LINEAR ALGEBRA

KIM, SUNGJIN

1. Preliminary

Lemma 1.1 (Smith Normal Form). *Let A be a nonzero $m \times n$ matrix over a principal ideal domain(PID) R. There exist invertible $m \times m$ and $n \times n$ matrices P, Q so that*

$$P^{-1}AQ^{-1} = Diag(\alpha_1, \cdots, \alpha_r), \tag{1}$$

where $\alpha_i \mid \alpha_{i+1}$ for $i < r$, here the last few terms can be 0.

The matrices P and Q may not be products of elementary matrices in general (see [1, p.23]). When the ring R is Euclidean, then it is possible to find P and Q through elementary row/column operations.

Lemma 1.2 (Structure Theorem over PID, Invariant factor decomposition). *Every finitely generated module M over a PID R is isomorphic to a unique one of the form*

$$R^f \bigoplus \oplus_{i=1}^{r} R/(d_i), \tag{2}$$

where $d_i \mid d_{i+1}$, and $d_i \neq (0)$. The summands are called the invariant factors.

Lemma 1.3 (Structure Theorem over PID, Primary decomposition). *Conditions are the same as above, M is isomorphic to a unique one of the form*

$$R^f \bigoplus \oplus_{i=1}^{s} R/(p_i^{r_i}), \tag{3}$$

where p_i are prime ideals.

2. Theorems

We regard a left-multiplication of an $n \times n$ matrix A over a field $\mathbb{F}$ as $\mathbb{F}[x]$-module element, namely x. Then $\mathbb{F}^n$ can be viewed as an $\mathbb{F}[x]$-module with $p(x) \in \mathbb{F}[x]$ acting as $p(A) \in M_{n\times n}(\mathbb{F})$, denoted as M^A. Note that for any field $\mathbb{F}$, the polynomial ring $\mathbb{F}[x]$ is an ED (Euclidean Domain), hence a PID. Our application of the structure theorem in invariant factor form is

Theorem 2.1 (Rational Canonical Form-Invariant factor form). *Let A be a $n \times n$ matrix over a field $\mathbb{F}$. Then A is similar to a block diagonal matrix of the form*

$$\oplus_{i=1}^{r} C(f_i), \tag{4}$$

where $f_i \mid f_{i+1}$, and $C(f_i)$ is the companion matrix associated to f_i. This form is unique up to rearrangement of blocks.

Using primary decomposition, we have

Theorem 2.2 (Rational Canonical Form-Primary decomposition). *Conditions are the same as above, A is similar to a block diagonal matrix of the form*

$$\oplus_{i=1}^{s} C(p_i^{r_i}), \tag{5}$$

where p_i are irreducible polynomials in $\mathbb{F}[x]$. This form is unique up to rearrangement of blocks.

For the proof, use structure theorem to the $\mathbb{F}[x]$-module M^A as described above.

If the ground field is algebraically closed, then we have Jordan Canonical Form.

Theorem 2.3 (Jordan Canonical Form). *Let A be a $n \times n$ matrix over an algebraically closed field $\overline{\mathbb{F}}$. Then A is similar to a block diagonal matrix of the form*

$$\oplus_{i=1}^{s} J(\lambda_i, r_i), \tag{6}$$

where λ_i are the eigenvalues of A, and $J(\lambda_i, r_i)$ is the Jordan block of diagonal λ_i and 1 directly below the main diagonal with size $r_i \times r_i$. This form is unique up to rearrangement of blocks.

Theorem 2.4 (Generalized Jordan Form). *Let A be a $n \times n$ matrix over a field $\mathbb{F}$. Then A is similar to a block diagonal matrix of the form*

$$\oplus_{i=1}^{s} J(p_i, r_i), \tag{7}$$

where p_i are the irreducible factors of the characteristic polynomial of A, and $J(p_i, r_i)$ is the generalized Jordan block of the form

$$\begin{pmatrix} C(p_i) & 0 & 0 & \cdots & 0 \\ U & C(p_i) & 0 & \cdots & 0 \\ 0 & U & C(p_i) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & U & C(p_i) \end{pmatrix}$$

with the matrix U whose sole nonzero entry 1 on the upper right corner, and r_i is the number of diagonal blocks. This form is unique up to rearrangement of blocks.

For the proof of Theorem 2.4, note that $J(p, r)$ is for the $\mathbb{F}[x]$-module $\mathbb{F}[x]/(p^r)$. Consider the expression

$$f(x) = a_0(x) + a_1(x)p(x) + \cdots + a_{r-1}(x)p(x)^{r-1} \in \mathbb{F}[x]/(p^r),$$

where $a_i(x) \in \mathbb{F}[x]$, $\deg(a_i) < \deg(p)$.

Now, the problem reduces to determining invariant factors. We use Smith Normal Form to do this.

Theorem 2.5 (Invariant Factors). *Let A be a $n \times n$ matrix over a field $\mathbb{F}$. Then invariant factors can be recovered from the Smith Normal Form of $xI - A$. More precisely, if $P^{-1}(xI - A)Q^{-1} = Diag(f_1, \cdots, f_n)$ for some invertible matrices P, Q and $f_i \mid f_{i+1}$, then f_i are the invariant factors of A.*

Here, first few terms can be 1. The proof starts from investigating the exact sequence

$$0 \longrightarrow \mathbb{F}[x]^n \xrightarrow{xI-A} \mathbb{F}[x]^n \xrightarrow{\pi} M^A \longrightarrow 0. \tag{8}$$

Then we see that

$$M^A \simeq \mathbb{F}[x]^n / \mathrm{Im}(xI - A).$$

Corollary 2.1 (Similarity of Transpose). *Let A be a $n \times n$ matrix over a field $\mathbb{F}$. Then A and its transpose A^T are similar.*

Proof. Write $xI - A = PDQ$ where P, Q are invertible in $M_{n\times n}(\mathbb{F}[x])$ and the Smith Normal Form D. Taking transpose, we have

$$xI - A^T = Q^T D^T P^T = Q^T D P^T.$$

Since Q^T, P^T are also invertible, we see that $xI - A$ and $xI - A^T$ have the same invariant factors. □

Corollary 2.2 (Computation of Similarity Transform for Transpose). *Let A be $n \times n$ matrix over a field $\mathbb{F}$. Then we are able to compute the nonsingular similarity transform X such that $XA = A^T X$.*

Proof. Let $xI - A = PDQ$ with intertible $P, Q \in M_{n\times n}(\mathbb{F}[x])$ and D be the Smith Normal Form of A. Note that the computation of Smith Normal Form is essentially elementary row/column operations. Thus, we keep track of row/column operations on $xI - A$, and P is obtained by a product of inverses of elementary matrices corresponding to the row operations, Q by a product of those inverses of column operations.

Then $xI - A = PDQ$ and $xI - A^T = Q^T D P^T$ yield an $\mathbb{F}[x]$-module isomorphism

$$\mathbb{F}[x]^n / \mathrm{Im}(xI - A) \xrightarrow{Q^T P^{-1}} \mathbb{F}[x]^n / \mathrm{Im}(xI - A^T).$$

We may regard the module isomorphism as a vector-space isomorphism between M^A and M^{A^T}. This isomorphism is obtained by writing $Q^T P^{-1} \in M_{n\times n}(\mathbb{F}[x])$ with x replaced by the left multiplication of A^T.

Write the resulting matrix as X. Then X is nonsingular since it is an isomorphism between n-dimensional vector spaces. Since $Q^T P^{-1}$ is an $\mathbb{F}[x]$-module isomorphism, it commutes with the action of x. Since the action of x in M^A is the left multiplication of A, and that in M^{A^T} is the left multiplication of A^T, we have $XA = A^T X$. ☐

Note that this method can be applied in finding a similarity transform to any given two similar matrices.

Example 2.1. Let $A = \begin{pmatrix} 2 & 3 \\ 5 & 7 \end{pmatrix}$. Then the elementary row/column operations on $xI - A$ yield the Smith Normal Form

$$P^{-1}(xI - A)Q^{-1} = \begin{pmatrix} 1 & \\ & x^2 - 9x - 1 \end{pmatrix}.$$

Here, we have

$$P = \begin{pmatrix} -3 & \\ x - 7 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} -\frac{x-2}{3} & 1 \\ \frac{1}{3} & \end{pmatrix}.$$

Then,

$$Q^T P^{-1} = \begin{pmatrix} \frac{2x-9}{9} & \frac{1}{3} \\ -\frac{1}{3} & \end{pmatrix} = x \begin{pmatrix} \frac{2}{9} & \\ & \end{pmatrix} + \begin{pmatrix} -1 & \frac{1}{3} \\ -\frac{1}{3} & \end{pmatrix}.$$

This gives the nonsingular matrix X,

$$X = \begin{pmatrix} 2 & 5 \\ 3 & 7 \end{pmatrix} \begin{pmatrix} \frac{2}{9} & \\ & \end{pmatrix} + \begin{pmatrix} -1 & \frac{1}{3} \\ -\frac{1}{3} & \end{pmatrix} = \begin{pmatrix} -\frac{5}{9} & \frac{1}{3} \\ \frac{1}{3} & \end{pmatrix}.$$

For this X, we have

$$XA = \begin{pmatrix} \frac{5}{9} & \frac{2}{3} \\ \frac{2}{3} & 1 \end{pmatrix} = A^T X.$$

Note that in this method, we do not need to fully produce a Smith Normal Form for A. We just need a diagonal D in the elementary row/column operations. This makes the calculation of P and Q simpler. With this simpler procedure, we have the following general result.

Example 2.2. In general, let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $b \neq 0$,

$$\begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \begin{pmatrix} 1 & \\ \frac{x-d}{b} & 1 \end{pmatrix} \begin{pmatrix} x - a & -b \\ -c & x - d \end{pmatrix} \begin{pmatrix} 1 & \\ \frac{x-a}{b} & 1 \end{pmatrix} = \begin{pmatrix} -c + \frac{(x-a)(x-d)}{b} & \\ & -b \end{pmatrix}.$$

With $P^{-1} = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \begin{pmatrix} 1 & \\ \frac{x-d}{b} & 1 \end{pmatrix}$ and $Q = \begin{pmatrix} 1 & \\ -\frac{x-a}{b} & 1 \end{pmatrix}$, we have $Q^T P^{-1} = \begin{pmatrix} \frac{a-d}{b} & 1 \\ 1 & \end{pmatrix}$.

Case 1: $b \neq 0$, we can take $X = \begin{pmatrix} \frac{a-d}{b} & 1 \\ 1 & \end{pmatrix}$. Then $XA = A^T X$.

Case 2: $b = 0$, $c \neq 0$, then we take $X = \begin{pmatrix} \frac{a-d}{c} & 1 \\ 1 & \end{pmatrix}$. Then $XA^T = AX$.

Case 3: $b = c = 0$, then A is diagonal, and $X = I$ works.

Corollary 2.3 (Similarity Preserved by Field Extension)**.** *Let A and B be $n \times n$ matrices over a field K. Let L be a field extension of K. Then A and B are similar over K if and only if they are similar over L.*

Proof. $\Rightarrow$) is obvious.
$\Leftarrow$) Let $\{A_i\}$ be the complete set of invariant factors of A, and $\{B_i\}$ that of B. Then we have

$$L \otimes_K (\oplus_i K[x]/(A_i)) = \oplus_i L[x]/(A_i),$$

and

$$L \otimes_K (\oplus_i K[x]/(B_i)) = \oplus_i L[x]/(B_i).$$

Since A and B are similar over L, we see that the RHS of the above formulas should be equal. Hence the sets of invariant factors $\{A_i\}$ and $\{B_i\}$ are identical, yielding that A and B are similar over K. ☐

Theorem 2.6 (Centralizer of a Matrix)**.** *Let A be a $n\times n$ matrix over $\mathbb{F}$. Let $C_A = \{B \in M_{n\times n}(\mathbb{F}) \mid AB = BA\}$. Then the minimal dimension of C_A over $\mathbb{F}$ is n, and this is obtained precisely when the minimal polynomial and characteristic polynomial of A coincide.*

The idea of proof is interpreting C_A as an $\mathbb{F}[x]$-endomorphism algebra of the $\mathbb{F}[x]$-module M^A (as described above). We use the Rational Canonical Form-Primary decomposition $M^A = \bigoplus_p \oplus_i \mathbb{F}[x]/(p^{\lambda_{p,i}})$. Then C_A can be written as

$$C_A \simeq \text{ End}_{\mathbb{F}[x]}M^A \simeq \bigoplus_p \oplus_{i,j} \text{Hom}_{\mathbb{F}[x]}(\mathbb{F}[x]/(p^{\lambda_{p,i}}), \mathbb{F}[x]/(p^{\lambda_{p,j}})).$$

where the first sum is over all irreducible polynomials p that divides the characteristic polynomial of A, and the indices i, j of second double sum is from the partition $\lambda_p = \sum_i \lambda_{p,i}$ that indicates the powers of p in p-primary part of M^A. We then have the formula for $\dim_{\mathbb{F}} C_A$

$$\dim_{\mathbb{F}} C_A = \dim_{\mathbb{F}} \text{ End}_{\mathbb{F}[x]}M^A = \sum_p (\deg p) \sum_{i,j} \min\{\lambda_{p,i}, \lambda_{p,j}\}. \tag{9}$$

The result we have on $\dim_{\mathbb{F}} C_A$ is a special case of Cecioni-Frobenius Theorem.

Theorem 2.7 (Cecioni-Frobenius)**.** *Let A be a $m \times m$ matrix, B be a $n \times n$ matrix over $\mathbb{F}$. Denote by $\nu_{A,B}$ the dimension of $C_{A,B}$ over $\mathbb{F}$ where*

$$C_{A,B} = \{X \in M_{m\times n}(\mathbb{F}) \mid AX - XB = 0\}.$$

Then we have $\nu_{A,B} = \sum_p (\deg p) \sum_{i,j} \min\{\lambda_{p,i}, \mu_{p,j}\}$.

Here the first sum is over all irreducible polynomials p which are common in the primary decompositions of M^A and M^B, and the indices i, j of second double sum is from the partition $\lambda_p = \sum_i \lambda_{p,i}$ that indicates the powers of p in p-primary part of M^A, $\mu_p = \sum_j \mu_{p,j}$ that of powers of p in p-primary part of M^B.

Proof. Note that $C_{A,B} = \text{Hom}_{\mathbb{F}[x]}(M^B, M^A)$. Then

$$\nu_{A,B} = \dim_{\mathbb{F}} C_{A,B} = \dim_{\mathbb{F}} \text{ Hom}_{\mathbb{F}[x]}(M^B, M^A) = \sum_p (\deg p) \sum_{i,j} \min\{\lambda_{p,i}, \mu_{p,j}\}.$$

□

An obvious application of Cecioni-Frobenius Theorem is

Theorem 2.8 (Sylvester Equation)**.** *Let A be a $m \times m$ matrix, B be a $n \times n$ matrix, and C be a $m \times n$ matrix over $\mathbb{F}$. Consider a matrix equation $AX - XB = C$. Then*

- *The matrix equation $AX - XB = C$ has a unique solution if and only if primary decompositions of M^A and M^B have no common irreducible polynomial.*
- *In case the equation does not have a unique solution, we have*

$$\nu_{A,B} = dim_{\mathbb{F}} C_{A,B} = dim_{\mathbb{F}}\{X \in M_{m\times n}(\mathbb{F}) \mid AX - XB = 0\} > 0.$$

We present a method for determining similarity of two matrices without module theory.

Definition 2.1 (Kronecker Product)**.** *Let $A \in M_m(\mathbb{F})$, $B \in M_n(\mathbb{F})$. The Kronecker Product of the matrices A and B is defined by*

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mm}B \end{pmatrix}.$$

Let $C \in M_{m\times n}(\mathbb{F})$. The vectorization of matrix C is written as $vec(C)$. This is a column vector in $\mathbb{F}^{mn}$ composed of columns of C.

Lemma 2.1. *Let $A \in M_m(\mathbb{F})$, $B \in M_n(\mathbb{F})$, and $X \in M_{m\times n}(\mathbb{F})$. Then*

$$vec(AXB) = (B^T \otimes A)vec(X).$$

By this lemma, Sylvester equation $AX - XB = C$ can be written as a linear system of mn variables.

$$(I_n \otimes A - B^T \otimes I_m)\mathrm{vec}(X) = \mathrm{vec}(C).$$

Then we have $\nu_{A,B} = \mathrm{Null}(I_n \otimes A - B^T \otimes I_m)$.

Theorem 2.9 (Byrnes-Gauger)**.** *Let $A \in M_m(\mathbb{F})$, $B \in M_n(\mathbb{F})$. Then we have*

$$\nu_{A,A} + \nu_{B,B} - 2\nu_{A,B} \geq 0.$$

The equality holds if and only if $m = n$ and the matrices A, B are similar.

Proof. This is a consequence of the following combinatorial inequality.

Let $m_1 \leq m_2 \leq \cdots \leq m_r$, $n_1 \leq n_2 \leq \cdots \leq n_s$ be integers. Then

$$\sum_{i,j} (\min(m_i, m_j) + \min(n_i, n_j) - 2\min(m_i, n_j)) \geq 0$$

with equality holds if and only if (m_i) and (n_j) are identical.

This inequality follows from considering the diagonal and off-diagonal pairs.

Case 1: $i = j$

We have $m_i + n_i - 2\min(m_i, n_i) \geq 0$ with equality holds if and only if $m_i = n_i$ for all i.

Case 2: The off-diagonal pairs (i, j) and (j, i) where $i < j$.

Subcase 2-1: The intervals $[m_i, m_j]$ and $[n_i, n_j]$ do not overlap.

Without loss of generality, assume $m_j < n_i$.

We have $m_i + n_i - 2m_i + m_i + n_i - 2m_j = 2n_i - 2m_j > 0$.

Subcase 2-2: The intervals $[m_i, m_j]$ and $[n_i, n_j]$ overlap.

Without loss of generality, assume $m_i \leq n_i \leq m_j$.

We have $m_i + n_i - 2m_i + m_i + n_i - 2n_i = 0$.

Thus, the off-diagonal pairs' contribution are always nonnegative. Considering diagonal contributions, it is easy to see that the equality holds if and only if the sequences are identical. □

The following criteria for similarity can be checked through elementary row operations on $n^2 \times n^2$ matrices.

Corollary 2.4. *Let $A, B \in M_n(\mathbb{F})$. Then A and B are similar if and only if $\nu_{A,A} = \nu_{B,B} = \nu_{A,B}$. That is,*

$$\mathrm{Null}(I_n \otimes A - A^T \otimes I_n) = \mathrm{Null}(I_n \otimes B - B^T \otimes I_n) = \mathrm{Null}(I_n \otimes A - B^T \otimes I_n).$$

Since A^T and A are similar, we also have by rank-nullity theorem,

Corollary 2.5. *Let $A, B \in M_n(\mathbb{F})$. Then A and B are similar if and only if*

$$\mathrm{rk}(I_n \otimes A - A \otimes I_n) = \mathrm{rk}(I_n \otimes B - B \otimes I_n) = \mathrm{rk}(I_n \otimes A - B \otimes I_n).$$

Theorem 2.10 (Symmetric Similarity Transform, [11])**.** *Let A be a $n \times n$ matrix over $\mathbb{F}$. Suppose also that the minimal polynomial and characteristic polynomial of A coincide. Then any invertible matrix X satisfying $XA = A^T X$ is symmetric.*

Proof. Consider the following system (Σ_A) of matrix equations.

$$XA = A^T X, \tag{10}$$

$$X = X^T. \tag{11}$$

Note that the below system is equivalent to (Σ_A).

$$XA = A^T X^T, \tag{12}$$

$$X = X^T. \tag{13}$$

The linear transform $X \mapsto (XA - A^T X^T,\ \ X - X^T)$ has rank at most $n^2 - n$. Thus, the solution space of the system (Σ_A) has dimension at least n.
Now, fix a non-singular transform X_0 such that $X_0 A = A^T X_0$. Then

$$XA = A^T X \ \text{ if and only if } \ X_0^{-1} XA = AX_0^{-1} X.$$

This yields an isomorphism $X \mapsto X_0^{-1} X$ between $\{X \mid XA = A^T X\}$ and $C_A = \{X' \mid X'A = AX'\}$. Since $\dim_{\mathbb{F}} C_A = n$, the solution space for (10) has dimension n. Since the solution space for (Σ_A) has dimension $\geq n$, the dimension must be exactly n. Hence, every matrix X satisfying (10) must also satisfy (11). □

Note that by Cecioni-Frobenius, it is clear that $C_{A^T,A}$ and C_A has the same dimension. There is no need for constructing isomorphism.

For the companion matrix, we have an explicit similarity transform (see [2, Proposition 5.4] and [8]).

Example 2.3. (Similarity Transform for Companion Matrices)

Let $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{F}[x]$ and

$$C(p) = \begin{pmatrix} 0 & 0 & \cdots & -a_0 \\ 1 & 0 & \cdots & -a_1 \\ 0 & 1 & \cdots & -a_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -a_{n-1} \end{pmatrix}$$

be the companion matrix. Then we have

$$C(p)Y = YC(p)^T = \begin{pmatrix} -a_0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & a_2 & a_3 & \cdots & a_{n-1} & 1 \\ 0 & a_3 & & \iddots & 1 & \\ \vdots & \vdots & \iddots & \iddots & & \\ \vdots & a_{n-1} & 1 & & & \\ 0 & 1 & & & & \end{pmatrix}$$

where

$$Y = \begin{pmatrix} a_1 & a_2 & \cdots & a_{n-2} & a_{n-1} & 1 \\ a_2 & a_3 & \cdots & a_{n-1} & 1 & \\ a_3 & a_4 & \cdots & 1 & & \\ \vdots & \vdots & \iddots & & & \\ a_{n-1} & 1 & & & & \\ 1 & & & & & \end{pmatrix}.$$

Theorem 2.11 (Symmetric Similarity Transform 2)**.** *Let A be a $n \times n$ matrix over a field $\mathbb{F}$. Then there exists an invertible symmetric matrix X such that $XA = A^T X$.*

Proof. Let X_0 be an invertible matrix over $\mathbb{F}$ such that $A = X_0^{-1} J X_0$ with J is consisted of blocks on diagonal, each block is a companion matrix of p^s for some irreducible polynomial p and $s \geq 1$. Say, $J = \text{diag}\{C_1, \ldots, C_r\}$. Now, we may have distinct blocks in J corresponding to the same irreducible polynomial. Consider

$$XA = A^T X \iff X_0^{-T} X X_0^{-1} J = J^T X_0^{-T} X X_0^{-1},$$

where I used the notation $X_0^{-T} = (X_0^{-1})^T$.

By Theorem 2.10 or Example 2.3, we can find a symmetric invertible matrix corresponding to each companion matrix in J. Put $X_0^{-T} X X_0^{-1} = \text{diag}\{Y_1, \ldots, Y_r\}$ such that $Y_i C_i = C_i^T Y_i$ and Y_i is symmetric invertible and of the same size with C_i for each $1 \leq i \leq r$.

Then we can put a symmetric invertible matrix

$$X = X_0^T \text{diag}\{Y_1, \ldots, Y_r\} X_0,$$

which satisfies $XA = A^T X$. □

Note that an invertible X with $XA = A^TX$ may not be symmetric if the characteristic polynomial and minimal polynomial of A do not coincide.

Theorem 2.12 (Double Commutant Theorem, [5])**.** *Let A, B be $n \times n$ matrix over a field $\mathbb{F}$ such that any matrix that commutes with A also commutes with B. Then $B = p(A)$ for some $p \in \mathbb{F}[x]$.*

Proof. We use rational canonical form-invariant factor form (Theorem 2.1). Then we have

$$M^A \simeq \mathbb{F}[x]/P_1 \oplus \cdots \oplus \mathbb{F}[x]/P_r,$$

where $P_i = (p_i)$, $p_i | p_{i+1}$. This gives invariant subspace decomposition,

$$M^A = \bigoplus_{i=1}^{r} M_i,$$

where $M_i \simeq \mathbb{F}[x]/P_i$.

Let $\pi_i : M^A \longrightarrow M_i$ be the projection, and $\pi_{ij} : M_i \longrightarrow M_j$ be the natural projection for $i > j$. Extend π_{ij} linearly to M^A by assigning 0 on all $M_k (k \neq i)$. Then all π_i and π_{ij} commute with A, thus commute with B. Therefore, each M_i is A-invariant, thus it is also B-invariant. Let $e_i \in M_i$ be the element corresponding to $1 + P_i \in \mathbb{F}[x]/P_i$.

We see that there is $p(x) \in \mathbb{F}[x]$ such that $Be_r = p(A)e_r$. We claim that $Be_i = p(A)e_i$ for all $i < r$, and hence $B = p(A)$.

$$Be_i = B\pi_{ri}e_r = \pi_{ri}Be_r = \pi_{ri}p(A)e_r = p(A)\pi_{ri}e_r = p(A)e_i.$$

This completes the proof of our claim. □

Example 2.4. (Variation of Parameters and Inverse of Confluent Vandermonde Matrix)

Let $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = \prod_{k=1}^{r}(x - \lambda_k)^{m_k} \in \mathbb{C}[x]$ with $\lambda_i \neq \lambda_j$ for $i \neq j$ and $A = C(p)^T$. Consider the differential equation (*1)

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1y' + a_0y = b(x).$$

We may convert this to a system of differential equations (*2)

$$\mathbf{y}' = A\mathbf{y} + \mathbf{b}(x)$$

where $\mathbf{y} = (y \ y' \ \cdots \ y^{(n-1)})^T$ and $\mathbf{b}(x) = (0 \ \cdots \ 0 \ b(x))^T = b(x)\mathbf{e_n}$.

We have $A = VJV^{-1}$ where $J = diag(J_1, \ldots, J_r)$ is Jordan form with $J_k = J^T(\lambda_k, m_k)$ having 1 above the main diagonal, and V is the confluent Vandermonde matrix $V = (V_1 \ \cdots \ V_r)$ with V_k is $n \times m_k$ matrix with entries

$$(V_k)_{ij} = \begin{cases} \binom{i-1}{j-1}\lambda_k^{i-j} & \text{if } i \geq j \\ 0 & \text{otherwise.} \end{cases}$$

Apply the matrix exponential to $A = VJV^{-1}$, then we have $e^{tA} = Ve^{tJ}V^{-1}$. Thus, $e^{tA}V = Ve^{tJ}$. The first row of Ve^{tJ} is

$$\begin{aligned} &(y_1, \ \cdots \ , y_n) \\ &= \left(e^{t\lambda_1}, \ e^{t\lambda_1}t, \ \cdots \ , e^{t\lambda_1}\frac{t^{m_1-1}}{(m_1-1)!}, \cdots , e^{t\lambda_r} \ , e^{t\lambda_r}t \ \cdots \ , e^{t\lambda_r}\frac{t^{m_r-1}}{(m_r-1)!}\right). \end{aligned}$$

Let $W(t) = e^{tA}V$. Then by $W'(t) = AW(t)$, $W(t)$ is the Wronskian matrix for the linearly independent solutions $y_1, \ldots, y_n$ of the homogeneous part of (*1). We apply the Variation of Parameters to (*2), then a particular solution has the form

$$\mathbf{y}(t) = W(t)\int_0^t W^{-1}(u)\mathbf{b}(u)du = W(t)\int_0^t b(u)e^{-uJ}V^{-1}\mathbf{e_n}\ du.$$

To complete the Variation of Parameters to (*1), we take the first row $(y_1 \ \cdots \ y_n)$ of $W(t)$ so that a particular solution has the form

$$(y_1 \ \cdots \ y_n)\int_0^t b(u)e^{-uJ}V^{-1}\mathbf{e_n}\ du.$$

Here, $V^{-1}\mathbf{e_n}$ is the last column of V^{-1}. This (see [4]) is obtained by the partial fraction

$$\frac{1}{p(x)} = \sum_{k=1}^{r}\sum_{j=1}^{m_k} \frac{K_{k,j}}{(x-\lambda_k)^j}.$$

Then $V^{-1}\mathbf{e_n} = (K_{1,1}\cdots K_{1,m_1}\ \cdots\ K_{r,1}\cdots K_{r,m_r})^T$. Note that $K_{k,j} = K_{k,j}^{(0)}$ can be computed from the power series expansion of $(x-\lambda_k)^{m_k}/p(x)$ at $x = \lambda_k$.

The column $n-j$ of V^{-1} is obtained from the column $n-j+1$ recursively. For each $k \le r$, obtain $(K_{k,1}^{(j)}, \cdots, K_{k,m_k}^{(j)})$ from $(K_{k,1}^{(j-1)}, \cdots, K_{k,m_k}^{(j-1)})$ by the following algorithm (see [4]):

- Apply the left shift to obtain $(K_{k,2}^{(j-1)}, \cdots, K_{k,m_k}^{(j-1)},\ 0)$,
- Add λ_k multiple of $(K_{k,1}^{(j-1)}, \cdots, K_{k,m_k}^{(j-1)})$,
- Add a_{n-j} multiple of $(K_{k,1}^{(0)}, \cdots, K_{k,m_k}^{(0)})$.

Then we have

$$\begin{aligned}\begin{pmatrix} K_{k,1}^{(j)} \\ \vdots \\ K_{k,m_k-1}^{(j)} \\ K_{k,m_k}^{(j)} \end{pmatrix} &= \begin{pmatrix} K_{k,2}^{(j-1)} \\ \vdots \\ K_{k,m_k}^{(j-1)} \\ 0 \end{pmatrix} + \lambda_k \begin{pmatrix} K_{k,1}^{(j-1)} \\ \vdots \\ K_{k,m_k-1}^{(j-1)} \\ K_{k,m_k}^{(j-1)} \end{pmatrix} + a_{n-j} \begin{pmatrix} K_{k,1}^{(0)} \\ \vdots \\ K_{k,m_k-1}^{(0)} \\ K_{k,m_k}^{(0)} \end{pmatrix} \\ &= \begin{pmatrix} K_{k,2}^{(j-1)} + \lambda_k K_{k,1}^{(j-1)} + a_{n-j}K_{k,1}^{(0)} \\ \vdots \\ K_{k,m_k}^{(j-1)} + \lambda_k K_{k,m_k-1}^{(j-1)} + a_{n-j}K_{k,m_k-1}^{(0)} \\ \lambda_k K_{k,m_k}^{(j-1)} + a_{n-j}K_{k,m_k}^{(0)} \end{pmatrix}.\end{aligned}$$

To see why this algorithm works, apply Laplace transform to the homogeneous part of the equation (*1). Let $y = c_{1,j}y_1 + \cdots + c_{n,j}y_n$ and $Y = \mathcal{L}(y)$. Since $\mathcal{L}(y^{(m)}) = s^mY - s^{m-1}y(0) - \cdots - y^{(m-1)}(0)$, we have

$$\begin{aligned} p(s)Y &- (s^{n-1} + a_{n-1}s^{n-2} + \cdots\cdots + a_1)y(0) \\ &- (s^{n-2} + a_{n-1}s^{n-3} + \cdots + a_2)y'(0) \\ &- \ \cdots\ \cdots\ \cdots \\ &- (s + a_{n-1})y^{(n-2)}(0) \\ &- y^{(n-1)}(0) = 0. \end{aligned}$$

If $(y(0)\ y'(0)\ \cdots\ y^{(n-1)}(0))^T = \mathbf{e}_j$, we have

$$Y = \frac{s^{n-j} + a_{n-1}s^{n-j-1} + \cdots + a_j}{p(s)}.$$

Thus, $\mathcal{L}(y) = c_{1,j}\mathcal{L}(y_1) + \cdots + c_{n,j}\mathcal{L}(y_n)$ must be the partial fraction decomposition of $\frac{s^{n-j}+a_{n-1}s^{n-j-1}+\cdots+a_j}{p(s)}$. As $W(0) = V$ and $V(c_{1,j}\ \cdots\ c_{n,j})^T = \mathbf{e}_j$, we obtain that $(c_{1,j}\ \cdots\ c_{n,j})^T$ is j-th column of V^{-1}.

For the recursive relation between consecutive columns, observe that for any root λ of $p(s) = 0$,

$$\frac{(s-\lambda+\lambda)(s^{n-j} + a_{n-1}s^{n-j-1} + \cdots + a_j) + a_{j-1}}{p(s)} = \frac{s^{n-j+1} + a_{n-1}s^{n-j} + \cdots + a_js + a_{j-1}}{p(s)}.$$

If λ-part of the partial fraction of $\frac{s^{n-j}+a_{n-1}s^{n-j-1}+\cdots+a_j}{p(s)}$ is

$$\frac{B_1}{s-\lambda} + \frac{B_2}{(s-\lambda)^2} + \cdots + \frac{B_m}{(s-\lambda)^m},$$

then λ-part of $\frac{s^{n-j+1}+a_{n-1}s^{n-j}+\cdots+a_js+a_{j-1}}{p(s)}$ is

$$B_1+\frac{B_2+\lambda B_1}{s-\lambda}+\frac{B_3+\lambda B_2}{(s-\lambda)^2}+\cdots+\frac{B_m+\lambda B_{m-1}}{(s-\lambda)^{m-1}}+\frac{\lambda B_m}{(s-\lambda)^m}$$
$$+\frac{a_{j-1}A_1}{s-\lambda}+\cdots+\frac{a_{j-1}A_m}{(s-\lambda)^m},$$

where $\frac{A_1}{s-\lambda}+\cdots+\frac{A_m}{(s-\lambda)^m}$ is λ-part of $1/p(s)$. The sum of the constant term B_1's over all λ's is zero if $j\geq 2$. For, observe that it is the sum of all residues of $\frac{s^{n-j}+a_{n-1}s^{n-j-1}+\cdots+a_j}{p(s)}$. Thus, its vanishing is shown by the following limit of the integral over the circle C_R of radius R centered at 0,

$$\lim_{R\to\infty}\int_{C_R}\frac{s^{n-j}+a_{n-1}s^{n-j-1}+\cdots+a_j}{p(s)}ds=0$$

Example 2.5. (Powers of a Matrix) Let $m_A(x)=x^m+a_{m-1}x^{m-1}+\cdots+a_1x+a_0=\prod_{k=1}^r(x-\lambda_k)^{m_k}\in\mathbb{C}[x]$ with $\lambda_i\neq\lambda_j$ for $i\neq j$, be the minimal polynomial of $A\in M_N(\mathbb{C})$. Then for any $n\geq m$, we write the powers of A symbolically as follows:

$$\begin{pmatrix}A^{n-m+1}\\A^{n-m+2}\\\vdots\\A^n\end{pmatrix}=M_A\begin{pmatrix}A^{n-m}\\A^{n-m+1}\\\vdots\\A^{n-1}\end{pmatrix}=M_A^{n-m+1}\begin{pmatrix}I\\A\\\vdots\\A^{m-1}\end{pmatrix}$$

where

$$M_A=C(m_A)^T=\begin{pmatrix}0&1&\cdots&\cdots&0\\0&0&1&\cdots&0\\\vdots&\vdots&\ddots&\ddots&\vdots\\0&0&\cdots&\cdots&1\\-a_0&-a_1&\cdots&\cdots&-a_{m-1}\end{pmatrix}.$$

By $M_A=VJV^{-1}$ where $J=diag(J_1,\ldots,J_r)$ is Jordan form with $J_k=J^T(\lambda_k,m_k)$ having 1 above the main diagonal, and V is the confluent Vandermonde matrix $V=(V_1\ \cdots\ V_r)$ with V_k is $m\times m_k$ matrix with entries

$$(V_k)_{ij}=\begin{cases}\binom{i-1}{j-1}\lambda_k^{i-j} & \text{if } i\geq j\\ 0 & \text{otherwise.}\end{cases}$$

We compute V^{-1} by the method of the previous example. The required powers of M_A are computed by $M_A^{n-m+1}=VJ^{n-m+1}V^{-1}$. Let $(c_{m,1}^{(n)},\cdots,c_{m,m}^{(n)})$ be the last row of M_A^{n-m+1}. Then

$$A^n=c_{m,1}^{(n)}I+c_{m,2}^{(n)}A+\cdots+c_{m,m}^{(n)}A^{m-1}.$$

Thus, each entry of A^n is a linear combination of $\lambda_k^n q_k(n)$ with $q_k\in\mathbb{C}[x]$ and $\deg q_k(n)\leq m_k-1$.

Example 2.6. (Stochastic Matrix-Perron-Frobenius Theorem) Let $A\in M_n(\mathbb{R})$ with nonnegative entries. A is called a *stochastic matrix* if each column sums to 1. Then each row of A^T sums to 1. The following properties hold for A^T.

- 1 is an eigenvalue of A^T with an eigenvector $\mathbf{1}=(1\ 1\cdots\ 1)^T$.
- For each $n\in\mathbb{N}$, $(A^T)^n$ is has each row sum to 1.

Proof. This follows from $(A^T)^n\mathbf{1}=\mathbf{1}$. □

- If $\lambda\in\mathbb{C}$ is any eigenvalue of A^T, then $|\lambda|\leq 1$.

Proof. To see this, let $x = (x_1 \ \cdots x_n)^T \in \mathbb{C}^n$ be an eigenvector associated to an eigenvalue λ. Let x_i be the entry of x with the largest absolute value. Dividing by x_i, we may assume that $x_i = 1$ and $|x_j| \leq 1$ for each j. Then i-th entry of $Ax = \lambda x$ gives (*)

$$a_{11}x_1 + \cdots + a_{i1} + \cdots + a_{in}x_n = \lambda$$

Then by Triangle Inequality and $|x_j| \leq 1$, it follows that

$$|\lambda| = |a_{11}x_1 + \cdots + a_{i1} + \cdots + a_{in}x_n| \leq a_{11} + \cdots + a_{in} = 1.$$

□

The above proof is in fact similar to that of Gershgorin Theorem (Theorem 3.5).

- Any Jordan block $J^T(1, m)$ of A^T corresponding to the eigenvalue 1, must satisfy $m = 1$.

Proof. Let X be an invertible matrix with $A^T = XJX^{-1}$ so that J is the Jordan form of A^T. Observe that $(A^T)^n$ remains bounded as $n \in \mathbb{N}$ varies. Then $X^{-1}(A^T)^n X = J^n$ is also bounded. If J has a block $J^T(1, m)$ with $m \geq 2$, then J^n has polynomial entries in n and J^n fails to be bounded. □

A stochastic matrix is said to be *positive* if all entries are positive. If $A \in M_n(\mathbb{R})$ is positive stochastic, then the following properties hold for A^T.

- Any eigenvalue λ of A^T with $|\lambda| = 1$ must be $\lambda = 1$. The eigenspace corresponding to the eigenvalue 1 is 1-dimensional, hence it is spanned by **1**.

Proof. If $|\lambda| = 1$, then inequality below must be an equality,

$$|\lambda| = |a_{i1}x_1 + \cdots + a_{i1} + \cdots + a_{in}x_n| \leq a_{i1} + \cdots + a_{in} = 1.$$

Thus, positivity of A shows that $|x_j| = 1$ for each j. Moreover, $|a_{i1}x_1 + \cdots + a_{i1} + \cdots + a_{in}x_n| = 1$ implies that $x_j = 1$ for each j. □

Since A and A^T are similar, they have equivalent Jordan forms.

- $\lim_{n\to\infty} A^n = (b \ \cdots \ b)$ where entries of $b \in \mathbb{R}^n$ are positive and sum to 1.

Proof. Let v be a real eigenvector of A corresponding to the eigenvalue 1. Let Y be an invertible matrix such that $A = YJY^{-1}$ with v the first column and J the Jordan form of A. Since any eigenvalue $\lambda \neq 1$ of A^T has $|\lambda| < 1$ and the eigenvalue 1 has 1-dimensional eigenspace, A also satisfies those properties. Then we see that J^n converges as $n \to \infty$. In fact,

$$\lim_{n\to\infty} J^n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Consequently,

$$\lim_{n\to\infty} A^n = Y \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} Y^{-1} = \begin{pmatrix} | & 0 & \cdots & 0 \\ v & \vdots & \ddots & \vdots \\ | & 0 & \cdots & 0 \end{pmatrix} Y^{-1} = vw^T$$

□

where $w^T \in M_{1\times n}(\mathbb{C})$ is the first row of Y^{-1}. Since A is stochastic, so is vw^T. Thus, there is $b \in \mathbb{R}^n$ with nonnegative entries sum to 1 such that all columns of vw^T are b. The vector b with nonnegative entries must in fact have positive entries due to $b = Ab$.

3. Spectral Theory

Definition 3.1. We say that a $n \times n$ complex matrix A is

- *hermitian* if $\overline{A}^T = A$,
- *skew-hermitian* if $\overline{A}^T = -A$,
- *unitary* if $A\overline{A}^T = I$,
- *normal* if $A\overline{A}^T = \overline{A}^T A$.

For simplicity, we write $\overline{A}^T = A^*$. Note that hermitian, skew-hermitian, and unitary matrices are normal. The eigenvalues of hermitian matrices are all real, those of skew-hermitian matrices are all pure imaginary, and those of unitary matrices are all on the unit circle.

Definition 3.2. *For any vectors $v, w \in \mathbb{C}^n$, the hermitian product of v and w is denoted by $\langle v, w\rangle$ and defined by*

$$\langle v, w\rangle = v^* w.$$

Theorem 3.1 (Common Eigenvector)**.** *Let A, B be $n \times n$ complex matrices such that $AB = BA$. Then there is a nonzero vector $v \in \mathbb{C}^n$ which is an eigenvector for both A and B.*

Proof. Let $\lambda \in \mathbb{C}$ be an eigenvalue of a matrix A and V_λ be the corresponding eigenspace. For any $x \in V_\lambda$, we have

$$ABx = BAx = B(\lambda x) = \lambda Bx.$$

This shows that $Bx \in V_\lambda$ for any $x \in V_\lambda$. Thus, V_λ is invariant under B. Then let $v \in V_\lambda$ be an eigenvector of B restriced to V_λ. This is a common eigenvector of both A and B. □

Proposition 3.1. *Let v be an common eigenvector of A and A^* for a normal matrix A. If $\langle v, w\rangle = 0$, then $\langle v, Aw\rangle = 0$.*

Proof. Let $A^*v = \lambda v$. Then $\langle v, Aw\rangle = \langle A^*v, w\rangle = \overline{\lambda}\langle v, w\rangle = 0$. □

Theorem 3.2 (Spectral Theorem)**.** *Let A be an $n \times n$ normal matrix. Then there exists a unitary matrix U and a diagonal matrix D such that $A = UDU^*$.*

As a corollary, a hermitian matrix A is diagonalizable through a unitary matrix U and a real diagonal matrix D. Denote by

$$\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$$

the real eigenvalues of a hermitian matrix A.

Theorem 3.3 (Courant-Fischer)**.** *Let A be a hermitian matrix. Then*

$$\lambda_i(A) = \sup_{\dim V = i} \inf_{\substack{v \in V \\ |v| = 1}} v^* A v,$$

$$\lambda_i(A) = \inf_{\dim V = n-i+1} \sup_{\substack{v \in V \\ |v| = 1}} v^* A v.$$

Proof. The second identity follows from the first by taking $-A$ in the first one. To prove the first identity, assume that the standard basis $e_1, \ldots, e_n$ are the eigenvectors of A. Take $V = \operatorname{span}\{e_1, \ldots, e_i\}$. Then $\dim V = i$ and

$$\inf_{\substack{v \in V \\ |v| = 1}} v^* A v = \lambda_i(A) \leq \sup_{\dim V = i} \inf_{\substack{v \in V \\ |v| = 1}} v^* A v.$$

For the reverse inequality, take any V with $\dim V = i$ and $W = \operatorname{span}\{e_i, \ldots, e_n\}$. Then $\dim(V \cap W) \geq 1$ by the dimension identity

$$\dim(V \cap W) = \dim V + \dim W - \dim(V + W).$$

We take a unit vector $v \in V \cap W$. Then $v^* A v \leq \lambda_i(A)$ since $v \in W$. Therefore, we have

$$\inf_{\substack{v \in V \\ |v| = 1}} v^* A v \leq \lambda_i(A)$$

since $v \in V$. Taking the supremum over $\dim V = i$, we have the result. $\square$

Corollary 3.1. *If A and B are both hermitian matrices, then*

$$|\lambda_i(A+B) - \lambda_i(A)| \leq \|B\|$$

where $\|B\|$ is the operator norm of B.

Proof. We use $v^*(A+B)v = v^*Av + v^*Bv$. Let $\dim V = i$. Find a unit vector $v \in V$ such that

$$v^*Av = \inf_{\substack{v \in V \\ |v|=1}} v^*Av.$$

Since $v^*Av \leq \lambda_i(A)$ and $v^*Bv \leq \|B\|$, we have

$$\inf_{\substack{v \in V \\ |v|=1}} v^*(A+B)v \leq \lambda_i(A) + \|B\|$$

Taking the supremum over $\dim V = i$, we have

$$\lambda_i(A+B) \leq \lambda_i(A) + \|B\|.$$

Similarly,

$$\lambda_i(A) \leq \lambda_i(A - (-B)) + \|-B\|.$$

Then the result follows. $\square$

Corollary 3.2 (Weyl's Inequality)**.** *If A and B are both hermitian matrices, then*

$$\lambda_{i+j-1}(A+B) \leq \lambda_i(A) + \lambda_j(B)$$

whenever $i, j \geq 1$ and $i+j-1 \leq n$.

Proof. Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis so that $Av_k = \lambda_k(A)v_k$ for all k, and $\{w_1, \ldots, w_n\}$ be an orthonormal basis so that $Bw_k = \lambda_k(B)w_k$ for all k. Consider $V = \operatorname{span}\{v_i, \ldots, v_n\}$ and $W = \operatorname{span}\{w_j, \ldots, w_n\}$. Then $\dim V = n-i+1$, $\dim W = n-j+1$, and $v^*Av \leq \lambda_i(A)$ for all unit vector $v \in V$, $w^*Bw \leq \lambda_j(B)$ for all unit vector $w \in W$. By the dimension identity, we have

$$\dim(V \cap W) \geq (n-i+1) + (n-j+1) - n = n - (i+j-1) + 1.$$

Consider a subspace V' of $V \cap W$ with $\dim V' = n-(i+j-1)+1$. For any unit vector $v \in V'$, we have

$$v^*(A+B)v = v^*Av + v^*Bv \leq \lambda_i(A) + \lambda_j(B).$$

Then by the second Courant-Fischer identity, the result follows. $\square$

Definition 3.3. *Let $A \in M_n(\mathbb{C})$. The spectral radius of A is defined as*

$$\rho(A) = \max\{|\lambda_1|, \ldots, |\lambda_n|\}$$

where $\lambda_1, \ldots, \lambda_n$ are eigenvalues of A.

If A is hermitian, we have $\rho(A) = \max\{|\lambda_1(A)|, |\lambda_n(A)|\}$.

Proposition 3.2. *The spectral radius function is continuous. That is $\rho : M_n(\mathbb{C}) \to \mathbb{R}$ is continuous.*

The proof relies on Rouche's theorem.

Lemma 3.1 (Fekete)**.** *Let (a_n) be a subadditive real sequence, that is, $a_n \in \mathbb{R}$ for all n, and $a_{m+n} \leq a_m + a_n$ for all m, n. Then $\lim(a_n/n) \in [-\infty, \infty)$ and*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n \in \mathbb{N}} \frac{a_n}{n}.$$

Proof. For any $n \in \mathbb{N}$, we have $a_n \leq na_1$. Thus, the sequence (a_n/n) is bounded above. Let $\alpha = \limsup(a_n/n) = \lim_{k\to\infty}(a_{n_k}/n_k)$ with increasing sequence of natural numbers (n_k). Fix any $m \in \mathbb{N}$. By Euclidean Division Algorithm, we have $n_k = mq_k + r_k$ for some $q_k \in \mathbb{Z} \cap [0, \infty)$ and $0 \leq r_k < m$. Then $q_k \to \infty$ as $k \to \infty$. We have

$$\frac{a_{n_k}}{n_k} = \frac{a_{mq_k+r_k}}{mq_k + r_k} \leq \frac{q_k a_m + a_{r_k}}{mq_k + r_k}.$$

Taking $k \to \infty$, we obtain $\alpha \leq a_m/m$. Thus, taking infimum over m, we have $\alpha \leq \inf(a_m/m)$. Hence, we must have $\lim(a_n/n) = \alpha$.

□

Theorem 3.4. *Let $A \in M_n(\mathbb{C})$. The following hold*

(1) $\rho(A) < 1$ *if and only if* $\lim_{k\to\infty} A^k = 0$.

(2) *For any $k \in \mathbb{N}$,* $\rho(A) \leq \left\|A^k\right\|^{1/k}$.

(3) (Gelfand) $\rho(A) = \lim_{k\to\infty} \left\|A^k\right\|^{1/k}$.

Proof of (1). Note that $\lim_{k\to\infty} A^k = 0$ if and only if $\lim_{k\to\infty} J^k = 0$ for any Jordan block of A. Also, $\rho(A) < 1$ if and only if any eigenvalues λ of A satisfy $|\lambda| < 1$. Then (1) follows from

$$\lim_{k\to\infty} J^k = 0 \text{ if and only if } J = J(\lambda, r) \text{ with } |\lambda| < 1.$$

□

Proof of (2). Let $Av = \lambda v$ for some $\lambda \in \mathbb{C}$ and $0 \neq v \in \mathbb{C}^n$. Then for any $k \in \mathbb{N}$,

$$|\lambda^k v| = |A^k v| \leq \|A^k\||v|.$$

Then $|\lambda^k| \leq \|A^k\|$. This gives $|\lambda| \leq \|A^k\|^{1/k}$.

□

Proof of (3). We have $\|A^k A^j\| \leq \|A^k\| \cdot \|A^j\|$ for all $k, j \geq 0$. If $\|A^k\| = 0$ for some $k \in \mathbb{N}$, then we have $\|A^{k+j}\| = 0$ for all $j \in \mathbb{N}$. Thus, we have $\lim_{k\to\infty} \|A^k\|^{1/k} = 0$. Assume $\|A\|^k > 0$ for all $k \in \mathbb{N}$. Then by Fekete's lemma, the sequence $(\frac{1}{k} \log \|A^k\|)$ converges. Thus, the sequence $(\|A^k\|^{1/k})$ converges. Consider the function $f : \{z \in \mathbb{C} | |z| > \rho(A)\} \to M_n(\mathbb{C})$ defined by $f(z) = (zI - A)^{-1}$. This is a vector-valued analytic function Laurant series is

$$\frac{1}{z}\frac{1}{1 - (A/z)} = \frac{1}{z}\sum_{k=0}^{\infty} \frac{A^k}{z^k}.$$

The series cannot converge for any $|z| < \rho(A)$, so its radius of convergence as a power series of $w = 1/z$ is $\rho(A)^{-1}$. By Cauchy-Hadamard's radius of convergence formula, we must have

$$\rho(A) = \limsup_{k\to\infty} \|A^k\|^{1/k} = \lim_{k\to\infty} \|A^k\|^{1/k}.$$

□

Note that this proof works when A is a bounded linear operator on a Banach space (see [7]).

The locations of eigenvalues depending on the entries of matrices can be found in Gershgorin's disk theorem.

Theorem 3.5 (Gershgorin's Disk Theorem)**.** *Let $A \in M_n(\mathbb{C})$. For each $i = 1, \ldots, n$, let $R_i = \sum_{j\neq i} |a_{ij}|$. Then the set of eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ of A satisfies*

$$\{\lambda_1, \ldots, \lambda_n\} \subseteq \cup_{i=1}^n D(a_{ii}, R_i).$$

Proof. Let λ be an eigenvalue of A. Choose an eigenvector x with $|x_i| = \max_{1\leq j\leq n} |x_j|$. Then by dividing x_i, we may impose $|x_j| \leq 1$ for all $j = 1, \ldots, n$, and $x_i = 1$. Since $Ax = \lambda x$, we have

$$\lambda = \lambda x_i = \sum_j a_{ij} x_j = a_{ii} + \sum_{j\neq i} a_{ij} x_j.$$

By triangle inequality, it follows that

$$|\lambda - a_{ii}| \leq R_i.$$

Thus, we have the result. □

The eigenvalue inequalities can be improved in case the matrix A is hermitian.

Theorem 3.6. *Let $A \in M_n(\mathbb{C})$ be hermitian. Let c_j be the j-th column of A with the j-th component is set to zero. Then for each $j = 1, \dots n$, there is an eigenvalue λ of A such that $|\lambda - a_{jj}| \le \|c_j\|_2$ where $\|c_j\|_2$ is ℓ_2-norm of c_j.*

Proof. Let $E_j = c_j e_j^T + e_j \overline{c_j}^T$, where e_j is the standard basis vector. Then a_{jj} is an eigenvalue of $A - E_j$. Since A and $A - E_j$ are both hermitian, their is an eigenvalue λ of A satisfies

$$|\lambda - a_{jj}| \le \|E_j\| = \|c_j\|_2$$

Thus, the result follows. □

This is a special case of Bauer-Fike theorem (see [9, Theorem 5.15]).

References

[1] P. Cohn, *On the Structure of the GL_2 of a Ring*, Publications mathématiques de l'I.H.É.S., tome 30(1966), 5-53.
[2] P. Fuhrmann, *A Polynomial Approach to Linear Algebra*, 2nd edition, Springer.
[3] R. Horn, C. Johnson, *Topics in Matrix Analysis*, Cambridge University Press.
[4] S. Hou, W. Pang, *Inversion of Confluent Vandermonde Matrices*, Computers and Mathematics with Applications, 43(2002) 1539-1547.
[5] P. Lagerstrom, *A Proof of Theorem on Commutative Matrices*, Bulletin of AMS, Volume 51, 535-536
[6] K. Matthews, *MP274 Lecture Notes 1991*, www.numbertheory.org/courses/MP274/applic.pdf
[7] M. Reed, B. Simon, *Method of Modern Mathematical Physics*, Academic Press.
[8] L. Solomon, *Similarity of the Companion Matrix and its Transpose*, Linear Algebra and its Applications 302–303 (1999) 555–561
[9] E. Suli, D. Mayers, *An Introduction to Numerical Analysis*, Cambridge University Press.
[10] T. Tao, *Topics in Random Matrix Theory*, AMS, GSM Volume 132
[11] O. Taussky, H. Zassenhaus, *On the Similarity Transform between a Matrix and its Transpose*, Pacific J. Math. Volume 9, Number 3 (1959), 893-896.