Lecture 24 : Diagonalizing Matrices

Recall that when $\vec{v} \in \mathbb{R}^n$ is nonzero, $\lambda \in \mathbb{R}$, and

$$A\vec{v} = \lambda \cdot \vec{v}, \qquad (A = n \times n \text{ matrix})$$

λ [resp. $\vec{v}$] is an eigenvalue [resp. eigenvector] of A.

- to find eigenvalues : solve $\det(A - \lambda I_n) = 0$
- to find eigenvectors : for each eigenvalue λ_0, find (a basis for)
$$E_{\lambda_0} = \mathrm{Nul}(A - \lambda_0 I_n)$$
by row-reduction. Its dimension is $n - \mathrm{rk}(A - \lambda_0 I_n)$, by R+N.
- to check if $\vec{v}_0$ is an eigenvector : Apply A to $\vec{v}$
- to check if λ_0 is an eigenvalue : see if $\mathrm{rk}(A - \lambda_0 I_n) < n$ by using row-reduction.
- the eigenvalues of an upper or lower-triangular matrix are the diagonal entries.

Now by the Fundamental Theorem of Algebra, the characteristic polynomial factors

$$(*) \qquad \det(A - \lambda I_n) = (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$$

where in general the $\{\lambda_i\}$ may be non-real (i.e. complex numbers) and may not be distinct. Assume for now that they <u>are</u> real.

Definition: The multiplicity of an eigenvalue of A is the number of times it appears in $(*)$. (If all multiplicities are 1, then A has n distinct eigenvalues.)

Lemma: If $\vec{v}_1, \ldots, \vec{v}_k$ are k eigenvectors of A with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, then they are linearly independent.

Proof: Use induction: this is clear for $k=1$, since $\vec{v}_1 \neq \vec{0}$ by definition. Assume it holds for $k-1$ eigenvectors with distinct eigenvalues, i.e. that $\vec{v}_1, \ldots, \vec{v}_{k-1}$ are independent, and let $\vec{v}_k$ be an eigenvector with a "new" eigenvalue.

Suppose $\vec{0} = c_1 \vec{v}_1 + \cdots + c_k \vec{v}_k$. (We must show the c_i's are all 0.) On one hand (multiplying by λ_k)

(1) $$\vec{0} = c_1 \lambda_k \vec{v}_1 + \cdots + c_k \lambda_k \vec{v}_k.$$

On the other hand (applying A)

(2) $$\vec{0} = c_1 A \vec{v}_1 + \cdots + c_k A \vec{v}_k = c_1 \lambda_1 \vec{v}_1 + \cdots + c_k \lambda_k \vec{v}_k.$$

Subtracting (1)-(2) gives

$$\vec{0} = c_1 \underbrace{(\lambda_k - \lambda_1)}_{\neq 0} \vec{v}_1 + \cdots + c_{k-1} \underbrace{(\lambda_k - \lambda_{k-1})}_{\neq 0} \vec{v}_{k-1} + c_k \cancelto{0}{(\lambda_k - \lambda_k)} \vec{v}_k$$

$\Longrightarrow c_1 = \cdots = c_{k-1} = 0$ (since $\vec{v}_1, \ldots, \vec{v}_{k-1}$ are L.I.).

But then the original equation reads $\vec{0} = c_k \vec{v}_k \rightarrow c_k = 0$. $\square$

Theorem: If the eigenvalues of A are distinct (and real), then $\underbrace{\text{a basis of } \mathbb{R}^n \text{ consisting of eigenvectors of } A}_{\text{"}A\text{-eigenbasis"}}$ exists.

Proof: For each of the n eigenvalues, there's an eigenvector. Apply the Lemma. $\square$

Remark: The existence of an A-eigenbasis is crucial for being able to write a given vector as a sum of eigenvectors of A, as part of solving systems of difference/differential equations, etc.

Given an eigenbasis $\vec{v}_1, \dots, \vec{v}_n$, write $P = \begin{pmatrix} \uparrow & & \uparrow \\ \vec{v}_1 & \cdots & \vec{v}_n \\ \downarrow & & \downarrow \end{pmatrix}$ and compute $AP = \begin{pmatrix} \uparrow & & \uparrow \\ A\vec{v}_1 & \cdots & A\vec{v}_n \\ \downarrow & & \downarrow \end{pmatrix} = \begin{pmatrix} \uparrow & & \uparrow \\ \lambda_1\vec{v}_1 & \cdots & \lambda_n\vec{v}_n \\ \downarrow & & \downarrow \end{pmatrix}$; assemble the eigenvalues into a diagonal matrix $D = \begin{pmatrix} \lambda_1 & & O \\ & \ddots & \\ O & & \lambda_n \end{pmatrix}$, so that $PD = \begin{pmatrix} \uparrow & & \uparrow \\ \lambda_1\vec{v}_1 & \cdots & \lambda_n\vec{v}_n \\ \downarrow & & \downarrow \end{pmatrix}$. Hence $AP = PD$ and

$$A = P \cdot D \cdot P^{-1} \qquad (\dagger)$$

(equivalently $P^{-1}AP = D$). We have diagonalized A.

Corollary 1: An $n \times n$ matrix with n distinct eigenvalues can be diagonalized (or "is diagonalizable").

Ex 1/ Diagonalize $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. From lecture 23,

A has

$$\left\{\begin{array}{l} \text{eigenvalues: } \overbrace{0,\ 0}^{\text{multiplicity } 2},\ 3 \\ \qquad\updownarrow \\ \text{eigenvectors: } \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \end{array}\right.$$

— recall characteristic polynomial was $-\lambda^2(\lambda-3)$.

So Lemma doesn't apply — in this case, check independence.

eigenbasis

(must be in the same order)

$$P = \begin{pmatrix} -1 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} \boxed{0} & 0 & 0 \\ 0 & \boxed{0} & 0 \\ 0 & 0 & \boxed{3} \end{pmatrix}$$

Therefore $A = \underbrace{\begin{pmatrix}-1 & -1 & 1\\ 1 & 0 & 1\\ 0 & 1 & 1\end{pmatrix}}_{P}\underbrace{\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 3\end{pmatrix}}_{D}\underbrace{\begin{pmatrix}-\frac{1}{3} & \frac{2}{3} & -\frac{1}{3}\\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{pmatrix}}_{P^{-1}}$.

So you don't <u>need</u> to have distinct eigenvalues in order to diagonalize. //

Ex 2/ Is $A = \begin{pmatrix}5 & -8 & -21\\ 0 & 0 & 7\\ 0 & 0 & -2\end{pmatrix}$ diagonalizable? If so, do it.

<u>eigenvalues</u>: $\overset{\lambda_1}{5}, \overset{\lambda_2}{0}, \overset{\lambda_3}{-2}$ (upper triangular) $\rightarrow$ distinct $\Rightarrow$ diagonalizable.

<u>eigenvectors</u>: row-reduce (if necessary) to find vector in null spaces of

$$\left.\begin{aligned}
A - 5I &= \begin{pmatrix}0 & -8 & -21\\ & -5 & 7\\ & & -7\end{pmatrix} \longrightarrow \vec{v}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}\\
A - 0I &= \begin{pmatrix}5 & -8 & -21\\ & 0 & 7\\ & & -2\end{pmatrix} \longrightarrow \vec{v}_2 = \begin{pmatrix}8\\5\\0\end{pmatrix}\\
A + 2I &= \begin{pmatrix}7 & -8 & -21\\ & 2 & 7\\ & & 0\end{pmatrix} \longrightarrow \vec{v}_3 = \begin{pmatrix}-1\\-7/2\\1\end{pmatrix}
\end{aligned}\right\} \Rightarrow \quad D = \begin{pmatrix}5 & & \\ & 0 & \\ & & -2\end{pmatrix}, \quad P = \begin{pmatrix}1 & 8 & -1\\ & 5 & -7/2\\ & & 1\end{pmatrix}$$

//

Ex 3/ What about $A = \begin{pmatrix}3 & 0 & 0\\ 0 & 2 & 0\\ 1 & 0 & 3\end{pmatrix}$? (lower triangular)

<u>eigenvalues</u>: $2, 3$ ← w/ multiplicity 2

<u>Find eigenvectors</u>: $A - 2I = \begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 1 & 0 & 1\end{pmatrix} \rightarrow \begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 1\end{pmatrix} \rightsquigarrow \begin{pmatrix}0\\1\\0\end{pmatrix}$ spans E_2.

$A - 3I = \begin{pmatrix}0 & 0 & 0\\ 0 & -1 & 0\\ 1 & 0 & 0\end{pmatrix} \rightarrow \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{pmatrix} \rightsquigarrow \begin{pmatrix}0\\0\\1\end{pmatrix}$ spans E_3.

Up to scale, there are only 2 eigenvectors. So there's <u>no</u> A-eigenbasis of $\mathbb{R}^3$, and A isn't diagonalizable. //

$\boxed{\text{Applications}}$

① The determinant of a diagonalizable $n \times n$ matrix is the product of the n (not necessarily distinct) eigenvalues:

$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n .$$

Why? $\det A = \det(PDP^{-1}) = \det P \cdot \det D \cdot \det P^{-1}$

$$= \det D = \det \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = \prod_{i=1}^{n} \lambda_i .$$

(In fact, this is still true even if A isn't diagonalizable: we still have $\det(A - \lambda I) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda)$ by (*), with constant term $\lambda_1 \cdots \lambda_n$. But another way to compute this constant term is by setting $\lambda = 0$, giving $\det(A - 0I) = \det(A)$. So $\det(A) = \lambda_1 \cdots \lambda_n$.)

② Compute $\underbrace{\begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}}_{A}^{10}$ by diagonalizing A:

$A = PDP$ where $D = \begin{pmatrix} -2 & 0 \\ 0 & 5 \end{pmatrix}$, $P = \begin{pmatrix} -1 & 3 \\ 1 & 4 \end{pmatrix}$

$$\Rightarrow A^{10} = (PDP^{-1})^{10} = PD\cancel{P^{-1} \cdot P}D\cancel{P^{-1} \cdot P}D\cancel{P^{-1} \cdot \ldots \cdot P}DP^{-1}$$

$$= PD^{10}P^{-1} = \begin{pmatrix} -1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} (-2)^{10} & 0 \\ 0 & 5^{10} \end{pmatrix} \begin{pmatrix} 4/7 & -3/7 \\ -1/7 & -1/7 \end{pmatrix}$$

$$= \frac{1}{7} \left(\begin{array}{c|c} -4 \cdot 2^{10} - 3 \cdot 5^{10} & 3 \cdot 2^{10} - 3 \cdot 5^{10} \\ \hline 4 \cdot 2^{10} - 4 \cdot 5^{10} & -3 \cdot 2^{10} - 4 \cdot 5^{10} \end{array} \right).$$

(3) Stochastic matrices

Let A be an $n \times n$ matrix whose columns sum to 1, some power of which has all positive entries.
(This is called a regular stochastic matrix.)

- A has a steady-state vector, i.e. eigenvector with eigenvalue 1:

The product of $(1\ 1 \cdots 1)$ with each column of A simply adds the entries of that column, and this is 1. So

$$(1 \cdots 1)\, A = (1 \cdots 1)$$

or equivalently ${}^tA \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ I'm writing transpose on the left so as not to confuse it with powers of A.

and $\vec{1} := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ is an eigenvector of ${}^tA$ with eigenvalue 1.

Conversely, if $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is another eigenvector of ${}^tA$ with eigenvalue 1, we have $x_i = \sum_j a_{ji} x_j$ (with $\sum_j a_{ji} = 1$) and each x_i is a weighted average of the x_j's. But if x_i is the largest entry of $\vec{x}$, this is impossible unless all the x_j's equal x_i: that is, $\vec{x}$ is a multiple of $\vec{1}$, and the eigenspace $E_1({}^tA) = \operatorname{span}\{\vec{1}\}$ is 1-dimensional. But then

$$1 = \dim E_1({}^tA) = \operatorname{nullity}({}^tA - I) = n - \operatorname{rank}({}^tA - I)$$
$$= n - \operatorname{rank}(A - I) = \operatorname{nullity}(A - I) = \dim E_1(A).$$

(rank is unchanged by transpose)

So there is an eigenvector $\vec{v}_1$ of A with eigenvalue 1 (this is not $\vec{1}$!), and it spans $E_1(A)$. Thus, if we normalize $\vec{v}_1$ so its entries sum to 1, it is unique!

- Any eigenvector with a different eigenvalue than 1
  - (a) must lie in the plane $x_1 + \cdots + x_n = 0$
  - (b) must have eigenvalue $\in (-1, 1)$

(b) is important for dynamical systems / Markov chains — if the initial state is $\vec{x}_0 = c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots$ and $\vec{v}_1$ = steady state vector, then

$$\vec{x}(t) = A^t \vec{x}_0 = c_1 \vec{v}_1 + c_2 \lambda_2^t \vec{v}_2 + \cdots \xrightarrow[t \to \infty]{\text{limit as}} c_1 \vec{v}_1$$

0 since $|\lambda_2| < 1$

EXAMPLE:

- 2×2 matrices: $A = \begin{pmatrix} a & 1-b \\ 1-a & b \end{pmatrix} \Rightarrow A - \lambda I_2 = \begin{pmatrix} a-\lambda & 1-b \\ 1-a & b-\lambda \end{pmatrix}$

$$\Rightarrow \det(A - \lambda I_2) = \lambda^2 - (a+b)\lambda + (a+b-1) = 0$$

$$\Rightarrow \lambda = \frac{a+b \pm \sqrt{(a+b)^2 - 4(a+b-1)}}{2} = \begin{cases} 1 \\ a+b-1 \end{cases}$$

Since $0 < a+b < 2$, $-1 < a+b-1 < 1$ as desired.