Northeastern University, Department of Mathematics

MATH G5110: Applied Linear Algebra and Matrix Analysis. (Fall 2020)

• Instructor: **He Wang** Email: **he.wang@northeastern.edu**

§8 Jordan Canonical Form

Contents

Not every square matrix is diagonalizable. However, we can block diagonalize it to be in Jordan canonical(normal, norm) form.

1. Block diagonal

An $n \times n$ matrix B is a **block diagonal matrix** if

$$B = \begin{bmatrix} B_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & B_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & B_m \end{bmatrix}$$

with the matrices on the diagonal. Block diagonal matrix B is also denoted as **direct sum:**

$$B = B_1 \oplus B_2 \oplus \cdots \oplus B_m.$$

Recall that given a linear transformation $T : \mathbb{F}^n \to \mathbb{F}^n$. A subspace $W \subseteq \mathbb{F}^n$ is said to be **invariant** under T if $T(\vec{w}) \in W$ whenever $w \in W$.

Theorem 1. *An $n \times n$ matrix A is similar to a block diagonal matrix B, (i.e., $A = PBP^{-1}$) if and only if there exists a decomposition of $\mathbb{F}^n = V_1 \oplus V_2 \oplus \cdots \oplus V_m$ such that V_i is invariant under T_A.*

Proof. Choose a basis $\mathscr{B}_i = \{\vec{v}_{i,1}, \dots, \vec{v}_{i,n_i}\}$ for each V_i. Denote matrix $P = [\vec{v}_{1,1} \dots \vec{v}_{1,n_1} \dots \vec{v}_{m,n_m}]$. By change of coordinate theorem, we know that $A = PBP^{-1}$ where matrix B is defined as $\vec{b}_{i,j} = [A\vec{v}_{i,j}]_{\mathscr{B}}$
Since V_i is invariant under T_A, then $A\vec{v}_{i,j} \in V_i$, hence $A\vec{v}_{i,j} = b_{i,1}\vec{v}_{i,1} + \cdots + b_{i,n_i}\vec{v}_{i,n_i}$. □

The following non-diagonalizable matrices are called **Jordan blocks** of size 1, 2, 3, 4, ...

$$J_{\lambda,1} = \begin{bmatrix}\lambda\end{bmatrix}, \quad J_{\lambda,2} = \begin{bmatrix}\lambda & 1\\ 0 & \lambda\end{bmatrix}, \quad J_{\lambda,3} = \begin{bmatrix}\lambda & 1 & 0\\ 0 & \lambda & 1\\ 0 & 0 & \lambda\end{bmatrix}, \quad J_{\lambda,4} = \begin{bmatrix}\lambda & 1 & 0 & 0\\ 0 & \lambda & 1 & 0\\ 0 & 0 & \lambda & 1\\ 0 & 0 & 0 & \lambda\end{bmatrix}, \dots$$

Definition 2. An $n \times n$ **Jordan normal matrix** (**Jordan normal form**) is a block diagonal matrix

$$J = \begin{bmatrix} J_{n_1}(\lambda_1) & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & J_{n_2}(\lambda_2) & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & J_{n_m}(\lambda_m) \end{bmatrix}$$

such that all diagonal matrices $J_{n_i}(\lambda_i)$ are of the form

$$J_{n_i}(\lambda_i) = \begin{bmatrix} \lambda_i & * & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \lambda_i & \ddots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \ddots & *\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \lambda_i \end{bmatrix}$$

where $* = 1$ or 0.

Remark: 1. $J_{n_i}(\lambda_i)$ is direct sum (block diagonal) of Jordan blocks J_{*,λ_i}. 2. $J_{n_i}(\lambda_i)$ is not uniquely determined by n_i and λ_i.

Our purpose in this section is to show the following theorem:

Theorem 3. *Every $n \times n$ matrix A with n eigenvalues in a field $\mathbb{F}$ is similar to a matrix J in Jordan normal matrix, that is $A = PJP^{-1}$.*

The **Jordan normal form** of A is unique up to the order of Jordan blocks.

2. Nilpotent matrix

Definition 4. An $n \times n$ matrix A is called **nilpotent of degree** m if $A^m = \mathbf{0}$ and $A^{m-1} \neq \mathbf{0}$ for some $m \geq 0$.

Proposition 5.
- *If A is nilpotent, then zero is the only eigenvalue of A.*
- *If A is nilpotent and diagonalizable, then $A = 0$.*

Proof. (1) If $\lambda \neq 0$ is an eigenvalue of A, then $A\vec{v} = \lambda\vec{v}$ with nonzero $\vec{v}$. So, $A^k\vec{v} = \lambda^k\vec{v}$ for any k. So A is not nilpotent.
(2) Suppose $A = PDP^{-1}$. From (1), we know that $D = 0$. So $A = 0$. □

Lemma 6.
- *$J_{0,k}$ is nilpotent of degree k.*
- *Suppose a Jordan matrix $J = J_n(\lambda)$ with the same entry λ on diagonal, then there exist a number m such that $(J - \lambda I_n)^m = \mathbf{0}$.*

Proof.

$$J_{0,k}\vec{x} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \\ x_k \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_k \\ 0 \end{bmatrix}$$

(1) Direct calculation $J_{0,k}^k = \mathbf{0}$ and $J_{0,k}^{k-1} \neq \mathbf{0}$.
(2) Let m be the size of the largest Jordan block in J. □

Some times, it is continent to describe a Jordan block as the sum of λI and a nilpotent block:

$$J_{\lambda,k} = \lambda I + J_{0,k}$$

Suppose A is similar to a Jordan block $J_{\lambda,n}$ (i.e., $A = PJ_{\lambda,n}P^{-1}$), then

$$AP = PJ_{\lambda,n}.$$

That is

$$[A\vec{w}_1 \ \cdots \ A\vec{w}_n] = [\vec{w}_1 \ \cdots \ \vec{w}_n] \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix} = [\lambda\vec{w}_1 \ \lambda\vec{w}_2 + \vec{w}_1 \ \ldots \ \lambda\vec{w}_n + \vec{w}_{n-1}]$$

Hence,

$$\begin{aligned} A\vec{w}_1 &= \lambda\vec{w}_1 \\ A\vec{w}_2 &= \lambda\vec{w}_2 + \vec{w}_1 \\ &\vdots \\ A\vec{w}_n &= \lambda\vec{w}_n + \vec{w}_{n-1} \end{aligned}$$

Equivalently,

$$\begin{aligned} (A - \lambda I)\vec{w}_1 &= \vec{0} \\ (A - \lambda I)\vec{w}_2 &= \vec{w}_1 \\ &\vdots \\ (A - \lambda I)\vec{w}_n &= \vec{w}_{n-1} \end{aligned}$$

Denote $N = A - \lambda I$, such a sequence of vectors $\{\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_n\} = \{N^{n-1}\vec{w}_n, N^{n-2}\vec{w}_n, \ldots, \vec{w}_n\}$ is called a **Jordan Chain**.

We also get

$$(A - \lambda I)^2 \vec{w}_2 = \vec{0}; (A - \lambda I)^3 \vec{w}_3 = \vec{0}; ...; (A - \lambda I)^n \vec{w}_n = \vec{0}$$

To get matrix $P = [\vec{w}_1 \ \vec{w}_2 \ \cdots \ \vec{w}_n]$, the key is obtain the vector $\vec{w}_n$.

Let $\vec{w}_n$ be the vector such that $\vec{w}_n \in \ker(A - \lambda I)^n$ and $\vec{w}_n \notin \ker(A - \lambda I)^{n-1}$.

Claim: $\{N^{n-1}\vec{w}_n, N^{n-2}\vec{w}_n, \ldots, \vec{w}_n\}$ is independent.

Definition 7. Let A be an $n \times n$ matrix. A non-zero vector $\vec{v}$ is called a **generalized eigenvector** of A if

$$(A - \lambda I)^k \vec{v} = \vec{0}$$

for some $k \geq 1$.

Remark:

(1) Any eigenvector is a generalized vector.
(2) A generalized vector can exist only for the regular eigenvalue λ. A generalized vector can exists if and only if $\det[(A - \lambda I)^k] = 0$, which only happen when $\det[(A - \lambda I)] = 0$.
(3) Let V_λ be the set of all generalized eigenvectors together with $\vec{0}$. Then V_λ is a subspace of $\mathbb{F}^n$.
(4) A Jordan chain is independent if and only if $\vec{v}_1 \neq \vec{0}$.
(5) A is similar to a Jordan block $J_{\lambda,n}$ if and only if there exists a Jordan Chain $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$ if and only if there exists a vector $\vec{v}_n$ such that $(A - \lambda I)^n \vec{v}_n = 0$ but $(A - \lambda I)^{n-1} \vec{v}_n \neq 0$.

We need to find the structure of a nilpotent matrix. We want to show that any nilpotent matrix is similar to $J_n(0)$. For example,

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdots$$

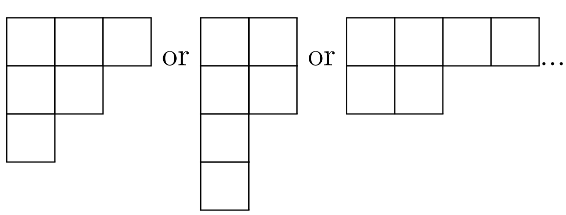
That is $J_3 \oplus J_2 \oplus J_1$, or $J_2 \oplus J_2 \oplus J_1 \oplus J_1$, or $J_4 \oplus J_2$...

There is a one-to-one corresponding between $J_n(0)$ and **partition** of n, $(n_1, n_2, \ldots, n_k)$ such that

$$n = n_1 + n_2 + \cdots + n_k \text{ and } n_1 \geq n_2 \geq \cdots \geq n_k \geq 1$$

In the above examples, the partition of 6 are $(3, 2, 1)$, or $(2, 2, 1, 1)$, or $(4, 2)$... (How many? 11)

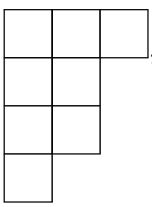
We can use the **Young diagram** to describe the partitions.



From the Young diagrams, we can easily find the dual partitions by summing the squares in another direction. In above examples, the dual partitions are $(3,2,1)$, $(4,2)$, $(2,2,1,1)$

For another example, the Jordan matrix corresponds to partition $(n_1, n_2, n_3) = (3,2,2,1)$ is

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

By Young diagram, , the dual partition is $(s_1, s_2, s_3) = (4,3,1)$.

Lemma 8. *Let N be an $n \times n$ nilpotent of degree r. Then we have strict inclusions*

$$\ker N \subset \ker N^2 \subset \cdots \subset \ker N^r = \mathbb{F}^n$$

Proof. If $\vec{v} \in \ker N^k$, then $N^k\vec{v} = \vec{0}$, hence $N^{k+1}\vec{v} = \vec{0}$, hence $\vec{v} \in \ker N^{k+1}$, hence $N^k \subseteq N^{k+1}$. Since N is nilpotent of degree r, there is a vector such that $\vec{v} \in \ker N^r = V$ but $\vec{v} \notin \ker N^{r-1}$. Then $N^{r-i}\vec{v} \in \ker N^i$ but $N^{r-i}\vec{v} \notin \ker N^{i-1}$. Hence each inclusion is strict. □

As for the above example B, $m_1 = \dim \ker N = 4$, $m_2 = \dim \ker N^2 = 7$, $m_3 = \dim \ker N^3 = 8$.

Notice that $m_1 = s_1$, $m_2 = m_1 + s_2$, $m_3 = m_2 + s_3$. Or $m_1 = s_1$, $m_2 = s_1 + s_2$, $m_3 = s_1 + s_2 + s_3$.

Theorem 9. *Let N be an $n \times n$ nilpotent matrix of degree r. Then there exist vectors $\vec{v}_1, \dots, \vec{v}_s$ and integers $n_1, \dots, n_s$ with $1 \le n_s \le \cdots \le n_1 = r$ such that $N^{n_i - 1}\vec{v}_i \neq \vec{0}$ and $N^{n_i}\vec{v}_i = \vec{0}$ for all $i = 1, 2, ..., s$ and vectors*

$$\begin{array}{c} N^{n_1-1}\vec{v}_1, \dots, \dots, \dots N\vec{v}_1,\ \vec{v}_1, \\ N^{n_2-1}\vec{v}_2 \dots, \dots, N\vec{v}_2,\ \vec{v}_2, \\ \vdots \\ N^{n_s-1}\vec{v}_s \dots, N\vec{v}_s,\ \vec{v}_s \end{array}$$

form a basis for $\mathbb{F}^n$.

Proof. By Lemma 8, there are strict inclusions

$$\ker N \subset \ker N^2 \subset \cdots \subset \ker N^r = \mathbb{F}^n$$

Hence, there exist direct decompositions

$$\ker N^i = \ker N^{i-1} \oplus W_i$$

Hence $\mathbb{F}^n = W_r \oplus W_{r-1} \oplus \cdots \oplus W_2 \oplus W_1$, where $W_1 = \ker N$.
Denote the dimension of each null space as $m_i = \dim\ker N^i$ for $i = 1, 2, ..., r$. Then denote $\dim W_i = s_i$ where $s_1 = m_1$, $s_2 = m_2 - m_1$, $s_3 = m_3 - m_2$,..., $s_r = m_r - m_{r-1}$.
Choose a basis $\{\vec{w}_{r,1}, .., \vec{w}_{r,s_r}\}$ for W_r. $\{\vec{w}_{r,1}, .., \vec{w}_{r,s_r}, ..., N^{r-1}\vec{w}_{r1}, .., N^{r-1}\vec{w}_{r,s_r}\}$ is independent. $N^i\vec{w}_{r,*} \in W_{r-i}$ for $i = 0, 1, 2, ..., r-1$.
Extend $\{N\vec{w}_{r,1}, .., N\vec{w}_{r,s_r}\}$ to be a basis for W_{r-1} by adding $\{\vec{w}_{r-1,1}, .., \vec{w}_{r-1,s_{r-1}-s_r}\}$.
Keep extending until to W_1, we extended $\{N^{r-1}\vec{w}_{r,1}, .., N^{r-1}\vec{w}_{r,s_r}, N^{r-2}\vec{w}_{r,1}, .., N^{r-2}\vec{w}_{r-2,s_r}, ...\}$ to be a basis for W_1 by adding $\{\vec{w}_{1,1}, .., \vec{w}_{1,s_1-s_2}\}$
Claim, the set

$$\begin{array}{lll}
N^{r-1}\vec{w}_{r1} \dots & N\vec{w}_{r,1}, & \vec{w}_{r,1}, \\
\vdots & & \\
N^{r-1}\vec{w}_{r,s_r} \dots & N\vec{w}_{r,s_r} & \vec{w}_{r,s_r} \\
N^{r-2}\vec{w}_{r-1,1}, \dots & \vec{w}_{r-1,1}, & \\
\vdots & & \\
N^{r-2}\vec{w}_{r-1,,s_{r-1}-s_r} \dots & \vec{w}_{r-1,s_{r-1}-s_r} & \\
\quad\vdots & & \\
\vec{w}_{1,1}, & & \\
\vdots & & \\
\vec{w}_{1,s_1-s_2} & &
\end{array}$$

is a basis for $\mathbb{F}^n$. □

Remark: The proof can also be done by induction on r or n. But our proof gives an algorithm of finding the basis.

In the example, if we want to fit the Young diagram, it is

$B^2\vec{v}_1$	$B\vec{v}_1$	$\vec{v}_1$
$B\vec{v}_2$	$\vec{v}_2$	
$B\vec{v}_3$	$\vec{v}_3$	
$\vec{v}_4$		

.

The first columns form basis for $\ker B$. The first two columns form basis for $\ker B^2$. All vectors form a basis for $\ker B^3 = \mathbb{F}^8$.

Remark: In the theorem, $(n_1, n_2, ..., n_s)$ is a partition of n corresponding the sizes of Jordan blocks. The dual partition is $(s_1, s_2, ..., s_r)$.

Denote the dimension of each null space as $m_i = \dim\ker N^i$ for $i = 1, 2, ..., r$. Then $m_1 = s_1$, $m_2 = m_1 + s_2$, $m_3 = m_2 + s_3$,..., $m_r = m_{r-1} + s_r$.

Denote the rank $c_i = \operatorname{rank} N^i$ for $i = 1, 2, ..., r$. Then $c_1 = n - s_1$, $c_2 = c_1 - s_2$, $c_3 = c_2 - s_3$,..., $c_r = c_{r-1} - s_r$.

So, the procedure of calculation is find m_i or c_i first, then s_i, then n_i.

Remark: In the theorem, $N^{n_1-1}\vec{v}_1$, $N^{n_2-1}\vec{v}_2$, ... $N^{n_s-1}\vec{v}_s$ form a basis for $\ker N$.

Corollary 10. *Let N be an $n \times n$ matrix. N is nilpotent if and only if N is similar to a Jordan canonical matrix $J_n(0)$.*

Proof. The forward direction ($\Rightarrow$) is by Theorem 9.
The backward direction ($\Leftarrow$) is from Lemma 6.

□

Corollary 11. *Let N be an $n \times n$ nilpotent matrix. Then $\lambda I + N$ is similar to a Jordan canonical matrix $J_n(\lambda)$.*

3. **Jordan Canonical Form**

Theorem 12. *Let A be an $n \times n$ matrix. If $\ker A \cap \operatorname{im} A = \{0\}$, then $\mathbb{F}^n = \ker A \oplus \operatorname{im} A$.*

Proof. We know that $\dim \ker A + \dim \operatorname{im} A = n$. Together with $\ker A \cap \operatorname{im} A = \{0\}$, we have the conclusion. □

Remark: (1) The assumption is needed. For example, $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. $\ker A = \operatorname{im} A = \operatorname{Span}\{\vec{e}_2\}$

(2) Notice that $T_A(\ker A) = \{\vec{0}\}$ and $T_A(\operatorname{im} A) \subset \operatorname{im} A$. So, both $\ker A$ are $\operatorname{im} A$ invariant under T_A.

Theorem 13. *Let A be an $n \times n$ matrix with an eigenvalue λ. Denote the set of all generalized eigenvectors of A corresponding to λ, together with $\{\vec{0}\}$ by V_λ. Then, there exists m such that*

$$V_\lambda = \ker(A - \lambda I)^m$$

and

$$\mathbb{F}^n = \ker(A - \lambda I)^m \oplus \operatorname{im}(A - \lambda I)^m.$$

Both $\ker(A - \lambda I)^m$ and $\operatorname{im}(A - \lambda I)^m$ are invariant under T_A.

Proof. The theorem can be proved by the following steps.
1. Verify V_λ is a subspace of $\mathbb{F}^n$. (Verify by definition.)
Let m be the (smallest) number that $(A - \lambda I)^m \vec{v} = \{\vec{0}\}$ for any $\vec{v} \in V_\lambda$. This can be done since V_λ is a finite-dimensional vector space. We only need to vanish the basis vectors.
2. It is clear that $V_\lambda = \ker(A - \lambda I)^m$.
3. $\ker(A - \lambda I)^m \cap \operatorname{im}(A - \lambda I)^m = \{\vec{0}\}$
Suppose $\vec{v} \in \ker(A - \lambda I)^m \cap \operatorname{im}(A - \lambda I)^m$, then $(A - \lambda I)^m \vec{v} = \vec{0}$ and $\vec{v} = (A - \lambda I)^m \vec{w}$. Then $(A - \lambda I)^{2m} \vec{w} = \vec{0}$. Then $\vec{w} \in V_\lambda$. Then $(A - \lambda I)^m \vec{w} = \vec{0}$. So, $\vec{v} = \vec{0}$.
4. Each space is invariant under $(A - \lambda I)$, hence also A. □

Theorem 14. *Let A be an $n \times n$ matrix with n eigenvalues. The distinct eigenvalues are $\lambda_1, \dots, \lambda_k$. Then, there exist numbers $m_1, m_2, \dots, m_k$ such that*

$$\mathbb{F}^n = \ker(A - \lambda_1 I)^{m_1} \oplus \cdots \oplus \ker(A - \lambda_k I)^{m_k}$$

and each $\ker(A - \lambda_i I)^{m_i}$ is invariant under T_A.

Proof. By induction on number of distinct eigenvalues. T_A has eigenvalues $\lambda_2, \dots, \lambda_k$ on $\text{im}(A - \lambda I)^m$. □

Remark: More generally, all properties in this section can be generalized to linear transformations T on a finite-dimensional vector space V. (We discussed a particular case when $V = \mathbb{F}^n$.)

Choose a basis for each subspace $\ker(A - \lambda_i I)^{m_i}$ and put them together we get a basis $\mathscr{B} = \{\vec{v}_1, \cdots, \vec{v}_n\}$ for $\mathbb{F}^n$. The $\mathscr{B}$-matrix for T_A is block diagonal $B = B_1 \oplus B_2 \oplus \cdots \oplus B_k$, since each space is invariant under T_A. For each matrix B_k we know that $(B_i - \lambda_i I)^{m_i} = \mathbf{0}$.

Hence we have showed that A is similar to a block diagonal matrix $B_1 \oplus B_2 \oplus \cdots \oplus B_k$ such that each B_i is $\lambda_i I$ plus a nilpotent matrix.

Theorem 15 (Block Diagonalization)**.** *Every $n \times n$ matrix A with n eigenvalues in a field $\mathbb{F}$ is similar to a block diagonal matrix, where each block has a single eigenvalue.*
More precisely, suppose $\lambda_1, \lambda_2, \dots, \lambda_k$ are the distinct eigenvalues of A. Then there is an invertible matrix $P \in \mathbb{F}^{n \times n}$ such that

$$P^{-1}AP = \text{diag}(B_1, B_2, \dots, B_k)$$

where the matrix $B_i - \lambda_i I$ is nilpotent for $i = 1, 2, \dots, k$.

Together with the result for nilpotent matrix, we have

Theorem 16. *Every $n \times n$ matrix A with n eigenvalues in a field $\mathbb{F}$ is similar to a matrix J in Jordan normal matrix, that is $A = PJP^{-1}$.*

4. **Algorithm and example**

Let A be an $n \times n$ matrix with distinct real eigenvalues $\lambda_1, \dots, \lambda_p$ such that

$$f_A(\lambda) = \det(A - \lambda I) = (\lambda_1 - \lambda)^{k_1}(\lambda_2 - \lambda)^{k_2} \cdots (\lambda_p - \lambda)^{k_p}.$$

Suppose $k_1 + k_2 + \cdots + k_p = n$. (This is always true if $\mathbb{F}$ is algebraic closed, e.g., when $\mathbb{F} = \mathbb{C}$).

Algorithm of computing Jordan Normal form of a matrix:

Step 1. Find all eigenvalues λ_i and their algebraic multiplicity $am(\lambda_i) = k_i$.
Step 2. For each eigenvalue λ_i, calculate $m_j = \dim\ker(A - \lambda_i I)^j$ for $j = 1, 2, ...$ until $\dim\ker(A - \lambda_i I)^s = k_i$.
Step 3. From $m_1, ..., m_s$ we can calculate $s_j = m_j - m_{j-1}$, then use Young diagram calculate $n_1, ..., n_t$. Now we have determined the Jordan normal form J.
Step 4. To calculate the matrix P such that $A = PJP^{-1}$, we calculate $\mathbf{rref}(A - \lambda I)^j$ for each $\lambda = \lambda_i$.
Step 5. Find vectors $\{\vec{w}_{r,1}, ...\vec{w}_{r,s_r}\}$,..., $\{\vec{w}_{1,1}, ...\vec{w}_{1,s_1 - s_2}\}$ such that

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & -8 & 4 & -3 & 1 & -3 \\ -3 & 13 & -8 & 6 & 2 & 9 \\ -2 & 14 & -7 & 4 & 2 & 10 \\ 1 & -18 & 11 & -11 & 2 & -6 \\ -1 & 19 & -11 & 10 & -2 & 7 \end{bmatrix}$$

Step 1, calculate all eigenvalues of A, which are $\lambda = 2$ with algebraic multiplicity 1 and $\lambda = -1$ with algebraic multiplicity 5. We know that the Jordan form looks like:

$$J = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & * & 0 & 0 & 0 \\ 0 & 0 & -1 & * & 0 & 0 \\ 0 & 0 & 0 & -1 & * & 0 \\ 0 & 0 & 0 & 0 & -1 & * \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Calculate $m_i = \dim \ker((A+I)^i)$ we have $m_1 = 2, m_2 = 4, m_3 = 5$ which is the algebraic multiplicity $am(-1)$. So, $s_1 = 2, s_2 = 2, s_3 = 1$ and by Young diagram

$$\begin{array}{|c|c|c|} \hline B^2\vec{v}_1 & B\vec{v}_1 & \vec{v}_1 \\ \hline B\vec{v}_2 & \vec{v}_2 \\ \cline{1-2} \end{array}.$$

$n_1 = 3, n_2 = 2$. So, $J = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$

To find matrix P such that $A = PJP^{-1}$, we need to calculate

$$\mathbf{rref}(A+I) = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & 0 & 1 & \frac{3}{2} \\ 0 & 0 & 1 & 0 & 2 & \frac{3}{2} \\ 0 & 0 & 0 & 1 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{rref}(A+I)^2 = \begin{bmatrix} 1 & 0 & -\frac{1}{2} & \frac{3}{2} & -2 & -\frac{5}{2} \\ 0 & 1 & -\frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\textbf{rref}(A+I)^3 = \begin{bmatrix} 0 & 1 & -\frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$\vec{v}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$ is the vector in $\ker(A+I)^3$ but not in $\ker(A+I)^2$

Calculate $(A+I)\vec{v}_1 = \begin{bmatrix} 1 & 0 & -3 & -2 & 1 & -1 \end{bmatrix}^T$ and $(A+I)^2\vec{v}_1 = \begin{bmatrix} 1 & -2 & -1 & 1 & -1 & 2 \end{bmatrix}^T$

$\vec{v}_2 = \begin{bmatrix} 0 & 1 & -2 & -2 & 3 & -3 \end{bmatrix}^T$ is the vector in $\ker(A+I)^2$ but not in $\ker(A+I)$ and not dependent on $\vec{v}_1$, $(A+I)\vec{v}_1$ and $(A+I)^2\vec{v}_1$

$\textbf{rref}(A+2I) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 & -\frac{2}{3} \\ 0 & 0 & 0 & 1 & 0 & -\frac{2}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ A basis for $\ker(A+2I)$ is $\begin{bmatrix} 0 & 1 & -2 & -2 & 3 & -3 \end{bmatrix}^T$ Hence matrix P is $P = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & -2 & 0 & 0 & 1 & 1 \\ -2 & -1 & -3 & 0 & -4 & 2 \\ -2 & 1 & -2 & 0 & -2 & 0 \\ 3 & -1 & 1 & 0 & 5 & 0 \\ -3 & 2 & -1 & 0 & -4 & 0 \end{bmatrix}$

Using Matlab directly A=sym(A) and [P, J] = jordan(A) will give us the result

$$J = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \quad P = \begin{bmatrix} 0 & -\frac{9}{2} & -7 & -7 & \frac{3}{2} & \frac{5}{2} \\ -1 & 9 & 3 & 1 & 0 & 0 \\ 2 & \frac{9}{2} & 18 & \frac{5}{2} & -\frac{9}{2} & -\frac{3}{2} \\ 2 & -\frac{9}{2} & \frac{17}{2} & 2 & -\frac{3}{2} & -1 \\ -3 & \frac{9}{2} & -6 & \frac{3}{2} & \frac{9}{2} & \frac{3}{2} \\ 3 & -9 & \frac{7}{2} & -\frac{3}{2} & -3 & -\frac{1}{2} \end{bmatrix}$$

Remark: The Jordan normal form is more useful in theory than in computation. There is a technical problem of Jordan Normal Form in numerical calculation. For example,

$$A = \begin{bmatrix} 1 & 1 \\ 0 & t \end{bmatrix}$$

Then when $t \neq 1$, the matrix A is diagonalizable with $D = \begin{bmatrix} 1 & 0 \\ 0 & t \end{bmatrix}$. However, when $t = 1$, the matrix A is not diagonalizable and the Jordan normal form is $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. This means that the calculation is not "continuous". A small floating approximation in computer calculation may give a huge mistake in the Jordan Normal Form calculation.

Definition 17. An **annihilating polynomial** for a square matrix A is a non-zero polynomial $p(t)$ such that $p(A) = 0$.

Theorem 18. *Then there exists an annihilating polynomial for any $n \times n$ matrix A.*

Proof. Suppose $\{\vec{v}_1, \dots, \vec{v}_n\}$ is a basis for $\mathbb{F}^n$, then, each $\{\vec{v}_i, A\vec{v}_i, \dots, A^n\vec{v}_i\}$ is dependent. ($n+1$ vectors) So, there exists a dependent relation

$$a_{i0}\vec{v}_i + a_{i1}A\vec{v}_i + \cdots + a_{in}A^n\vec{v}_i = \vec{0}$$

Denote the polynomial $p_i(t) = a_{i0} + a_{i1}t + \cdots + a_{in}t^n$ So $p(A) = \prod_{i=1}^n p_i(A)$ sent a basis of $\mathbb{F}^n$ to zero. So, $P(A) = \vec{0}$. □

Remark: Another way to prove the theorem is using vector spaces $F^{n\times n}$ with $\dim(F^{n\times n}) = n^2$. So, A is a vector in $F^{n\times n}$. So $n+1$ vectors in $F^{n\times n}$ is dependent. So I, A, A^2, ..., A^{n^2} is dependent. So, there exists a polynomial annihilating A.

The degree of the annihilating polynomial is n^2. In fact, the degree can be smaller.

Theorem 19 (**Cayley-Hamilton Theorem**)**.** *If $f(t)$ is the characteristic polynomial of A, then $f(A) = \mathbf{0}$.*

Proof. Suppose $f_A(t) = \det(A - tI) = (\lambda_1 - t)^{k_1}(\lambda_2 - t)^{k_2}\cdots(\lambda_p - t)^{k_p}$.
If A is diagonalizable, (i.e., $A = PDP^{-1}$), the proof is easy. Since f is a polynomial, $f(A) = Pf(D)P^{-1} = P\begin{bmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(\lambda_p) \end{bmatrix} P^{-1} = P\mathbf{0}P^{-1} = \mathbf{0}$.
In general, we use Jordan normal forms decomposition $A = PJP^{-1}$. We only need to show that $f(J) = \mathbf{0}$.

$$f(J) = \begin{bmatrix} f(J_{\lambda_1}(k_1)) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & f(J_{\lambda_2}(k_2)) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & f(J_{\lambda_m}(k_p)) \end{bmatrix}$$

Each matrix $f(J_{\lambda_i}(k_i)) = (\lambda_1 I - J_{\lambda_i}(k_i))^{k_1}\cdots(\lambda_i I - J_{\lambda_i}(k_i))^{k_i}\cdots(\lambda_p I - J_{\lambda_i}(k_i))^{k_p} = \mathbf{0}$, since $(\lambda_i I - J_{\lambda_i}(k_i))^{k_i} = \mathbf{0}$ by Lemma 6 □

Wrong proof: $f(t) = \det(A - tI)$. So, $f(A) = \det(A - AI) = \det(0) = 0$. (Why?)

Application to computing powers A^k of matrix A using linear combinations of I, A, ..., A^{n-1}.

6. **Minimal polynomial**

By Cayley-Hamilton Theorem, we know that we can find annihilating polynomial of A with degree $\leq n$.

Definition 20. The smallest degree annihilating polynomial of A is called the **minimal polynomial** of A.

Theorem 21 (Minimal Polynomial Theorem)**.** *Consider $\mathbb{F} = \mathbb{C}$. The eigenvalues of A are the roots of the minimal polynomial $f(t)$ of A.*

Corollary 22. *The minimal polynomial $f(t)$ of A has the form*

$$f(t) = (t - \lambda_1)^{p_1}(t - \lambda_2)^{p_2} \cdots (t - \lambda_m)^{p_m}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_m$ be the distinct eigenvalues of A and the exponents p_k is the largest block size for each eigenvalue..