11 Lecture 11

11.1 Cayley-Hamilton theorem

Proposition 11.1. *Let $A \in M_{n\times n}(F)$ and v be an eigenvector with eigenvalue λ.*

(i) For any $f(x) \in F[x]$, $f(A)(v) = f(\lambda)v$ thus $f(\lambda)$ is an eigenvalue of $f(A)$ and v is an eigenvector with eigenvalue $f(\lambda)$. In particular, $f(\sigma(A)) \subset \sigma(f(A))$.

(ii) If p_A splits in F, then $\sigma(f(A)) = f(\sigma(A))$.

Proof. (i) $A^k(v) = A^{k-1}(A(v)) = \lambda A^{k-1}(v) = \cdots = \lambda^k v$. Suppose $f(x) = a_n x^n + \cdots + a_0$. Then $f(A)v = (a_n A^n + \cdots + a_0 I)v = (a_n\lambda^n + \cdots + a_0)v = f(\lambda)v$.

(ii) Let $\mu \in \sigma(f(A))$, i.e. $f(A)v = \mu v$ for $0 \neq v \in F^n$. Consider the polynomial $f(x) - \mu$. Let E be a field extension of F such that $f(x) - \mu = (x - \lambda_1)\cdots(x - \lambda_m)$. Then $0 = (f(A) - \mu I)v = (A - \lambda_1 I)\cdots(A - \lambda_m I)v$. The matrix $(A - \lambda_1 I)\cdots(A - \lambda_m I)$ is singular since it maps a nonzero vector v to 0. So there must be $1 \leq i \leq m$ such that $A - \lambda_i I$ is singular. Thus λ_i is an eigenvalue of A. However, since $p_A(x)$ splits in F, $\lambda_i \in F$ and we have $f(\lambda_i) - \mu = 0$ that is $f(\lambda_i) = \mu$. □

Corollary 11.2. *Let $A \in M_{n\times n}(F)$ and $f(x) \in F[x]$. If $f(A) = 0$, then for any eigenvalue λ of A, we have $f(\lambda) = 0$.*

Example 11.3. If $A^2 = I$, then the eigenvalue of A is either 1 or -1.

If $A^2 = A$ then the eigenvalue of A is either 1 or 0.

If $A^m = 0$ then the eigenvalue of A is 0.

Example 11.4. If p_A does not splits in F, then there are $A \in M_{n\times n}(F)$ and $f \in F[x]$ such that $f(\sigma(A)) \subsetneq \sigma(f(A))$. Let $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in M_{2\times 2}(\mathbb{R})$ and $f(x) = x^2$. Then $f(A) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$. We have $p_A(x) = x^2 + 1$ and $\sigma(A) = \emptyset$ and $\sigma(f(A)) = \{-1\}$. Hence $f(\sigma(A)) \subsetneq \sigma(f(A)) = \{-1\}$.

Theorem 11.5 (Cayley-Hamilton theorem)**.** *Let $A \in M_{n\times n}(F)$. Then $p_A(A) = 0$.*

Remark 11.6. Wrong but not very wrong proof: $p_A(A) = \det(AI - A) = \det 0 = 0$. This is wrong because plugging a matrix in $\det(xI - A)$ does not make sense and also $p_A(A)$ should be a matrix while in the above proof $\det 0$ is a scalar. However, with some efforts, one can make this argument to work. See Chapter 6, Theorem 5 of Linear algebra and its applications by Lax.

Proof when A is diagonalizable. Since A is diagonalizable, $P^{-1}AP = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then $p_A(A) = P p_A(\operatorname{diag}(\lambda_1, \ldots, \lambda_n))P^{-1} = P \operatorname{diag}(p_A(\lambda_1), \ldots, p_A(\lambda_n))P^{-1} = 0$ since λ_i are roots of p_A. □

To deal with the case where A is not diagonalizable, we need the following theorem.

Theorem 11.7. *Let $A \in M_{n\times n}(F)$, p_A splits as in $F[x]$. Then there exists $P \in M_{n\times n}(F)$ invertible, such that*

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A (with multiplicity).

Proof. We prove the statement by induction on n.

The base case $n = 1$ is trivial.

Assume the statement is true for all matrices of size $(n-1)\times(n-1)$ whose characteristic polynomial splits in F. Since p_A splits in F, A has at least one eigenvalue $\lambda_1 \in F$. Choose a nonzero eigenvector $v_1 \in F^n$ with $Av_1 = \lambda_1 v_1$. Extend $\{v_1\}$ to a basis $(v_1, v_2, \ldots, v_n)$ of F^n and let $P_1 = (v_1\ v_2\ \cdots\ v_n)$ be the change-of-basis matrix. In this basis the linear map A has matrix

$$P_1^{-1}AP_1 = \begin{pmatrix} \lambda_1 & * \\ 0 & B \end{pmatrix},$$

where $B \in M_{(n-1)\times(n-1)}(F)$. By Proposition 8.14, $p_A(x) = (x-\lambda_1)p_B(x)$. Since $p_A(x)$ splits in F, we have $p_B(x)$ also splits in F. By the induction hypothesis there exists an $Q \in M_{(n-1)\times(n-1)}(F)$ invertible such that $Q^{-1}BQ$ is upper triangular with eigenvalues $\lambda_2, \ldots, \lambda_n$ on the diagonal. Put

$$P = P_1 \begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix}.$$

Then a direct block multiplication gives

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

which completes the induction and the proof. □

Lemma 11.8. *If* $A = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$ *and* $B = \begin{pmatrix} \mu_1 & * & \cdots & * \\ 0 & \mu_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix}$ *then*

$$A + B = \begin{pmatrix} \lambda_1+\mu_1 & * & \cdots & * \\ 0 & \lambda_2+\mu_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n+\mu_n \end{pmatrix}$$

and

$$AB = \begin{pmatrix} \lambda_1\mu_1 & * & \cdots & * \\ 0 & \lambda_2\mu_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n\mu_n \end{pmatrix}.$$

Note that $*$ may be different in different matrices.

In particular, we have for any integer $k \geq 1$ and $f(x) \in F[x]$

$$A^k = \begin{pmatrix} \lambda_1^k & * & \cdots & * \\ 0 & \lambda_2^k & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{pmatrix}$$

and

$$f(A) = \begin{pmatrix} f(\lambda_1) & * & \cdots & * \\ 0 & f(\lambda_2) & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(\lambda_n) \end{pmatrix}.$$

Proof. Directly verify using matrix addition and multiplication. □

Lemma 11.9 (Linear transformation given by an upper triangular matrix)**.** *Let $T : V \to V$ be a linear transformation and $\beta = (v_1, \ldots, v_n)$ be a basis of V. If $A = [T]_\beta^\beta$ is an upper triangular matrix then $T(E_j) \subset E_j$ for all $j = 1, \ldots, n$ where $E_j = \text{span}(v_1, \ldots, v_j)$. Moreover, if $a_{jj} = \lambda_j$, then $(T - \lambda_j \text{id})(E_j) \subset E_{j-1}$.*

Proof. Since $A = [T]_\beta^\beta$ is upper triangular, for any $1 \leq j \leq n$, $T(v_j) = a_{1j}v_1 + \cdots + a_{jj}v_j \in E_j$. Thus $T(E_j) \subset E_j$.

Since $[T - \lambda_j \text{id}]_\beta^\beta = A - \lambda_j I$ is upper triangular, we have $(T - \lambda_j \text{id})(E_{j-1}) \subset E_{j-1}$. On the other hand, $(T - \lambda_j \text{id})(v_j) = a_{1j}v_1 + \cdots + a_{jj}v_j - \lambda_j v_j = a_{1j}v_1 + \cdots + a_{(j-1)j}v_{j-1} \in E_{j-1}$. Since $(T - \lambda_j \text{id})$ maps a basis of E_j into E_{j-1}, we have $(T - \lambda_j \text{id})(E_j) \subset E_{j-1}$. □

Proof. If $p_A(x) = (x - \lambda_1) \cdots (x - \lambda_n)$ splits, then by Theorem 11.7 there is $P \in M_{n \times n}(F)$ invertible such that

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Define the subspaces $E_j = \text{span}(v_1, v_2, \ldots, v_j)$, where $v_1, v_2, \ldots, v_n$ are the columns of P.

Take an arbitrary vector $v \in F^n = E_n$. Then we have $p_A(A)v = (A - \lambda_1 I) \cdots (A - \lambda_n I)v$. Then we have

$$\begin{aligned}
&(A - \lambda_n I)v \in E_{n-1} \\
&(A - \lambda_{n-1} I)(A - \lambda_n I)v \in E_{n-2} \\
&\vdots \\
&(A - \lambda_2 I) \cdots (A - \lambda_n I)v \in E_1.
\end{aligned}$$

The last inclusion implies that $(A - \lambda_2 I) \cdots (A - \lambda_n I)v = cv_1$ for some $c \in F$. But $(A - \lambda_1 I)v_1 = 0$, and hence

$$0 = (A - \lambda_1 I)(cv_1) = (A - \lambda_1 I)(A - \lambda_2 I) \cdots (A - \lambda_n I)v.$$

Therefore, $p_A(A)v = 0$ for all $v \in F^n$, which means exactly that $p_A(A) = 0$.

If p_A does not splits in F, then by Proposition 9.17, there is a field extension E of F such that $p_A(x)$ splits in E. We now view A as a matrix over E. Then the previous argument shows $p_A(A) = 0$. Since p_A is the same polynomial and $p_A(A)$ is the same matrix regardless of whether you think of them as over E or F, we finish the proof. □

11.2 Minimal polynomial

Theorem 11.10. *There exists a unique polynomial such that*

(i) $m_A(x)$ is monic.

(ii) $m_A(A) = 0$

(iii) For any $f(x) \in F[x]$ such that $f(A) = 0$, we have $m_A(x) \mid f(x)$.

In fact, $m_A(x)$ is the monic nonzero polynomial with least degree such that $m_A(A) = 0$.

Proof. Consider the set $\mathcal{I}_A = \{f(x) \in F[x] : f(A) = 0\}$. By the Cayley-Hamilton theorem, the characteristic polynomial $p_A(x) = \det(xI - A) \in \mathcal{I}_A$, since $p_A(A) = 0$. So $\mathcal{I}_A \neq \{0\}$. Let m_A be a nonzero polynomial in $\mathcal{I}_A$ of least degree. We scale m_A such that it is monic.

Since $m_A \in \mathcal{I}_A$, we have $m_A(A) = 0$.

For any $f \in \mathcal{I}_A$ we have $f(A) = 0$. By Theorem 9.6, $f(x) = m_A(x)q(x) + r(x)$ where $\deg r < \deg m_A$. Since $0 = f(A) = m_A(A)q(A) + r(A) = r(A)$. We must have $r = 0$ since otherwise $r \neq 0$ and we have $r \in \mathcal{I}_A$. $\deg r < \deg m_A$. This contradicts m_A has least degree in $\mathcal{I}_A$. Hence $m_A(x) \mid f(x)$.

To show uniqueness, suppose that m_1, m_2 are two polynomials satisfying (i), (ii) and (iii). Then since $m_1(A) = 0$, $m_2(x) \mid m_1(x)$ by (iii) of $m_2(x)$. Since $m_2(A) = 0$, $m_1(x) \mid m_2(x)$ by (iii) of m_1. Thus $m_1 = cm_2$ for some $c \in F$. Since m_1 and m_2 are both monic, we have $m_1 = m_2$. □

Definition 11.11. The polynomial $m_A(x)$ is called the **minimal polynomial** of A.

Example 11.12. A $n \times n$ **Jordan block** with diagonal λ is a matrix of the form

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} \in M_{n\times n}(F).$$

Then $p_{J_n(\lambda)}(x) = (x-\lambda)^n$ and by Propostion 11.13 $m_{J_n(\lambda)}(x) = (x-\lambda)^k$ for some $1 \leq k \leq n$. Let $N = J_n(\lambda) - \lambda I$. We have

$$\begin{aligned} &N(e_1) = 0, N(e_j) = e_{j-1} \text{ for any } j = 2, \ldots, n \\ &N^2(e_1) = N^2(e_2) = 0, N^2(e_j) = e_{j-2} \text{ for any } j = 3, \ldots, n \\ &\cdots \\ &N^{n-1}(e_1) = \cdots = N^{n-1}(e_{n-1}) = 0, N^{n-1}(e_n) = e_1 \\ &N^n(e_1) = \cdots = N^n(e_n) = 0. \end{aligned}$$

Thus $m_{J_n(\lambda)}(x) = (x-\lambda)^n$.

Proposition 11.13. *If the characteristic polynomial*

$$p_A(x) = (x-\lambda_1)^{k_1}(x-\lambda_2)^{k_2}\cdots(x-\lambda_m)^{k_m}$$

splits over F so the $\lambda_i \in F$ are the distinct eigenvalues, then the minimal polynomial $m_A(x)$ has the same distinct linear factors, i.e.

$$m_A(x) = (x-\lambda_1)^{l_1}(x-\lambda_2)^{l_2}\cdots(x-\lambda_m)^{l_m} \tag{5}$$

with $1 \leq l_i \leq k_i$. In other words, every factor $(x-\lambda_i)$ that appears in p_A also appears in m_A.

Proof. By Theorem 11.7, there exists $P \in M_{n\times n}(F)$ invertible such that

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

By Lemma 11.8,

$$m_A(A) = P \begin{pmatrix} m_A(\lambda_1) & * & \cdots & * \\ 0 & m_A(\lambda_2) & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_A(\lambda_n) \end{pmatrix} P^{-1} = 0.$$

Thus $m_A(\lambda_i) = 0$ for all $1 \leq i \leq n$. By Proposition 9.10, $(x-\lambda_i) \mid m_A(x)$ for any eigenvalue λ_i of A. □

Theorem 11.14. *Let $A \in M_{n\times n}(F)$. Then A is diagonalizable over F if and only if the minimal polynomial $m_A(x)$ splits over F as a product of distinct linear factors, i.e.*

$$m_A(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_m)$$

where $\lambda_1, \ldots, \lambda_m$ are distinct. In other words, $l_1 = \cdots = l_m = 1$ in (5).

Proof. "$\Longrightarrow$" If A is diagonalizable over F then there exists an invertible P with $P^{-1}AP = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, where each $\lambda_j \in F$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of A. We define $f(x) = (x-\lambda_1)\cdots(x-\lambda_m) \in F[x]$. Then we have $f(A) = Pf(\mathrm{diag}(\lambda_1, \ldots, \lambda_n))P^{-1} = 0$. Thus $m_A(x) \mid f(x)$. By Proposition 11.13, we have $f(x) \mid m_A(x)$. Thus $m_A(x) = f(x)$.

"$\Longleftarrow$" Conversely, suppose $m_A(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_m)$ with distinct $\lambda_i \in F$. We define $p_i(x) = (x-\lambda_1)\cdots(x-\lambda_{i-1})(x-\lambda_{i+1})\cdots(x-\lambda_n)$. By Lagrange interpolation theorem (Homework 9 Problem 4),

$$1 = \sum_{i=1}^{m} c_i p_i(x), \quad c_i = \frac{1}{(\lambda_i-\lambda_1)\cdots(\lambda_i-\lambda_{i-1})(\lambda_i-\lambda_{i+1})\cdots(\lambda_i-\lambda_n)}.$$

Substituting A for x gives a decomposition of the identity matrix as a sum of matrices

$$I = \sum_{i=1}^{m} P_i, \qquad P_i := c_i p_i(A).$$

For any $v \in F^n$ we have $v = \sum_{i=1}^m P_i v$. For each $1 \le i \le m$,

$$(A-\lambda_i I)P_i v = c_i(A-\lambda_i I)p_i(A)v = c_i m_A(A)v = 0v = 0.$$

Thus $P_i v \in E(\lambda_i, A)$. Hence F^n decomposes as a direct sum of eigenspaces $F^n = E(\lambda_1, A) + \cdots + E(\lambda_m, A) = E(\lambda_1, A) \oplus \cdots \oplus E(\lambda_m, A)$. Therefore A is diagonalizable over F by Theorem 10.11. □

Example 11.15. If $A^2 = A$ then A is diagonalizable with eigenvalue 1 or 0. This is because $f(x) = x^2 - x \in \mathcal{I}_A$. Then $m_A(x) \mid f(x)$. Since $f(x)$ splits as distinct linear factors, $m_A(x)$ splits as distinct linear factors. Similarly, if $A^2 = I$, then A is diagonalizable with eigenvalue 1 or -1.