Lecture XXXIII: §4.5 Eigenvectors & eigenspaces
Defective matrices

§1 More on characteristic polynomials

Recall: Given an $n\times n$ matrix A, we defined the characteristic polynomial of A as $P_A(t) = \det(A - tI_n)$.

It satisfies:
- $P_A(t)$ is a degree n polynomial in t & $P_A(0) = \det(A)$
- The eigenvalues of A are the roots (or zeroes) of $P_A(t)$

Properties: If λ is an eigenvalue of A then

(1) λ^k is an eigenvalue of A^k for each $k = 2, 3, 4, \ldots$

(2) If A is invertible, then $\lambda \neq 0$ & $\frac{1}{\lambda}$ is an eigenvalue of A^{-1}.

(3) For any μ in $\mathbb{R}$: $(\lambda + \mu)$ is an eigenvalue of $A + \mu I_n$.

[Moreover, we can choose the same eigenvector $\vec{w}$ of A to work for all other matrices used above]

Q: More properties?

Theorem: $P_A(t) = P_{A^T}(t)$ as polynomials in t

In particular, A & A^T have the same eigenvalues (but typically different eigenvectors!)

Proof: Use $\det(B) = \det(B^T)$ for any $n\times n$ matrix B.

$$P_{A^T}(t) = \det(\underbrace{A^T - tI_n}_{=B}) = \det((A^T - tI_n)^T)$$

$$= \det((A^T)^T - (tI_n)^T) \underset{(A^T)^T = A \text{ \& } I_n^T = I_n}{=} \det(A - tI_n) = P_A(t)$$

Since eigenvalues are roots of characteristic polynomials, the second part of the statement follows directly from here.

TODAY: Focus on eigenvectors! (Real case, next time: over $\mathbb{C}$)

§2. Eigenvectors:

Recall: A vector $\vec{w}$ in $\mathbb{R}^n$, $\vec{w} \neq \vec{0}$ is an eigenvector for λ if

$$A\vec{w} = \lambda \vec{w}$$

Def: $E_\lambda = \mathrm{Span}(\{\vec{w} : \vec{w} \text{ eigenvector for } \lambda\})$ Eigenspace for λ

Q: How do we compute E_λ? First, we input an eigenvalue

Ans: $A\vec{w} = \lambda\vec{w}$ means $(A - \lambda I_n)\vec{w} = \vec{0}$

So $E_\lambda = \mathcal{N}(A - \lambda I_n) = \left(\begin{matrix}\text{all eigenvectors}\\ \text{for } \lambda\end{matrix}\right)$ plus $\vec{0}$.

nullspace of the matrix $A - \lambda I_n$

Let's see some examples:

$\boxed{\text{EXAMPLE 1}}$ $A = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$ $\rightsquigarrow$ $\begin{cases} 1^{st}: \text{compute Eigenvalues} \\ 2^{nd}: \text{compute eigenspaces.} \end{cases}$

$P_A(t) = \det\begin{pmatrix} -1-t & 1 \\ 0 & -1-t \end{pmatrix} = (-1-t)^2 = (t+1)^2 \rightsquigarrow \lambda = -1$ is only eigenvalue

$E_{-1} = \mathcal{N}(A - (-1)I_2) = \mathcal{N}(A + I_2) = \mathcal{N}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$

$\mathrm{rank}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = 1$, so by Rank-Nullity, $\dim E_{-1} = 2 - 1 = 1$

We have $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ so e_1 is in E_{-1}.

By dimension reasons: $E_{-1} = \mathrm{Sp}(e_1)$.

Summary: • $\lambda = -1$ was a double root of $P_A(t)$ (multiplicity two) because

$$P_A(t) = (t - (-1))^{2}$$

- $\dim E_{-1} = 1 < 2$
- $\mathbb{R}^2$ does not have a basis of eigenvectors of A.

$\boxed{\text{EXAMPLE 2}}$: $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$

$$P_A(t) = \det \begin{pmatrix} 2-t & 0 & 0 \\ 0 & -1-t & 1 \\ 0 & 0 & -1-t \end{pmatrix} = (2-t)(-1-t)^2 = (2-t)(t+1)^2$$

$\rightsquigarrow$ 2 eigenvalues: $\lambda = 2$ (multiplicity 1)

$\lambda = -1$ (—— 2)

• $E_{-1} = \mathcal{N}(A + I_2) = \mathcal{N}\begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \mathcal{N}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$

$x_1 = 0$
$x_3 = 0$
x_2 any

indep

So $E_{-1} = \operatorname{Sp}\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right)$

gen. soln: $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ x_2 \\ 0 \end{bmatrix} = x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

• $E_2 = \mathcal{N}(A - 2I_2) = \mathcal{N}\begin{pmatrix} 0 & 0 & 0 \\ 0 & -3 & 1 \\ 0 & 0 & -3 \end{pmatrix}$

Use Gauss-Jordan elimination:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & -3 & 1 \\ 0 & 0 & -3 \end{bmatrix} \xrightarrow[\substack{R_2 \to R_2/-3 \\ R_3 \to R_3/-3}]{} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1/3 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[R_2 \to R_2 + \frac{1}{3} R_3]{} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\substack{R_1 \leftrightarrow R_3 \\ R_1 \leftrightarrow R_2}]{} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

indep

x_1 any
$x_2 = 0$
$x_3 = 0$

$\underline{x} = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

So $E_2 = \operatorname{Sp}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right)$

<u>Summary</u>: • Two eigenvalues with $\operatorname{mult}(-1, P_A(t)) = 2$, $\operatorname{mult}(2, P_A(t)) = 1$

• $\dim E_{-1} = \dim E_2 = 1$

• $\dim E_{-1} = 1 < 2 = \operatorname{mult}(-1, P_A(t))$ $\dim E_2 = 1 = \operatorname{mult}(2, P_A(t))$

• Once more: $\mathbb{R}^3$ does not have a basis of eigenvectors of A.

$\boxed{\text{EXAMPLE 3:}}$ $A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & -1 & 8 \\ 0 & 0 & 4 \end{bmatrix}$

$$P_A(t) = \det\begin{pmatrix} 2-t & 0 & 1 \\ 0 & -1-t & 8 \\ 0 & 0 & 4-t \end{pmatrix} = (2-t)(-1-t)(4-t) = (-1)^3(t-2)(t+1)(t-4)$$

so we have 3 simple roots

3 eigenvalues: $\lambda = 4, -1, 2$ (all of multiplicity 1)

- $E_{-1} = \mathcal{N}\begin{pmatrix} 3 & 0 & 1 \\ 0 & 0 & 8 \\ 0 & 0 & 3 \end{pmatrix}$. has $\dim = 3 - \operatorname{rank}\begin{pmatrix} 3 & 0 & 1 \\ 0 & 0 & 8 \\ 0 & 0 & 3 \end{pmatrix} = 3-2 = 1$

  . $\begin{pmatrix} 3 & 0 & 1 \\ 0 & 0 & 8 \\ 0 & 0 & 3 \end{pmatrix}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

  So $E_{-1} = Sp\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right)$

- $E_2 = \mathcal{N}\begin{pmatrix} 0 & 0 & 1 \\ 0 & -3 & 8 \\ 0 & 0 & 2 \end{pmatrix}$ has $\dim = 3 - \operatorname{rank}\begin{pmatrix} 0 & 0 & 1 \\ 0 & -3 & 8 \\ 0 & 0 & 2 \end{pmatrix} = 3-2 = 1$

  $\begin{pmatrix} 0 & 0 & 1 \\ 0 & -3 & 8 \\ 0 & 0 & 2 \end{pmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

  So $E_2 = Sp\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right)$

- $E_4 = \mathcal{N}\begin{pmatrix} -2 & 0 & 1 \\ 0 & -5 & 8 \\ 0 & 0 & 0 \end{pmatrix}$ $\begin{bmatrix} -2 & 0 & 1 \\ 0 & -5 & 8 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[R_2 \to R_2/-5]{R_1 \to R_1/-2} \begin{bmatrix} 1 & 0 & -1/2 \\ 0 & 1 & -8/5 \\ 0 & 0 & 0 \end{bmatrix}$ $\begin{matrix} x_1 = +\frac{1}{2}x_3 \\ x_2 = \frac{8}{5}x_3 \end{matrix}$

  indep

  So general solution: $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}x_3 \\ \frac{8}{5}x_3 \\ x_3 \end{bmatrix} = \frac{x_3}{10}\begin{bmatrix} 5 \\ 16 \\ 10 \end{bmatrix}$

  Conclude: $E_3 = Sp\left(\begin{bmatrix} 5 \\ 16 \\ 10 \end{bmatrix}\right)$

$B = \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 16 \\ 10 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^3$ consisting of 3 eigenvectors (we have 3 & they are l.i. vectors)

Summary:
- A has 3 distinct eigenvectors, all of multiplicity 1
- $\dim E_{-1} = \dim E_2 = \dim E_4 = 1$
- $\mathbb{R}^3$ has a basis B consisting of eigenvectors for A.

§ 3. Eigenspaces

Fix A of size $n \times n$

Define: $E_\lambda = \mathcal{N}(A - \lambda I_n)$ is always a subspace of $\mathbb{R}^n$

- If λ is an eigenvalue, $E_\lambda \neq \{\vec{0}\}$. Moreover:

$$E_\lambda = \{\vec{0}\} \cup \{\vec{w} : \vec{w} \text{ is an eigenvector for the eigenvalue } \lambda\} \quad (\dim E_\lambda > 0)$$

- If λ is not an eigenvalue, then $E_\lambda = \{\vec{0}\}$ $\quad (\dim E_\lambda = 0)$

This gives a third way of characterizing eigenvalues, namely, as those λ in $\mathbb{R}$ with $\dim E_\lambda > 0$.

Q: What is $\dim E_\lambda$ when λ is an eigenvalue?

Def: $\dim E_\lambda$ is called the geometric multiplicity of λ

This name is chosen to contrast with the algebraic multiplicity $=$ multiplicity of λ as a root of $P_A(t)$

- Let's check the earlier examples:

EXAMPLE 1: $A = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$ $\quad \lambda = -1$ only eigenvalue & $\dim E_{-1} = 1$

algebraic mult $= 2$ but geometric multiplicity $= 1$

EXAMPLE 2 $\quad A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$ $\quad \lambda = 2$ & $\lambda = -1$ only eigenvalues

$P_A(t) = (2-t)(t+1)^2$, $\dim E_{-1} = \dim E_2 = 1$

$\lambda = 2$ has algebraic multiplicity 1 & geometric multiplicity 1.

$\lambda = -1$ ——— 2 & ——— 1.

EXAMPLE 3 $\quad A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & -1 & 8 \\ 0 & 0 & 4 \end{bmatrix}$ $\quad P_A(t) = -(t-4)(t+1)(t-2) \rightsquigarrow$ 3 eigenvalues

$\dim E_{-1} = \dim E_4 = \dim E_2 = 1$

$\lambda = 2, -1$ & 4 all have algebraic multiplicity 1 & also geometric multiplicity 1.

- In all examples above: ① alg mult$(\lambda) \geq 1$, geom mult$(\lambda) \geq 1$
  ② alg mult$(\lambda) \geq$ geom. mult(λ)

Both statements are always true. ① follows from the definitions ($E_\lambda \neq \vec{0}$, $\dim E_\lambda \geq 1$)
② is a hard theorem. We'll take it for a fact.

§4 Defective matrices: Fix A of size n×n

- We are interested in situations when

(*) $\boxed{\text{alg mult}(\lambda) = \text{geom mult}(\lambda) \text{ for all eigenvalues } \lambda \text{ of } A.}$

Why? Because in that situation we'll be able to find a basis B for $\mathbb{R}^n$ consisting of eigenvectors for A & we'll be able to a matrix representation for the transformation $T: \mathbb{R}^n \longrightarrow \mathbb{R}^n$, $\vec{v} \longmapsto A\vec{v}$ that is diagonal, namely $[T]_{BB}$. This will be known as "diagonalizing the matrix A". We'll discuss it in §4.7.

- When (*) fails for A, we say A is a <u>defective matrix</u>. The definition is as follows (Here, we are using alg mult$(\lambda) \geq$ geom mult(λ) for any eigenvalue λ)

<u>Def</u>: If A has a (real or complex) eigenvalue with
$$\text{alg mult}(\lambda) > \text{geom mult}(\lambda),$$
then we say A is a <u>defective matrix</u>.

<u>Earlier examples</u>: Ex 1 & Ex 2 were defective (both cases: $\lambda = -1$), Ex 3 was not.

<u>Prop</u>: If alg mult$(\lambda) = 1$ for all eigenvalues λ, then A is <u>not</u> defective.

<u>Proof</u>: We know geom mult$(\lambda) \geq 1$ because $E_\lambda \neq \{\vec{0}\}$ so it's dimension is at least 1.

This ensures that we don't have an eigenvalue λ with
$$\text{alg mult}(\lambda) > \text{geom mult}(\lambda) = \dim E_\lambda$$
So A cannot be defective, by definition of a defective matrix. □

<u>Obs</u>: This was the situation for Example 3 (all alg mult$(\lambda) = 1$).

To finish, we will show that when A of size $n\times n$ has n distinct eigenvalues then $\mathbb{R}^n$ has a basis of eigenvectors (one for each eigenvalue).

To this end, we'll show that vectors of different eigenspaces will be automatically l i (whenever they are not $\vec{0}$). This will be a very useful result, so we write the general statement as a theorem:

Theorem: Given an $n\times n$ matrix A and a list of distinct eigenvalues $\lambda_1, \ldots \lambda_k$, pick a nonzero eigenvector $\vec{v}_j$ for each (that is, $\vec{v}_j \neq \vec{0}$ & $A\vec{v}_j = \lambda_j \vec{v}_j$)

Then $S = \{v_1, \ldots, v_k\}$ is a linearly independent set in $\mathbb{R}^n$

Proof: If $k=1$ the result is true because $\{\vec{v}\}$ is l i whenever $\vec{v}\neq\vec{0}$.

- Assume $k>1$. We will argue by contradiction. We assume S is l d and we'll reach a false statement.

Since each $\{\vec{v}_i\}$ is l i but the whole S is l d, we can find some m with $1<m\leq k$ for which $\{v_1, \ldots v_{m-1}\}$ is l i but $\{v_1, \cdots v_m\}$ is l d.
(This is because we have to move from a l i set to a l d subset, so we can find m first spot where the transition happens)

In particular, we have $\boxed{c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}}$ & $c_m \neq 0$
(otherwise the linear combination will involve $\{\vec{v}_1, \ldots \vec{v}_{m-1}\}$ only & it would be automatically trivial)

Next, we multiply both sides by A on the left & use that $Av_j = \lambda_j v_j$

$$A(c_1\vec{v}_1 + \cdots + c_m\vec{v}_m) = A\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}$$

$$c_1 A\vec{v}_1 + \cdots + c_m A\vec{v}_m = \begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}$$

$$\boxed{c_1\lambda_1\vec{v}_1 + \cdots + c_m\lambda_m\vec{v}_m = \begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}}$$

Next, we take the 2 boxed expressions $\boxed{}$ & do:

$$-\begin{array}{l} c_1\lambda_1\vec{v}_1 + \cdots + c_{m-1}\lambda_{m-1}\vec{v}_{m-1} + c_m\lambda_m\vec{v}_m = \vec{0} \\ \lambda_m(c_1\vec{v}_1 + \cdots + c_{m-1}\vec{v}_{m-1} + c_m\vec{v}_m = \vec{0}) \end{array}$$

$$\underbrace{c_1(\lambda_1-\lambda_m)}_{=d_1}\vec{v}_1 + \cdots + \underbrace{c_{m-1}(\lambda_{m-1}-\lambda_m)}_{d_{m-1}}\vec{v}_{m-1} + \vec{0} = \vec{0}$$

So we get a linear combination involving $\{\vec{v}_1, \ldots \vec{v}_{m-1}\}$ (because we cancelled the terms involving $\vec{v}_m$ in both expressions)

$d_1\vec{v}_1 + \cdots + d_{m-1}\vec{v}_{m-1} = \vec{0}$ forces $d_1 = \cdots = d_{m-1} = 0$ because $\{\vec{v}_1, \ldots \vec{v}_{m-1}\}$ is l.i.

So
$$\begin{cases} d_1 = c_1(\lambda_1 - \lambda_m) = 0 \\ d_2 = c_2(\lambda_2 - \lambda_m) = 0 \\ \vdots \\ d_{m-1} = c_{m-1}(\lambda_{m-1} - \lambda_m) = 0 \end{cases}$$

But each identity forces either $c_i = 0$ or $\lambda_i - \lambda_m = 0$ / $\lambda_i = \lambda_m$

But all the eigenvalues where different & $i < m$, so we conclude $c_1 = c_2 = \cdots = c_{m-1} = 0$

Going back to our original boxed expression:

$$(*) \quad \underbrace{\underbrace{c_1}_{=0}\vec{v}_1 + \underbrace{c_2}_{=0}\vec{v}_2 + \cdots + \underbrace{c_{m-1}}_{=0}\vec{v}_{m-1}}_{\vec{0}} + c_m\vec{v}_m = \vec{0}$$

$$+ c_m\vec{v}_m = \vec{0}$$

And $\vec{v}_m \neq \vec{0}$ forces $c_m = 0$. This contradicts our original assumption that $(*)$ was a nontrivial linear combination!

Only way out to the contradiction: S is l.i. (This is what we wanted to show!)

Consequence 1: If A is of size $n \times n$ & has n distinct real eigenvectors, then $\mathbb{R}^n$ has a basis of eigenvectors

Why? $\{\lambda_1, \ldots, \lambda_m\}$ distinct eigenvalues, pick $\vec{v}_1, \ldots \vec{v}_n$ eigenvectors, with $\vec{v}_j \neq \vec{0}$ (by definition) & $A\vec{v}_j = \lambda_j \vec{v}_j$

Then, the theorem guarantees that $B = \{\vec{v}_1, \ldots \vec{v}_n\}$ is l.i.

Since we have n elements & $\dim \mathbb{R}^n = n$, we know that B is automatically a basis. $\square$

- We will see other consequences of this in § 3.7. Upshot. The same will hold if we pick S_{λ_i} subsets in each E_{λ_i} with $|S_{\lambda_i}| \geq 1$.

Namely, grouping these sets into a big set $S = S_{\lambda_1} \cup S_{\lambda_2} \cup \cdots \cup S_{\lambda_k}$ will also be l.i.