straightforward by $(A+B)^* = A^* + B^* = A + B$ and $(\lambda A)^* = \overline{\lambda} A^* = \lambda A$ for any A, B self-adjoint operators and $\lambda \in \mathbb{R}$. □

Theorem 7.1.9. *Eigenvalues of self-adjoint operators are real.*

Proof. Suppose T is a self-adjoint operator on V. Let λ be an eigenvalue of T, and let v be an eigenvector. Then

$$\lambda \|v\|^2 = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\|v\|^2.$$

Therefore $\lambda = \overline{\lambda}$. Then λ is real. □

Theorem 7.1.10. *Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Suppose $\langle Tv, v\rangle = 0$ for all $v \in V$. Then $T = 0$.*

Proof. Choose a Jordan basis $\{e_1, \ldots, e_n\}$ for T. Then the basis vector has two types: eigenvector and generalized eigenvector. If e_k is an eigenvector, then $Te_k = \lambda e_k$. If e_k is an generalized eigenvector, then $Te_k = \lambda e_k + e_{k-1}$.

If e_k is an eigenvector with eigenvalue λ, then $\langle Te_k, e_k\rangle = 0$. So $\langle \lambda e_k, e_k\rangle = 0$. Since $e_k \neq 0$, λ has to be 0. Therefore all eigenvalues of T are 0's.

If e_k is a generalized eigenvector, since its eigenvalue is 0, we have $Te_k = e_{k-1}$. Consider a Jordan block of size r, and rename the Jordan basis vectors associated to this block $\{e_1, \ldots, e_r\}$. Assume that $r \geq 2$. Then $Te_1 = 0$ and $Te_i = e_{i-1}$ for $i = 2, \ldots, r$. Then $0 = \langle Te_2, e_2\rangle = \langle e_1, e_2\rangle$. Therefore by $e_2 \neq 0$, we have $\langle T(e_1 + e_2), e_1 + e_2\rangle = \langle e_2, e_1 + e_2\rangle = \langle e_2, e_2\rangle \neq 0$. This is a contradiction. Then $r = 1$. So all Jordan blocks are of size 1. Then the matrix is a diagonal matrix. Then the linear operator T has to be 0. □

Remark 7.1.11. The above proof only works for complex numbers since we use Jordan canonical form. In the real case there are counter examples. For example, consider $\mathbb{R}^2$ and the standard inner product on it. Let R be rotating about the origin in $\mathbb{R}^2$ by $90°$. Then $\langle Rv, v\rangle = 0$ for any v, but $R \neq 0$.

Theorem 7.1.12. *Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Then T is self-adjoint if and only if $\langle Tv, v\rangle \in \mathbb{R}$ for every $v \in V$.*

Proof. ($\Rightarrow$)**:** Since $\langle Tv, v\rangle \in \mathbb{R}$ for any $v \in V$, $\langle Tv, v\rangle = \overline{\langle v, Tv\rangle} = \langle v, Tv\rangle = \langle v, (T^*)^*v\rangle = \langle T^*v, v\rangle$. Then $\langle (T - T^*)v, v\rangle = 0$ for any $v \in V$. Then $T = T^*$. So T is self-adjoint.

($\Leftarrow$)**:** If T is self-adjoint, $T = T^*$. Then $\langle (T - T^*)v, v\rangle = 0$ for any $v \in V$. Then $\langle Tv, v\rangle = \langle T^*v, v\rangle = \overline{\langle v, T^*v\rangle} = \overline{\langle Tv, v\rangle}$. Therefore we have $\langle Tv, v\rangle = \overline{\langle Tv, v\rangle}$ for any $v \in V$. Then $\langle Tv, v\rangle \in \mathbb{R}$ for any $v \in V$.

□

The above two theorems apply to complex vector spaces only. The next one can be extended to real vector spaces.

Theorem 7.1.13. *Let V be a real vector space. Suppose T is a self-adjoint operator on V such that $\langle Tv, v\rangle = 0$ for all $v \in V$. Then $T = 0$.*

Proof. Since T is self-adjoint, $\langle Tv, w\rangle = \langle v, Tw\rangle = \langle Tw, v\rangle$ for any $v, w \in V$. Then for any $v, w \in V$,

$$\begin{aligned}
&\langle T(v+w), v+w\rangle - \langle T(v-w), v-w\rangle \\
&= \langle Tv, v\rangle + \langle Tv, w\rangle + \langle Tw, v\rangle + \langle Tw, w\rangle - (\langle Tv, v\rangle - \langle Tv, w\rangle - \langle Tw, v\rangle + \langle Tw, w\rangle) \\
&= 2\langle Tv, w\rangle + 2\langle Tw, v\rangle = 4\langle Tv, w\rangle .
\end{aligned}$$

Since $\langle T(v+w), v+w\rangle = \langle T(v-w), v-w\rangle = 0$ for any $v, w \in V$, $\langle Tv, w\rangle = 0$ for any $v, w \in V$. Then $T = 0$. □

7.1.3. *Normal operators.*

Definition 7.1.14. An operator on an inner product space is called **normal** if it commutes with its adjoint.

In other words, $T \in \mathcal{L}(V)$ is normal if $TT^* = T^*T$.

Theorem 7.1.15. *An operator $T \in \mathcal{L}(V)$ is normal if and only if $\|Tv\| = \|T^*v\|$ for all $v \in V$.*

Proof. **($\Rightarrow$):** Since T is normal, $TT^* = T^*T$. By $(T^*)^* = T$, we have

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle v, T^*Tv\rangle = \langle v, TT^*v\rangle = \langle v, (T^*)^*T^*v\rangle = \langle T^*v, T^*v\rangle = \|T^*v\|^2.$$

Then $\|Tv\| = \|T^*v\|$.

($\Leftarrow$): By the above computation, we have $\|Tv\|^2 = \langle v, T^*Tv\rangle$ and $\|T^*v\|^2 = \langle v, TT^*v\rangle$. If $\|Tv\| = \|T^*v\|$ for any $v \in V$, $\langle v, T^*Tv\rangle = \langle v, TT^*v\rangle$ for any v. So for any $v \in V$, $\langle v, (T^*T - TT^*)v\rangle = 0$. By Theorem 7.1.10

□

Theorem 7.1.16. *Suppose $T \in \mathcal{L}(V)$ is normal and $v \in V$ is an eigenvector of T with eigenvalue λ. Then v is also an eigenvector of T^* with eigenvalue $\overline{\lambda}$.*

Proof. Since $Tv = \lambda v$, then $(T - \lambda I)v = 0$. So $\|(T - \lambda I)v\| = 0$. Then since T is normal, $T - \lambda I$ should also be normal. Then $0 = \|(T - \lambda I)v\| = \|(T - \lambda I)^*v\|$. So $(T - \lambda I)^*v = 0$. Then $T^*v = \overline{\lambda}I$. So v is an eigenvector of T^* with eigenvalue $\overline{\lambda}$. □

Theorem 7.1.17. *Suppose $T \in \mathcal{L}(V)$ is normal. Then eigenvectors of T corresponding to distinct eigenvalues are orthogonal.*

Proof. Suppose $Tv = \lambda v$ and $Tw = \mu w$ while $\lambda \neq \mu$. Then

$$\langle T^*Tv, w\rangle = \langle v, T^*Tw\rangle,$$

$$\langle T^*Tv, w\rangle = \left\langle |\lambda|^2 v, w\right\rangle = |\lambda|^2 \langle v, w\rangle,$$

$$\langle v, T^*Tw\rangle = \left\langle v, |\mu|^2 w\right\rangle = |\mu|^2 \langle v, w\rangle.$$

So $|\lambda|^2 \langle v, w\rangle = |\mu|^2 \langle v, w\rangle$. Since $\lambda \neq \mu$, $\langle v, w\rangle = 0$. □

7.2. **The Spectral Theorem.**

Theorem 7.2.1 (Complex Spectral Theorem)**.** *Suppose* $\mathbb{F} = \mathbb{C}$ *and* $T \in \mathcal{L}(V)$*. Then* T *is normal if and only if* T *has a diagonal matrix with respect to some orthonormal basis of* V*.*

Proof of the Spectral Theorem.

(⇐): Let $\mathcal{B}$ be the orthonormal basis such that $\Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}}$ is diagonal, then $\Big[T^*\Big]_{\mathcal{B}\leftarrow\mathcal{B}} = \Big[T\Big]^H_{\mathcal{B}\leftarrow\mathcal{B}}$ is also diagonal. Since diagonal matrices commute, $\Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}}\Big[T^*\Big]_{\mathcal{B}\leftarrow\mathcal{B}} = \Big[T^*\Big]_{\mathcal{B}\leftarrow\mathcal{B}}\Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}}$. So $TT^* = T^*T$. Then T is normal.

(⇒): By Theorem 6.2.14, there exists an orthonormal basis $\mathcal{B} = \{e_1, \ldots, e_n\}$ such that $\Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}}$ is upper-triangular. That is to say,

$$Te_1 = a_{11}e_1,$$

$$Te_2 = a_{12}e_1 + a_{22}e_2,$$

$$\ldots\ldots$$

$$Te_n = a_{1n}e_1 + a_{2n}e_2 + \ldots + a_{nn}e_n.$$

Since $\Big[T^*\Big]_{\mathcal{B}\leftarrow\mathcal{B}} = \Big[T\Big]^H_{\mathcal{B}\leftarrow\mathcal{B}}$, we have

$$T^*e_1 = \overline{a_{11}}e_1 + \overline{a_{12}}e_2 + \ldots + \overline{a_{1n}}e_n,$$

$$T^*e_2 = \overline{a_{22}}e_2 + \overline{a_{23}}e_3 + \ldots + \overline{a_{2n}}e_n,$$

$$\ldots\ldots$$

$$T^*e_n = \overline{a_{nn}}e_n.$$

Then since $\{e_1, \ldots, e_n\}$ is an orthonormal basis, for any $k = 1, \ldots, n$,

$$\|Te_k\|^2 = \|a_{1k}e_1 + a_{2k}e_2 + \ldots + a_{kk}e_k\|^2 = \|a_{1k}\|^2 + \ldots + \|a_{kk}\|^2,$$

and

$$\|T^*e_k\|^2 = \|\overline{a_{kk}}e_k + \overline{a_{k,k+1}}e_{k+1} + \ldots + \overline{a_{kn}}e_n\|^2 = \|a_{kk}\|^2 + \ldots + \|a_{kn}\|^2.$$

Since T is normal, $T^*T = TT^*$. Then

$$\|Te_k\|^2 = \langle Te_k, Te_k\rangle = \langle e_k, T^*Te_k\rangle = \langle e_k, TT^*e_k\rangle = \langle T^*e_k, T^*e_k\rangle = \|T^*e_k\|^2.$$

Then for any $k = 1, \ldots, n$,

$$\|a_{1k}\|^2 + \ldots + \|a_{kk}\|^2 = \|a_{kk}\|^2 + \ldots + \|a_{kn}\|^2.$$

($k = 1$)**:** $\|a_{11}\|^2 = \|a_{11}\|^2 + \|a_{12}\|^2 + \ldots + \|a_{1n}\|^2$. Then $\|a_{12}\|^2 + \ldots + \|a_{1n}\|^2 = 0$. Since all terms are non-negative real numbers, all of them has to be 0. Then $a_{12} = a_{13} = \ldots = a_{1n} = 0$.

($k = 2$)**:** $\|a_{12}\|^2 + \|a_{22}\|^2 = \|a_{22}\|^2 + \ldots + \|a_{2n}\|^2$. Then since $a_{12} = 0$, $\|a_{23}\|^2 + \ldots + \|a_{2n}\|^2 = 0$. Since all terms are non-negative real numbers, all of them has to be 0. Then $a_{23} = a_{24} = \ldots = a_{2n} = 0$.

Keep repeating this, we can get that all $a_{ij} = 0$ for $i < j$. Then this means that the matrix $\big[T\big]_{\mathcal{B}\leftarrow\mathcal{B}}$ is a diagonal matrix. To sum up, T is diagonalizable by an orthonormal basis.

□

Proposition 7.2.2. *Let V be an inner product space. Let $T \in \mathcal{L}(V)$ be a normal operator. Then there exists $R, M \in \mathcal{L}(V)$ be two self-adjoint operators such that $T = R + iM$ and $RM = MR$. Conversely, any two self-adjoint operators R and M which commute can make a normal operator by $R + iM$.*

Proof. ($\Rightarrow$)**:** T is normal. Then there is an orthonormal basis $\mathcal{B}$ such that

$$\big[T\big]_{\mathcal{B}\leftarrow\mathcal{B}} = \begin{bmatrix} \alpha_1 + i\beta_1 & & \\ & \ddots & \\ & & \alpha_n + i\beta_n \end{bmatrix}.$$

Then let $R, M \in \mathcal{L}(V)$ be defined by

$$\big[R\big]_{\mathcal{B}\leftarrow\mathcal{B}} = \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{bmatrix}, \quad \big[M\big]_{\mathcal{B}\leftarrow\mathcal{B}} = \begin{bmatrix} \beta_1 & & \\ & \ddots & \\ & & \beta_n \end{bmatrix}.$$

Then

$$\begin{aligned}\Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}} &= \Big[R\Big]_{\mathcal{B}\leftarrow\mathcal{B}} + i\Big[M\Big]_{\mathcal{B}\leftarrow\mathcal{B}},\\ \Big[R\Big]_{\mathcal{B}\leftarrow\mathcal{B}} &= \Big[R\Big]^H_{\mathcal{B}\leftarrow\mathcal{B}},\\ \Big[M\Big]_{\mathcal{B}\leftarrow\mathcal{B}} &= \Big[M\Big]^H_{\mathcal{B}\leftarrow\mathcal{B}},\\ \Big[R\Big]_{\mathcal{B}\leftarrow\mathcal{B}}\Big[M\Big]_{\mathcal{B}\leftarrow\mathcal{B}} &= \Big[M\Big]_{\mathcal{B}\leftarrow\mathcal{B}}\Big[R\Big]_{\mathcal{B}\leftarrow\mathcal{B}}.\end{aligned}$$

Then $T = R + iM$, $R = R^*$, $M = M^*$ and $RM = MR$.

($\Leftarrow$): Since $R = R^*$, $M = M^*$, and $RM = MR$, then $T^* = (R + iM)^* = R^* + (iM)^* = R^* - iM^* = R - iM$. Then

$$TT^* = (R + iM)(R - iM) = R^2 - iRM + iMR + M^2 = R^2 + m^2,$$

$$T^*T = (R - iM)(R + iM) = R^2 + iRM - iMR + M^2 = R^2 + m^2.$$

Then $TT^* = T^*T$. So T is normal.

□

Remark 7.2.3. From the Proposition, normal operators and self-adjoint operators can be treated as generalizations of complex numbers and real numbers. From this point of view, conjugate transpose / adjoint can be viewed as the generalization of conjugate in complex numbers.

Corollary 7.2.4. *Let $T \in \mathcal{L}(V)$ be a normal operator on a complex inner product space V. If all eigenvalues of T are real, then T is self-adjoint.*

Proof. Since T is normal, there is an orthonormal basis $\mathcal{B}$ such that the matrix $\Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}}$ is diagonal. Then since the diagonal are all real, $\Big[T\Big]^H_{\mathcal{B}\leftarrow\mathcal{B}} = \Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}}$. So $T = T^*$. Then T is self-adjoint. □

7.3. **Exercises.**

Exercise 7.1. Consider $\mathbb{C}^n$ with the dot product. Define $T \in \mathbb{C}^n$ by

$$T\left(\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}\right) = \begin{bmatrix} 0 \\ z_1 \\ \vdots \\ z_{n-1} \end{bmatrix}.$$

Find a formula for $T^*\left(\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}\right)$. You cannot directly use Theorem 7.1.6.

Exercise 7.2. Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$ is a normal operator such that $T^9 = T^8$. Prove that T is self-adjoint and $T^2 = T$.

Exercise 7.3. Let $S, T \in \mathcal{L}(V)$ be self-adjoint. Show that ST is self-adjoint if and only if $ST = TS$.

Exercise 7.4. Suppose V is a complex inner product space with $V \neq \{0\}$. Show that the set of self-adjoint operators on V is not a subspace of $\mathcal{L}(V)$.

Exercise 7.5. Give an example of an operator T on a complex vector space such that $T^9 = T^8$ but $T^2 \neq T$.

Exercise 7.6. Let V be a finite-dimensional complex vector space. Suppose that T is a normal operator on V and that 3 and 4 are eigenvalues of T. Prove that there exists a vector $v \in V$ such that $\|v\| = \sqrt{2}$ and $\|Tv\| = 5$.

Exercise 7.7. Give an example of an operator $T \in \mathcal{L}(\mathbb{C}^4)$ such that T is normal but not self-adjoint.

Exercise 7.8. Let $T \in \mathcal{L}(V, W)$. Please prove that

(1) T is injective if and only if T^* is surjective.

(2) T is surjective if and only if T^* is injective.

Exercise 7.9. Consider $\mathbb{C}^3$ with the dot product. Let $\mathcal{E}$ be the standard basis. Let $T \in \mathcal{L}(\mathbb{C}^3)$ be defined by

$$\Big[T\Big]_{\mathcal{E}\leftarrow\mathcal{E}} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Please find an orthonormal basis such that the matrix of T is diagonal, or prove that such a basis doesn't exist.

The homework is due on Jun. 7.

8.1. Theory of null spaces.

Lemma 8.1.1 (Sequence of increasing null spaces)**.** *Suppose $T \in \mathcal{L}(V)$. Then*

$$\{0\} = \mathtt{Nul}(T^0) \subset \mathtt{Nul}(T^1) \subset \ldots \subset \mathtt{Nul}(T^k) \subset \mathtt{Nul}(T^{k+1}) \subset \ldots.$$

Proof. Suppose k is a non-negative integer and $v \in \mathtt{Nul}(T^k)$. Then $T^k v = 0$, and hence $T^{k+1}v = T(T^k v) = T(0) = 0$. Thus $v \in \mathtt{Nul}(T^{k+1})$. Hence $\mathtt{Nul}(T^k) \subset \mathtt{Nul}(T^{k+1})$, as desired. □

Lemma 8.1.2 (Equality in the sequence of null spaces)**.** *Suppose $T \in \mathcal{L}(V)$. Suppose m is a non-negative integer such that $\mathtt{Nul}(T^m) = \mathtt{Nul}(T^{m+1})$. Then*

$$\mathtt{Nul}(T^m) = \mathtt{Nul}(T^{m+1}) = \mathtt{Nul}(T^{m+2}) = \mathtt{Nul}(T^{m+3}) = \ldots.$$

Proof. Let k be a positive integer. We want to prove that $\mathtt{Nul}(T^{m+k}) = \mathtt{Nul}(T^{m+k+1})$. We already know that $\mathtt{Nul}(T^{m+k}) \subset \mathtt{Nul}(T^{m+k+1})$. Then we just need to show that $\mathtt{Nul}(T^{m+k+1}) \subset \mathtt{Nul}(T^{m+k})$. For any $v \in \mathtt{Nul}(T^{m+k+1})$, $T^{m+k+1}(v) = 0$. Then $T^{m+1}(T^k(v)) = 0$. So $T^k(v) \in \mathtt{Nul}(T^{m+1}) = (T^m)$. Then $T^m(T^k(v)) = 0$. So $T^{m+k}(v) = 0$. Then $v \in \mathtt{Nul}(T^{m+k})$. □

Lemma 8.1.3 (Null spaces stop growing)**.** *Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then*

$$\mathtt{Nul}(T^n) = \mathtt{Nul}(T^{n+1}) = \mathtt{Nul}(T^{n+2}) = \ldots.$$

Proof. Assume this is not true. Then by Lemma 8.1.1 and Lemma 8.1.2, it has to be

$$\{0\} = \mathtt{Nul}(T^0) \subsetneq \mathtt{Nul}(T^1) \subsetneq \mathtt{Nul}(T^2) \subsetneq \ldots \subsetneq \mathtt{Nul}(T^n) \subsetneq \mathtt{Nul}(T^{n+1}).$$

Then this means that

$$0 = \dim \mathtt{Nul}(T^0) < \dim \mathtt{Nul}(T^1) < \dim \mathtt{Nul}(T^2) < \ldots < \dim \mathtt{Nul}(T^n) < \dim \mathtt{Nul}(T^{n+1}).$$

Then $\dim \mathtt{Nul}(T^{n+1}) \geq n + 1$, which is greater than $\dim V$. This is a contradiction. Then $\mathtt{Nul}(T^n) = \mathtt{Nul}(T^{n+1})$. □

Theorem 8.1.4 (Direct sum decomposition of V)**.** *Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then $V = \texttt{Nul}(T^n) \oplus \texttt{im}(T^n)$.*

Proof. Let $v \in \texttt{Nul}(T^n) \cap \texttt{im}(T^n)$. Then $T^n(v) = 0$ and there exists $w \in V$ such that $v = T^n(w)$. Then $T^n(T^n(w)) = 0$. So $w \in \texttt{Nul}(T^{2n}) = \texttt{Nul}(T^n)$ by Lemma 8.1.3. So $v = T^n(w) = 0$. Then $\texttt{Nul}(T^n) \cap \texttt{im}(T^n) = \{0\}$. Then $\texttt{Nul}(T^n) + \texttt{im}(T^n) = \texttt{Nul}(T^n) \oplus \texttt{im}(T^n) \subset V$. Then by $\dim V = \dim \texttt{Nul}(T^n) + \dim \texttt{im}(T^n) = \dim(\texttt{Nul}(T^n) \oplus \texttt{im}(T^n))$, $V = \texttt{Nul}(T^n) \oplus \texttt{im}(T^n)$. □

Remark 8.1.5. It is easy to see that $\texttt{Nul}(T^n)$ and $\texttt{im}(T^n)$ are both T-invariant (Exercises). Then this theorem gives a direct sum decomposition into invariant subspaces.

Example 8.1.6. Let $T \in \mathbb{C}^3$ be defined by $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. Then $T^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $T^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Let $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

(1) $\texttt{Nul}(T^1) = \texttt{Span}(e_1)$, $\texttt{im}(T^1) = \texttt{Span}(e_1, e_2)$.

(2) $\texttt{Nul}(T^2) = \texttt{Span}(e_1, e_2)$, $\texttt{im}(T^2) = \texttt{Span}(e_1)$.

(3) $\texttt{Nul}(T^3) = \texttt{Span}(e_1, e_2, e_3)$, $\texttt{im}(T^3) = \{0\}$.

Then $\mathbb{C}^3 \neq \texttt{Nul}(T^1) \oplus \texttt{im}(T^1)$, $\mathbb{C}^3 \neq \texttt{Nul}(T^2) \oplus \texttt{im}(T^2)$ and $\mathbb{C}^3 = \texttt{Nul}(T^3) \oplus \texttt{im}(T^3)$.

8.2. **Generalized Eigenvectors.**

Definition 8.2.1 (Generalized eigenvector)**.** Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. A vector $v \in V$ is called a **generalized eigenvector** of T corresponding to λ if $v \neq 0$ and

$$(T - \lambda I)^j v = 0$$

for some positive integer j.

Example 8.2.2. Let $T = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$, $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Then we have

$$(T - 2I)^1(e_1) = 0, \quad (T - 2I)^2(e_2) = 0, \quad (T - 4I)^1(e_3) = 0.$$

Therefore e_1 is an eigenvector (also generalized) of T corresponding to 2, e_2 is a generalized eigenvector of T corresponding to 2, and e_3 is an eigenvector of T corresponding to 4.

Definition 8.2.3 (Generalized eigenspace)**.** Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. The **generalized eigenspace** of T corresponding to λ, denoted $V^G_{\lambda,T}$, is defined to be the set of all generalized eigenvectors of T corresponding to λ, along with the 0 vector.

Remark 8.2.4. Eigenvectors corresponding to λ are also generalized eigenvectors corresponding to λ. Therefore $V_\lambda \subset V^G_\lambda$.

Theorem 8.2.5. *Suppose* $T \in \mathcal{L}(V)$ *and* $\lambda \in \mathbb{F}$. *Let* $\dim V = n$. *Then* $V^G_\lambda = \texttt{Nul}\,((T - \lambda I)^n)$.

Proof. From the definition of generalized eigenvectors, $\texttt{Nul}\,((T - \lambda I)^n) \subset V^G_\lambda$. Also by Lemma 8.1.1 and Lemma 8.1.3, $V^G_\lambda \subset \texttt{Nul}\,((T - \lambda I)^n)$. Then $V^G_\lambda = \texttt{Nul}\,((T - \lambda I)^n)$. ☐

Remark 8.2.6. Then Theorem 8.1.4 can be rewritten as $V = V^G_\lambda \oplus \texttt{im}\,((T - \lambda I)^n)$.

Theorem 8.2.7 (Linearly independent generalized eigenvectors)**.** *Let* $T \in \mathcal{L}(V)$. *Suppose* $\lambda_1, \ldots, \lambda_m$ *are distinct eigenvalues of* T *and* $v_1, \ldots, v_m$ *are corresponding generalized eigenvectors. Then* $v_1, \ldots, v_m$ *is linearly independent.*

Proof. Suppose $a_1, \ldots, a_m \in \mathbb{F}$ such that $a_1v_1 + \ldots + a_mv_m = 0$. Let k be the largest non-negative integer such that $(T - \lambda_1 I)^k v_1 \neq 0$. Let $w = (T - \lambda_1 I)^k v_1$. Then

$$(T - \lambda_1 I)(w) = (T - \lambda_1 I)^{k+1}(v) = 0.$$

So $T(w) = \lambda_1 w$. Then $(T - \lambda I)w = T(w) - \lambda w = (\lambda_1 - \lambda)w$ for any $\lambda \in \mathbb{F}$. Therefore $(T - \lambda I)^n w = (\lambda_1 - \lambda)^n w$ for any $\lambda \in \mathbb{F}$.

Now consider an operator $L_1 = (T - \lambda_1 I)^k (T - \lambda_2 I)^n \ldots (T - \lambda_m I)^n \in \mathcal{L}(V)$. The order of each factor of the operator L_1 can be changed by Proposition 5.3.19. Since $(T - \lambda_i I)^n(v_i) = 0$ for $i = 2, \ldots, m$ and $(T - \lambda_1 I)^k(v_1) = w$, we have for $i = 2, \ldots, m$

$$L_1(v_i) = (T - \lambda_1 I)^k (T - \lambda_2 I)^n \ldots (T - \lambda_{i-1} I)^n (T - \lambda_{i+1} I)^n \ldots (T - \lambda_m I)^n (T - \lambda_i I)^n (v_i) = 0,$$

and

$$\begin{aligned} L_1(v_1) &= (T - \lambda_2 I)^n \ldots (T - \lambda_m I)^n (T - \lambda_1 I)^k (v_1) = (T - \lambda_2 I)^n \ldots (T - \lambda_m I)^n w \\ &= (\lambda_1 - \lambda_2)^n \ldots (\lambda_1 - \lambda_m)^n w. \end{aligned}$$

Then since $0 = a_1 v_1 + \ldots + a_m v_m$, we have

$$0 = L_1(0) = L_1(a_1 v_1 + \ldots + a_m v_m) = a_1 (\lambda_1 - \lambda_2)^n \ldots (\lambda_1 - \lambda_m)^n w.$$

Then $a_1 (\lambda_1 - \lambda_2)^n \ldots (\lambda_1 - \lambda_m)^n = 0$. Since all λ_i's are distinct, $(\lambda_1 - \lambda_2)^n \ldots (\lambda_1 - \lambda_m)^n \neq 0$. So $a_1 = 0$.

Use the similar method to construct operators L_i for $i = 2, \ldots, m$. Then we have $a_i = 0$ for $i = 2, \ldots, m$. Then $\{v_1, \ldots, v_m\}$ is linearly independent. □

8.3. **Decomposition of a Space with an Operator.**

Lemma 8.3.1. *Suppose $T \in \mathcal{L}(V)$ and $p \in P(\mathbb{F})$. Then $\texttt{Nul}(p(T))$ and $\texttt{im}(p(T))$ are invariant under T. In particular, $\texttt{Nul}\left((T - \lambda I)^n\right)$ and $\texttt{im}\left((T - \lambda I)^n\right)$ are T-invariant.*

Proof. For $v \in \texttt{Nul}(p(T))$, $p(T)(v) = 0$. Then

$$p(T)(T(v)) = (p(T)T)(v) = (Tp(T))(v) = T(p(T)(v))0.$$

So $T(v) \in \texttt{Nul}\,(p(T))$. Then $\texttt{Nul}\,(p(T))$ is invariant under T.

For $w \in \texttt{im}\,(p(T))$, there exists $u \in V$ such that $p(T)(z) = w$. Then

$$T(w) = T(p(T)(z)) = (Tp(T))(z) = (p(T)T)(z) = p(T)(T(z)).$$

So $T(w) \in \texttt{im}\,(p(T))$. Then $\texttt{im}\,(p(T))$ is invariant under T. □

Theorem 8.3.2. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Then $V = V^G_{\lambda_1} \oplus \dots \oplus V^G_{\lambda_m}$.*

Proof. Apply complete induction on the $\dim V$. $\dim V = 1$ case is obvious. Assume the decomposition is true for any vector spaces of dimension $\leq n$. Then for a vector space V of dimension $n+1$, since it has an eigenvalue λ, there is a generalized eigenspace V^G_λ. So $V = V^G_\lambda \oplus \texttt{im}\,((T-\lambda I)^{n+1})$. Let $\texttt{im}\,((T-\lambda I)^{n+1}) = W$. By induction, since $\dim W < \dim V = n+1$, $W = W^G_{\lambda_1} \oplus \dots W^G_{\lambda_m}$ for $\lambda_1, \dots, \lambda_m$ being distinct eigenvalues of $T|_W$ and $W^G_{\lambda_i}$ being corresponding generalized eigenspace in W. Therefore the rest is to prove that λ_i's are also eigenvalues of $T \in \mathcal{L}(V)$ and $W^G_{\lambda_i} = V^G_{\lambda_i}$.

- λ_i is an eigenvalue means that $\exists w \in W$ such that $T|_W(w) = \lambda_i w$. Since $w \in W \subset V$, $T|_W(w) = T(w)$. So $T(w) = \lambda_i w$. Then λ_i is an eigenvalue of T and $w \in V$ is a corresponding eigenvector of T in V.
- For any $w \in W^G_{\lambda_i}$, there exists k such that $(T|_W - \lambda_i I_W)^k(w) = 0$. Since $w \in W \subset V$, $T|_W(w) = T(w)$. Then $(T-\lambda_i I)^k(w) = 0$. So $w \in V^G_{\lambda_i}$. Then $W^G_{\lambda_i} \subset V^G_{\lambda_i}$.

  On the other side, for any $v \in V^G_{\lambda_i}$, by $V = V^G_\lambda \oplus W = V^G_\lambda \oplus W^G_{\lambda_1} \oplus \dots \oplus W^G_{\lambda_m}$, there exists $u \in V^G_\lambda$, $w_i \in W^G_{\lambda_i}$ such that $v = u + w_1 + \dots + w_m$. Then since $v \in V^G_{\lambda_i}$, $u \in V^G_\lambda$, $w_i \in W^G_{\lambda_i} \subset V^G_{\lambda_i}$ for any $i = 1, \dots, m$, and by that generalized eigenvectors corresponding to distinct eigenvalues are linearly independent, v has to be w_i. Then $v \in W^G_{\lambda_i}$. Then $V^G_{\lambda_i} \subset W^G_{\lambda_i}$.

  To sum up, $V^G_{\lambda_i} = W^G_{\lambda_i}$.

Therefore $V = V^G_\lambda \oplus V^G_{\lambda_1} \oplus \dots \oplus V^G_{\lambda_m}$. By complete induction, the Theorem holds. □

Corollary 8.3.3. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then there is a basis of V consisting of generalized eigenvectors of T.*

Remark 8.3.4. Generalized eigenspace decomposition is a refined result of the upper-triangular matrix chain of basis. At least this time we get a direct sum decomposition of invariant subspaces. Can we decompose generalized eigenspace further? That is the next topic.

Corollary 8.3.5. *Suppose V is a complex vector space, $T \in \mathcal{L}(V)$, and V_λ^G is a generalized eigenspace corresponding to the eigenvalue λ. Then there exists a basis of V_λ^G such that the matrix $T|_{V_\lambda^G}$ is upper-triangular, and the diagonal entries are all λ.*

Proof. The upper-triangular assertion is obvious. Now we use the construction of basis which make $T|_{V_\lambda^G}$ upper-triangular matrix to prove that the matrix has a constant diagonal. Let the basis be $\{v_1, \dots, v_n\}$. Then we have

$$T(v_1) = a_{11}v_1,$$

$$T(v_2) = a_{12}v_1 + a_{22}v_2,$$

$$\dots\dots$$

$$T(v_n) = a_{1n}v_1 + a_{2n}v_2 + \dots + a_{nn}v_n.$$

What we want to show is that all $a_{ii} = \lambda$ for $i = 1, \dots, n$. We use contradiction. Assume that for some k, $a_{kk} \neq \lambda$. Let $W = \texttt{Span}(v_1, \dots, v_{k-1})$, $U = \texttt{Span}(v_1, \dots, v_k)$. Then both W and U are T-invariant. Then there is a quotient operator $\overline{T} \in \mathcal{L}(U/W)$. The basis of U/W is $\left[v_k\right]$ such that $\overline{T}(\left[v_n\right]) = a_{kk}\left[a_k\right]$. Then $(\overline{T} - \lambda\overline{I})^n(\left[a_k\right]) = (a_{kk} - \lambda)^n\left[v_k\right]$. Since $a_{kk} \neq \lambda$, $(a_{kk} - \lambda)^n\left[v_k\right] \neq \left[0\right]$. Then $(\overline{T} - \lambda\overline{I})^n(\left[v_k\right]) \neq \left[0\right]$.

However since $v_k \in V_\lambda^G$, $(T - \lambda I)^n(v_k) = 0$. Then put them into the quotient space, we have

$$(\overline{T} - \lambda\overline{I})^n(\left[v_k\right]) = \overline{(T - \lambda I)^n}(\left[v_k\right]) = \left[(T - \lambda I)^n(v_k)\right] = \left[0\right].$$

This is a contradiction. Then all $a_{kk} = \lambda$ for any $k = 1, \dots, n$. □

Remark 8.3.6. Due to the previous Corollary, we sometimes call λ in V_λ^G the generalized eigenvalue.

8.4. **Jordan Form.**

8.4.1. *Characteristic and Minimal Polynomials.*

Proposition 8.4.1. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\mathcal{E}$ and $\mathcal{F}$ be two bases of V. The matrix $A = \Big[T\Big]_{\mathcal{E}\leftarrow\mathcal{E}}$ and $B = \Big[T\Big]_{\mathcal{F}\leftarrow\mathcal{F}}$. Then $\det(A - \lambda I) = \det(B - \lambda I)$ as polynomials of λ.*

Proof. There is an invertible matrix P such that $A = P^{-1}BP$. Then

$$\begin{aligned}\det(A - \lambda I) &= \det\big(P^{-1}BP - \lambda P^{-1}P\big) = \det\big(p^{-1}(B - \lambda I)P\big)\\ &= \det\big(P^{-1}\big)\det(B - \lambda I)\det(P) = \det(B - \lambda I).\end{aligned}$$

□

Definition 8.4.2 (Characteristic polynomial)**.** Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\mathcal{B}$ be a basis of V. The matrix $A = \Big[T\Big]_{\mathcal{B}\leftarrow\mathcal{B}}$. Then the **characteristic polynomial** of T is defined to be $\det(A - \lambda I)$. It is proved by the previous proposition that it is independent of choice of bases.

Definition 8.4.3. Suppose $T \in \mathcal{L}(V)$. The **multiplicity** of an eigenvalue λ of T is defined to be the dimension of the corresponding generalized eigenspace V_λ^G. In other words, the multiplicity of λ is $\dim \texttt{Nul}(T - \lambda I)^{\dim V}$.

Theorem 8.4.4. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. Then the characteristic polynomial of T is $(x - \lambda_1)^{d_1} \ldots (x - \lambda_m)^{d_m}$.*

Proof. By Theorem 8.3.2 and Corollary 8.3.5, after writing a matrix into an upper-triangular matrix, the diagonal are all eigenvalues, and the number of each distinct eigenvalue λ is the dimension of V_λ^G. Then the characteristic polynomial can be computed directly. □

Remark 8.4.5. The Theorem can be used in the other way. We can now first compute the characteristic polynomial of a matrix, and then use the multiplicity of the roots to get the multiplicity of the eigenvalues.

Theorem 8.4.6. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then*

(1) the characteristic polynomial of T has degree $\dim V$*.*

(2) the zeros of the characteristic polynomial of T are the eigenvalues of T.

Theorem 8.4.7 (Cayley-Hamilton Theorem)**.** *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.*

Proof. From Theorem 8.4.4, let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. Then the characteristic polynomial of T is $(x - \lambda_1)^{d_1} \ldots (x - \lambda_m)^{d_m}$. Then since for any $v_k \in V^G_{\lambda_k}$, $(T - \lambda_k I)^{d_k}(v_k) = 0$, and we can rearrange the order of the characteristic polynomial $q(x) = (x - \lambda_1)^{d_1} \ldots (x - \lambda_m)^{d_m}(x - \lambda_k)^{d_k}$, then we have

$$q(T)(v_k) = (T - \lambda_1 I)^{d_1} \ldots (T - \lambda_m I)^{d_m}(T - \lambda_k I)^{d_k}(v_k) = 0.$$

Since this holds for any $k = 1, \ldots, m$, and $V = V^G_{\lambda_1} \oplus \ldots \oplus V^G_{\lambda_m}$, then $q(T)(v) = 0$ for any $v \in V$. Then $q(T) = 0$. □

8.4.2. *Minimal polynomials.*

Definition 8.4.8. Let V be a vector space and $T \in \mathcal{L}(V)$. A polynomial $p(x)$ with the highest degree coefficient $= 1$ (monic) of smallest degree such that $p(T) = 0$ is called the **minimal polynomial** of T.

Theorem 8.4.9. *Suppose $T \in \mathcal{L}(V)$. The minimal polynomial of T exists and it is unique.*

Proof. The uniqueness part is easy. Let $p_1(x) = x^k + a_{k-1}x^{k-1} + \ldots + a_1 x + a_0$ and $p_2(x) = x^k + b_{k-1}x^{k-1} + \ldots + b_1 x + b_0$ be two minimal polynomials. Then $p_3(x) = p_1(x) - p_2(x)$ be a polynomial of degree $< k$ and $p_3(T) = p_1(T) - p_2(T) = 0$. So if $p_3 \neq 0$, it contradicts the facts that both p_1 and p_2 are minimal polynomials. So $p_3 = 0$. Then $p_1 = p_2$.

We collect all monic polynomials $p(x)$ such that $p(T) = 0$ together and consider the set S of their degrees. S is a set of natural numbers, and it is not empty due to the characteristic polynomial. Then by Well-Ordering principle, there is a minimum degree, and the polynomial of that degree is the minimal polynomial we want. □

Theorem 8.4.10. *Suppose $T \in \mathcal{L}(V)$ and $q \in P(\mathbb{F})$. Then $q(T) = 0$ if and only if q is a polynomial multiple of the minimal polynomial p of T.*

Proof. (⇒)**:** Suppose $q(T) = 0$. By the division theorem for polynomials, there exist polynomial s, r such that $q = ps + r$ and $\deg r < \deg p$. Then $0 = q(T) = p(T)s(T) + r(T) = r(T)$. Then r has to be 0 otherwise it will violate that p is the minimal polynomial.

(⇐)**:** If q is a polynomial multiple of p, then $q(x) = p(x)s(x)$. Then $q(T) = p(T)s(T) = 0$.

□

Corollary 8.4.11. *Characteristic polynomial is a multiple of minimal polynomial.*

Corollary 8.4.12. *Eigenvalues are the zeros of the minimal polynomial.*

8.4.3. *Nilpotent operator.*

Definition 8.4.13. An operator is celld **nilpotent** if some power of it equals 0.

Example 8.4.14. Let $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Since $A^2 = 0$, and $B^k = B$ for any $k \geq 0$, A is a nilpotent operator and B is not.

Proposition 8.4.15. *Suppose $N \in \mathcal{L}(V)$ is nilpotent and let $n = \dim V$. Then $N^n = 0$.*

Proof. Since N is nilpotent, there is a power k that $N^k = 0$. Then for any $v \in V$, $(N - 0I)^k v = 0$. So $v \in V_0^G$. Then $V = V_0^G = \texttt{Nul}\,((N - 0I)^n) = \texttt{Nul}\,(N^n)$. So $N^n = 0$. □

Lemma 8.4.16 (Matrix of a nilpotent operator)**.** *Suppose $N \in \mathcal{L}(V)$ is nilpotent. There is a basis of V such that the matrix of N corresponding to the basis is a strictly upper-triangular matrix.*

Proof. By Proposition 8.4.15, $N^n = 0$. Then by Theorem 8.4.10 and Corollary 8.4.12, 0 is the only eigenvalue of N. Then by Corollary 8.3.5 the diagonal entries of the upper-triangular matrix of N are all 0. □

8.4.4. *Basis corresponding to a nilpotent operator.*

Theorem 8.4.17. *Let V be a complex vector space of dimension n. Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then there exist vectors $v_1, \ldots, v_n \in V$ and non-negative integers $m_1, \ldots, m_n$ such that*

(1) $\{N^{m_1}v_1, \ldots, Nv_1, v_1, N^{m_2}v_2, \ldots, Nv_2, v_2, \ldots, N^{m_n}v_n, \ldots, Nv_n, v_n\}$ *is a basis of V.*
(2) $N^{m_1+1}v_1 = N^{m_2+1}v_2 = \ldots = N^{m_n+1}v_n = 0$.

Proof. Use complete induction on $\dim V = n$. $n = 1$ case is obvious. Assume that the statement holds for any complex vector space of dimension $\leq k$. Let V be a complex vector space of dimension $k+1$. Let N be a nilpotent operator on V. There are two possibilities:

(1) if V can be decomposed into a direct sum of generalized eigenspaces, then by induction the result holds.
(2) if V cannot be decomposed, then choose a basis $\{v_1, \ldots, v_{n+1}\}$ such that under the basis N is an upper-triangular matrix. Now construct a new basis in the following way:
   - Start from $w_{n+1} = v_{n+1}$.
   - If $N(w_k) \neq 0$, then $w_{k-1} = N(w_k)$.
   - If $k > 1$ and $N(w_k) = 0$, then $w_{k-1} = v_{k-1}$.
   - Stop after we get w_1.

   It is easy to see that $\{w_1, \ldots, w_{n+1}\}$ is a basis. What's more, if $N(w_k) = 0$, then $\texttt{Span}(w_n, \ldots, w_k)$ and $\texttt{Span}(w_{k-1}, \ldots, w_1)$ are both N-invariant. Then $V = \texttt{Span}(w_n, \ldots, w_k) \oplus \texttt{Span}(w_{k-1}, \ldots, w_1)$. Since V is indecomposable, this cannot happen. So there are no w_k such that $k > 1$ and $N(w_k) = 0$. Then the basis satisfies that $Nw_k = w_{k-1}$ for $k = 2, \ldots, n+1$. Also since $N^{n+1} = 0$, we have $N^{n+1}(w_n) = 0$. This proves the statement.

By induction, we can always find a basis in the given form. ∎

Theorem 8.4.18. *Under the basis described in the previous Theorem, the matrix of the nilpotent operator is a block diagonal matrix where the block always look like*

$$\begin{bmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}$$

8.4.5. *Jordan basis.*

Definition 8.4.19. A matrix of the form

$$J_\lambda = \begin{bmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$$

is called a **Jordan block** corresponding to λ. A matrix is in **Jordan canonical form** if the matrix is a block diagonal matrix

$$A = \begin{bmatrix} J_{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & J_{\lambda_m} \end{bmatrix}$$

where the diagonal are all Jordan blocks.

Definition 8.4.20. Let $T \in \mathcal{L}(V)$. A basis of V is called a **Jordan basis** for T if under this basis the matrix of T is in Jordan canonical form.

Remark 8.4.21. The basis in Theorem 8.4.17 is a Jordan basis for the nilpotent operator N.

Theorem 8.4.22 (Jordan form)**.** *Suppose V is a complex vector space. If $T \in \mathcal{L}(V)$, then there is a basis of V that is a Jordan basis for T.*

Proof. First consider the generalized eigenspaces decomposition. $V = V^G_{\lambda_1} \oplus \dots V^G_{\lambda_m}$. On each $V^G_{\lambda_i}$, $(T - \lambda_i I)|_{V^G_{\lambda_i}}$ is nilpotent. Then there is a basis on $V^G_{\lambda_i}$ such that the matrix of $(T - \lambda_i I)|_{V^G_{\lambda_i}}$

is in Jordan canonical form:

$$\begin{bmatrix} J_0^1 & & 0 \\ & \ddots & \\ 0 & & J_0^r \end{bmatrix}$$

where the upper index suggests that there might be multiple Jordan blocks. Then under the same basis, the matrix of $T|_{V_{\lambda_i}^G}$ is

$$\begin{bmatrix} J_{\lambda_i}^1 & & 0 \\ & \ddots & \\ 0 & & J_{\lambda_i}^r \end{bmatrix}$$

which is in Jordan canonical form.

We collect these Jordan bases from each generalized eigenspace and put them together to form a basis of V. This is the Jordan basis for T on V. □

Remark 8.4.23. To find the Jordan basis, the key point lies in the idea that the basis look like $v, Nv, N^2v, \ldots$. Therefore here is the steps:

(1) Each eigenvector is corresponding to a Jordan block.

(2) Find a eigenvector v_1 by solving the equation $(A - \lambda I)X = 0$.

(3) The next basis vector v_2 is the solution to the equation $(A - \lambda I)X = v_1$.

(4) The next basis vector v_3 is the solution to the equation $(A - \lambda I)X = v_2$.

(5) Repeat until you find all basis related to this Jordan block.

(6) Repeat until you run through all Jordan blocks.

Definition 8.4.24. The change-of-basis matrix of Jordan basis is also called the transformation matrix. Using the transformation matrix, the original matrix is similar to its Jordan canonical form.

Example 8.4.25. Find the Jordan basis of the matrix $\begin{bmatrix} 2 & 2 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$

Solve. First find the characteristic polynomial:

$$\det(A - xI) = \det\left(\begin{bmatrix} 2-x & 2 & 1 \\ 0 & 2-x & 1 \\ 0 & 0 & 3-x \end{bmatrix}\right) = (2-x)^2(3-x).$$

Then the matrix has eigenvalue 2 with multiplicity 2 and eigenvalue 3 with multiplicity 1.

Then try $(2I - A)(3I - A)$:

$$(2I - A)(3I - A) = \begin{bmatrix} 0 & -2 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} 1 & -2 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq 0.$$

Therefore the minimal polynomial is not $(2-x)(3-x)$. So the minimal polynomial has to be the characteristic polynomial. Then the Jordan canonical form of the matrix A is

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

To find the transformation matrix P, we need to find a basis consisting of eigenvectors of eigenvalue 2 and 3 and a generalized eigenvector of eigenvalue 2.

Eigenvalue 2: Solve $(A - 2I)X = 0$:

$$\begin{bmatrix} 0 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} X = 0.$$

The solution is $X \in \texttt{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right)$. So the first basis vector can be $v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

Generalized eigenvector of eigenvalue 2: Solve $(A-2I)X = v_1$:

$$\begin{bmatrix} 0 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} X = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

The solution is $X \in \texttt{Span}\left(\begin{bmatrix} 0 \\ \frac{1}{2} \\ 0 \end{bmatrix}\right)$. We can choose the second basis vector to be $v_2 = \begin{bmatrix} 0 \\ \frac{1}{2} \\ 0 \end{bmatrix}$.

Eigenvalue 3: Solve $(A-3I)X = 0$:

$$\begin{bmatrix} -1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix} X = 0.$$

The solution is $X \in \texttt{Span}\left(\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}\right)$. We can choose the third basis vector to be $v_3 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$.

Therefore $\mathcal{B} = \{v_1, v_2, v_3\}$ is a Jordan basis. The transformation matrix is

$$P = \begin{bmatrix} 1 & 0 & 3 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

We may jusitify our answer by computing

$$P^{-1}AP = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

$\square$

8.4.6. *The relation between minimal polynomials, characteristic polynomials and Jordan forms.*

- The characteristic polynomial can be factorized into $(x-\lambda_1)^{d_1}\ldots(x-\lambda_m)^{d_m}$. This is directly related to the decomposition $V = V^G_{\lambda_1}\oplus\ldots\oplus V^G_{\lambda_m}$ where $\dim V^G_{\lambda_i} = d_i$ for $i = 1,\ldots,m$. Here we know that $d_1+d_2+\ldots+d_m = \dim V$.
- The minimal polynomial can be factorized into $(x-\lambda_1)^{r_1}\ldots(x-\lambda_m)^{r_m}$. Here $r_i \le d_i$. Each $(x-\lambda_i)^{r_i}$ is related to a Jordan block J_{λ_i} of the size r_i.
- The Jordan block we read from the minimal polynomial is the biggest Jordan block of the given eigenvalue. We might have many other Jordan blocks of the given eigenvalue with smaller sizes.
- We we only know the characteristic polynomial and the minimal polynomial, the Jordan canonical form of the linear operator is partially determined. Roughly speaking, how many Jordan canonical form we can get depends on how many more possibilities we can find after fixing the biggest Jordan block of a given eigenvalue and the total dimension of the generalized eigenspace of the given eigenvalue.

Example 8.4.26. A is a 3×3 matrix, with the characteristic polynomial $(x-2)^3$ and the minimal polynomial $(x-2)$. Since the biggest Jordan block has size 1, the matrix has to be similar to
$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Example 8.4.27. A is a 3×3 matrix, with the characteristic polynomial $(x-2)^3$ and the minimal polynomial $(x-2)^2$. Since the biggest Jordan block has size 2, then the only possible combination is a block of size 1 and a block of size 2. Then the matrix has to be similar to
$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Example 8.4.28. A is a 3×3 matrix, with the characteristic polynomial $(x-2)^3$ and the minimal polynomial $(x-2)^3$. Since the biggest Jordan block has size 3, the matrix has to be similar to $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$.

Example 8.4.29. A is a 4×4 matrix, with the characteristic polynomial $(x-2)^4$ and the minimal polynomial $(x-2)^3$. The biggest Jordan block has size 3. Then the only possible combination is a block of size 3 and a block of size 1. Then the matrix has to be similar to $\begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$.

Example 8.4.30. A is a 4×4 matrix, with the characteristic polynomial $(x-2)^4$ and the minimal polynomial $(x-2)^2$. The biggest Jordan block has size 2. Then the possible combinations are twos blocks of size 2, or a block of size 2 with two blocks of size 1. Then the matrix has to be similar to $\begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$ or $\begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$.

8.5. **Exercises.**

Exercise 8.1. Let V be a finite-dimensional complex vector space of dimension n, and $T \in \mathcal{L}(V)$. Please show that $\texttt{Nul}(T^n)$ and $\texttt{im}(T^n)$ are all invariant under T.

Exercise 8.2. Suppose $T \in \mathcal{L}(V)$ and m is a non-negative integer. Show that $\texttt{Nul}(T^m) = \texttt{Nul}(T^{m+1})$ if and only if $\texttt{im}(T^m) = \texttt{im}(T^{m+1})$.

Exercise 8.3. Let $T \in \mathcal{L}(\mathbb{C}^2)$ be defined in the following ways. Find all the generalized eigenspaces.

(1) $T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} -b \\ a \end{bmatrix}$.

(2) $T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} b \\ 0 \end{bmatrix}$.

Exercise 8.4. Prove or give a counterexample: If V is a complex vector space and $\dim V = n$ and $T \in \mathcal{L}(V)$, then T^n is diagonalizable.

Exercise 8.5. Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Prove that V has a basis consisting of eigenvectors of T if and only if every generalized eigenvector of T is an eigenvector of T.

The homework is due on May 10.

Exercise 8.6. Suppose $S, T \in \mathcal{L}(V)$ and ST is nilpotent. Prove that TS is nilpotent.

Exercise 8.7. Prove or give a counterexample: The set of nilpotent operators on V is a subspace of $\mathcal{L}(V)$.

Exercise 8.8. Give an example of an operator T on a finite-dimensional real vector space such that 0 is the only eigenvalue of T but T is not nilpotent.

Exercise 8.9. Suppose $T \in \mathcal{L}(\mathbb{C}^4)$ is such that the eigenvalues of T are 3, 5, 8. Prove that $(T-3I)^2(T-5I)^2(T-8I)^2 = 0$.

Exercise 8.10.

(1) Give an example of an operator on $\mathbb{C}^4$ whose characteristic polynomial equals $(x-1)(x-3)^3$ and whose minimal polynomial equals $(x-1)(x-3)^2$.

(2) Give an example of an operator on $\mathbb{C}^4$ whose characteristic and minimal polynomials both equal $x(x-1)^2(x-3)$.

Exercise 8.11. Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Prove that V has a basis consisting of eigenvectors of T if and only if the minimal polynomial of T has no repeated zeros.

Exercise 8.12. Let $N = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Please find the characteristic polynomial and minimal polynomial of N.

Exercise 8.13. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$. Please find its Jordan canonical form C and find the transformation matrix P such that $C = P^{-1}AP$.

The homework is due on May 17.