Notes for CIE 6002 Matrix Analysis

Binghao He

March 9, 2021

Contents

Lecture 2. Fundamental and Basic Operations

Subspace

Definition

The set $S \subseteq \mathbb{R}^m$ is a subspace, if for $x, y \in S$

$$\alpha x + \beta y = S, \quad \alpha, \beta \in \mathbb{R}^m$$

This indicates

$$x_i \in S, \quad i = 1, ..., N \quad \Longrightarrow \quad \sum_{i=1}^{N} \alpha_i x_i \in S$$

Some important subspaces

1. Span: a collection of vectors
$$span\{a_1, ..., a_n\} = \{y \in \mathbb{R}^m | y = \sum_{i=1}^{n} \alpha_i a, \alpha_i \in \mathbb{R}\}$$
2. Orthogonal complement: given a subspace $S \subseteq \mathbb{R}^m$
$$S_\perp = \{y \in \mathbb{R}^m | y^T x = 0, \forall x \in S\}$$
3. Range space of A (linear comb. of col. vectors)
$$\begin{aligned} R(A) &= span\{a_1, a_2, ..., a_n\} \\ &= \{y \in \mathbb{R}^m | y = Ax, x \in \mathbb{R}^n\} \end{aligned}$$
4. Null space of A (linear comb. of col. vectors)
$$Null(A) = \{x \in \mathbb{R}^n | Ax = 0\}$$

Linear independence

Given a set $\{a_1, a_2, ..., a_n\}$, we say they are linear independent if

$$\sum_{i=1}^{n} \alpha_i a_i = 0 \Longrightarrow \alpha_i = 0, \forall i = 1, ..., n$$

Basis

$\{b_1, b_2, ..., b_k\} \subseteq \mathbb{R}^m$ is a basis for the subspace $S \subseteq \mathbb{R}^m$ if $S = span\{b_1, b_2, ..., b_k\}$ and $\{b_1, b_2, ..., b_k\}$ is linear independent.

The number of the vectors of the basis for S

$$k = dim(S)$$

1. k: max number of linear indep. vectors in S
2. k: min number of vectors that can span S

Rank

For a matrix $A \in \mathbb{R}^{m \times n}$

$$dim(R(A)) \leq min\{m, n\}$$

Rank of a matrix (?)

$$\begin{aligned} rank(A) &= dim(R(A)) \\ &= \text{max number of linearly indep. columns(rows) of } A \\ &= rank(A^T) \end{aligned}$$

And

$$Null(A) = (R(A^T))_\perp$$

Nonsingularity/Invertibility

A symmetric matrix $A \in \mathbb{R}^{m \times m}$ is non-singular if

$$Ax = 0 \quad \Longrightarrow \quad x = 0$$

$\Longrightarrow$ columns of A are linearly indep.
$\Longrightarrow \exists A^{-1} \in \mathbb{R}^{m \times m}$ that $AA^{-1} = A^{-1}A = I_m$
Some properties

1. $(AB)^{-1} = B^{-1}A^{-1}$
2. $(A^{-1})^T = (A^T)^{-1}$

Determinant

For a matrix $A \in \mathbb{R}^{m \times m}$,
Inductive

1. If $m = 1$, $det(A) = A$
2. If $m > 1$, Define: $A_{ij} \in \mathbb{R}^{(m-1) \times (m-1)}$ is a submatrix that removes ith row and jth column of A.
Define: $c_i j = (-1)^{i+j} det(A_{ij})$.
$$\begin{aligned} det(A) &= \textstyle\sum_{j=1}^m a_{ij} c_{ij} \quad \forall i = 1, ..., m \\ &= \textstyle\sum_{i=1}^m a_{ij} c_{ij} \quad \forall j = 1, ..., m \end{aligned}$$

For $A, B \in \mathbb{R}^{m \times m}$

- $det(AB) = det(A) det(B)$
- $det(A) = det(A^T)$
- $det(\alpha A) = \alpha^m det(A)$
- $det(A) = 0 \Longleftrightarrow A$ is singular
- $det(A^{-1}) = 1/det(A)$ if A is invertible

Vector Norm

A mapping $f : \mathbb{R}^{⋗} \to \mathbb{R}$ is a norm if it satisfies that $\forall x, y \in \mathbb{R}^m$

- $f(x) \geq 0$
- $f(x) = 0$ *iff.* $x = 0$
- $f(x + y) \leq f(x) + f(y)$
- $f(\alpha x) = |\alpha| f(x)$

Examples

- $\|x\|_2 = \sqrt{\sum_{i=1}^m |x_i|^2}$ (Euclidean norm)
- $\|x\|_1 = \sum_{i=1}^m |x_i|$
- $\|x\|_\infty = \max_{i=1,...,m} |x_i|$
- $\|x\|_p = \sqrt[p]{\sum_{i=1}^m |x_i|^p}$

Inner Product

$$< x, y >= x^T y$$

$$< x, x >= \|x\|_2^2$$

Cauchy-Schwarz Inequality

$$|x^T y| \leq \|x\|_2 \cdot \|y\|_2$$

Holder Inequality

$$|x^T y| \leq \|x\|_p \cdot \|y\|_q$$

where

$$1/p + 1/q = 1 \quad , \quad p \cdot q \geq 1$$

Projection onto a Subspace

Projection

Given a set $S \subseteq \mathbb{R}^m$ and a point $y \in \mathbb{R}^m$, the projection of y onto S is

$$y_S = \arg\min_{z \in S} \|z - y\|_2^2$$

> **Theorem 1**
> Suppose S is a subspace,
>
> 1. y_S exists and is unique
> 2. Let $y_S = \Pi_S(y)$ *iff.*
> $z^T(y_S - y) = 0 \ \forall z \in S$

Orthogonal complement

$S_\perp = \{y \in \mathbb{R}^m | y^T x = 0, \forall x \in S\}$ is the orthogonal complement of subspace S.

- Orthogonal complement is a subspace
- $S \cap S_\perp$

> **Theorem 2**
> Let $S \subseteq \mathbb{R}^m$ be a subspace. For any $y \in \mathbb{R}^m$
>
> $$y = \Pi_S(y) + \Pi_{S_\perp}(y)$$

Proof

$$y = \Pi_S(y) + \Pi_{S_\perp}(y)$$

$$\begin{aligned} \iff \quad & y - \Pi_S(y) = \Pi_{S_\perp}(y) \\ \iff \quad & z^T(y - y + \Pi_S(y)) = 0 \quad \forall z \in S_\perp \\ \iff \quad & z^T \Pi_S(y) = 0 \quad \forall z \in S_\perp \end{aligned}$$

$$\because \quad \Pi_S(y)) \in S$$

$$\therefore z^T \Pi_S(y)) = 0 \quad \forall z \in S_\perp$$

□

Sum of subspaces

For two sets $X, Y \subseteq \mathbb{R}^m$,

$$X + Y = \{x + y \in \mathbb{R}^m\}$$

For a subspace $S \subseteq \mathbb{R}^m$

1. $S + S_\perp = \mathbb{R}^m$
2. $dim(S) + dim(S_\perp) = m$
3. $(S_\perp)_\perp = S$

Proof

1. Comes from Theorem 2
2. Let $\{a_1, a_2, ..., a_k\}$ be a basis for S ($dim(S) = k$) and $\{b_1, b_2, ..., b_\ell\}$ be a basis for $S\perp$ ($dim(S_\perp) = \ell$).
   $\{a_1, a_2, ..., a_k\} \cup \{b_1, b_2, ..., b_\ell\}$ is
   (a) linear independent
   (b) Span $S + S_\perp = \mathbb{R}^m$
   Therefore, $dim(S) + dim(S_\perp) = m$
3. Obvious

□

Orthogonal Set

$\{a_1, a_2, ..., a_n\}$ is orthogonal/orthonormal set if

1. $\|a_i\|_2 = 1 \quad \forall i = 1, ..., n$
2. $a_i^T a_j = 0 \quad \forall i \neq j$

The orthogonal set is linear independent.

$$y = \sum_{i=1}^{n} \alpha_i a_i \implies \alpha_i = a_i^T y = \alpha \|a_i\|_2^2$$

Given a linear independent set $\{a_1, a_2, ..., a_n\}$. The procedure of finding an orthogonal set $\{q_1, q_2, ..., 1_n\}$ so that $span\{a_1, a_2, ..., a_n\} = span\{q_1, q_2, ..., q_n\}$ is called "Gram-Schimit" procedure.

Orthogonal/Unitary Matrix

$Q = [q_1, q_2, ..., q_m] \in \mathbb{R}^{n \times m}$, $q_i, i = 1, ..., m$ are orthogonal.

- If $m = n$,
  Q is orthogonal matrix $\implies Q^T Q = I_m = QQ^T$.
- If $n > m$,
  semi-orthogonal $\implies Q^T Q = I_m \neq QQ^T$.

$Q = [q_1, q_2, ..., q_m] \in \mathbb{C}^{n \times m}$, $q_i, i = 1, ..., m$ are orthogonal.

- If $m = n$,
  Q is unitary matrix $\implies Q^H Q = I_m = QQ^H$.
- If $n > m$,
  semi-unitary $\implies Q^H Q = I_m \neq QQ^H$.

If $Q \in \mathbb{R}^{n\times m}$ $(n > m)$ is semi-orthogonal, $\exists \tilde{Q} \in \mathbb{R}^{n\times(n-m)}$ so that $\begin{bmatrix} Q & \tilde{Q} \end{bmatrix}$ is orthogonal.

Dimensional Theorem

For a matrix $A \in \mathbb{R}^{m\times n}$, it holds

$$dim(R(A)) + dim(Null(A)) = n$$

Proof

$$dim(R(A^T)) + dim(R(A^T)_\perp) = n$$
$$dim(R(A^T)) = rank(A^T) = rank(A) = dim(R(A))$$
$$dim(R(A^T)_\perp) = Null(A) = n$$

$\square$

Complexity of Matrix Computation

Flops

Arithmetic operations including addition, substraction, multiplication, division.

Big O notation

Big O notation: "order" of the complexity.

$$f(n) = O(g(n))$$

for some $g(n)$ if $\exists$ a constant $C > 0$ and n_0 so that for $n > n_0$

$$|f(n)| \leq C \cdot |g(n)|$$

Examples: for $x \in \mathbb{R}^m$, $y \in \mathbb{R}^m$, $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{m\times p}$

$x + y$	m flops	$O(m)$
$x^T y$	$2m - 1$ flops	$O(m)$
Ax	$n \cdot (2m - 1)$ flops	$O(nm)$
AB	$p \cdot n \cdot (2m - 1)$ flops	$O(pnm)$

Lecture 3. Least Square Problem (LSP)

Formulation

LS problem

Given $y \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$

$$\min_{x\in\mathbb{R}^n} \|y - Ax\|_2^2$$

Linear representation of signals

$$y = Ax$$

In practice,

$$y = Ax + v$$

where

- y is the observation (Given)
- A is the model (Given)
- x is the parameter to be estimated/determined
- v is the noice follow multi-variate Gaussian distribution

$$\mathrm{pdf}(v) \propto e^{-\|v\|^2}$$

$\Longrightarrow$ Maximum likelihood estimator

$$\max_x P(y|x) \propto \min_x \|y - Ax\|_2^2$$

AR and Polynomial Models

Autoregressive (AR) model

Given a time series y_t, $t = 0, 1, 2, ...,$

$$\underbrace{y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p}}_{\text{AR model}} + \underbrace{v_t}_{\text{noise}}$$

If $\alpha_1, \alpha_2, ..., \alpha_p$ are known, estimate y_t by

$$\hat{y}_t = \underbrace{\alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p}}_{\text{historical observation}}$$

Collect

$$\underbrace{\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_T \end{bmatrix}}_{y\in\mathbb{R}^T} = \underbrace{\begin{bmatrix} y_0 & & & \\ y_1 & y_0 & & \\ \vdots & \ddots & \ddots & \\ y_{p-1} & \cdots & \cdots & y_0 \\ \vdots & \ddots & \ddots & \vdots \\ y_{T-1} & \cdots & \cdots & y_T \end{bmatrix}}_{A\in\mathbb{R}^{T\times p}} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}}_{x\in\mathbb{R}^p} + \underbrace{\begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_T \end{bmatrix}}_{v\in\mathbb{R}^T}$$

$\Longrightarrow$ Solve LS problem to obtain α_i, $i = 1, ..., p$

Polynomial model

$$\underbrace{y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_p t^p}_{\text{polynomial model}} + \underbrace{v_t}_{\text{noise}}$$

Collect

$$\underbrace{\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_T \end{bmatrix}}_{y\in\mathbb{R}^T} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 \\ & \vdots & & \\ 1 & t & \cdots & t^p \\ & \vdots & & \\ 1 & T-1 & \cdots & (T-1)^p \end{bmatrix}}_{A\in\mathbb{R}^{T\times p}} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}}_{x\in\mathbb{R}^p} + \underbrace{\begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_T \end{bmatrix}}_{v\in\mathbb{R}^T}$$

$\Longrightarrow$ Solve LS problem to obtain α_i, $i = 1, ..., p$

Basis Representation

Basis matrix

For $y \in \mathbb{R}^m$, given a set of basis vectors (can be given or learned)

$$\Phi = \{\phi_1, \phi_2, ..., \phi_n\} \in \mathbb{R}^{m\times n}$$

We model

$$y = \sum_{i=1}^{n} \phi_i x_i = \Phi x$$

- $m \gg n$, assume data lies in a low-dimensional subspace
- $m \ll n$, dictionary $\Longrightarrow$ x is sparse

Fourier basis

Discrete Fourier Transform (DFT)
Inverse Discrete Fourier Transform (IDFT)

$$\Phi = [\phi_1, ..., \phi_n] \in \mathbb{C}^{n\times n}$$

$$\phi_i = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 \\ e^{j\frac{2\pi}{n}(i-1)} \\ e^{j\frac{2\pi}{n}(i-1)\times 2} \\ \vdots \\ e^{j\frac{2\pi}{n}(i-1)\times n-1} \end{bmatrix} \in \mathbb{R}^n$$

- Φ is IDFT matrix, Φ^H is DFT matrix.
- Φ is unitary matrix.

LTI System

Lienar time invariant system

– Linearity
$$\begin{cases} x_t^1 \rightarrow \boxed{\text{LTI}} \rightarrow y_t^1 \\ x_t^2 \rightarrow \boxed{\text{LTI}} \rightarrow y_t^2 \\ \alpha x_t^1 + \beta x_t^2 \rightarrow \boxed{\text{LTI}} \rightarrow \alpha y_t^1 + \beta y_t^2 \end{cases}$$

– Time Invariant $\quad x_{t-d} \rightarrow \boxed{\text{LTI}} \rightarrow y_{t-d}$

Impulse $\delta(t)$
$$\begin{cases} \delta(t) = 0, t \neq 0 \\ \int_{-\infty}^{+\infty} \delta(t)\, dt \end{cases}$$

LTI system is characterized by an impulse response
$$h_t, \quad t = 0, ..., p-1$$

$\Longrightarrow$ x_t and y_t is related through "convolution"
$$\begin{aligned} y_t &= \sum_{i=1}^{p} h_i x_{t-i} \\ &= h_t * x_t \end{aligned}$$

$$x_t \rightarrow \overbrace{\boxed{h_t}}^{\text{LTI}} \rightarrow y_t$$

System identification

Given $\{y_t\}_{t=0}^{T-1}$ and $\{x_t\}_{t=0}^{T-1}$, find the impulse response $\{h_t\}_{t=0}^{p}$

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_p \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} x_0 & & & \\ x_1 & x_0 & & \\ \vdots & \ddots & \ddots & \\ x_p & \cdots & \cdots & x_0 \\ \vdots & \ddots & \ddots & \vdots \\ x_{T-1} & \cdots & \cdots & x_{T-p} \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_p \end{bmatrix} + \text{noise}$$

$\Longrightarrow$ Solve LS problem to obtain $\{h_t\}_{t=0}^{p}$

Deconvolution

Given $\{y_t\}_{t=0}^{T-1}$ and $h_{t\,t=0}^{\,p}$, find out $\{x_t\}_{t=0}^{T-1}$
Put them into a linear system of equations

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_p \\ \vdots \\ y_T \end{bmatrix} = \underbrace{\begin{bmatrix} h_0 & 0 & & \cdots & & & 0 \\ h_1 & h_0 & 0 & \cdots & & \ddots & \\ \vdots & \vdots & & \ddots & & & \vdots \\ h_p & h_{p-1} & & \cdots & & h_1 & h_0 \\ \vdots & \ddots & \ddots & & & \vdots & \vdots \\ 0 & \cdots & h_p & h_{p-1} & \cdots & h_1 & h_0 \end{bmatrix}}_{A \in \mathbb{R}^{T \times T}} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_p \\ \vdots \\ x_T \end{bmatrix} + \text{noise}$$

$\Longrightarrow$ Solve LS problem to obtain $\{x_t\}_{t=0}^{T-1}$

Toeplitz matrix

$$A = \begin{bmatrix} a_0 & a_{-1} & \cdots & \cdots & a_{-n+1} \\ a_1 & a_0 & a_{-1} & \cdots & a_{-n+2} \\ a_2 & a_1 & a_0 & a_{-1} & \ddots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ a_{n-1} & \cdots & a_2 & a_1 & a_0 \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Circulant Matrix

Circulant matrix is a special case of Toeplitz matrix

$$H = \begin{bmatrix} h_0 & h_{n-1} & \cdots & \cdots & h_1 \\ h_1 & h_0 & h_{n-1} & \cdots & h_2 \\ h_2 & h_1 & h_0 & h_{n-1} & \ddots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ h_{n-1} & \cdots & h_2 & h_1 & h_0 \end{bmatrix} \in \mathbb{R}^{n \times n}$$

$$H \overbrace{\Longrightarrow}^{O(n \log n)} H^{-1}$$

The key property:
$$H \cdot \underbrace{\phi_i}_{\text{IDFT}} = d_i \cdot \phi_i \quad , d_i \in \mathbb{R}$$

Let
$$D = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}$$

Then
$$\begin{aligned} H \cdot \Phi &= H \cdot [\phi_1, ..., \phi_n] \\ &= [d_1 \phi_1, ..., d_n \phi_n] \\ &= \Phi \cdot D \end{aligned}$$

$\Longrightarrow$
$$\underbrace{H = \Phi \cdot D \cdot \Phi^H}_{\text{eigen-decomposition of } H} \quad \Longleftrightarrow \quad \Phi^H H \Phi = D$$

Lecture 4. Eigenvalues and Eigenvectors

> **Definition (Eigenvalues and Eigenvector)**
> Let $A \in \mathbb{R}^{m\times m}$, if a scalar $\lambda \in \mathbb{C}$ and a nonzero vector $v \in \mathbb{C}^m$ satisfy the equation
> $$Av = \lambda v$$
> then λ is called the eigenvalue of A and v is called the eigenvector of A associated with λ.

$\Longrightarrow$

$$\underbrace{(A-\lambda I)}_{\text{singular}} v = 0$$

$P_A(\lambda) \triangleq det(A-\lambda I) = \prod_{i=1}^{m}(\lambda-\lambda_i)$ is a polynomial of λ with degree m. The equation $P_A(\lambda) \triangleq det(A-\lambda I) = 0$ is called the characteristic euqation of A.

For each λ_i find v_i such that

$$(A-\lambda_i I)v_i = 0$$

$v_i \in Null(A-\lambda_i I)$ and $Null(A-\lambda_i I)$ is the eigenspace for λ_i

- Even if $A \in R^{m\times m}$ is real-valued, λ_i and v_i could be complex. *e.g.*
$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \Longrightarrow \quad \lambda = \pm j$$
- But if $\lambda_i \in \mathbb{R}$ is real-valued, $v_i \in \mathbb{R}^m$ should be in $\mathbb{R}^m$.

Diagonalizability

Similar matrices

> **Definition (Similar matrices)**
> $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{m\times m}$ are similar to each other if there exists a nonsingular $S \in \mathbb{R}^{m\times m}$ such that
> $$B = SAS^{-1}$$

If A and B are similar to each other, then they have the same eigenvalues.
Proof

$$\begin{aligned} P_B(\lambda) &= det(B-\lambda I) \\ &= det(SAS^{-1}-\lambda I) \\ &= det(S(A-\lambda I)S^{-1}) \\ &= det(S)det(A-\lambda I)det(S^{-1}) \\ &= det(A-\lambda I) \\ &= P_A(\lambda) \end{aligned}$$

□

Diagonalizable matrix

> **Definition (Diagonalizable)**
> A matrix $A \in \mathbb{R}^{m\times m}$ is diagonalizable if A is similar to a diagonal matrix.

$\Longrightarrow$ (**Eigenvalue decomposition**)
There exists an invertible matrix $V \in \mathbb{R}^{m\times m}$ and a diagonal matrix $D \in \mathbb{R}^{m\times m}$ s.t.

$$\underbrace{A = VDV^{-1}}_{\text{eigenvalue decomposition (EVD)}}$$

$\Longrightarrow$

$$AV = VD$$

$$V = [v_1, v_2, ..., v_m] \quad , \quad D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix}$$

$\Longleftrightarrow$

$$Av_i = \lambda_i v_i, \quad \forall i = 1, ..., m$$

$\Longrightarrow$
A is diagonalizable *iff.* there is a set of linear indep. vectors which are engenvectors of A. **Properties**
If a matrix $A \in \mathbb{R}^{m\times m}$ is diagonalizable

1. $det(A) = \prod_{i=1}^{m} \lambda_i$
2. $Tr(A) = \sum_{i=1}^{m} \lambda_i$
3. $rank(A) =$ the number of non-zero eigenvalues

Proof

1. $det(A) = det(VDV^{-1}) = det(V)det(D)det(V^{-1}) = det(D) = \prod_{i=1}^{m} \lambda_i$
2. $Tr(A) = Tr(VDV^{-1}) = Tr(DV^{-1}V) = Tr(D) = \sum_{i=1}^{m} \lambda_i$

□

When Diagonalizable

Distinct eigenvalues

Let $\{\lambda_1, \lambda_2, ..., \lambda_k\}_{k\in M}$ $(k \geq 2)$ be a subset of eigenvalues of A and $\lambda_i \neq \lambda_j \ \forall i \neq j$. Then the corresponding eigenvectors, denoted with $v_1, ..., v_k$ respectively, are linearly independent.

If A has $\lambda_1 \neq \lambda_2 \neq \cdots \neq \lambda_k$, then the corresponding eigenvectors $v_1, ..., v_k$ are linear independent.
Proof

Suppose $v_1, ..., v_k$ are linearly **dependent**. Then there exists $\alpha_1, \alpha_2, ..., \alpha_k$ not all zero (without losing generality, assume $\alpha_1 \neq 0$) s.t.

$$\sum_{i=1}^{k} \alpha_i v_i = 0$$

Then

$$A \cdot \left(\sum_{i=1}^{k} \alpha_i v_i\right) = \sum_{i=1}^{k} \alpha_i A v_i = \sum_{i=1}^{k} \alpha_i \lambda_i v_i = 0$$

$$\sum_{i=1}^{k-1} \alpha_i \lambda_i v_i + \alpha_k \lambda_k v_k = 0 \tag{1}$$

On the other hand,

$$\lambda_k \cdot \left(\sum_{i=1}^{k} \alpha_i v_i \right) = \sum_{i=1}^{k} \alpha_i \lambda_k v_i = 0$$

$$\sum_{i=1}^{k-1} \alpha_i \lambda_k v_i + \alpha_k \lambda_k v_k = 0 \tag{2}$$

(1) − (2):

$$\sum_{i=1}^{k-1} \alpha_i (\lambda_i - \lambda_k) v_i = 0$$

Then

$$\begin{aligned} A \cdot \left(\textstyle\sum_{i=1}^{k-1} \alpha_i (\lambda_i - \lambda_k) v_i \right) &= \textstyle\sum_{i=1}^{k-1} \alpha_i (\lambda_i - \lambda_k) A v_i \\ &= \textstyle\sum_{i=1}^{k-1} \alpha_i (\lambda_i - \lambda_k) \lambda_i v_i \\ &= 0 \end{aligned}$$

$$\sum_{i=1}^{k-2} \alpha_i (\lambda_i - \lambda_k) \lambda_i v_i + \alpha_{k-1} (\lambda_{k-1} 0 \lambda_k) \lambda_{k-1} v_{k-1} = 0 \tag{3}$$

On the other hand

$$\lambda_{k-1} \cdot \left(\sum_{i=1}^{k-1} \alpha_i (\lambda_i - \lambda_k) v_i \right) = \sum_{i=1}^{k-1} \alpha_i (\lambda_i - \lambda_k) \lambda_{k-1} v_i = 0$$

$$\sum_{i=1}^{k-2} \alpha_i (\lambda_i - \lambda_k) \lambda_{k-1} v_i + \alpha_{k-1} (\lambda_{k-1} 0 \lambda_k) \lambda_{k-1} v_{k-1} = 0 \tag{4}$$

(3) − (4):

$$\sum_{i=1}^{k-2} \alpha_i (\lambda_i - \lambda_k)(\lambda_i - \lambda_{k-1}) v_i = 0$$

Repeat until $i = 1$

$$\alpha_1 \left(\prod_{i=1}^{k} (\lambda_1 - \lambda_i) \right) v_1 = 0$$

Since $\alpha_1 \neq 0$ and $\prod_{i=1}^{k} (\lambda_1 - \lambda_i) \neq 0$, so $v_1 = 0$ (Contradict!)
Therefore, $v_1, ..., v_k$ are linearly **independent** □

If $A \in \mathbb{R}^{m \times m}$ has m distinct eigenvalues, then A is diagonalizable.

Repeated eigenvalues

Let λ be an eigenvalue of A repeated r times. Define

- Algebra multiplicity: r.
- Geometric multiplicity: $s = dim(Null(A - \lambda I))$

Then

$$s \leq r$$

Proof
Let $dim(Null(A - \bar{\lambda} I)) = s$ and $\{q_1, ..., q_s\}$ be an orthogonal basis for $Null(A - \bar{\lambda} I)$.
Then $Q_1 = [q_1, ..., q_s] \in \mathbb{R}^{m \times s}$ is semi-unitary and $\exists Q_2 \in \mathbb{R}^{m \times (m-s)}$ s.t. $Q \triangleq [Q_1, Q_2]$ is unitary.
Consider

$$Q^H A Q = \begin{bmatrix} Q_1^H A Q_1 & Q_1^H A Q_2 \\ Q_2^H A Q_1 & Q_2^H A Q_2 \end{bmatrix} = \begin{bmatrix} \bar{\lambda} I_s & Q_1^H A Q_2 \\ 0 & Q_2^H A Q_2 \end{bmatrix}$$

Then

$$\begin{aligned} P_A(\lambda) &= P_{Q^H A Q}(\lambda) \\ &= det(Q^H A Q - \lambda I) \\ &= det\left(\begin{bmatrix} (\bar{\lambda} - \lambda) I_s & Q_1^H A Q_2 \\ 0 & Q_2^H A Q_2 - \lambda I_{m-s} \end{bmatrix} \right) \\ &= \underbrace{(\bar{\lambda} - \lambda)^s}_{\bar{\lambda} \text{ repeats } s \text{ times}} \quad det(Q_2^H A Q_2 - \lambda I_{m-s}) \\ &= 0 \end{aligned}$$

So

$$r \geq s$$

□

If $s = r$, A is diagonalizable.

Symmetric/Hermitian Matrices

- **Symmetric** $A = A^T$ $A \in \mathbb{R}^{m \times n}$
- **Hermitian** $A = A^H$ $A \in \mathbb{C}^{m \times n}$

Property 1: The eigenvalues of a Hermitian matrix $A \in \mathbb{C}^{m \times n}$ are real-valued.
Proof
Let λ be the eigenvalue of A, v be the eigenvector of A and $\|v\|_2 = 1$.
$\Longrightarrow Av = \lambda v$
$\Longrightarrow v^H A v = \lambda v^H v = \lambda$
$\overset{\text{take}H}{\Longrightarrow} v^H A^H v = \lambda^*$
Since A is Hermitian, $A = A^H$ and hence $\lambda = \lambda^*$
So λ is real-valued □

Property 2: If $A \in \mathbb{C}^{m \times n}$ is Hermitian and that

- $\lambda_i \neq \lambda_j$ be two eigenvalues,
- v_i, v_j be the corresponding eigenvectors.

Then the eigenvectors are mutually orthogonal.
Proof

$$\begin{cases} v_i^H A v_j = v_i^H (\lambda_j v_j) = \lambda_j v_i^H v_j \\ v_j^H A v_i = v_j^H (\lambda_i v_i) = \lambda_i v_j^H v_i \overset{\text{take}H}{\Longrightarrow} v_i^H A^H v_j = \lambda_i v_i^H v_j \end{cases}$$

Since $v_i^H A^H v_j = v_i^H A v_j$, so we have

$$(\lambda_i - \lambda_j) v_i^H v_j = 0 \Longrightarrow v_i^H v_j = 0$$

□

Unitarily diagonalizable

If Hermitian/symmetric A is diagonalizable

$$A = VDV^{-1}$$

where

- V contains orthogonal eigenvectors and is unitary matrix,
- D real-valued diagonal eigenvalues.

Therefore, $VV^H = I$ and $V^{-1} = V^H$. Thus,

$$A = VDV^H$$

And we say A is unitarily diagonalizable.

Schur triangularization

Any $A \in \mathbb{C}^{m\times m}$ can be decomposed as an upper triangular matrix.

> **Theorem 3 (Schur triangularization)**
> Given $A \in \mathbb{C}^{m\times m}$ with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_m \in \mathbb{C}$ there exists a unitary matrix $U \in \mathbb{C}^{m\times m}$ such that
> $$U^H AU = T \iff A = UTU^H$$
> where T is an upper triangular matrix with $[T]_{ii} = \lambda_i$, $i = 1, 2, ..., m$

If $A \in \mathbb{C}^{m\times m}$, by Schur triangularization, there exists unitary $U \in \mathbb{C}^{m\times m}$ s.t.

$$U^H AU = T$$

Then by taking H, we have

$$\underset{(=U^H AU)}{U^H A^H U} = T^H$$

$\Longrightarrow$

$$T = T^H$$

Since T is an upper triangular matrix, so T^H is a lower triangular matrix. Therefore, T is diagonal. $\Longrightarrow$

$$U^H AU = T$$

Thus, A is diagonalizable.

Proof for Schur Triangularization (**Theorem 3**)
This proof is also an algorithm.
Let $v_1 \in \mathbb{R}^m$ be the eigenvalue of A associated with λ_1, i.e., $Av_1 = \lambda_1 v_1$. Without losing generality, we can assume $v_1^H v_1 = 1$ and $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$

Construct a unitary matrix U_1

$$U_1 = \begin{bmatrix} v_1 & V_1 \end{bmatrix} \in \mathbb{C}^{m\times m}$$

where $V_1 \in \mathbb{C}^{m\times(m-1)}$. U_1 being unitary indicates

$$v_1^H V_1 = \vec{0}^T$$

Then we have

$$U_1^H AU_1 = \begin{bmatrix} v_1^H \\ V_1^H \end{bmatrix} A \begin{bmatrix} v_1 & V_1 \end{bmatrix} = \begin{bmatrix} v_1^H Av_1 & v_1^H AV_1 \\ V_1^H Av_1 & V_1^H AV_1 \end{bmatrix}$$

where

$$\begin{cases} v_1^H Av_1 = \lambda_1 \\ v_1^H AV_1 \neq \vec{0} \\ V_1^H Av_1 = \vec{0} \end{cases}$$

So,

$$U_1^H AU_1 = \begin{bmatrix} \lambda_1 & \times \\ 0 & V_1^H AV_1 \end{bmatrix}$$

Define

$$A_1 \triangleq V_1^H AV_1 \in \mathbb{C}^{(m-1)\times(m-1)}$$

Then

$$\begin{aligned} P_A(\lambda) &= det(A - \lambda I) \\ &= det(U_1^H AU_1 - \lambda I)^\dagger \\ &= det\left(\begin{bmatrix} \lambda_1 & \times \\ 0 & A_1 \end{bmatrix} - \lambda I\right) \\ &= det\left(\begin{bmatrix} \lambda_1 - \lambda & \times \\ 0 & A_1 - \lambda I \end{bmatrix}\right) \\ &= (\lambda_1 - \lambda)\underbrace{det(A_1 - \lambda I)}_{=P_{A_1}(\lambda)} \end{aligned}$$

$\dagger$: $A = U_1(U_1^H AU_1)U_1^{-1}$ so A is similar to $U_1^H AU_1$ and hence A and $U_1^H AU_1$ have same eigenvalues.

Therefore, A_1 has eigenvalues $\lambda_2, \lambda_3, ..., \lambda_m$
Let $v_2 \in \mathbb{C}^{m-1}$ be the eigenvector of A_1 associated with λ_2.
Construct a unitary matrix U_2

$$U_2 = \begin{bmatrix} v_2 & V_2 \end{bmatrix} \in \mathbb{C}^{(m-1)\times(m-1)}$$

where $V_2 \in \mathbb{C}^{(m-1)\times(m-2)}$. U_2 being unitary indicates

$$v_2^H V_2 = \vec{0}^T$$

$\Longrightarrow$

$$U_2^H A_1 U_2 = \begin{bmatrix} \lambda_2 & \times \\ 0 & V_2^H A_1 V_2 \end{bmatrix}$$

Let $\tilde{U}_2 = \begin{bmatrix} 1 & 0^T \\ 0 & U_2 \end{bmatrix}$, then $\tilde{U}_2$ is unitary.

So,

$$\begin{aligned} \tilde{U}_2^H [U_1^H AU_1]\tilde{U}_2 &= \tilde{U}_2^H \begin{bmatrix} \lambda_1 & \times \\ 0 & A_1 \end{bmatrix} \tilde{U}_2 \\ &= \begin{bmatrix} \lambda_1 & \times \\ 0 & U_2^H A_1 U_2 \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \times \\ 0 & \begin{bmatrix} \lambda_2 & \times \\ 0 & V_2^H A_1 V_2 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \times & \times \\ 0 & \lambda_2 & \times \\ 0 & 0 & V_2^H A_1 V_2 \end{bmatrix} \end{aligned}$$

Repeat this operations and we can finally get

$$\tilde{U}_m^H \cdots \tilde{U}_3^H \tilde{U}_2^H (U_1 A \underbrace{U_1)\tilde{U}_2\tilde{U}_3 \cdots \tilde{U}_m}_{\triangleq U \text{ unitary}}$$

$$= \begin{bmatrix} \lambda_1 & \times & \times & \times \\ & \lambda_2 & \times & \times \\ & & \ddots & \times \\ & & & \lambda_m \end{bmatrix} \triangleq T$$

Therefore,

$$U^H A U = T$$

□

Variational Characterization of Eigenvalues

Rayleigh quotient

Definition (Rayleigh quotient)

$$R(A,x) = \frac{x^H A x}{\|x\|_2^2} = \left(\frac{x}{\|x\|_2}\right)^H A \left(\frac{x}{\|x\|_2}\right)$$

Theorem 4 (Rayleigh-Ritz theorem)
Let $A \in \mathbb{C}^{m\times m}$ be Hermitian, then

$$\begin{aligned} \lambda_{\max}(A) &= \max_{x\in\mathbb{C}^m} \tfrac{x^H A x}{\|x\|_2^2} \\ &= \max_{x\in\mathbb{C}^m} x^H A x \quad (\|x\|_2 = 1) \\ \lambda_{\min}(A) &= \min_{x\in\mathbb{C}^m} \tfrac{x^H A x}{\|x\|_2^2} \end{aligned}$$

and

$$\lambda_{\min}(A)\cdot\|x\|_2^2 \le x^H A x \le \lambda_{\max}(A)\cdot\|x\|_2^2$$

Proof
According to the EVD of A,

$$A = V\Lambda V^H = \sum_{i=1}^m (\lambda_i \cdot v_i v_i^H)$$

$\Longrightarrow$

$$\begin{aligned} x^H A x &= x^H V\Lambda V^H x \\ &= \textstyle\sum_{i=1}^m (\lambda_i \cdot x^H v_i v_i^H x) \\ &= \textstyle\sum_{i=1}^m (\lambda_i \cdot |v_i^H x|^2) \\ &\le \lambda_{\max}(A)\cdot \textstyle\sum_{i=1}^m |v_i^H x|^2 \end{aligned}$$

Since

$$\sum_{i=1}^m |v_i^H x|^2 = x^H V V^H x = \|x\|_2^2$$

So,

$$x^H A x \le \lambda_{\max}(A)\cdot\|x\|_2^2$$

$\Longrightarrow$

$$\frac{x^H A x}{\|x\|_2^2} \le \lambda_(\max)(A)$$

The equality holds only when $x = v_1$, which is an eigenvector associated with $\lambda_{\max}(A)$ and is called the principal eigenvector. □

Courant-Fischer theorem

Theorem 5 (Courant-Fischer theorem)
Let $A \in \mathbb{C}^{m\times m}$ be Hermitian and $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m$ are ordered eigenvalues. Then

$$\begin{aligned} \lambda_k &= \min_{\omega_1,\omega_2,\ldots,\omega_{k-1}} \left\{ \max_{x\in\mathbb{C}^m, x\perp\omega_1,\omega_2,\ldots,\omega_{k-1}} \tfrac{x^H A x}{\|x\|_2^2} \right\} \\ &= \max_{\omega_{k+1},\omega_{k+2},\ldots,\omega_m} \left\{ \min_{x\in\mathbb{C}^m, x\perp\omega_{k+1},\omega_{k+2},\ldots,\omega_m} \tfrac{x^H A x}{\|x\|_2^2} \right\} \end{aligned}$$

Proof ideas:
1.

$$\max_{x\in\mathbb{C}^m, x\perp v_1,v_2,\ldots,v_{k-1}} \frac{x^H A x}{\|x\|_2^2} = \lambda_k$$

where v_k is the eigenvector associated with λ_k of A
2.

$$\max_{x\in\mathbb{C}^m, x\perp\omega_1,\omega_2,\ldots,\omega_{k-1}} \frac{x^H A x}{\|x\|_2^2} \ge \lambda_k$$

1+2 $\Longrightarrow$

$$\min_{\omega_1,\omega_2,\ldots,\omega_{k-1}} \left\{ \max_{x\in\mathbb{C}^m, x\perp\omega_1,\omega_2,\ldots,\omega_{k-1}} \frac{x^H A x}{\|x\|_2^2} \right\} = \lambda_k$$

attained when $\omega_i = v_i$ $(i = 1, ..., k-1)$

Proof
1.
Since $A = V\Lambda V^H = \sum_{i=1}^m \lambda_i v_i v_i^H$,

$$\begin{aligned} &\max_{x\in\mathbb{C}^m, x\perp v_1,v_2,\ldots,v_{k-1}} \tfrac{x^H A x}{\|x\|_2^2} \\ &= \max_{x\in\mathbb{C}^m, x\perp v_1,v_2,\ldots,v_{k-1}} \tfrac{\sum_{i=1}^m \lambda_i |v_i^H x|^2}{\|x\|_2^2} \\ &= \max_{x\in\mathbb{C}^m, x\perp v_1,v_2,\ldots,v_{k-1}} \tfrac{\sum_{i=k}^m \lambda_i |v_i^H x|^2}{\|x\|_2^2} \\ &\le \max_{x\in\mathbb{C}^m} \tfrac{\sum_{i=k}^m \lambda_i |v_i^H x|^2}{\|x\|_2^2} \quad \text{(Relax constraint)} \\ &\le \max_{x\in\mathbb{C}^m} \tfrac{\lambda_k \sum_{i=k}^m |v_i^H x|^2}{\|x\|_2^2}{}^{\dagger} \\ &\le \lambda_k \quad \text{(equality attained if } x = v_k) \end{aligned}$$

†: $\sum_{i=k}^m |v_i^H x|^2 = x^H V V^H x = \|x\|_2^2$.

Therefore,

$$\max_{x\in\mathbb{C}^m, x\perp v_1,v_2,\ldots,v_{k-1}} \frac{x^H A x}{\|x\|_2^2} = \lambda_k$$

2.
Let $y = V^H x$

$$\begin{aligned} &\max_{x\perp\omega_1,\ldots,\omega_{k-1}} \tfrac{x^H A x}{\|x\|_2^2} \\ &= \max_{x\perp\omega_1,\ldots,\omega_{k-1}} \tfrac{x^H V\Lambda V^H x}{\|x\|_2^2} \\ &= \max_{y\perp V^H\omega_1,\ldots,V^H\omega_{k-1}} \tfrac{y^H \Lambda y}{\|y\|_2^2} \\ &\ge \max_{y\perp V^H\omega_1,\ldots,V^H\omega_{k-1}, y_{k+1}=y_{k+2}=\cdots=y_m=0, \|y\|_2=1} y^H \Lambda y \\ &= \max_{y\perp V^H\omega_1,\ldots,V^H\omega_{k-1}, |y_1|^2+|y_2|^2+\cdots+|y_k|^2=1} \textstyle\sum_{i=1}^k \lambda_i |y_i|^2 \\ &\ge \lambda_k \end{aligned}$$

□

Weyl's inequality

For $A, B \in \mathbb{C}^{m\times m}$ be two Hermitian matrices.
Then

1. $\lambda_k(A) + \lambda_m(B) \le \lambda_k(A+B) \le \lambda_k(A) + \lambda_1(B)$
2. $\lambda_{k+1}(A) \le \lambda_k(X \pm zz^H) \le \lambda_{k-1}(A)$
3. $\lambda_{k+r}(A) \le \lambda_k(A+B) \le \lambda_{k-r}(A)$ where $rank(B) \le r$
4. $\lambda_{j+k-1}(A+B) \le \lambda_j(A) + \lambda_k(B) \le \lambda_{j+k-m}(A+B)$
5. Let A_I be the principal submatrix of A associated with subset $I = \{i_1, i_2, ..., i_r\} \subseteq \{1, 2, ..., m\}$
$$\lambda_{k+m-r}(A) \le \lambda_k(A_I) \le \lambda_k(A)$$
6. Let $U \in \mathbb{C}^{m\times r]}$ be semiunitary
$$\lambda_{k+m-r} \le \lambda_k(U^H AU) \le \lambda_k(A)$$

Proof 1.

$$\lambda_k(A+B) = \min_{\omega_1,...,\omega_{k-1}} \left\{ \max_{x\in\mathbb{C}^m, x\perp\omega_1,...,\omega_{k-1}} \frac{x^H(A+B)x}{\|x\|_2^2} \right\}$$

where

$$\frac{x^H(A+B)x}{\|x\|_2^2} = \frac{x^H Ax}{\|x\|_2^2} + \underbrace{\frac{x^H Bx}{\|x\|_2^2}}_{\le\lambda_1(B)}$$

So,

$$\begin{aligned}&\lambda_k(A+B)\\ &\le \underbrace{\min_{\omega_1,...,\omega_{k-1}} \left\{ \max_{x\in\mathbb{C}^m, x\perp\omega_1,...,\omega_{k-1}} \frac{x^H Ax}{\|x\|_2^2} \right\}}_{=\lambda_k(A)} + \lambda_1(B)\end{aligned}$$

Therefore,

$$\lambda_k(A+B) \le \lambda_k(A) + \lambda_1(B)$$

□

Proof 2.

$$\begin{aligned}&\lambda_k(A \pm zz^H)\\ &= \min_{\omega_1,...,\omega_{k-1}} \left\{ \max_{x\in\mathbb{C}^m, x\perp\omega_1,...,\omega_{k-1}} \frac{x^H(A\pm zz^H)x}{\|x\|_2^2} \right\}\\ &\ge \min_{\omega_1,...,\omega_{k-1}} \left\{ \max_{x\in\mathbb{C}^m, x\perp\omega_1,...,\omega_{k-1}, \boxed{z}} \frac{x^H(A\pm zz^H)x}{\|x\|_2^2} \right\}\\ &= \min_{\omega_1,...,\omega_{k-1}, \omega_k = z} \left\{ \max_{x\in\mathbb{C}^m, x\perp\omega_1,...,\omega_{k-1},\omega_k} \frac{x^H Ax}{\|x\|_2^2} \right\}\\ &\ge \min_{\omega_1,...,\omega_{k-1}, \omega_k} \left\{ \max_{x\in\mathbb{C}^m, x\perp\omega_1,...,\omega_{k-1},\omega_k} \frac{x^H Ax}{\|x\|_2^2} \right\}\\ &= \lambda_{k+1}(A)\end{aligned}$$

□

Proof 3.

$$\because B = \sum_{i=1}^{r} \lambda_i(B) u_i u_i^H$$

$$\therefore A + B = A + \sum_{i=1}^{r-1} \lambda_i(B) u_i u_i^H + \lambda_r(B) u_r u_r^H$$

Define $B^{r-k} \triangleq \sum_{i=1}^{r-k} \lambda_i(B) u_i u_i^H \ \ k = 0, 1, ..., r$
$\lambda_{k+1}(A+B^{r-1}) \le \lambda_k(A+B^{r-1}+\lambda_r(B)u_r u_r^H) \le \lambda_{k-1}(A+B^{r-1})$
Do again and we get

$$\begin{aligned}\lambda_{k+2}(A+B^{r-2}) &\le \lambda_{k+1}(A+B^{r-2}+\lambda_{r-1}(B)u_{r-1}u_{r-1}^H)\\ &= \lambda_{k+1}(A+B^{r-1})\end{aligned}$$

Repeat r times, we can get

$$\lambda_{k+r}(A) \le \lambda_k(A+B) \le \lambda_{k-r}(A)$$

where $rank(B) \le r$ □

Proof 4.
Define $A_{j-1} = \sum_{i=1}^{j-1} \lambda_i(A) v_i v_i^H$ (if $A = V\Lambda V^H$).
Define $B_{k-1} = \sum_{i=1}^{k-1} \lambda_i(B) u_i u_i^H$ (if $B = U\Lambda U^H$).
Then

$$\begin{aligned}\lambda_j(A) &= \lambda_1(A - A_{j-1})\\ \lambda_k(B) &= \lambda_1(B - B_{k-1})\end{aligned}$$

$\Longrightarrow$

$$\begin{aligned}\lambda_j(A) + \lambda_k(B) &= \lambda_1(A - A_{j-1}) + \lambda_1(B - B_{k-1})\\ &\ge \lambda_1(A + B - (A_{j-1} + B_{k-1}))\\ &\ge \lambda_{j+k-1}(A+B)^\dagger\end{aligned}$$

†: $rank((A_{j-1} + B_{k-1}) \le j + k - 2$ and from 3. □

Proof 5.

$$\begin{aligned}&\lambda_k(A)\\ &= \min_{\omega_1,...,\omega_{k-1}} \left\{ \max_{x\perp\omega_1,...,\omega_{k-1}, \|x\|_2=1} x^H Ax \right\}\\ &\ge \min_{\omega_1,...,\omega_{k-1}} \left\{ \max_{x\perp\omega_1,...,\omega_{k-1}, \|x\|_2=1, x\in span\{e_{i_1},...,e_{i_r}\}} x^H Ax \right\}^\dagger\\ &= \min_{v_1,...,v_{k-1}} \left\{ \max_{y\perp v_1,...,v_{k-1}, \|y\|_2=1\}} y^H Ay \right\}^\ddagger\\ &= \lambda_k(A_I)\end{aligned}$$

†: $e_i = \begin{bmatrix} \mathbf{0} \\ 1 \\ \mathbf{0} \end{bmatrix}$, only the ith entry equal to 1.

‡: Define $y \in \mathbb{C}^r$, $y_\ell = x_{i_\ell}$ ($\ell = 1, ..., r$) and take $v_i \in \mathbb{C}^r$, $(v_i)_\ell = (\omega_i)_{i_\ell}$ □

Proof 6.
Let $\tilde{U} = [\underbrace{U}_{\in\mathbb{C}^{m\times r}}, \underbrace{\bar{U}}_{\in\mathbb{C}^{m\times r}}]$ be unitary. Then

$$\tilde{U}^H A\tilde{U} = \begin{bmatrix} U^H AU & \times \\ \times & \times \end{bmatrix}$$

where $U^H AU$ is the principal submatrix of $\tilde{U}^H A\tilde{U}$ with $I = 1, 2, ..., r$
Therefore,

$$\lambda_k(A) = \lambda_k(\tilde{U}^H A\tilde{U}) \ge \lambda_k(U^H AU)$$

□

The Power Method

The power method is a simple method to obtain eigenvector/eigenvalue of a matrix, particularly for large scale problem.

Algorithm 1 (Power Method)
Given a matrix $A \in \mathbb{C}^{m\times m}$ and $v^0 \in \mathbb{C}^m$

$$v^{(k+1)} \leftarrow \frac{A \cdot v^{(k)}}{\|A \cdot v^{(k)}\|_2}$$

until a stopping condition is met.

$\Longrightarrow v^{(k+1)}$ takes as an estimate of the **principle eigenvector** of A.
Assumption:

- A is diagonalizable (in particular, A is Hermitian (symmetric)). $A = V\Lambda V^H$.
- $|\lambda_1| \geq |\lambda_2| \geq ... \geq |\lambda_m|$ and $|\lambda_1| > |\lambda_2|$

$v^{(k+1)} = \frac{A \cdot v^{(k)}}{\|A\cdot v^{(k)}\|_2} = \frac{A^{k+1}\cdot v^{(0)}}{\|A^{k+1}\cdot v^{(0)}\|_2}$

Consider $A^k x$, x satisfies $[V^H x]_1 \neq 0$. (x is $v^{(0)}$)
Let $\alpha = V^H x$

$$A^k \cdot x = (V\Lambda^k V^H)\cdot x = V\Lambda^k\alpha$$

Let $V = [v_1, v_2, ..., v_m]$

$$\begin{aligned} A^k \cdot x &= V\Lambda^k \alpha \\ &= \textstyle\sum_{i=1}^m v_i \cdot \lambda_i^k \cdot \alpha_i \\ &= v_1(\lambda_1\alpha_1) + \textstyle\sum_{i=2}^m \alpha_i \lambda_i^k v_i \\ &= \lambda_1^k \alpha_1 (v_1 + \underbrace{\sum_{i=2}^m \frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i)}_{\triangleq r_k \text{(residule)}} \end{aligned}$$

$\Longrightarrow$

$$\begin{aligned} \|r_k\|_2 &= \|\textstyle\sum_{i=2}^m \frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i\|_2 \\ &\leq \textstyle\sum_{i=1}^m \frac{\alpha_i}{\alpha_1}\cdot\left(\frac{\lambda_i}{\lambda_1}\right)^k \cdot \underbrace{\|v_i\|_2}_{=1} \\ &\leq \underbrace{|\frac{\lambda_2}{\lambda_1}|}_{<1}{}^k \cdot \textstyle\sum_{i=1}^m |\frac{\alpha_i}{\alpha_1}| \to 0 \text{ as } k\to\infty \end{aligned}$$

Consider $c_k = \frac{|\alpha_1||\lambda_1^k|}{\alpha_1\lambda_1^k}$ and $\lim_{k\to\infty} c_k \cdot \frac{A^k\cdot x}{\|A^k \cdot x\|_2}$

$$\begin{aligned} & c_k \cdot \frac{A^k\cdot x}{\|A^k\cdot x\|_2} \\ &= \frac{|\alpha_1||\lambda_1^k|}{\alpha_1\lambda_1^k}\cdot\frac{1}{\|A^k\cdot x\|_2^\dagger}\cdot\lambda_1^k\alpha_1\left(v_1 + \textstyle\sum_{i=2}^m \frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i\right) \\ &= \frac{|\alpha_1||\lambda_1^k|}{\sqrt{\sum_{i=1}^m(\lambda_i^k)^2\alpha_i^2}}^{\ddagger}\cdot\left(v_1 + \underbrace{\sum_{i=2}^m \frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i}_{\to 0 \text{ as } k\to\infty}\right) \end{aligned}$$

$\dagger$: $\|A^k x\|_2 = \|V\Lambda^k V^H x\|_2 = \|V\Lambda^k\alpha\|_2 = \|\Lambda^k\alpha\|_2 = \sqrt{\sum_{i=1}^m(\lambda_i^k\alpha_i)^2}$

$\ddagger$: $$\begin{aligned} \frac{|\alpha_1||\lambda_1^k|}{\sqrt{\sum_{i=1}^m(\lambda_i^k)^2\alpha_i^2}} &= \frac{|\alpha_1||\lambda_1^k|}{\sqrt{(\lambda_1^k)^2\alpha_1^2 + \sum_{i=2}^m(\lambda_i^k)^2\alpha_i^2}} \\ &= \frac{1}{\sqrt{1+\sum_{i=2}^m\left(\frac{\lambda_i}{\lambda_1}\right)^{2k}\left(\frac{\alpha_i}{\alpha_1}\right)^2}} \end{aligned}$$

Since $\sum_{i=2}^m \left(\frac{\lambda_i}{\lambda_1}\right)^{2k}\left(\frac{\alpha_i}{\alpha_1}\right)^2 \to 0$ as $k\to\infty$, so $\frac{|\alpha_1||\lambda_1^k|}{\sqrt{\sum_{i=1}^m(\lambda_i^k)^2\alpha_i^2}} \to 1$ as $k \to\infty$.

Therefore, $\lim_{k\to\infty} c_k \cdot \frac{A^k\cdot x}{\|A^k\cdot x\|_2} \to v_1$.

Deflation:

$$\begin{aligned} A &= \textstyle\sum_{i=1}^m \lambda_i v_i v_i^H \\ &= \lambda_1 v_1 v_1^H + \textstyle\sum_{i=2}^m \lambda_i v_i v_i^H \end{aligned}$$

Obtain $A^{(1)} = A - \lambda_1 v_1 v_1^H$ and apply power method to obtain v_2
Obtain $A^{(2)} = A - \lambda_2 v_2 v_2^H$ and apply power method to obtain v_3
...

Lecture 5. PageRank

PageRank Model

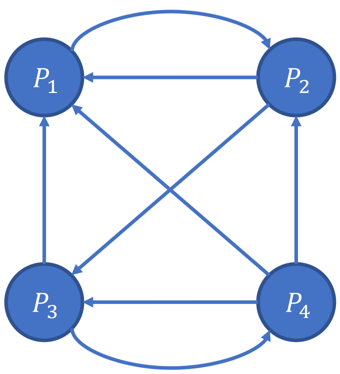
Importance score

- v_i: importance score of page i
- c_j: number of outgoing links of page j
- $\mathcal{L}_i$: set of pages that refers to page i

$$\begin{gathered} v_i = |\mathcal{L}_i| \\ \Downarrow \\ v_i = \sum_{j \in \mathcal{L}_i} v_j \\ \Downarrow \\ v_i = \sum_{j \in \mathcal{L}_i} \frac{v_j}{c_j} \end{gathered}$$

The model

e.g.



$v_1 = 1 \cdot v_2 + \frac{1}{2} v_3 + \frac{1}{3} v_4$
$v_2 = 1 \cdot v_1 + \frac{1}{3} v_4$
...
$\Longrightarrow$

$$\underbrace{\begin{bmatrix} 0 & 1/2 & 1/2 & 1/3 \\ 1 & 0 & 0 & 1/3 \\ 0 & 1/2 & 0 & 1/3 \\ 0 & 0 & 1/2 & 0 \end{bmatrix}}_{\text{link matrix}} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

$\Longrightarrow$

$$Av = v \quad , \quad v \geq 0$$

PageRank Problem

Find solution $v \geq 0$ satisfying $Av = v$
The foure questions:

1. Does A admit a solution of $Av = v$?
   (Does A has eigenvalue equal to 1?)
2. Does $Av = v$ admit a non-negative solution?
   ($v \geq 0$ or $v \leq 0$)
3. Is the solution unique?
   (unique ranking)
4. Can we obtain the solution efficiently?
   (Power method applicable?)

Answer to Question 1

Does A admit a solution of $Av = v$? (Does A has eigenvalue equal to 1?)

Note that $A \geq 0$ and column-sum eqaul to 1 (column-stochastic matrix).
$\Longrightarrow \mathbf{1}^T A = \mathbf{1}^T$
$\Longrightarrow A^T \cdot \mathbf{1} = \mathbf{1}$
$\Longrightarrow A^T$ has an eigenvalue equal to 1
Since A and A^T have the same set of eigenvalues
$\Longrightarrow A$ has an eigenvlue equal to 1

> **Theorem 6**
> Let A be non-negative and column(row)-stochastic, then
>
> 1. $\lambda = 1$ is an eigenvlue of A
> 2. $|\lambda| \leq 1$ for all eigenvalues of A

Proof Let $Av = \lambda v$ for some $|\lambda| \neq 1$.
Let $|v_k| = \max\{|v_1|, |v_2|, ..., |v_m|\}$.

$$\lambda v_k = \sum_{i=1}^{m} [A]_{ki} v_i$$

$\Longrightarrow$

$$\begin{aligned} |\lambda||v_k| &= |\textstyle\sum_{i=1}^m [A]_{ki} v_i| \\ &\leq \textstyle\sum_{i=1}^m |[A]_{ki}| \cdot |v_i| \\ &\leq |v_k| \cdot \textstyle\sum_{i=1}^m |[A]_{ki}| \\ &= |v_k| \end{aligned}$$

$\Longrightarrow$

$$|\lambda| \leq 1$$

$\square$

Disjoint page networks

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$$

$$A_1 v_1 = v_1$$
$$A_2 v_2 = v_2$$

$$A\left(\alpha_1 \begin{pmatrix} v_1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} v_1 \\ 0 \end{pmatrix}\right) = \left(\alpha_1 \begin{pmatrix} v_1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} v_1 \\ 0 \end{pmatrix}\right)$$

Modified model

$$M \triangleq (1-\beta)A + \beta\left(\frac{11^T}{m}\right) > 0$$

This is positive column-stochastic

$$1^T M = (1-\beta)1^T A + \beta 1^T \left(\frac{11^T}{m}\right) = 1^T$$

Answer to Question 2

Does $Av = v$ admit a non-negative solution? ($v \geq 0$ or $v \leq 0$)

> **Theorem 7**
> For positive and column-stochastic matrix M, any eigenvector associated with λ_1 has either all positive or all negative components.

Proof Note that for $y \in \mathbb{R}^m$, $|\sum_{i=1}^m y_i| < \sum_{i=1}^m |y_i|$ if $\{y_i\}$ have mixed signs.
Suppose the above theorem is not true. Then we have

$$|v_i| = |\sum_{j=1}^m [M]_{ij} v_j| < \sum_{j=1}^m [M]_{ij}|v_j|$$

$\Longrightarrow$

$$\sum_{i=1}^m |v_i| < \sum_{i=1}^m \sum_{j=1}^m [M]_{ij} v_j = \sum_{j=1}^m \sum_{i=1}^m [M]_{ij} v_j = \sum_{j=1}^m |v_j|$$

Contradict! So the above theorem is true. □

Answer to Question 3

Is the solution unique? (unique ranking)
Since $Mv = v \Longrightarrow (M - I)v = 0$, show

$$dim(Null(M - I)) = 1$$

i.e., there is a unique ranking.

Proof

Simple lemma: Let a and b be linearly independent, then $\exists x \in span\{a, b\}$ which has mixed sign.
Proof for this lemma:
If $1^T q = 0 \Longrightarrow q$ has mixed sign.
If $1^T q \neq 0$, set $\alpha_1 = \frac{-1^T b}{1^T q}$, $\alpha_2 = 1$ and let $x = \alpha_1 \cdot a + \alpha_2 \cdot b$
$\Longrightarrow x$ satisfies $1^T x = 0$, so x has mixed sign.

Suppose $dim(Null(M - I)) > 1$, then for $q_1, q_2 \in Null(M - I)$ that are linear independent. $\Longrightarrow \exists v = \alpha_1 q_1 + \alpha_2 q_2 \in Null(M - I)$
$\Longrightarrow Mv = v$ which has mixed sign. (Contradict!)
$\Longrightarrow dim(Null(M - I)) = 1$ □

Answer to Question 4

Can we obtain the solution efficiently? (Power method applicable?)
Property: Let M be positive and column-stochastic.
Let $S = \{v \in \mathbb{R}^m | 1^T v = 0\}$.
Then,

1. $Mv \in S$ if $v \in S$
2. $\|Mv\|_1 \leq c\|v\|_1$ with $c < 1$ $\forall$ any $v \in S$
3. the eigenvector associated with $\lambda = 1$, denoted v_1, can be obtainable by the power method

Proof
1) $v \in S \Longleftrightarrow 1^T v = 0$
2) □

Lecture 6. Positive Semidefinite Matrix

For $A \in \mathbb{C}^{m\times m}$, if

- A is symmetric, we denote $A \in S^m$
- A is Hermitian, we denote $A \in H^m$

> **Definition**
> We say $A \in H^m$ is positive semidefinite (PSD) if
> $$x^H Ax \geq 0 \quad \forall x \in \mathbb{C}^m$$
> We say $A \in H^m$ is positive definite (PD) if
> $$x^H Ax > 0 \quad \forall x \in \mathbb{C}^m, x \neq 0$$
> We say $A \in H^m$ is indefinite if A is not PSD.

Properties of PSD Matrices

Propeties

Property 1: If A is PSD then any principal submatrix of A (A_I) is PSD.
Proof
Let $x \in \mathbb{C}^m$ with $x_{i_k} \neq 0$ for all $i_k \notin I$.
Let $(x_I)_k = x_{i_k}$ for $i_k \in I$.
Then
$$x^H Ax = x_I^H A_I x_I \geq 0, \quad \forall x_I \in \mathbb{C}^{|I|}$$
$\square$

By Property 1, we know
1)
$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$
$$A_{11} \in H^{k\times k}, A_{22} \in H^{(m-k)\times(m-k)}$$
If A is PSD(PD) $\Longrightarrow A_{11}$, A_{22} are PSD(PD). 2)
If A is PSD(PD), then $A_{ii} \geq 0 (> 0)$

Eigenvalues of PSD matrices

> **Theorem 1**
> $A \in H^m$ is PSD(PD) if and only if $\lambda_1, \lambda_2, ..., \lambda_3 \geq 0 (> 0)$

Proof
Consider EVD of A as
$$A = V\Lambda V^H, \quad \Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_m)$$
Consider $x^H Ax = x^H V\Lambda V^H x = z^H \Lambda z$.

A is PSD $\Longleftrightarrow x^H Ax \geq 0 \; \forall x \in \mathbb{C}^m \Longleftrightarrow z^H\Lambda z \geq 0 \; \forall z \in \mathbb{C}^m$.

$z^H \Lambda z = \sum_{i=1}^m \lambda_i |z_i|^2 \geq 0 \; \forall z \in \mathbb{C}^m \Longleftrightarrow \lambda_i \geq 0 \; \forall i = 1, ..., m$ $\square$

If A is PSD(PD),

- $Tr(A) \geq 0 (> 0)$
- $det(A) \geq 0 (> 0)$

If A is PD, then A is invertible.
$$A = V\Lambda V^H$$
$$A^{-1} = V\Lambda^{-1}V^H$$

Property 2: $A \in H^m$ can be factorized as $A = BB^H$ for some matrix B if and only if A is PSD.
Proof
$\Longrightarrow$:
$$x^H Ax = x^H BB^H x = \|B^H x\|^2 \geq 0 \Longrightarrow A\text{is PSD}$$
$\Longleftarrow$:
If A is PSD, define $\Lambda^{1/2} = diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_m})$
$$\begin{aligned} A &= V\lambda V^H \\ &= V\Lambda^{1/2}\Lambda^{1/2}V^H \\ &= V\Lambda^{1/2}Q^H Q\Lambda^{1/2}V^H \end{aligned}$$
where Q is a Hermitian matrix.
Define $B \triangleq V\Lambda^{1/2}Q^H$. Then, $A = BB^H$. $\square$

> **Theorem 2**
> Consider $A \in H^m$, $B \in \mathbb{C}^{m\times n}$ and $C = B^H AB$
>
> - If A is PSD, then C is PSD.
> - If A is PD, then C is PD if and only if $rank(B) = n$
> - If B is square and invertible, then A is PSD(PD) if and only if C is PSD(PD)

Proof of 2)
($\Longrightarrow$) Show C is PSD $\Longrightarrow rank(B) = n$
Assume B is not full column rank.
Consider $x^H Cx = x^H B^H ABx$, then there exists $x \neq 0$ s.t. $Bx = 0 \Longrightarrow x^H Cx = 0$ (Contradict!) ($\Longleftarrow$) Show $rank(B) = n \Longrightarrow$ Show C is PSD
$rank(B) = n \Longrightarrow Bx = 0$ iff $x = 0$. So $\forall x \neq 0 \Longrightarrow Bx \neq 0$.
$\Longrightarrow x^H Cx = x^H B^H ABx > 0$ as A is PD. So C is PD.
$\square$

Property 3: Let $A \in \mathbb{C}^{m\times k}$ and $B \in \mathbb{C}^k$ and B has full row rank ($rank(B) = k$). Then
$$Range(A) = Rnage(AB)$$
Proof
$Range(B) = \mathbb{C}^k$ as B is full row rank.
$$\begin{aligned} Range(AB) &= \{y|y = ABx, x \in \mathbb{C}^n\} \\ &= \{y|y = Az, \underbrace{z = Bx, x \in \mathbb{C}^n}_{z\in Range(B)=\mathbb{C}^k}\} \\ &= \{y|y = Az, z \in \mathbb{C}^k\} \\ &= Range(A) \end{aligned}$$
$\square$

Property 4: Let $B \in \mathbb{C}^{m\times k}$ and $C \in \mathbb{C}^{m\times k}$ be two matrices with full column rank. Then

$$BB^H = CC^H$$

if and only if $C = BQ$ for some unitary Q. *Proof*
($\Longleftarrow$): obvious.
($\Longrightarrow$): $(C^H)^+ \triangleq C(C^HC)^{-1}$. Since $rank(C^HC) = rank(C)$, C^HC is invertible †.
†: As C has full column rank, then $Cx = 0 \Longleftrightarrow x = 0$. $Cx = 0 \Longrightarrow C^HCx = 0 \Longrightarrow x^HC^HCx = 0 \Longleftrightarrow \|Cx\|^2 = 0 \Longrightarrow x = 0$. Therefore, $C^HCx = 0 \Longrightarrow x = 0$ and hence $rank(C^HC) = rank(C)$.

$$\begin{aligned} BB^H(C^H)^+ &= CC^H(C^H)^+ \\ &= CC^HC(C^HC)^{-1} \\ &= C \end{aligned}$$

$\Longrightarrow$

$$\begin{aligned} C &= BB^H(C^H)^+ \\ &= B\underbrace{B^HC(C^HC)^{-1}}_{\triangleq Q(\text{square})} \end{aligned}$$

$$\begin{aligned} Q^HQ &= (C^HC)^{-H}C^HBB^HC(C^HC)^{-1} \\ &= \underbrace{(C^HC)^{-H}C^HC}_{=I}\underbrace{C^HC(C^HC)^{-1}}_{=I} \\ &= I \end{aligned}$$

Therefore, Q is unitary. $\square$