Chapter 2

November 15, 2023

1 Direct Sum

Definition *Let V be a vector space and let M_1, M_2, ..., M_k are subspaces of V. The **sum** of these subspaces M is defined as*

$$M = \{m \in V : \quad m = m_1 + m_2 + \ldots + m_k \quad \textit{where } m_i \in M_i, \quad i = 1, 2, \ldots, k\}.$$

Theorem *The sum of subspaces is also a subspace of V.*

Proof:

Definition *Let M_1, M_2, …, M_k be subspaces of a vector space V. These subspaces are said to be* ***linearly independent*** *if,*

$$m_1 + m_2 + \ldots + m_k = 0 \quad \textit{implies}$$

$$m_1 = m_2 = \ldots = m_k = 0, \quad \textit{where } m_i \in M_i \quad \textit{for } i = 1, 2, \ldots, k.$$

Definition *Let M_1, M_2, …, M_k be subspaces of a vector space and also let*

- $M = M_1 + M_2 + \ldots + M_k$
- *M_1, M_2, …, M_k are linearly independent*

Then M is said to be the ***direct sum*** *of subspaces M_1, M_2, …, M_k and denoted by $M = M_1 \oplus M_2 \oplus \ldots \oplus M_k$*

Example: Let V=$\mathbb{R}^4$, $x = [x_1, x_2, x_3, x_4]^T \in \mathbb{R}^4$.

Definition *If $M = V$ (the linear space itself) then $V = M_1 \oplus M_2 \oplus \ldots \oplus M_k$ is called the direct sum decomposition of V.*

Remark: Let $V = M_1 \oplus M_2 \oplus \ldots \oplus M_k$, and let $\mathbf{x} \in V$. Then the decomposition of $\mathbf{x}$ over M_i's is **unique**.

Remark:

Definition *Let V be an inner product space. Two subspaces M_1 and M_2 are said to be* ***orthogonal*** *if,*

$$\langle m_1, m_2 \rangle = 0 \quad \forall m_1 \in M_1, m_2 \in M_2.$$

Orthogonality is denoted as $M_1 \perp M_2$

Definition *Let $M = M_1 \oplus M_2 \oplus \ldots \oplus M_k$ and let $M_i \perp M_j$ for all $i \neq j$. Then M is said to be* ***orthogonal direct sum*** *of subspaces $M_1, M_2, \ldots, M_k$.*

Symbolically, we write $M = M_1 \overset{\perp}{\bigoplus} M_2 \overset{\perp}{\bigoplus} \ldots \overset{\perp}{\bigoplus} M_k$

Definition *Let M be a subspace of an inner product space V. The* ***orthogonal complement*** *$M^{\perp}$ of the subspace M is defined as*

$$M^{\perp} := \{x \in V : \langle x, m \rangle = 0 \quad \forall m \in M\}.$$

Theorem *$M^\perp$ is itself a subspace.*

Proof:

Example: $V = \mathbb{R}^3$, $M = Span(\left\{\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\})$, $\quad M^\perp = ?$

Theorem *Let V be an inner product space and M is a subspace of V. V can always be written as the direct sum of a subspace and its orthogonal complement, i.e., we always have* $\quad V = M \oplus M^\perp$.

Proof:

2 Projection Theorem

Theorem *"Projection Theorem"*

Let H be a Hilbert space (inner product space, complete w.r.t the norm induced by the inner product) and let M be a finite dimensional subspace of H. For any $x \in H$, the following minimization problem has a solution.

$$\min_{m \in M} \|x - m\|$$

(i.e., we can find the closest vector to x lying in the subspace M).

Proof:

Remark: $m^* = x_1$ can be interpreted as the "best approximation" of x chosen from the vectors in M. Vector x_2 can be interpreted as the "error in the approximation". This error must be orthogonal to the subspace.

Example:

Example:

Example:

Example: Suppose we are given a basis for M. That is,

$M = Span(\{v_1, v_2, \ldots v_k\})$. Given $x \in H \supset M$, we want to figure out $x_1 \in M$, where $x = x_1 + x_2$ for $x_2 \in M^\perp$.

Example: Let H be the space of square integrable functions with domain $[-\pi, \pi]$ with inner product $\langle f_1, f_2 \rangle = \int_{-\pi}^{\pi} f_1(t)\overline{f_2(t)}dt$. Let M be the subspace $M = Span\{\frac{e^{jkt}}{\sqrt{2\pi}}\}_{k=-N}^{N}$. Note that dimension of M is $2N+1$ and the basis set is orthonormal.

$$\langle f_n, f_m \rangle =$$

Now, let $g \in H$ be an arbitrary vector (a function). Then $g = g_1 + g_2$, where $g_1 \in M$ and $g_2 \in M^{\perp}$.

Note that, $g_1(t)$ is the **best approximation** to $g(t)$ within the subspace M. $g_1(t)$ turns out to be the finite Fourier series representation of $g(t)$. As $N \to \infty$ we obtain the **Fourier series** representation.

Application of the projection theorem in $\mathbb{C}^n$: Let $\{m_1, m_2, \ldots, m_k\}$, $k < n$ be a basis for a subspace M of $\mathbb{C}^n$. That is, $M = Span(\{m_1, m_2, \ldots, m_k\})$. Given an arbitrary vector $x \in \mathbb{C}^n$, we know that $x = x_1 + x_2$ with $x_1 \in M$, $x_2 \in M^{\perp}$. We also know that x_1 and x_2 are unique. Let $x_1 = \sum_{i=1}^{k} \alpha_i m_i$. Define matrix $B = [m_1 \ m_2 \ \ldots m_k]$ whose columns are basis vectors. Then we can write $x_1 = B\alpha$ for $\alpha = [\alpha_1, \ldots, \alpha_k]^T$.

Remark: In $\mathbb{C}^n$ the standard inner product is $\langle x, y \rangle = y^* x$. In $\mathbb{R}^n$, this boils down to $\langle x, y \rangle = y^T x$

<u>Remark:</u> An orthogonal projection matrix $P \in \mathbb{C}^{k \times k}$ satisfies:

- $P^* = P$
- $P^2 = P \Rightarrow P^i = P$ for all $i = 1, 2, \ldots$

Example: "Orthogonal projection" Find the orthogonal projection of the vector $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ onto the subspace spanned by $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$

3 Projection Theorem & Solution of Linear Equations

consider the linear equation expressed as

$$Ax = b \quad \text{where } A \in \mathbb{C}^{m\times n} \quad \& \quad b \in \mathbb{C}^{m\times 1} \quad \& \quad x \in \mathbb{C}^{n\times 1}.$$

Is there a solution to x? If the answer is yes, is it unique?

Remark:

- A solution exists if and only if $b \in R(A)$.
- A solution is unique if and only if $N(A) = \{0\}$.

Example: Let $A = [1\,1\,1\,1]^T$ and $b = [2.2 \quad 1.9 \quad 2.1 \quad 1.8]^T$.

When there is no exact solution, one can try to find the "‘best approximation"’ to a solution. An approximation can be found by minimizing the norm of the error,

$$\min_{x\in\mathbb{C}^n} \|Ax - b\|^2$$

If a solution exists, then $\|Ax - b\|^2 = 0$, otherwise we can find an approximate solution such that $\hat{x} = \arg\min_{x \in \mathbb{C}^n} \|Ax - b\|^2$.

Example: The length x of a metal rod is inaccurately measured four times and the results are recorded as l_1, l_2, l_3, and l_4. What is the best approximation to x?

Example: Consider the following scenario:

Remark: Suppose $\hat{x}$ is a solution of $Ax = b_1$. Suppose that m is any vector in $N(A)$, then $\hat{x} + m$ is another solution.

In the case of non-uniqueness, we are going to look for a solution with the minimum norm.

$$\min_{Ax=b_1} \|x\|$$

Let $\hat{x}^a$ and $\hat{x}^b$ be two solutions to $Ax = b_1$. We can decompose both solutions uniquely as:

Theorem $N(A)^{\perp} = R(A^*)$.

Proof:

Example: Consider the previous example:

Summary:

4 Special Cases of $Ax = b$

4.1 Columns of A form a linearly independent set

A is full-column rank.

4.2 Rows of A form a linearly independent set

A is full-row rank.

4.3 Both rows & columns of A form a linearly independent set

A is invertible.

5 Spectral Analysis of Linear Operators

Definition *Let $A : V \to V$ be a linear transformation defined over the vector space V. A subspace M of V is said to be* ***invariant*** *under A if $A(x) \in M$ for all $x \in M$.*

Example: R(A) is invariant under A.

Example: N(A) is invariant under A.

Definition *Powers of a linear operator are defined as,*

$$A^k(x) = \underbrace{A(A(\ldots A(x)\ldots))}_{A \text{ applied } k \text{ times}}$$

By using the above definition polynomials of A can be constructed as linear combinations of powers of A.

Example: $P(A) = \alpha_0 A^n + \alpha_1 A^{n-1} + \ldots + \alpha_{n-1} A + \alpha_n I$.

Exercise: Show that $R(P(A))$ and $N(P(A))$ are invariant under A.

Definition *Let A denote the matrix representation of a linear operator from V to V (A is a square matrix). The* ***eigenvalues*** *of A, denoted by λ_i, are defined as the n (n = dim V) roots of the equation det(sI-A)=0, where det(sI-A) is known as the* ***characteristic polynomial of*** *A.*

Definition *Vector(s) $e_i \in V$ satisfying $e_i \neq 0$ and $Ae_i = \lambda_i e_i$ is called the **eigenvector(s)** of A corresponding to eigenvalue λ_i.*

Example: Let $A : \mathbb{C}^n \to \mathbb{C}^n$ and λ_i be an eigenvalue of A. $N(A - \lambda_i I)$ is invariant under A.

Proof:

Theorem *Let $A \in \mathbb{C}^{n\times n}$ be the matrix representation of a linear transformation $T : \mathbb{C}^n \to \mathbb{C}^n$ with respect to the canonical basis. Suppose that,*

- $\mathbb{C}^n = M_1 \oplus M_2 \oplus \ldots \oplus M_k$
- *Each subspace M_i is invariant under T.*

Let $dim(M_i) = n_i$ and M_i has a basis set $\{b_1^i, b_2^i, \ldots, b_{n_i}^i\} =: B_i$. Then with respect to basis $\{b_1^1, b_2^1, \ldots, b_{n_1}^1;\quad b_1^2, b_2^2, \ldots, b_{n_2}^2;\quad \ldots\quad ;b_1^k, b_2^k, \ldots, b_{n_k}^k\}$, transformation T has a block *diagonal matrix representation.*

$$\bar{A} = \begin{bmatrix} \bar{A}_1 & 0 & \dots & 0 \\ 0 & \bar{A}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \bar{A}_k \end{bmatrix},$$

where $\bar{A}_i \in \mathbb{C}^{n_i \times n_i}$*. Furthermore,* $\bar{A} = B^{-1}AB$*, where* $B \in \mathbb{C}^{n \times n}$ *is given by* $B = [B_1, B_2, \dots, B_{n_i}]$

Proof:

Example: Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & -5 & -4 \end{bmatrix}$, $M_1 = Span\left\{\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}$, $M_2 = Span\left\{\begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}\right\}$

i) Is M_1 invariant under A?

ii) Is M_2 invariant under A?

iii) Find $\bar{A}$?

Let A be an $n \times n$ matrix with n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ Let $e_1, e_2, \ldots, e_n$ be the eigenvectors corresponding to these eigen values, i.e., $Ae_i = \lambda_i e_i$, $i = 1, 2, \ldots, n$.

Claim: The set of eigenvectors $\{e_1, \quad e_2, \quad \ldots, \quad e_n\}$ form a linearly independent set. Furthermore, $N(A - \lambda_i I) = Span(e_i)$ for all i.

Proof:

Theorem *"'Cayley-Hamilton"' Every $n \times n$ matrix satisfies its own characteristic equation, i.e., $d(A) = 0_{n \times n}$.*

Example:

Proof:

Proof (..continued)

Fact: Let $M \in \mathbb{C}^{n\times n}$. For each $\delta > 0$, there exists a matrix $\tilde{M} \in \mathbb{C}^{n\times n}$ with distinct eigenvalues satisfying $\left\|M - \tilde{M}\right\| < \delta$. This is equivalent to stating: "‘Matrices with distinct eigenvalues form a dense subset of $\mathbb{C}^{n\times n}$ "’.

Fact: "‘*An example*"’, Assume $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has repeated eigenvalues, i.e., $\lambda_1 = \lambda_2$. Then for each $\delta > 0$ one can find numbers $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ satisfying $|\epsilon_i| < \delta$ such that $A_\delta = \begin{bmatrix} a+\epsilon_1 & b+\epsilon_2 \\ c+\epsilon_3 & d+\epsilon_4 \end{bmatrix}$ has distinct eigenvalues, i.e., $\lambda_1^\delta \neq \lambda_2^\delta$

Let us choose a sequence $\{A_n\}_{n=1}^{\infty}$ satisfying $\|A - A_n\| < \frac{1}{n}$ and A_n has distinct eigenvalues. Let us define $d_n(s) := det(sI - A_n)$. Since $det(.)$ is a continuous function one can write $d(A) = \lim_{n\to\infty} d_n(A_n)$.

Note that $d_n(A_n) = 0$ for all n. Hence d(A)=0.

Example:

Example:

6 Minimal Polynomial

Definition *For an $n \times n$ matrix A, the* ***minimal polynomial*** *$m(s)$ is the monic polynomial with smallest degree such that $m(A) = 0_{n \times n}$*

Remark: A monic polynomial has unity as the coefficient of its highest order term.

Theorem *Given $A \in \mathbb{C}^{n \times n}$, let $m(s)$ be its minimal polynomial.*

- *$m(s)$ is unique;*
- *$m(s)$ divides $d(s)$, i.e., there exist a $q(s)$ such that $d(s) = m(s)q(s)$;*
- *Every root of $d(s)$ is also a root of $m(s)$.*

Proof:

Example: $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

Example: "‘repeated eigenvalues"’

i) $A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

ii) $A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$

Remark:

$$d(s) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2}\dots(s-\lambda_\sigma)^{r_\sigma},$$

$$m(s) = (s-\lambda_1)^{m_1}(s-\lambda_2)^{m_2}\dots(s-\lambda_\sigma)^{m_\sigma},$$

where $m_i \leq r_i$ and $i = 1, 2, \dots, \sigma$.

r_i: algebraic multiplicity of λ_i.

m_i: geometric multiplicity of λ_i.

Let $N_i := N(A - \lambda_i I)^{m_i}$. Then,

$$\mathbb{C}^n = N_1 \oplus N_2 \oplus \dots N_\sigma$$

Furthermore, $dim(N_i) = r_i$ hence $n = r_1 + r_2 + \dots + r_\sigma$.

Theorem $N(A - \lambda_i I) \subset_{\neq} N(A - \lambda_i I)^2 \subset_{\neq} \dots \subset_{\neq} N(A - \lambda_i I)^{k_i} = N(A - \lambda_i I)^{k_i+1}$ *for some* $k_i \geq 1$.

Theorem .

- $k_i = m_i$
- $dim(N(A - \lambda_i I)^{m_i}) = r_i$

Example: $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

Example: $A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

Example:

Remark: Let A be an $n \times n$ matrix and $\bar{A}$ be its Jordan canonical form.

- $\bar{A} = B^{-1}AB$, where B is invertible.
- $\mathrm{rank}(A) = \mathrm{rank}\ (BA) = \mathrm{rank}\ (AB) = \mathrm{rank}(\bar{A})$
- $dim(N(A - \lambda_i I)) = dim(N(\bar{A} - \lambda_i I))$

  where the last remark follows from the previous ones and the following:

$$B^{-1}(A - \lambda_i I)B = B^{-1}AB - \lambda_i I B^{-1}B = \bar{A} - \lambda_i.$$

We already know that $\mathbb{C}^n = N((A - \lambda_1 I)^{m_1}) \oplus N((A - \lambda_2 I)^{m_2}) \oplus \ldots \oplus N((A - \lambda_\sigma I)^{m_\sigma})$.

The transformation matrix B can be written as $B = [B_1 \quad B_2 \quad \ldots \quad B_\sigma]$, where columns of B_i span $N((A - \lambda_i I)^{m_i})$.

Our next aim is to construct $B_i \in \mathbb{C}^{n \times r_i}$ whose columns span $N((A - \lambda_i I)^{m_i})$ and $r_i \times\ r_i$ block $\bar{A}_i$ satisfies $AB_i = B_i \bar{A}_i$.

Let $M_i := A - \lambda_i I$, and let's choose a vector x such that $x \in N(M_i^{m_i})$, but $x \notin N(M_i^{m_i - 1})$.

Now consider the chain of vectors:

$$\{M_i^{m_i-1}x, \quad M_i^{m_i-2}x, \quad \ldots, \quad M_i x, \quad x\}$$

Claim: The set $\{M_i^{m_i-1}x, \quad M_i^{m_i-2}x, \quad \ldots, \quad M_i x, \quad x\}$ is linearly independent.

Proof:

Example:

Special cases:

i) A has a single eigenvalue λ_i, and $m_i = r_i$.

ii) A has a single eigenvalue λ_i, and $m_i = 1$.

Example:

Example:

Example:

7 Hermitian Matrices

Definition *An $n \times n$ complex matrix A is called **Hermitian** if $A^* = A$, i.e., its conjugate transpose is equal to itself. If A is a real matrix then $A^* = A^T$.* [1]

Hermitian matrices enjoys important properties.

Theorem *Let A be Hermitian, then $\langle x, Ax \rangle$ is real for all $x \in \mathbb{C}^n$*

Proof:

Theorem *All eigenvalues of a Hermitian matrix are real.*

Proof:

[1] In some books, conjugate transpose is denoted by A^H instead of A^*.

Theorem *Eigenvectors of Hermitian matrices are orthogonal. Let A be Hermitian and λ_i, λ_j be two distinct ($\lambda_i \neq \lambda_j$) eigenvalues with eigenvectors e_i, e_j, then $\langle e_i, e_j \rangle = 0$.*

Proof:

Theorem *Let A be Hermitian. Then its minimal polynomial is*

$$m(s) = (s - \lambda_1)(s - \lambda_2) \dots (s - \lambda_\sigma).$$

That is, $m_i = 1$ for all eigenvalues of Hermitian matrices.

Proof:

Therefore for a Hermitian matrice A with characteristic polynomial $d(s) = (s - \lambda_1)^{r_1}(s - \lambda_2)^{r_2} \dots (s - \lambda_\sigma)^{r_\sigma}$, we can write

$$\mathbb{C}^n = \underset{dim=r_1}{N(A - \lambda_1 I)} \overset{\perp}{\oplus} \underset{dim=r_2}{N(A - \lambda_2 I)} \overset{\perp}{\oplus} \dots \overset{\perp}{\oplus} \underset{dim=r_\sigma}{N(A - \lambda_\sigma I)}\,.$$

Theorem *Let A be a Hermitian matrice with characteristic polynomial $d(s) = (s-\lambda_1)^{r_1}(s-\lambda_2)^{r_2}\dots(s-\lambda_\sigma)^{r_\sigma}$. Then there exist a unitary matrix P, i.e., $P^{-1} = P^*$ such that $P^*AP = \Lambda$ where*

$$\Lambda = \begin{bmatrix} \Lambda_1 & 0 & \dots & 0 \\ 0 & \Lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \Lambda_\sigma \end{bmatrix} \text{ and each } \quad \Lambda_i \quad \text{is } r_i \times r_i, \quad \Lambda_i = \begin{bmatrix} \lambda_i & 0 & \dots & 0 \\ 0 & \lambda_i & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \lambda_i \end{bmatrix}$$

Proof:

Theorem *Let A be an $n \times n$ Hermitian matrix with eigen values $\lambda_1, \lambda_2, \dots, \lambda_\sigma$. Let $\lambda_{min} := \min_i \lambda_i$ and $\lambda_{max} := \max_i \lambda_i$. Then for all $x \in \mathbb{C}^n$ we have,*

$$\lambda_{min} \langle x, x \rangle \le \langle x, Ax \rangle \le \lambda_{max} \langle x, x \rangle$$

Proof:

Recall that $\mathbb{C}^n = N(A-\lambda_1 I) \overset{\perp}{\oplus} N(A-\lambda_2 I) \overset{\perp}{\oplus} \dots \overset{\perp}{\oplus} N(A-\lambda_\sigma I)$. Then for a given x we can write $x = x_1 + x_2 + \ldots + x_\sigma$ with $x_i \in N(A - \lambda_i I)$. Then

$$\begin{aligned}
\langle x, Ax\rangle &= \left\langle \sum_{i=1}^{\sigma} x_i, A\sum_{j=1}^{\sigma} x_j \right\rangle \\
&= \left\langle \sum_{i=1}^{\sigma} x_i, \sum_{j=1}^{\sigma} Ax_j \right\rangle \\
&= \left\langle \sum_{i=1}^{\sigma} x_i, \sum_{j=1}^{\sigma} \lambda_j x_j \right\rangle \\
&= \sum_{i=1}^{\sigma} \left\langle x_i, \sum_{j=1}^{\sigma} \lambda_j x_j \right\rangle \\
&= \sum_{i=1}^{\sigma} \langle x_i, \lambda_i x_i\rangle \\
&= \sum_{i=1}^{\sigma} \lambda_i \langle x_i, x_i\rangle \\
&\Rightarrow \lambda_{min} \langle x, x\rangle \leq \langle x, Ax\rangle \leq \lambda_{max} \langle x, x\rangle
\end{aligned}$$

Definition *A Hermitian matrix A is said to be **positive definite** if $\langle x, Ax\rangle > 0$ for all $x \neq 0$.* [2]

Theorem *If A is a positive definite Hermitian matrix, then all of its eigenvalues are positive.*

Proof:

[2] A Hermitian matrix A is said to be **positive semi-definite** if $\langle x, Ax\rangle \geq 0$ for all $x \neq 0$.

8 Functions of a Matrix

The basic motivation to study matrix-valued functions comes from the differential equations describing linear systems[3]

$$\dot{x}(t) = Ax(t),$$

and its solution

$$x(t) = e^{At}x(0).$$

Definition *Consider a scalar valued function $f(s)$ with the following power series expansion:*

$$f(s) = \sum_{i=0}^{\infty} \alpha_i s^i$$

The matrix valued function $f(A)$ is defined as,

$$f(A) := \sum_{i=0}^{\infty} \alpha_i A^i,$$

which is another matrix with the ***same size*** *as A.*

Example:

$$e^t = \sum_{i=0}^{\infty} \frac{t^i}{i!} \qquad \Rightarrow \quad e^A := \sum_{i=0}^{\infty} \frac{A^i}{i!}$$

By using Cayley Hamlton theorem, we can express n^{th} or higher orders of an $n \times n$ matrix as a linear combination of its lower powers: $I, \quad A, \quad A^2, \ldots, A^{n-1}$. Then e^A can be expressed as,

$$e^A = c_0 + c_1 A + c_2 A^2 + \ldots + c_{n-1}A^{n-1}.$$

[3]Motivation to pass this course is neglected in this statement.

Similarly, one can use the minimal polynomial of a matrix to express the l^{th} power of an $n \times n$ matrix as a linear combination of its lower powers: $I, \quad A, \quad A^2, \ldots, A^{l-1}$, where l is the order of its minimal polynomial. In that case we can write,

$$e^A = c_0 + c_1 A + c_2 A^2 + \ldots + c_{l-1} A^{l-1}.$$

Since $l \leq n$, in general it is easier to find the l coefficients of the above equation.

Next, we will deal with the problem of finding the unknown coefficients.

8.1 First Method

Let

$$f(s) = \sum_{i=0}^{\infty} \alpha_i s^i,$$

and

$$f(A) = \sum_{i=0}^{\infty} \alpha_i A^i.$$

Let us define $p(s)$ and $p(A)$ as follows,

$$p(s) = c_0 + c_1 s + c_2 s^2 + \ldots + c_{l-1} s^{l-1},$$

$$p(A) = c_0 + c_1 A + c_2 A^2 + \ldots + c_{l-1} A^{l-1}.$$

Then we have the equality

$$f(A) = p(A).$$

Case 1: Matrix A is diagonalizable. Suppose $m(s) = (s - \lambda_1)(s - \lambda_2) \ldots (s - \lambda_\sigma)$, i.e., $l = \sigma$ and $m_1 = m_2 = \ldots = m_\sigma = 1$.

$\Rightarrow$ we have $f(\lambda_j) = p(\lambda_j)$ for $j = 1, \ldots, l$ which results in l equations for l unknowns $c_0, c_1, \ldots, c_{l-1}$.

Example: $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, find e^A and $\log(A)$.

Case 2: Matrix A is not diagonalizable.

Consider the following example. Let $A \in \mathbb{R}^3$ and $m(s) = (s-\lambda_1)^2(s-\lambda_2)$. Let the Jordan canonical form of A be equal to $J = \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{bmatrix}$. Then,

$$f(A) = c_0 + c_1 A + c_2 A^2 = p(A)$$

$$f(\lambda_1) = p(\lambda_1)$$

$$f(\lambda_2) = p(\lambda_2)$$

These two equations are not enough to find the three unknowns c_0, c_1, and c_2.

Consider the matrix P, which transforms the matrix A into its Jordan canonical form, i.e., $J = P^{-1}AP$. We know that P has the following form: $P = [\ \underbrace{e_1 \quad f_1}_{chain for \lambda_1} \quad \underbrace{e_2}_{chain for \lambda_2}\ ]$, where e_1, e_2 are eigenvectors corresponding to λ_1 and λ_2 respectively and f_1 is a generalized eigenvector for λ_1. Notice that,

$$[e_1 \quad f_1 \quad e_1] \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{bmatrix} = A[e_1 \quad f_1 \quad e_1]$$

$$
\begin{aligned}
\Rightarrow \qquad Af_1 &= \lambda_1 f_1 + e_1 \\
A^2 f_1 &= \lambda_1 A f_1 + A e_1 \\
&= \lambda_1(\lambda_1 f_1 + e_1) + \lambda_1 e_1 = \lambda_1^2 f_1 + 2\lambda_1 e_1 \\
A^3 f_1 &= \lambda_1^2 A f_1 + 2\lambda_1 A e_1 = \lambda_1^3 f_1 + 3\lambda_1^2 e_1 \\
&\vdots \\
A^i f_1 &= \lambda_1^i f_1 + i\lambda_1^{i-1} e_1
\end{aligned}
$$

Returning back to the equation,

$$
\begin{aligned}
f(A) &= p(A) \\
\sum_{i=0}^{\infty} \alpha_i A^i &= \sum_{i=0}^{l-1} c_i A^i
\end{aligned}
$$

and multiplying both sides by f_1 from right results,

$$
\begin{aligned}
\sum_{i=0}^{\infty} \alpha_i A^i f_1 &= \sum_{i=0}^{l-1} c_i A^i f_1 \\
\sum_{i=0}^{\infty} \alpha_i (\lambda_1^i f_1 + i\lambda_1^{i-1} e_1) &= \sum_{i=0}^{l-1} c_i (\lambda_1^i f_1 + i\lambda_1^{i-1} e_1) \\
\Rightarrow \qquad f(\lambda_1) f_1 + f'(\lambda_1) e_1 &= p(\lambda_1) f_1 + p'(\lambda_1) e_1.
\end{aligned}
$$

Since $f(\lambda_1) = p(\lambda_1)$, we have

$$f'(\lambda_1) e_1 = p'(\lambda_1) e_1.$$

Since $e_i \neq 0$ we have

$$f'(\lambda_1) = p'(\lambda_1),$$

which is the additional equation needed to find the coefficients $c_0, c_1, \ldots, c_{l-1}$ of $p(A)$.

General case:

Let $m(s) = (s-\lambda_1)^{m_1}(s-\lambda_2)^{m_2}\dots(s-\lambda_\sigma)^{m_\sigma}$, we have the following set of equations to find the coefficients $c_0, c_1, \dots, c_{l-1}$ of $p(A)$:

$$f^{(t)}(\lambda_j) = p^{(t)}(\lambda_j), \quad \text{for} \quad j = 1, \dots, \sigma \qquad t = 0, \dots, m_j - 1,$$

where t denotes the derivative order.

Example: $A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$. Find $\sin(\pi A)$.

Remark:

$f(A)$ does not exist when $f^{(t)}(\lambda_j) \quad j = 1, \ldots, \sigma, \qquad t = 0, \ldots, m_j - 1$, does not exist.

Example: $A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Consider $f_1(s) = \log(s)$ and $f_2(s) = (1 - s)^{-1}$.

$\lambda_1 = 0$, $\lambda_2 = 1$, $\Rightarrow m(s) = s(s - 1)$

$\log(A)$ and $(I - A)^{-1}$ do not exist since $f_1(\lambda_1)$ and $f_2(\lambda_2)$ do not exist.

9 Function of a Matrix Given Its Jordan Form

Example:

Example: