

12 Lecture 12

12.1 Commuting operators

Theorem 12.1. If $A, B \in M_{n \times n}(F)$ are both diagonalizable. Then $AB = BA$ if and only if there exists $P \in M_{n \times n}(F)$ invertible such that $P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n)$ and $P^{-1}BP = \text{diag}(\mu_1, \dots, \mu_n)$.

Proof. \Leftarrow : If $P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n)$ and $P^{-1}BP = \text{diag}(\mu_1, \dots, \mu_n)$, then $AB = P \text{diag}(\lambda_1\mu_1, \dots, \lambda_n\mu_n)P^{-1} = BA$.

\Rightarrow : Since A is diagonalizable over F , $F^n = E(\lambda_1, A) \oplus \dots \oplus E(\lambda_m, A)$.

If $v \in E(\lambda_i, A)$ then $A(Bv) = B(Av) = B(\lambda_i v) = \lambda_i(Bv)$ so $Bv \in E(\lambda_i, A)$. Thus $B(E(\lambda_i, A)) \subset E(\lambda_i, A)$ for any $1 \leq i \leq m$. Thus if we let $Q \in M_{n \times n}(F)$ such that its columns are basis of $E(\lambda_i, A)$ for $1 \leq i \leq m$, then

$$Q^{-1}AQ = \text{diag}(\lambda_1, \dots, \lambda_n), \quad Q^{-1}BQ = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_m \end{pmatrix}$$

where $B_i \in M_{k_i \times k_i}(F)$ for $k_i = \dim E(\lambda_i, A)$.

Since $m_B(B) = 0$, we have $m_{B_i}(x) \mid m_B(x)$. Since B is diagonalizable, $m_B(x)$ has no repeated factors. Hence $m_{B_i}(x)$ also has no repeated factors. Thus by Theorem 11.14 there exists a basis β_i of $E(\lambda_i, A)$ such that B_i is diagonalized.

By Theorem 10.9, $\beta = \beta_1 \cup \dots \cup \beta_m$ is a basis of F^n . Since β_i is a basis of $E(\lambda_i, A)$, A is diagonalized under β . Since B_i is diagonalized under β_i , B is diagonalized under β . If P is the change of basis matrix then

$$P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n), \quad P^{-1}BP = \text{diag}(\mu_1, \dots, \mu_n).$$

□

Remark 12.2. The theorem is saying that there is a basis consisting of vectors that are eigenvectors of both A and B . This is not saying that any eigenvector of A is an eigenvector of B . We take $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ as an example. Clearly $AB = BA$ and they are already in diagonalized under the standard basis e_1, e_2 . However, since $A = I$, any nonzero vector is an eigenvectors of A but the eigenvectors of B are e_1 with eigenvalue 1 and e_2 with eigenvalue 2. Thus $e_1 + e_2$ is an eigenvector of A but not an eigenvector of B .

Remark 12.3. Theorem 12.1 also appears in quantum mechanics. If the operators corresponding to two physical quantities commute, then their values can be measured simultaneously. If they do not commute, the famous uncertainty principle states that their values cannot be simultaneously determined. In fact, the uncertainty principle has a qualitative version: we fix $\psi \in \mathbb{C}^n$. For any $A \in M_{n \times n}(\mathbb{C})$, we write

$\langle A \rangle = \bar{\psi}^t A \psi \in \mathbb{C}$. Let $A, B \in M_{n \times n}(\mathbb{C})$ with $\bar{A}^t = A$ and $\bar{B}^t = B$. Then we have $|(A - \langle A \rangle)^2 \psi|^2 |(B - \langle B \rangle)^2 \psi|^2 \geq |\frac{1}{2i} \langle [A, B] \rangle|^2$. The term $|(A - \langle A \rangle)^2 \psi|^2$ represents the variance of A in state ψ and the inequality is saying the product of variance is bounded below by the mean of the commutator so that if the commutator isn't 0, then lowering the variance of A will increase the variance of B .

12.2 A brief introduction to Jordan canonical form

In Lecture 9, we raised the question: given $T : V \rightarrow V$ a linear transformation, can we find a basis β such that $[T]_\beta^\beta$ is diagonal? We have seen in Example 9.27 and 11.12 that there are matrices that cannot be diagonalized. So we step back and ask what is the simplest form that $[T]_\beta^\beta$ can take? In Theorem 11.7, we see that we could make $[T]_\beta^\beta$ an upper triangular matrix. However, the issue with Theorem 11.7 is we have no information about the entries that are above the diagonal. The following theorem is the ultimate resolution to the question above.

Theorem 12.4. *Let $A \in M_{n \times n}(F)$ be such that $p_A(x)$ splits in F . Then there exists $P \in M_{n \times n}(F)$ invertible such that*

$$P^{-1}AP = \begin{pmatrix} J_{k_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{k_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{k_m}(\lambda_m) \end{pmatrix}$$

where $J_{k_i}(\lambda_i)$ is a Jordan block. This is called the **Jordan canonical form** of A .

Moreover, The Jordan canonical form for A unique up to a permutation of the Jordan blocks along the diagonal.

Remark 12.5. The λ_i here need not be distinct. E.g. we could have a Jordan canonical form like $\begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}$.

Remark 12.6. In Example 11.12, we see that a Jordan block is diagonalizable if and only if it is 1×1 . In general, A is diagonalizable if and only if all its Jordan block are 1×1 , i.e. its Jordan canonical form is diagonal.

Moreover, two matrices with different Jordan canonical form are not similar to each other. So we have:

Theorem 12.7. *A, B are similar if and only if their Jordan canonical form are the same up to permutation.*

The two theorems above are among the most difficult results to prove in linear algebra, and we unfortunately do not have time to cover their proofs in class. There are several different approaches in the literature, and I personally prefer the following two.

The first proof, from [1, Chapter 9], is completely elementary, and anyone with the background provided in these lecture notes should be able to follow it. The second proof, from [2, Chapter 12], is based on the structure theorem for finitely generated modules over a PID. I find the first proof more accessible for students encountering these ideas for the first time, while the second presents linear algebra from a more advanced, conceptual viewpoint and connects naturally with the broader theory underlying the Jordan canonical form.

In this note, we prove the simpler case for 2×2 matrices.

Theorem 12.8 (Jordan canonical form for 2×2 matrices). *Let $A \in M_{2 \times 2}(F)$ be such that $p_A(x)$ splits in F . Then there exists $P \in M_{2 \times 2}(F)$ invertible such that either $P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ or $P^{-1}AP = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$.*

Proof. If A has two distinct eigenvalues then A is diagonalizable by Corollary 10.12.

If A has one eigenvalue and $\dim E(\lambda, A) = 2$, then $A = \lambda I$ is diagonalizable.

If A has one eigenvalue and $\dim E(\lambda, A) = 1$, then by rank-nullity Theorem 4.14, $\text{rank}(\lambda I - A) = 1$. Let $0 \neq v \in \text{im}(A - \lambda I)$. Then $v = (A - \lambda I)w$ for $0 \neq w \in F^2$ i.e. $Aw = \lambda w + v$. The characteristic polynomial of A is $(x - \lambda)^2$. Then by Cayley-Hamilton theorem, $(A - \lambda I)^2 = 0$. Then we have $(A - \lambda I)v = (A - \lambda I)^2w = 0$ i.e. $Av = \lambda v$. We claim that v, w are linearly independent. If $av + bw = 0$, then applying $(A - \lambda I)$ on both sides $0 = (A - \lambda I)(av + bw) = bv$. Since $v \neq 0$, we have $b = 0$. Thus $av = 0$ and hence $a = 0$. Thus v, w are linearly independent. We let $P = (v, w)$. Then P is invertible, and $P^{-1}AP = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ \square

The proof provides a good way to compute the Jordan canonical form for 2×2 matrices

Example 12.9. We recall Example 9.27 where $A = \begin{pmatrix} -1 & -4 \\ 1 & 3 \end{pmatrix}$ and $p_A(x) = (x - 1)^2$. Then $\lambda = 1$ is the only eigenvalue of A . Since $A - I = \begin{pmatrix} -2 & -4 \\ 1 & 2 \end{pmatrix}$, $\dim \ker(A - I) = 1 < 2$. $\text{im}(A - I) = \text{span} \begin{pmatrix} -2 \\ 1 \end{pmatrix}$. Let $v = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$. Solving $(A - I)w = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$, we get $w = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. We set $P = \begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix}$, then we have $P^{-1}AP = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

We have seen in Example 9.23 that a real matrix can have eigenvalues not in \mathbb{R} . Of course we can view it as a complex matrix and apply Theorem 12.8. However, we would like to know what is the simplest form of it under similarity by $P \in M_{2 \times 2}(\mathbb{R})$ invertible.

Theorem 12.10 (Real canonical form for 2×2 matrices over \mathbb{R}). *Let $A \in M_{2 \times 2}(\mathbb{R})$. Then there exists $P \in M_{2 \times 2}(\mathbb{R})$ invertible such that either $P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ or $P^{-1}AP = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ or $P^{-1}AP = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ where $a \pm bi$ are the eigenvalues of A in \mathbb{C} .*

Proof. If $p_A(x)$ splits in \mathbb{R} , then the theorem follows from Theorem 12.8.

We deal with the case where A has no real eigenvalues. Then $p_A(x) = x^2 - \text{tr}(A)x + \det(A)$ has complex roots $a \pm bi$ for $b \neq 0$. Let $v \in \mathbb{C}^2$ be an eigenvector with eigenvalue $a + bi$. Then $Av = (a + bi)v$. Let $v = v_1 + v_2i$ where $v_1 = \text{Re } v \in \mathbb{R}^2$ and $v_2 = \text{Im } v \in \mathbb{R}^2$. We write $P = (v_1, v_2) \in M_{2 \times 2}(\mathbb{R})$. Then $Av_1 + Av_2i = A(v_1 + v_2i) = (a + bi)(v_1 + v_2i) = (av_1 - bv_2) + (av_2 + bv_1)i$. Thus $Av_1 = av_1 - bv_2$ and $Av_2 = av_2 + bv_1$. Then $P^{-1}AP = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ provided P is invertible which can be seen as follows:

By Remark 9.25, v, \bar{v} are the eigenvector with distinct eigenvalues $a + bi$ and $a - bi$. Thus by Theorem 10.9 v and \bar{v} are linearly independent. That v_1, v_2 are linearly independent follows from the following lemma.

Lemma 12.11. *Let $v \in \mathbb{C}^n$. If v, \bar{v} are linearly independent, then $\text{Re } v$ and $\text{Im } v$ are linearly independent.*

Proof. We have $\text{Re } v = \frac{1}{2}(v + \bar{v})$ and $\text{Im } v = \frac{1}{2i}(v - \bar{v})$. Suppose $a \text{Re } v + b \text{Im } v = 0$ for $a, b \in \mathbb{C}$. Then $(\frac{1}{2}a + \frac{1}{2i}b)v + (\frac{1}{2}a - \frac{1}{2i}b)\bar{v} = 0$. Since v, \bar{v} are linearly independent, we have

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2i} \\ \frac{1}{2} & -\frac{1}{2i} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Since $\det \begin{pmatrix} \frac{1}{2} & \frac{1}{2i} \\ \frac{1}{2} & -\frac{1}{2i} \end{pmatrix} = -\frac{1}{2i} \neq 0$, we have $a = b = 0$. □

□

Example 12.12. We recall Example 9.24 where $A = \begin{pmatrix} -1 & 2 \\ -1 & 1 \end{pmatrix} \in M_{2 \times 2}(\mathbb{C})$ and $p_A(x) = x^2 + 1 = (x - i)(x + i)$. We have $v_1 = \begin{pmatrix} 1-i \\ 1 \end{pmatrix}$ is an eigenvector with eigenvalue i . Then $\text{Re } v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\text{Im } v_1 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$. If $P = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$ then $P^{-1}AP = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

12.3 Convergence of a sequence of matrices

Definition 12.13. A sequence of matrices $\{A_k = (a_{ij}^{(k)})\} \subset M_{m \times n}(\mathbb{C})$ converges to a matrix $A = (a_{ij}) \in M_{m \times n}(\mathbb{C})$ if $a_{ij}^{(k)} \rightarrow a_{ij}$ for any $i = 1, \dots, m, j = 1, \dots, n$ as $k \rightarrow \infty$. We write

$$\lim_{k \rightarrow \infty} A_k = A.$$

Definition 12.14. Let $A(t) = (a_{ij}(t)) \in M_{m \times n}(\mathbb{C})$ be a family of matrix depending on $t \in (a, b)$. Then we define the derivative at t to be

$$A'(t) = \lim_{h \rightarrow 0} \frac{A(t+h) - A(t)}{h}.$$

Lemma 12.15. *If $A(t) = (a_{ij}(t))$ where a_{ij} are differentiable at t then $A'(t) = (a'_{ij}(t))$.*

Proof. Since $a'_{ij}(t) = \lim_{h \rightarrow 0} \frac{a_{ij}(t+h) - a_{ij}(t)}{h}$ exists, we have $A'(t)$ exists and $A'(t) = (a'_{ij}(t))$. \square

12.4 Matrix function

Definition 12.16. Let $f(x) = \sum_{k=0}^{\infty} c_k x^k$ with $c_k \in \mathbb{C}$ be a function defined by a power series. We define

$$f(A) = \sum_{k=0}^{\infty} c_k A^k = \lim_{N \rightarrow \infty} \sum_{k=0}^N c_k A^k$$

for $A \in M_{n \times n}(\mathbb{C})$ if the right hand side converges.

Proposition 12.17. Let $f(x) = \sum_{k=0}^{\infty} c_k x^k$ with $c_k \in \mathbb{C}$ be a function defined by a power series and $r = \frac{1}{\limsup_{k \rightarrow \infty} |c_k|^{1/k}}$ be the radius of convergence of the power series f .

- (i) If $f(A)$ exists, then $f(P^{-1}AP)$ also exists and $f(P^{-1}AP) = P^{-1}f(A)P$.
- (ii) If $|\lambda| < r$ then $f(J_n(\lambda))$ converges and we have

$$f(J_n(\lambda)) = \begin{pmatrix} f(\lambda) & \frac{f'(\lambda)}{1!} & \frac{f''(\lambda)}{2!} & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} & \frac{f^{(n)}(\lambda)}{n!} \\ 0 & f(\lambda) & \frac{f'(\lambda)}{1!} & \cdots & \frac{f^{(n-2)}(\lambda)}{(n-2)!} & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ 0 & 0 & f(\lambda) & \cdots & \frac{f^{(n-3)}(\lambda)}{(n-3)!} & \frac{f^{(n-2)}(\lambda)}{(n-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & f(\lambda) & \frac{f'(\lambda)}{1!} \\ 0 & 0 & 0 & \cdots & 0 & f(\lambda) \end{pmatrix}.$$

(iii) Let $A \in M_{n \times n}(\mathbb{C})$ be such that $|\lambda| < r$ for each eigenvalue λ of A . Then $f(A) = \sum_{k=0}^{\infty} c_k A^k$ exists and if $P \in M_{n \times n}(\mathbb{C})$ is invertible such that

$$P^{-1}AP = \begin{pmatrix} J_{k_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{k_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{k_m}(\lambda_m) \end{pmatrix}$$

then

$$f(A) = P \begin{pmatrix} f(J_{k_1}(\lambda_1)) & 0 & \cdots & 0 \\ 0 & f(J_{k_2}(\lambda_2)) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(J_{k_m}(\lambda_m)) \end{pmatrix} P^{-1}$$

where $f(J_{k_i}(\lambda_i))$ is given by the formula in (ii).

Proof. (i) Let $f(x) = \sum_{k=0}^{\infty} c_k x^k$ and assume that the series $f(A) = \sum_{k=0}^{\infty} c_k A^k$ converges. Let $B = P^{-1}AP$ for some invertible matrix P . Then we have

$$\sum_{k=0}^l c_k B^k = \sum_{k=0}^l c_k (P^{-1}AP)^k = \sum_{k=0}^l c_k P^{-1}A^k P = P^{-1} \left(\sum_{k=0}^l c_k A^k \right) P.$$

Taking $l \rightarrow \infty$, we have $f(P^{-1}AP) = P^{-1}f(A)P$.

(ii) We write $J_n(\lambda) = \lambda I + N$, where $N = J_n(0)$ satisfies $N^n = 0$ as in Example 11.12. For any integer $k \geq 0$ the binomial expansion (finite because N is nilpotent) yields

$$\begin{aligned} J_n(\lambda)^k &= (\lambda I + N)^k = \sum_{j=0}^{n-1} \binom{k}{j} \lambda^{k-j} N^j = \sum_{j=0}^{n-1} \frac{k!}{j!(k-j)!} \lambda^{k-j} N^j \\ &= \sum_{j=0}^{n-1} \frac{1}{j!} k(k-1)\cdots(k-j+1) \lambda^{k-j} N^j. \end{aligned}$$

Then

$$\begin{aligned} f(J_n(\lambda)) &= \sum_{k=0}^{\infty} c_k J_n(\lambda)^k = \sum_{k=0}^{\infty} c_k \left(\sum_{j=0}^{n-1} \frac{1}{j!} k(k-1)\cdots(k-j+1) \lambda^{k-j} N^j \right) \\ &= \sum_{j=0}^{n-1} \frac{1}{j!} N^j \left(\sum_{k=0}^{\infty} c_k k(k-1)\cdots(k-j+1) \lambda^{k-j} \right). \end{aligned}$$

Since $|\lambda| < r$, we can differentiate f term-by-term and get $f^{(j)}(\lambda) = \sum_{k=0}^{\infty} c_k k(k-1)\cdots(k-j+1) \lambda^{k-j}$ where the series converges uniformly in λ . Thus $f(J_n(\lambda)) = \sum_{j=0}^{n-1} \frac{f^{(j)}(\lambda)}{j!} N^j$. Since N^j is a matrix with 1's on the j -th superdiagonal and 0 elsewhere we get the result.

(iii) By the Jordan canonical form Theorem 12.4, there exists an invertible $P \in M_{n \times n}(\mathbb{C})$ such that

$$P^{-1}AP = \begin{pmatrix} J_{k_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{k_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{k_m}(\lambda_m) \end{pmatrix}.$$

Since $|\lambda_i| < r$ for all $1 \leq i \leq m$, by (ii) $f(J_{k_i}(\lambda_i))$ converges. By (i) we have $f(A)$ converges and

$$f(A) = P \begin{pmatrix} f(J_{k_1}(\lambda_1)) & 0 & \cdots & 0 \\ 0 & f(J_{k_2}(\lambda_2)) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(J_{k_m}(\lambda_m)) \end{pmatrix} P^{-1}.$$

□

Corollary 12.18. For any $A \in M_{n \times n}(\mathbb{C})$, $\exp(A) = e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$ exists.

Proof. The radius of convergence of $f(x) = e^x = \sum_{j=0}^{\infty} \frac{x^j}{j!}$ is ∞ . □

$$\text{Example 12.19. } \exp(J_n(\lambda)) = \begin{pmatrix} e^\lambda & \frac{e^\lambda}{1!} & \frac{e^\lambda}{2!} & \cdots & \frac{e^\lambda}{(n-1)!} & \frac{e^\lambda}{n!} \\ 0 & e^\lambda & \frac{e^\lambda}{1!} & \cdots & \frac{e^\lambda}{(n-2)!} & \frac{e^\lambda}{(n-1)!} \\ 0 & 0 & e^\lambda & \cdots & \frac{e^\lambda}{(n-3)!} & \frac{e^\lambda}{(n-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & e^\lambda & \frac{e^\lambda}{1!} \\ 0 & 0 & 0 & \cdots & 0 & e^\lambda \end{pmatrix}$$

References

- [1] Linear algebra done wrong, by Sergei Treil 2025,
- [2] Dummit, David Steven, and Richard M. Foote. Abstract algebra. Vol. 3. Hoboken: Wiley, 2004.