Math 346 Lecture #32
12.2 Generalized Eigenvectors

We saw last time in Section 12.1 that a simple linear operator $A \in M_n(\mathbb{C})$ has the *spectral decomposition*

$$A = \sum_{i=1}^{n} \lambda_i P_i$$

where $\lambda_1, \ldots, \lambda_n$ are the distinct eigenvalues of A and P_i is the eigenprojection onto the eigenspace $\mathscr{N}(\lambda_i I - A)$. Something similar holds for semisimple A.

When A is not semisimple, there are not enough eigenvectors to form an eigenbasis. We must look for generalized eigenspaces that contains the eigenspaces in order to find a spectral decomposition of A.

Throughout we assume that V is a finite dimensional vector space over $\mathbb{F}$, which we know means that V is isomorphic to $\mathbb{F}^n$ for $n = \dim(V)$. When we speak of a linear operator A on V we will mean a linear operator on $\mathbb{F}^n$, i.e., $A \in M_n(\mathbb{F})$.

12.2.1 The Index of an Operator

Recall from Exercise 2.8 that for a linear operator B on any vector space (this includes infinite dimensional) we have the increasing sequence or *ascending chain* of subspaces

$$\mathscr{N}(B) \subset \mathscr{N}(B^2) \subset \cdots \subset \mathscr{N}(B^k) \subset \cdots .$$

When V is finite dimensional, the ascending chain stabilizes, i.e., there exists $K \in \mathbb{N}$ such that for all $k \geq K$ there holds $\mathscr{N}(B^k) = \mathscr{N}(B^{k+1})$, because the the nondecreasing sequence of dimensions $(\dim(\mathscr{N}(B^l)))_{l=0}^{\infty}$ is bounded above by $\dim(V)$ (proof of this upper bound is HW Exercise 12.6), where we understand $B^0 = I$.

Definition 12.2.1. The index of $B \in M_n(\mathbb{F})$, denoted by $\operatorname{ind}(B)$, is the smallest $k \in \{0, 1, 2, 3, \ldots\}$ such that $\mathscr{N}(B^k) = \mathscr{N}(B^{k+1})$.

Example 12.2.2. If $B \in M_n(\mathbb{F})$ is invertible, i.e., $\det(B) \neq 0$, then $\mathscr{N}(B^l) = \{0\}$ for all $l = 0, 1, 2, 3, \ldots$. Thus for invertible B we have $\operatorname{ind}(B) = 0$. To get a positive index requires that B is not invertible.

We show that the value k of the index of B is the integer K at which the ascending chain stabilizes, that is it not possible for $\mathscr{N}(B^l) = \mathscr{N}(B^{l+1})$ for some $l < \operatorname{ind}(B)$.

Theorem 12.2.3. If $\operatorname{ind}(B) = k$, then for all $l \geq k$ there holds $\mathscr{N}(B^l) = \mathscr{N}(B^{l+1})$, and each of the inclusions $\mathscr{N}(B^l) \subset \mathscr{N}(B^{l+1})$ is proper for all $l = 0, \ldots, k-1$.

Proof. The finite dimensionality of $\mathbb{F}^n$ implies that only finite many of the inclusions in the ascending chain

$$\mathscr{N}(B) \subset \mathscr{N}(B^2) \subset \cdots \subset \mathscr{N}(B^k) \subset \cdots$$

can be proper.

You showed in Exercise 2.12, that if $\mathscr{N}(B^l) = \mathscr{N}(B^{l+1})$ for some $l = 0, 1, 2, 3, \ldots$, then $\mathscr{N}(B^j) = \mathscr{N}(B^{j+1})$ for all $j \geq l$.

Thus with $k = \text{ind}(B)$ being the smallest value for which $\mathscr{N}(B^k) = \mathscr{N}(B^{k+1})$, we obtain for all $l \geq k$ that $\mathscr{N}(B^l) = \mathscr{N}(B^{l+1})$, and for all $l = 0, \ldots, k-1$ that the inclusions $\mathscr{N}(B^l) \subset \mathscr{N}(B^{l+1})$ are proper. □

Example (in lieu of 12.2.4). For the matrix

$$B = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

we have

$$\mathscr{N}(B) = \text{span}\{e_2\} \text{ and } \mathscr{R}(B) = \text{span}\{e_1, e_2, e_3\}.$$

Since

$$B^2 = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

we have

$$\mathscr{N}(B^2) = \text{span}\{e_2, e_3\} \text{ and } \mathscr{R}(B^2) = \text{span}\{e_1, e_2\}.$$

Since

$$B^3 = \begin{bmatrix} 9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 27 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

we have

$$\mathscr{N}(B^3) = \text{span}\{e_2, e_3, e_4\} \text{ and } \mathscr{R}(B^3) = \{e_1\}.$$

Since for all $l \geq 3$ we have

$$B^l = \begin{bmatrix} 3^l & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

for $l \geq 3$ we have

$$\mathscr{N}(B^l) = \text{span}\{e_2, e_3, e_4\} \text{ and } \mathscr{R}(B^l) = \text{span}\{e_1\}.$$

This gives $\text{ind}(B) = 3$.

Notice also that $\mathscr{N}(B^j)$ and $\mathscr{R}(B^j)$ intersect nontrivially when $j = 1, 2$, but that these subspaces intersect trivially when $j \geq 3$. This is not a coincidence.

Theorem 12.2.5. For $B \in M_n(\mathbb{C})$, if $k \geq \text{ind}(B)$, then $\mathbb{C}^n = \mathscr{R}(B^k) \oplus \mathscr{N}(B^k)$.

Proof. Suppose $k \geq \text{ind}(B)$ and let $x \in \mathscr{R}(B^k) \cap \mathscr{N}(B^k)$.

Then there exists $y \in \mathbb{C}^n$ such that $x = B^k y$, and $B^k x = 0$.

Hence $0 = B^k x = B^k(B^k y) = B^{2k} y$.

This implies that $y \in \mathscr{N}(B^{2k})$.

Since $k \geq \text{ind}(B)$ we have $\mathscr{N}(B^{2k}) = \mathscr{N}(B^k)$, so that $y \in \mathscr{N}(B^k)$.

With $x = B^k(y)$ and $y \in \mathscr{N}(B^k)$ we get $x = 0$.

Thus $\mathscr{R}(B^k) \cap \mathscr{N}(B^k) = \{0\}$.

This means that the subspace $\mathscr{R}(B^k)+\mathscr{N}(B^k)$ of $\mathbb{C}^n$ is a direct sum (by Definition 1.3.6).

By the Rank-Nullity Theorem, we have $\dim(\mathscr{R}(B^k)) + \dim(\mathscr{N}(B^k)) = n$, whence we have $\mathbb{C}^n = \mathscr{R}(B^k) \oplus \mathscr{N}(B^k)$. □

Corollary 12.2.6. For $B \in M_n(\mathbb{C})$, if $k = \text{ind}(B)$, then for all $m \geq k$ there holds $\mathscr{R}(B^m) = \mathscr{R}(B^k)$.

Proof. For $m \geq k$, we have $\mathscr{R}(B^m) \subset \mathscr{R}(B^k)$ by Exercise 2.8.

It suffices to show $\dim(\mathscr{R}(B^m)) = \dim(\mathscr{R}(B^k))$ because this implies that $\mathscr{R}(B^m) = \mathscr{R}(B^k)$ by Exercise 1.21.

By Theorem 12.2.3 we have $\dim(\mathscr{N}(B^m)) = \dim(\mathscr{N}(B^k))$.

By the Rank-Nullity Theorem we have $\dim(\mathscr{R}(B^m)) = \dim(\mathscr{R}(B^k))$. □

We present an important observation in the finite dimensional case about the vectors obtained by repeated powers of a linear operator acting on a given vector.

Proposition 12.2.7. For $B \in M_n(\mathbb{C})$ and $x \in \mathbb{C}^n$, if there exists $m \in \mathbb{N}$ such that $B^m x = 0$ and $B^{m-1}x \neq 0$, then the set $\{x, Bx, \ldots, B^{m-1}x\}$ is linearly independent.

Proof. Suppose by way of contradiction that there exist constants $a_0, a_1, \ldots, a_{m-1} \in \mathbb{F}$, not all zero, such that

$$a_0 x + a_1 Bx + \cdots + a_{m-1}B^{m-1}x = 0.$$

Let $i \in \{0, 1, \ldots, m-1\}$ be the smallest for which $a_i \neq 0$.

Then the equation becomes

$$a_i B^i x + a_{i+1}B^{i+1}x + \cdots + a_{m-1}B^{m-1}x = 0.$$

Multiplying the this equation through by B^{m-i-1} gives

$$a_i B^{m-1}x + a_{i+1}B^m x + \cdots + a_{m-1}B^{2m-i-2}x = 0.$$

Since $B^m x = 0$, then $B^{m+j}x = 0$ for all $j \geq 1$, and the equation reduces to $a_i B^{m-1}x = 0$.

Since $a_i \neq 0$, this contradicts $B^{m-1}x \neq 0$. □

12.2.2 Generalized Eigenspaces

Recall that the eigenspace of a linear operator $A \in M_n(\mathbb{C})$ associated to one of its eigenvalues λ is the subspace

$$\Sigma_\lambda = \mathscr{N}(\lambda I - A),$$

where the dimension of this subspace is the geometric multiplicity of λ.

If $A \in M_n(\mathbb{C})$ is semisimple (which includes the simple case) with spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_r\}$ (the distinct eigenvalues of A), then there holds

$$\mathbb{C}^n = \mathscr{N}(\lambda_1 I - A) \oplus \mathscr{N}(\lambda_2 I - A) \oplus \cdots \oplus \mathscr{N}(\lambda_r I - A),$$

where the geometric multiplicity of each eigenspace equals the algebraic multiplicity of the corresponding eigenvalue. Using the union of the bases for the eigenspaces of a semisimple operator A results in a diagonal matrix where the diagonal entries are the eigenvalues of A appearing according to their multiplicity.

When A is not diagonalizable, we do not have an eigenbasis for $\mathbb{C}^n$. But for each eigenvalue $\lambda \in \sigma(A)$ the ascending chain

$$\mathscr{N}(\lambda I - A) \subset \mathscr{N}((\lambda I - A)^2) \subset \cdots \subset \mathscr{N}((\lambda I - A)^l) \subset \cdots$$

for the noninvertible linear operator $\lambda I - A$ stabilizes when $l = \text{ind}(\lambda I - A)$. We will show that the subspaces $\mathscr{N}((\lambda I - A)^{\text{ind}(\lambda I - A)})$, $\lambda \in \sigma(A)$, do give a direct sum decomposition of $\mathbb{F}^n$ and the linear operator in the corresponding basis is a block diagonal matrix.

Definition 12.2.8. For $A \in M_n(\mathbb{C})$ and $\lambda \in \sigma(A)$, the subspace

$$\mathscr{E}_\lambda = \mathscr{N}((\lambda I - A)^{\text{ind}(\lambda I - A)})$$

is called the generalized eigenspace of A corresponding to λ. Every nonzero vector in $\mathscr{E}_\lambda$ is called a generalized eigenvector of A corresponding to λ.

Through the next four lemmas we develop the theory needed to prove that the generalized eigenspaces of a linear operator on a finite dimensional vector space do indeed give a direct sum decomposition of the the vector space.

Lemma 12.2.9. For $A \in M_n(\mathbb{C})$ and $\lambda \in \sigma(A)$, the generalized eigenspace $\mathscr{E}_\lambda$ is A-invariant.

Proof. Any subspace W is invariant under the linear operator λI, i.e., if $\mathrm{x} \in W$, then $\lambda I \mathrm{x} = \lambda \mathrm{x} \in W$.

We can write $A = \lambda I - (\lambda I - A)$.

We show that the subspace $\mathscr{E}_\lambda$ is invariant under A if and only if it is invariant under $(\lambda I - A)$.

Suppose that $\mathscr{E}_\lambda$ is invariant under A, i.e., $A(\mathscr{E}_\lambda) \subset \mathscr{E}_\lambda$.

Since $\mathscr{E}_\lambda$ is invariant under λI, it follows that for each $\mathrm{x} \in \mathscr{E}_\lambda$ that $\lambda I \mathrm{x} \in \mathscr{E}_\lambda$.

Since $A\mathrm{x} \in \mathscr{E}_\lambda$, we have that

$$(\lambda I - A)\mathrm{x} = \lambda \mathrm{x} - A\mathrm{x} \in \mathscr{E}_\lambda.$$

This says that $\mathscr{E}_\lambda$ is invariant under $\lambda I - A$.

Now suppose that $\mathscr{E}_\lambda$ is invariant under $\lambda I - A$.

Since $\mathscr{E}_\lambda$ is invariant under λI, it follows for each $\mathrm{x} \in \mathscr{E}_\lambda$ that $\lambda I \mathrm{x} \in \mathscr{E}_\lambda$.

Since $(\lambda I - A)\mathrm{x} \in \mathscr{E}_\lambda$, we have that

$$A\mathrm{x} = (\lambda I + \lambda I - A)\mathrm{x} = \lambda I\mathrm{x} + (\lambda I - A)\mathrm{x} \in \mathscr{E}_\lambda.$$

We now show that $\mathscr{E}_\lambda$ is invariant under $\lambda I - A$.

For $\mathrm{x} \in \mathscr{E}_\lambda$, $\mathrm{y} = (\lambda I - A)\mathrm{x}$, and $k = \operatorname{ind}(\lambda I - A)$, we have

$$(\lambda I - A)^k \mathrm{y} = (\lambda I - A)^{k+1}\mathrm{x} = 0.$$

Thus $y \in \mathscr{E}_\lambda$, and hence that $\mathscr{E}_\lambda$ is invariant under $\lambda I - A$. □

Example (in lieu of 12.2.10). The matrix

$$A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

has two distinct eigenvalues $\lambda_1 = 2$ of algebraic multiplicity 3 and $\lambda_2 = 5$ of algebraic multiplicity 1.

Since

$$\lambda_1 I - A = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & -3 \end{bmatrix},$$

there is one eigenvector e_1 of A corresponding to $\lambda_1 = 2$.

Now as

$$(\lambda_1 I - A)^2 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & -3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 9 \\ 0 & 0 & 0 & 9 \end{bmatrix}$$

and

$$(\lambda_1 I - A)^3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 9 \\ 0 & 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & -9 \\ 0 & 0 & 0 & -27 \\ 0 & 0 & 0 & -27 \end{bmatrix},$$

the generalized eigenspace $\mathscr{E}_{\lambda_1}$ has a basis of $\{\mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3\}$.

Since

$$\lambda_2 I - A = \begin{bmatrix} 3 & -1 & 0 & 0 \\ 0 & 3 & -1 & 0 \\ 0 & 0 & 3 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

the (generalized) eigenspace $\mathscr{E}_{\lambda_2}$ has a basis of $\begin{bmatrix} 1 & 3 & 9 & 9 \end{bmatrix}^{\mathrm{T}}$.

Notice that $\mathscr{E}_{\lambda_1} \cap \mathscr{E}_{\lambda_2} = \{0\}$. This is not a coincidence.

Lemma 12.2.11. If λ and μ are distinct eigenvalues of $A \in M_n(\mathbb{C})$, then $\mathscr{E}_\lambda \cap \mathscr{E}_\mu = \{0\}$.

Proof. Suppose there is $\mathrm{x} \in \mathscr{E}_\mu \cap \mathscr{E}_\lambda$ such that $\mathrm{x} \neq 0$.

Set $k = \mathrm{ind}(\lambda I - A)$ and $l = \mathrm{ind}(\mu I - A)$.

Let $j \in \{1, \ldots, l\}$ be the unique value for which $(\mu I - A)^{j-1}\mathrm{x} \neq 0$ and $(\mu I - A)^j \mathrm{x} = 0$.

Set $\mathrm{y} = (\mu I - A)^{j-1}\mathrm{x}$.

Then $\mathrm{y} \neq 0$ because $\mathrm{y} = (\mu I - A)^{j-1}\mathrm{x} \neq 0$, and $\mathrm{y} \in \mathscr{N}(\mu I - A)$ because $(\mu I - A)\mathrm{y} = (\mu I - A)^j \mathrm{x} = 0$.

Since $\mathrm{x} \in \mathscr{E}_\mu$, the invariance of $\mathscr{E}_\mu$ under $\mu I - A$ (as shown in the proof of Lemma 12.2.9) implies that

$$\mathrm{y} = (\mu I - A)^{j-1}\mathrm{x} \in \mathscr{E}_\mu.$$

On the other hand, with $\mathrm{x} \in \mathscr{E}_\lambda$ we have $(\lambda I - A)^k \mathrm{x} = 0$.

The operators $(\lambda I - A)^k$ and $(\mu I - A)^{j-1}$ commute because both are polynomials in A.

The vector y belongs to $\mathscr{E}_\lambda$ because

$$(\lambda I - A)^k \mathrm{y} = (\lambda I - A)^k (\mu I - A)^{j-1}\mathrm{x} = (\mu I - A)^{j-1}(\lambda I - A)^k \mathrm{x} = 0.$$

Thus we have a nonzero vector $\mathrm{y} \in \mathscr{N}(\mu I - A) \cap \mathscr{E}_\lambda$.

From $\mathrm{y} \in \mathscr{N}(\mu I - A)$ we obtain $A\mathrm{y} = \mu \mathrm{y}$.

Applying the Binomial Theorem to $(\lambda I - A)^k \mathrm{y} = 0$ and using $A\mathrm{y} = \mu\mathrm{y}$ gives

$$(\lambda - \mu)^k \mathrm{y} = 0.$$

[You have it as HW (Exercise 12.9) to carry out the expansion of $(\lambda I - A)^k$ using the Binomial Theorem and then to replace $A^m\mathrm{y}$ with $\mu^m\mathrm{y}$ and simplify to get $(\lambda - \mu)^k \mathrm{y} = 0$.]

Since $\lambda \neq \mu$, the equation $(\lambda - \mu)^k \mathrm{y} = 0$ implies that $\mathrm{y} = 0$.

But this is a contradiction to $\mathrm{y} \neq 0$, which gives $\mathrm{x} = 0$. □

Lemma 12.2.12. For $A \in M_n(\mathbb{C})$, suppose W_1 and W_2 are A-invariant subspaces of $\mathbb{C}^n$ with $W_1 \cap W_2 = \{0\}$. If, for $\lambda \in \sigma(A)$, the generalized eigenspace $\mathscr{E}_\lambda$ satisfies $\mathscr{E}_\lambda \cap W_i = \{0\}$ for all $i = 1, 2$, then

$$\mathscr{E}_\lambda \cap (W_1 \oplus W_2) = \{0\}.$$

Proof. Suppose $\mathrm{x} \in \mathscr{E}_\lambda \cap (W_1 \oplus W_2)$.

Then there are unique $\mathrm{x}_i \in W_i$, $i = 1, 2$, such that $\mathrm{x} = \mathrm{x}_1 + \mathrm{x}_2$.

For $k = \mathrm{ind}(\lambda I - A)$ we have

$$0 = (\lambda I - A)^l \mathrm{x} = (\lambda I - A)^k \mathrm{x}_1 + (\lambda I - A)^k \mathrm{x}_2$$

where the first equality holds because $\mathrm{x} \in \mathscr{E}_\lambda$.

Thus $(\lambda I - A)^k \mathrm{x}_1 = -(\lambda I - A)^k \mathrm{x}_2$.

As shown in the proof of Lemma 12.2.9, the A-invariance of W_1 and W_2 implies the $(\lambda I - A)$-invariance of W_1 and W_2.

Thus $\mathrm{x}_i \in W_i$ implies $(\lambda I - A)^k \mathrm{x}_i \in W_i$ for both $i = 1, 2$.

Since $(\lambda I - A)^k \mathrm{x}_1 = -(\lambda I - A)^k \mathrm{x}_2$ we obtain that

$$(\lambda I - A)^k \mathrm{x}_1 = -(\lambda I - A)^k \mathrm{x}_2 \in W_1 \cap W_2 = \{0\}.$$

This implies that $\mathrm{x}_i \in \mathscr{N}((\lambda I - A)^k) = \mathscr{E}_\lambda$ for both $i = 1, 2$.

Thus we have $\mathrm{x}_i \in \mathscr{E}_\lambda \cap W_i = \{0\}$ for both $i = 1, 2$.

This gives $\mathrm{x}_i = 0$ for both $i = 1, 2,$, and hence that $\mathrm{x} = 0$. $\square$.

Lemma 12.2.13. For $A \in M_n(\mathbb{C})$ and $\lambda \in \sigma(A)$, the dimension of the generalized eigenspace $\mathscr{E}_\lambda$ equals the algebraic multiplicity m_λ of λ.

Proof. The argument follows that in the proof of Theorem 4.4.5.

By Schur's Lemma we can assume that A is upper triangular, having the block form

$$A = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}.$$

Here T_{11} and T_{22} are upper triangular $m_\lambda \times m_\lambda$ and $(n - m_\lambda) \times (n - m_\lambda)$ matrices respectively where the diagonal entries of T_{11} are all λ, and none of the diagonal entires of T_{22} are λ.

The upper triangular matrix $\lambda I - T_{11}$ has zeros on its diagonals, while the upper triangular matrix $\lambda I - T_{22}$ has nonzero entries for all of its diagonals.

Thus $(\lambda I - T_{11})^{m_\lambda} = 0$ and $(\lambda I - T_{22})^k$ is nonsingular for all $k \in \mathbb{N}$.

Therefore the dimension of $\mathscr{E}_\lambda$ is the dimension of $\mathscr{N}((\lambda I - A)^{m_\lambda})$ which is the dimension of $\mathscr{N}((\lambda I - T_{11})^{m_\lambda})$ which is m_λ. $\square$

Theorem 12.2.14. For each $A \in M_n(\mathbb{C})$ there is decomposition of $\mathbb{C}^n$ into a direct sum of A-invariant subspaces

$$\mathbb{C}^n = \bigoplus_{\lambda \in \sigma(A)} \mathscr{E}_\lambda.$$

Proof. First, we show, for a fixed but arbitrary $\lambda \in \sigma(A)$, and any nonempty subset $M \subset \sigma(A) \setminus \{\lambda\}$, that

$$\mathscr{E}_\lambda \cap \bigoplus_{\mu \in M} \mathscr{E}_\mu = \{0\}.$$

We use strong induction on the cardinality of M.

If $|M| = 1$, then $M = \{\mu\}$ and $\mathscr{E}_\lambda \cap \mathscr{E}_\mu$ follows by Lemma 12.2.11.

Now for $m \geq 2$ suppose that

$$\mathscr{E}_\lambda \cap \bigoplus_{\mu \in M} \mathscr{E}_\mu = \{0\}$$

holds when $1 \leq |M| \leq m$ and $M \subset \sigma(A) \setminus \{\lambda\}$.

Let $M' = M \cup \{\nu\}$ where $|M| = m$ and $\nu \in \sigma(A) \setminus M$.

Then $|M'| = m + 1$.

Set $W_1 = \mathscr{E}_\nu$, and

$$W_2 = \bigoplus_{\mu \in M} \mathscr{E}_\mu.$$

For the set $\tilde{M} = \{\lambda, \nu\}$ we have by the strong induction hypothesis that $\mathscr{E}_\lambda \cap W_1 = \{0\}$.

Also by the strong induction hypothesis we have $\mathscr{E}_\lambda \cap W_2 = \{0\}$.

We get $W_1 \cap W_2 = \{0\}$ by applying again the strong induction hypothesis to the set M and the eigenvalue ν replacing λ.

The condition $W_1 \cap W_2 = \{0\}$ implies that $W_1 + W_2 = W_1 \oplus W_2$, i.e.,

$$\mathscr{E}_\nu + \left(\bigoplus_{\mu \in M} \mathscr{E}_\mu \right) = \mathscr{E}_\nu \oplus \left(\bigoplus_{\mu \in M} \mathscr{E}_\mu \right) = \bigoplus_{\mu \in M'} \mathscr{E}_\mu.$$

The spaces $\mathscr{E}_\lambda$, W_1, and W_2 now satisfy the hypotheses of Lemma 12.2.12 which gives

$$\{0\} = \mathscr{E}_\lambda \cap (W_1 \oplus W_2) = \mathscr{E}_\lambda \cap \bigoplus_{\mu \in M'} \mathscr{E}_\mu.$$

By strong induction we obtain for a fixed by arbitrary $\lambda \in \sigma(A)$ and for any nonempty $M \subset \sigma(A) \setminus \{\lambda\}$ that

$$\mathscr{E}_\lambda \cap \bigoplus_{\mu \in M} \mathscr{E}_\mu = \{0\}.$$

By the definition of a direct sum (see Definition 1.3.6) we obtain a subspace

$$W = \sum_{\mu \in \sigma(A)} \mathscr{E}_\mu = \bigoplus_{\mu \in \sigma(A)} \mathscr{E}_\mu \subset \mathbb{C}^n.$$

By Lemma 12.2.13, the dimension of each $\mathscr{E}_\mu$ is the algebraic multiplicity m_μ of μ, and the sum of the algebraic multiplicities of A add to n.

Therefore $W = \mathbb{C}^n$, and since each $\mathscr{E}_\mu$ is A-invariant by Lemma 12.2.9, we obtain the desired decomposition. □

Remark 12.2.15. Theorem 12.2.14 implies that every $A \in M_n(\mathbb{C})$ is similar to a block diagonal matrix where each block is the representation of A on the A-invariant $\mathscr{E}_\lambda$. There exists a basis for each block in which the block matrix is upper triangular with the eigenvalue in each diagonal entry and either zeros or ones on the super diagonal. With each block put into this form, we obtain what is known as the Jordan Canonical Form of A. Although useful in theory, the Jordan Canonical Form is poorly conditioned, meaning small errors in the floating-point arithmetic can compound into large errors in the final result.