Hello everyone! I am Osamu Kada.

In the previous video, we proved Cayley-Hamilton Theorem from the approach using the notion of matrices whose componets are polynomials in the matrix.

Today I'm going to prove the theorem from the approach using the notion of R-module, here R is the polynomial ring $\mathbb{R}[t]$, or $M_n(\mathbb{R}[t])$. The theorem states as follows.

Theorem (Cayley-Hamilton)
Let A be an $n \times n$ matrix whose components are real numbers (or elements of a commutative ring R), and let $P_A(t) := \det(tI - A) \in \mathbb{R}[t]$ be the characteristic polynomial of A, here I is the identity matrix. Then substituting A for t, we have $P_A(A) = \mathrm{O}$, the zero matrix.
Here, denoting $P_A(t) = \sum_{i=0}^{n} a_i t^i,\ a_n = 1, a_{n-1} = -\mathrm{tr}(A), \ldots, a_0 = (-1)^n \det A$, $P_A(A) = \sum_{i=0}^{n} a_i A^i = \mathrm{O}$, here $A^0 = I$. That is, every square matrix satisfies its own characteristic equation.

Now we prove the theorem.

Proof. Let $B(t) := tI - A^T \in M_n(\mathbb{R}[t])$, here A^T is the transposed matirx of A. Then $\det B(t) = \det(tI - A^T) = \det(tI - A) = P_A(t)$.
Let $\widetilde{B(t)} \in M_n(\mathbb{R}[t])$ be the adjugate matrix of $B(t)$. Then we have

$$B(t)\widetilde{B(t)} = \widetilde{B(t)}B(t) = (\det B(t))I = P_A(t)I.$$

Let $R := \mathbb{R}[t],\ V := \mathbb{R}^n$, and consider V as R-module by the following:
for $f(t) = \sum_{i=0}^{n} a_i t^i \in \mathbb{R}[t]$ and $x \in V$,

$$f(t)\cdot x := f(A)x = \sum_{i=0}^{n} a_i A^i x \in V.$$

Here we recall the notion of R-modeule M.

Let R be a ring, and M an abelian group. Then M is said to be an R-module if for $r \in R$ and $x \in M$, $r\cdot x \in M$ is defined, and the following conditons are satisfied:for $r, s \in R, x, y \in M$,

$$\begin{aligned} &\text{(i) } (rs)\cdot x = r\cdot(s\cdot x),\\ &\text{(ii) } 1\cdot x = x\\ &\text{(iii) } r\cdot(x+y) = r\cdot x + r\cdot y,\\ &\text{(iv) } (r+s)\cdot x = r\cdot x + s\cdot x. \end{aligned}$$

In our case, R is the polynomial ring $\mathbb{R}[t],\ M = V = \mathbb{R}^n$, and the action of R to M is defined as above. The condition (i) is satisfied because

$$\begin{aligned} \text{(i)} \qquad (a_l t^l b_m t^m)\cdot x &= (a_l b_m t^{l+m})\cdot x \\ &= a_l b_m A^{l+m} x, \\ (a_l t^l) \cdot (b_m t^m \cdot x) &= (a_l t^l)\cdot(b_m A^m x) \\ &= a_l A^l (b_m A^m x) \\ &= a_l b_m A^{l+m} x, \text{ and} \\ \text{(ii)} \qquad 1 \cdot x &= A^0 x = Ix = x. \end{aligned}$$

Let $R' := M_n(R)$ and consier $V^n = (\mathbb{R}^n)^n$ as R'-module as usual, that is, matrix times vector.

Since $B(t) = tI - A^T \in R' = M_n(\mathbb{R}[t])$ and $\begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} \in V^n = (\mathbb{R}^n)^n$, here $e_1, \ldots, e_n$ is the canonical basis of $V = \mathbb{R}^n$, we have

$$B(t) \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = (tI - A^T) \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \left(\begin{pmatrix} t & & \\ & \ddots & \\ & & t \end{pmatrix} - \begin{pmatrix} a_{11} & \ldots & a_{n1} \\ \vdots & & \vdots \\ a_{1n} & \ldots & a_{nn} \end{pmatrix} \right) \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

$$= \begin{pmatrix} t{\cdot}e_1 \\ \vdots \\ t{\cdot}e_n \end{pmatrix} - \begin{pmatrix} a_{11}{\cdot}e_1 + \cdots + a_{n1}{\cdot}e_n \\ \vdots \\ a_{1n}{\cdot}e_1 + \cdots + a_{nn}{\cdot}e_n \end{pmatrix} = \begin{pmatrix} Ae_1 \\ \vdots \\ Ae_n \end{pmatrix} - \begin{pmatrix} \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \vec{0} \\ \vdots \\ \vec{0} \end{pmatrix}.$$

Hence, we have that

$$\begin{pmatrix} \vec{0} \\ \vdots \\ \vec{0} \end{pmatrix} = \widetilde{B(t)} B(t) \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} P_A(t) & & \\ & \ddots & \\ & & P_A(t) \end{pmatrix} \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} P_A(t){\cdot}e_1 \\ \vdots \\ P_A(t){\cdot}e_n \end{pmatrix} = \begin{pmatrix} P_A(A)e_1 \\ \vdots \\ P_A(t)e_n \end{pmatrix},$$

which implies that $P_A(A) = \mathrm{O}$. □

Consider the following example.

Let $A = \begin{pmatrix} & & -1 \\ & 0 & \\ 2 & & \end{pmatrix} = (2e_3,\ \vec{0},\ -e_1)$. Then

$$B(t) := tI - A^T = \begin{pmatrix} t & & \\ & t & \\ & & t \end{pmatrix} - \begin{pmatrix} & & 2 \\ & 0 & \\ -1 & & \end{pmatrix} = \begin{pmatrix} t & & -2 \\ & t & \\ 1 & & t \end{pmatrix},$$

$$\widetilde{B(t)} = \begin{pmatrix} t^2 & & 2t \\ & t^2 + 2I & \\ -t & & t^2 \end{pmatrix}, P_A(t) = \det B(t) = t^3 + 2t,$$

$$P_A(A) = A^3 + 2A,$$

$$\widetilde{B(t)} B(t) = \begin{pmatrix} t^2 & & 2t \\ & t^2 + 2I & \\ -t & & t^2 \end{pmatrix} \begin{pmatrix} t & & -2 \\ & t & \\ 1 & & t \end{pmatrix} = \begin{pmatrix} t^3 + 2t & & \\ & t^3 + 2t & \\ & & t^3 + 2t \end{pmatrix},$$

$$B(t) \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = \begin{pmatrix} t & & \\ & t & \\ & & t \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} - \begin{pmatrix} & & 2 \\ & 0 & \\ -1 & & \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

$$= \begin{pmatrix} t{\cdot}e_1 \\ t{\cdot}e_2 \\ t{\cdot}e_3 \end{pmatrix} - \begin{pmatrix} 2e_3 \\ 0 \\ -e_1 \end{pmatrix} = \begin{pmatrix} Ae_1 \\ Ae_2 \\ Ae_3 \end{pmatrix} - \begin{pmatrix} 2e_3 \\ 0 \\ -e_1 \end{pmatrix} = \begin{pmatrix} \vec{0} \\ \vec{0} \\ \vec{0} \end{pmatrix}.$$

So that

$$\begin{pmatrix} \vec{0} \\ \vec{0} \\ \vec{0} \end{pmatrix} = \widetilde{B(t)} B(t) \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = \begin{pmatrix} t^3 + 2t & & \\ & t^3 + 2t & \\ & & t^3 + 2t \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

$$= \begin{pmatrix} (t^3 + 2t) \cdot e_1 \\ (t^3 + 2t) \cdot e_2 \\ (t^3 + 2t) \cdot e_3 \end{pmatrix} = \begin{pmatrix} (A^3 + 2A) e_1 \\ (A^3 + 2A) e_2 \\ (A^3 + 2A) e_3 \end{pmatrix},$$

which implies that $A^3 + 2A = P_A(A) = \mathrm{O}$.