Lecture 23: Cayley-Hamilton

Suppose an $n \times n$ matrix A has $\lambda_0 = 3$ as an eigenvalue: then $(3I_n - A)\vec{v}_0 = \vec{0}$ for the corresponding eigen<u>vector</u> $\vec{v}_0$.

So: plugging A into the polynomial* $3-x$ yields a matrix annihilating $\vec{v}_0$. Is there a polynomial into which plugging A gives a matrix annihilating <u>all</u> vectors? — i.e., the zero matrix?

Consider $M_{n,n}$ as a vector space over $\mathbb{R}$ of dimension n^2.

Clearly $I, A, A^2, \ldots, A^{n^2}$ are dependent (there are n^2+1 of them)

$\Rightarrow \exists\, \alpha_0, \ldots, \alpha_{n^2} \in \mathbb{R}$ s.t. $\sum_{k=0}^{n^2} \alpha_k A^k = 0$. So indeed, writing $q(x) := \sum_{k=0}^{n^2} \alpha_k x^k$, $q(A) = 0$. But maybe we can do better, i.e. succeed with a polynomial of smaller degree?

If A is diagonalizable with eigenvalues $\lambda_1, \ldots, \lambda_n$, then $(\lambda_1 I - A)(\lambda_2 I - A) \cdots (\lambda_n I - A)\vec{v} = 0$ for every $\vec{v}$, by writing $\vec{v} = \sum_{m=1}^{n} a_m \vec{v}_m$ as a linear combination of eigenvectors. (Use the fact that the $(\lambda_m I - A)$ all commute.) So then writing $p(x) = (\lambda_1 - x) \cdots (\lambda_n - x)$, we get $p(A) = [0]$.

But this p is the characteristic polynomial!

---

* small point: constants c become $c \cdot I_n \; (= c \cdot A^0)$

Perhaps plugging A into its characteristic polynomial always gives zero? We should be careful — diagonalizable matrices are not representative of the 'general' case. Also, the first idea that comes to mind for proving such a thing — $\det(A\mathbb{I} - A) = \det(0) = 0$ — is wrong: substituting in A before you take the determinant is not the same as doing so afterwards! On the other hand, here's a nondiagonalizable $A := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$, with characteristic poly.

$$p_A(\lambda) := \det \begin{pmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ 0 & 0 & \lambda \end{pmatrix} = \lambda^3,$$ and indeed $A^3 = 0$. So perhaps the following isn't surprising after all:

<u>Theorem (Cayley-Hamilton)</u>: Writing $p_A(\lambda) := \det(\lambda \mathbb{I}_n - A)$, we have $p_A(A) = [0]$.

<u>Proof</u>: We work with matrices whose entries are polynomials in λ. So we can't <u>divide</u> by λ, or polynomials in λ, like the determinant of such a matrix. But

$(i,j)^{\text{th}}$ entry of the matrix product.

$$M \cdot \mathrm{adj}(M) = \left[\sum_{k=1}^{n} m_{ik} \cdot (-1)^{k+j} \det M_{\hat{j}\hat{k}} \right]$$
$$= \left[\det(M)\, \delta_{ij} \right]$$
$$= \det(M)\, \mathbb{I}_n$$

By Laplace expansion, this is the determinant of the matrix obtained by replacing the j^{th} row of M by the i^{th} row. So get $\det M$ if $i=j$, 0 if $i \neq j$.

is true for such matrices. Applying this to $M = \lambda \mathbb{I}_n - A$,

$$(\lambda \mathbb{I}_n - A) \cdot \operatorname{adj}(\lambda \mathbb{I}_n - A) = \det(\lambda \mathbb{I}_n - A) \cdot \mathbb{I}_n = p_A(\lambda)\, \mathbb{I}_n .$$

Now decompose* $\operatorname{adj}(\lambda \mathbb{I}_n - A) = \sum_{k=0}^{n-1} \lambda^k S_k$ why? and write

$p_A(\lambda) = \sum_{j=0}^{n} a_j \lambda^j$. We have

$$(\lambda \mathbb{I} - A) \sum_{k=0}^{n-1} \lambda^k S_k = \sum_{j=0}^{n} a_j \lambda^j \mathbb{I}_n \implies$$

$$-AS_0 + \lambda(S_0 - AS_1) + \lambda^2(S_1 - AS_2) + \cdots + \lambda^{n-1}(S_{n-2} - AS_{n-1}) + \lambda^n S_{n-1}$$

$$= a_0 \mathbb{I} + \lambda(a_1 \mathbb{I}) + \lambda^2(a_2 \mathbb{I}) + \cdots + \lambda^{n-1}(a_{n-1}\mathbb{I}) + \lambda^n(a_n \mathbb{I})$$

equate coeffs. of like powers of λ

$$\implies a_0 \mathbb{I} = -AS_0, \quad a_1 \mathbb{I} = S_0 - AS_1, \quad a_2 \mathbb{I} = S_1 - AS_2,$$

$$\ldots, \quad a_{n-1}\mathbb{I} = S_{n-2} - AS_{n-1}, \quad a_n \mathbb{I} = S_{n-1}$$

$$\implies p_A(A) = \sum_{j=0}^{n} a_j A^j = a_0 \mathbb{I} + A(a_1 \mathbb{I}) + A(a_2 \mathbb{I}) + \cdots + A(a_{n-1}\mathbb{I}) + A(a_n \mathbb{I})$$

$$= -AS_0 + A(S_0 - AS_1) + A^2(S_1 - AS_2) + \cdots + A^{n-1}(S_{n-2} - AS_{n-1}) + A^n S_{n-1}$$

$$= -AS_0 + AS_0 - A^2 S_1 + A^2 S_1 - A^3 S_2 + \cdots + A^{n-1} S_{n-2} - A^n S_{n-1} + A^n S_{n-1}$$

$$= 0. \qquad \square$$

---

* We can decompose any matrix with entries polynomial in λ into powers of λ: for instance, $\begin{pmatrix} \lambda^2 + \lambda & \lambda^2 \\ 1 & 1+\lambda \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\lambda + \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\lambda^2$.

So, you may be wondering: exactly what does this have to do with computing e^{tA}, our topic from yesterday?

Suppose A is $n \times n$, and has <u>one</u> eigenvalue η (with multiplicity n). The characteristic polynomial is $p_A(\lambda) = (\lambda - \eta)^n$, so by Cayley-Hamilton

$$0 = p_A(A) = (A - \eta I)^n \Rightarrow (A - \eta I)^k = 0 \text{ for } k \geq n$$

$$\Rightarrow e^{tA} = e^{\eta t I + t(A - \eta I)} = e^{\eta t I} e^{t(A - \eta I)}$$

commute

$$= e^{\eta t} \sum_{k=0}^{\infty} \frac{t^k}{k!} (A - \eta I)^k = e^{\eta t} \sum_{k=0}^{n-1} \frac{t^k}{k!} (A - \eta I)^k.$$

Ex/ Suppose we want to solve

$$(\dagger) \begin{cases} f'''(t) - 6f''(t) + 12f'(t) - 8f(t) = 0 \\ f(0) = 1, \quad f'(0) = 0, \quad f''(0) = -1. \end{cases}$$

Writing $\begin{cases} x_1(t) = f(t) \\ x_2(t) = f'(t) \\ x_3(t) = f''(t) \end{cases}$, $(\dagger)$ is equivalent to

$$\begin{cases} x_1'(t) = x_2(t) \\ x_2'(t) = x_3(t) \\ x_3'(t) = 8x_1(t) - 12x_2(t) + 6x_3(t) \end{cases} + \vec{x}(0) = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix},$$

which is the same as $\begin{cases} \vec{x}'(t) = A\vec{x}(t) \\ \vec{x}(0) = \vec{b} \end{cases}$

where $\vec{b} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 8 & -12 & 6 \end{pmatrix}$. The characteristic

polynomial is $P_A(\lambda) = \det\begin{pmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ -8 & 12 & \lambda-6 \end{pmatrix} = \lambda^3 - 6\lambda^2 + 12\lambda - 8 = (\lambda-2)^3$

— same as that of the original DE $(\dagger)$, as it must be.

(Can you prove that?) So the above calculation gives

$$e^{tA} = e^{2t}\sum_{k=0}^{2}\frac{t^k}{k!}(A-2I)^k$$

$$\boxed{A - 2I = \begin{pmatrix} -2 & 1 & 0 \\ 0 & -2 & 1 \\ 8 & -12 & 4 \end{pmatrix} \qquad (A-2I)^2 = \begin{pmatrix} 4 & -4 & 1 \\ 8 & -8 & 2 \\ 16 & -16 & 4 \end{pmatrix}}$$

$$= e^{2t}\mathbb{I} + te^{2t}\begin{pmatrix} -2 & 1 & 0 \\ 0 & -2 & 1 \\ 8 & -12 & 4 \end{pmatrix} + \frac{t^2}{2}e^{2t}\begin{pmatrix} 4 & -4 & 1 \\ 8 & -8 & 2 \\ 16 & -16 & 4 \end{pmatrix}.$$

The solution is $\vec{x}(t) = e^{tA}\vec{b} = e^{2t}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + te^{2t}\begin{pmatrix} -2 \\ -1 \\ 4 \end{pmatrix} + \frac{t^2}{2}e^{2t}\begin{pmatrix} 3 \\ 6 \\ 12 \end{pmatrix}$,

which is to say $f(t)\ (= x_1(t)) = e^{2t} - 2te^{2t} + \frac{3}{2}t^2e^{2t}$.

(Check that this solves $(\dagger)$.) //

Here is how to do the other "non-diagonalizable" possibility for 3×3 matrices: $\lambda_1 = \mu$, $\lambda_2 = \lambda_3 = \eta$. We have

$$P_A(\lambda) = (\lambda-\mu)(\lambda-\eta)^2 \overset{\text{C-H}}{\Longrightarrow} 0 = (A-\mu I)(A-\eta I)^2$$

$$\Longrightarrow (A-\eta I)^k = [(A-\mu I) + (\mu-\eta)I]^{k-2}(A-\eta I)^2$$

$$(k\geq 2) \quad = (\mu-\eta)^{k-2}(A-\eta I)^2$$

$$
\begin{aligned}
\Longrightarrow e^{tA} &= e^{\eta t I + t(A-\eta I)} \\
&= e^{\eta t} \sum_{k=0}^{\infty} \frac{t^k}{k!} (A-\eta I)^k \\
&= e^{\eta t} I + t e^{\eta t} (A-\eta I) + e^{\eta t} \sum_{k=2}^{\infty} \frac{t^k}{k!} (\mu-\eta)^{k-2} (A-\eta I)^2 \\
&= e^{\eta t} I + t e^{\eta t} (A-\eta I) + \frac{e^{\eta t}}{(\mu-\eta)^2} \left\{ e^{(\mu-\eta)t} - 1 - (\mu-\eta)t \right\} (A-\eta I)^2,
\end{aligned}
$$

a formula which looks more complicated than it is.

Notice that the functions of t that appear in e^{tA}, hence in any solution $e^{tA}\vec{b}$ to $\vec{x}\,'(t) = A\vec{x}(t)$, are simply $e^{\eta t}$, $te^{\eta t}$, and $e^{\mu t}$.