Linear Algebra  Jordan Canonical Form  Lecture 38

Last Time: Characteristic Polynomial

This Time: Matrices that are not diagonalizable

Today \( \mathbb{F} = \mathbb{C} \)

Since \( \mathbb{C} \) is algebraically closed every \( A \in M_n(\mathbb{C}) \)

is similar to an upper triangular matrix (Schur decomposition).

We know that a matrix is unitarily diagonalizable

if and only if the matrix is normal (i.e., it commutes with its adjoint).

If we don't require orthonormal bases,

where is a matrix diagonalizable?

Given \( A \in M_n(\mathbb{C}) \) we may as well assume that it is

upper triangular, \( A = \begin{pmatrix} \lambda_1 & * & * \\ 0 & \ddots & \ddots \\ 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} \lambda_1 & * & * \\ 0 & \ddots & \ddots \\ 0 & \cdots & \lambda_n \end{pmatrix} \)

The elements on the diagonal are the eigenvalues of \( A \)

& they are repeated with (algebraic) multiplicity.

Let \( m_A(\lambda) \) = algebraic multiplicity of \( \lambda \) as an eigenvalue of \( A \)

So if \( \lambda_1, \ldots, \lambda_k \) are the distinct eigenvalues of \( A \)

then \( m_A(\lambda_1) + \cdots + m_A(\lambda_k) = n \).

E.g., \( A = \begin{bmatrix} 3 & 7 \\ 0 & 3 \end{bmatrix} \) is the only eigenvalue of \( A \) & it has

algebraic multiplicity 2.

Its eigenspace is \( \operatorname{Eyg}_3(A) = \ker (A - 3 I_2) = \ker \left( \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix} \right) = \operatorname{span} \{ e_1 \} \)

has dimension 1.
We always have \( \dim \ker (A-\lambda I_n) \leq m_A(\lambda) \).
Indeed, as both sides are invariant we may assume that
\( A \) is upper triangular, then
\( \text{rank} \ (A-\lambda I_n) \geq \# \text{non-zero diagonal entries} \)
so \( \text{null} \ (A-\lambda I_n) = n - \# \text{non-zero diagonal entries} = \# \text{zeros on diagonal} - m_A(\lambda) \).

Proof: \( A \in M_n(\mathbb{C}) \) is diagonalizable \( \iff \dim \ker (A-\lambda I_n) = m_A(\lambda) \)
for all eigenvalues \( \lambda \) of \( A \).

\( \iff A \) is diagonalizable \( \iff \) there is a basis of \( \mathbb{C}^n \) consisting of
eigenvectors of \( A \)
\( \iff \dim \ker (A-\lambda I_n) + \ldots + \dim \ker (A-\lambda_k I_n) = n = m_A(\lambda_1) + \ldots + m_A(\lambda_k) \)
\( \iff \dim \ker (A-\lambda_j I_n) = m_A(\lambda_j) \) for all \( \lambda_j \).

If \( A \) is upper triangular & not diagonalizable
then there's an eigenvalue \( \lambda \) that occurs on the diagonal
more times than there are lin. ind. eigenvectors.

Eg., \( A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix} \), \( A-2I_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \), \( \text{null} (A-2I_3) \leq 3 \),
\( (A-2I_3)^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \), \( \text{null} (A-2I_3)^2 \leq 3 \),
\( (A-2I_3)^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \), \( \text{null} (A-2I_3)^3 = 3 = m_A(2) \).
For any eigenvalue $\lambda$ of $A$ we have
\[\{0\} \subseteq \text{Ker} (A-\lambda I_n) \subseteq \text{Ker} (A-\lambda I_n)^2 \subseteq \ldots \subseteq \text{Ker} (A-\lambda I_n)^k\]
If $\text{Ker} (A-\lambda I_n)^j = \text{Ker} (A-\lambda I_n)^j$
then $\text{Ker} (A-\lambda I_n)^j = \text{Ker} (A-\lambda I_n)^{j+1} \neq \emptyset$.
So this chain of spaces has to stop at some point
in fact $\text{Ker} (A-\lambda I_n)^m = \text{Ker} (A-\lambda I_n)^{m+1} \neq \emptyset$.

**Def.** If $\lambda \in \mathbb{C}$ is an eigenvalue of $A \in M_n(\mathbb{C})$
then the generalized eigenspace of $\lambda$ is $\text{GEig}_\lambda(A) = \text{Ker} (A-\lambda I_n)^n$.
Non-zero elements of $\text{GEig}_\lambda(A)$ are referred to as
generalized eigenvectors of $A$.

**Theorem.** Let $A \in M_n(\mathbb{C})$ have distinct eigenvalues $\lambda_1, \ldots, \lambda_k$.

1) Multiplication by $A$ preserves $\text{GEig}_{\lambda_j}(A) \neq \emptyset$.
2) If $i \neq j$ then $\text{GEig}_{\lambda_i}(A) \cap \text{GEig}_{\lambda_j}(A) = \{0\}$.
3) If $v_1, \ldots, v_m$ are generalized eigenvectors corresponding to
   distinct eigenvalues then $\{v_1, \ldots, v_m\}$ is lin. indep.
4) $\mathbb{C}^n = \text{GEig}_{\lambda_1}(A) \oplus \ldots \oplus \text{GEig}_{\lambda_k}(A)$
5) For all $\lambda_j$, $\dim \text{GEig}_{\lambda_j}(A) = m_A(\lambda_j)$.

**Proof.**

Given $v \in \text{GEig}_\lambda(A)$ we want to show that $A v \in \text{GEig}_\lambda(A)$
We know that $(A-\lambda I_n)^n v = 0$, want to show $(A-\lambda I_n)^m(Av) = 0$,
but $(A-\lambda I_n)^m(Av) = (A-\lambda I_n)^m(A-\lambda I_n)v + \lambda I_n v$
\[= (A-\lambda I_n)^{m+1} v + \lambda (A-\lambda I_n)^m v = 0.\]
iii) Suppose \( i \neq j \) and \( \lambda \in \text{GEig}_k(A) \cap \text{GEig}_{k_j}(A) \)

We know that there are \( l_i, l_j \in \mathbb{N} \) s.t.

\[
(A - \lambda I)^{l_i} \omega = 0 = (A - \lambda_j I)^{l_j} \omega
\]

but \( (A - \lambda I)^{l_i - 1} \omega \neq 0, (A - \lambda_j I)^{l_j - 1} \omega \neq 0 \). Proceed by induction on \( l_j \).

If \( l_j = 1 \), so \( \omega = \lambda_j \omega, \lambda_n (A - \lambda I) \omega = (\lambda - \lambda_j) \omega \)

and since \( \lambda_j - \lambda \neq 0 \) this vanishes \( \implies \omega = 0 \).

For the general case, let \( \tilde{\omega} = (A - \lambda_j I)^{l_j - 1} \omega \neq 0 \) and \( (A - \lambda I) \tilde{\omega} = 0 \)

so by the previous case \( \tilde{\omega} = 0 \). \( \checkmark \)

iv) Suppose \( v_j \in \text{GEig}_{k_j}(A) \) for \( j \in \{1, \ldots, N\} \)

we want to show \( \{v_1, \ldots, v_N\} \) is lin. indep.

The case \( N = 2 \) is (iii). Assume inductively that this is known for sets of size \( N - 1 \) & that

\[
a_1 v_1 + \cdots + a_N v_N = 0
\]

Applying \( (A - \mu N I)^n \) to both sides, we get

\[
a_1 (A - \mu I)^n v_1 + \cdots + a_N (A - \mu I)^n v_N = 0
\]

by inductive hypothesis we can conclude that \( a_1 = 0, \ldots, a_N = 0 \).

Plugging this into \( a_1 v_1 + \cdots + a_N v_N = 0 \) we see that \( a_1 = 0 \).

v) \( C^n = \text{GEig}_{\lambda_1}(A) \oplus \cdots \oplus \text{GEig}_{\lambda_k}(A) \)

By induction on \( n \).

If \( n = 1 \) then every non-zero vector is an eigenvector of \( A \).

If inductively we knew this for complex vector spaces of dimension \( \text{dim} A \) then given \( A \) it has an eigenvalue \( \lambda \Delta \)

\( C^n = \text{GEig}_{\lambda}(A) \oplus \text{Range} (A - \lambda I)^n \)
The range of \((A-\lambda I)^n\) is preserved by multiplication by \(A\) if \(A\) has dimension less than \(n\) so we can apply our inductive hypothesis.

v) Since \(C\) is a basis of \(A_{\text{dim}}(C)\)
\[ \dim GE_{\lambda_1}(A) + \ldots + \dim GE_{\lambda_k}(A) = n = m_1(\lambda_1) + \ldots + m_k(\lambda_k) \]
then \(\dim GE_{\lambda_j}(A) = m_j(\lambda_j)\).

Given a matrix \(A \in M_n(C)\). Since \(C^n = GE_{\lambda_1}(A) \oplus \ldots \oplus GE_{\lambda_k}(A)\) is a basis of \(C^n\) consisting of generalized eigenvectors of \(A\), then \(A\) is similar to a "block diagonal" matrix.

ii. \(A\) has the form \(\begin{pmatrix} \mathcal{A}_1 & 0 & \cdots & 0 \\ 0 & \mathcal{A}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{A}_k \end{pmatrix}\) with \(\mathcal{A}_j \in M_{m_j(\lambda_j)}(C)\).

We can choose the bases so that each \(\mathcal{A}_j\) is upper triangular,
\[ A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix} \text{ is similar to a matrix of the form } \begin{pmatrix} \mathcal{A} & 0 \\ 0 & \mathcal{B} \end{pmatrix} \]

We can improve this further and get a block diagonal matrix whose diagonal blocks are Jordan blocks.

\[ J(\lambda, 1) = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \quad J(\lambda, 2) = \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}, \quad \text{ etc.} \]

\[ J(\lambda, m) \in M_m(C), \quad J(\lambda, m) = \lambda I_m + N_m, \quad [N_m]_{ij} = \begin{cases} 1 & \text{if } j = i+1 \\ 0 & \text{else} \end{cases} \]

\[ J(\lambda, m) = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{pmatrix}, \quad N_m, N_m^2, \ldots, N_m^{m-1} \text{ not zero, let } N_m^m = 0. \]
Let’s construct a basis of $\mathbb{C}^3$, starting with a vector in $\ker (A-2I)^2$ but not in $\ker (A-2I)^3$.

- $v_3 = e_3$
- Let the set $v = (A-2I) v_3 = \begin{bmatrix} 4 \\ 3 \\ 0 \end{bmatrix}$
- $v_1 = (A-2I) v_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

We get a basis $B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$ of $\mathbb{C}^3$.

and, in this basis,

$$
\begin{bmatrix}
[A(v)]_{v_3} & [A(v)]_{v_1} & [A(v)]_{v_2}
\end{bmatrix}
$$

$$
A v_1 = (A-2I) v_1 + 2 v_1 = 2 v_1
$$

$$
A v_2 = (A-2I) v_2 + 2 v_2 = v_1 + 2 v_2
$$

$$
A v_3 = (A-2I) v_3 + 2 v_3 = v_2 + 2 v_3
$$

**Lemma:** If $(A-\lambda I)^n v = 0 \neq (A-\lambda I)^{n-1} v \neq 0$

then $\{ v, (A-\lambda I)v, \ldots, (A-\lambda I)^{n-1} v \}$ is lin. indep.

**Proof:** Assume $\sum a_j (A-\lambda I)^{j-1} v = 0$.

Applying $(A-\lambda I)^{k-1}$ to both sides, we get $a_k (A-\lambda I)^{k-1} v = 0 = a_k v = 0$.

Inductively assume we’ve shown $a_1, \ldots, a_{k-1}$ are zero.

So apply $(A-\lambda I)^{k-n}$ to both sides, we get $a_k (A-\lambda I)^{k-n} v = 0 = a_k v = 0$.

So by induction $a_1 = 0, \ldots, a_k = 0$. \(\blacksquare\)
Conclusion: each \( cA_3 \), the block corresponding to \( A_3 \), is in an appropriate basis of the form

\[
\begin{bmatrix}
J(\lambda_3, k_1) & 0 & \cdots & 0 \\
0 & J(\lambda_4, k_4) & 0 & \cdots \\
0 & 0 & \ddots & \vdots \\
0 & 0 & \cdots & J(\lambda_N, k_N)
\end{bmatrix}
k_1 + \cdots + k_N = m_A(A_3)
\]

Any \( A \in M_n(\mathbb{C}) \) with distinct eigenvalues \( \lambda_1, \ldots, \lambda_N \) is similar to a matrix of the form:

\[
\begin{bmatrix}
J(\lambda_1, k_1) & 0 & \cdots & 0 \\
0 & J(\lambda_2, k_2) & 0 & \cdots \\
0 & 0 & \ddots & \vdots \\
0 & 0 & \cdots & J(\lambda_N, k_N)
\end{bmatrix}
\]

\text{eg. } A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.

Consider \( \mathbb{T} : V \rightarrow V \), let

\[N(\mathbb{T}) = \{ x \in V : \mathbb{T}^kx = 0 \text{ for some } k > 0 \} = \bigcup \ker(\mathbb{T}^k)\]

\[N(\mathbb{T}^t) = \{ x \in V : (\mathbb{T}^t)^kx = 0 \text{ for any } k > 0 \} = \bigcap \ker(\mathbb{T}^k)\]

\[\exists m \text{ s.t. } N(\mathbb{T}) = \ker(\mathbb{T}^m), R(\mathbb{T}) = \mathbb{T}^m(V)\]

\underline{Lemma} \( V = N(\mathbb{T}) \oplus R(\mathbb{T}) \)
\[ T|_{\text{ker}} : \text{ker}(T) \rightarrow \text{ker}(T) \text{ is invertible} \]

(\text{ker} T \text{ is not } \emptyset \text{ so surjective})

\[ \forall x \in V \text{ let } y = (T|_{\text{ker}})^{-1} T^{-1} x \quad \text{and } \quad z = \overline{x - y} \]

Clearly \( x = y + z \), \( y \in \text{ker}(T) \) and \( \overline{T^{-1} z} = T^{-1} x - T^{-1} y = 0 \) \( \forall x \in \text{ker}(T) \)