

CIE 6002 Matrix Analysis

Binghao He

February 26, 2021

Contents

Lecture 2. Fundamental and Basic Operations	3
Subspace	3
Definition	3
Some important subspaces	3
Linear independence	3
Basis	3
Rank	3
Nonsingularity/Invertibility	3
Determinant	3
Vector Norm	3
Inner Product	4
Projection onto a Subspace	4
Projection	4
Orthogonal complement	4
Sum of subspaces	4
Orthogonal Set	4
Orthogonal/Unitary Matrix	4
Dimensional Theorem	5
Complexity of Matrix Computation	5
Flops	5
Big O notation	5
Lecture 3. Least Square Problem (LSP)	6
Formulation	6
LS problem	6
Linear representation of signals	6
AR and Polynomial Models	6
Autoregressive (AR) model	6
Polynomial model	6
Basis Representation	6
Basis matrix	6
Fourier basis	6
LTI System	7
Lienar time invariant system	7
System identification	7
Deconvolution	7
Toeplitz matrix	7

Circulant Matrix	7
Lecture 4. Eigenvalues and Eigenvectors	8
Diagonalizability	8
Similar matrices	8
Diagonalizable matrix	8
When Diagonalizable	8
Distinct eigenvalues	8
Repeated eigenvalues	9
Symmetric/Hermitian Matrices	9
Unitarily diagonalizable	10
Schur triangularization	10
Variational Characterization of Eigenvalues	11
Rayleigh quotient	11
Courant-Fischer theorem	11
Weyl's inequality	12
The Power Method	12
Lecture 5. PageRank	14
PageRank Model	14
Importance score	14
The model	14
PageRank Problem	14
Answer to Question 1	14
Answer to Question 2	15

Lecture 2. Fundamental and Basic Operations

Subspace

Definition

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

This indicates

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

Some important subspaces

1. Span: a collection of vectors

$$\text{span}\{a_1, \dots, a_n\} = \{y \in \mathbb{R}^m | y = \sum_{i=1}^n \alpha_i a_i, \alpha_i \in \mathbb{R}\}$$

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)

$$\begin{aligned} R(A) &= \text{span}\{a_1, a_2, \dots, a_n\} \\ &= \{y \in \mathbb{R}^m | y = Ax, x \in \mathbb{R}^n\} \end{aligned}$$

4. Null space of A (linear comb. of col. vectors)

$$\text{Null}(A) = \{x \in \mathbb{R}^n | Ax = 0\}$$

Linear independence

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

$$\sum_{i=1}^n \alpha_i a_i = 0 \implies \alpha_i = 0, \forall i = 1, \dots, n$$

Basis

$\{b_1, b_2, \dots, b_k\} \subseteq \mathbb{R}^m$ is a basis for the subspace $S \subseteq \mathbb{R}^m$ if $S = \text{span}\{b_1, b_2, \dots, b_k\}$ and $\{b_1, b_2, \dots, b_k\}$ 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 (?)

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

And

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

Nonsingularity/Invertibility

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

$$Ax = 0 \implies x = 0$$

\implies columns of A are linearly indep.

$\implies \exists A^{-1} \in \mathbb{R}^{m \times m}$ 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 sub-matrix that removes i th row and j th column of A .

Define: $c_{ij} = (-1)^{i+j} \det(A_{ij})$.

$$\begin{aligned} \det(A) &= \sum_{j=1}^m a_{ij} c_{ij} \quad \forall i = 1, \dots, m \\ &= \sum_{i=1}^m a_{ij} c_{ij} \quad \forall j = 1, \dots, m \end{aligned}$$

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

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

Vector Norm

A mapping $f : \mathbb{R}^> \rightarrow \mathbb{R}$ 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)
- $\|x\|_1 = \sum_{i=1}^m |x_i|$
- $\|x\|_{\infty} = \max_{i=1, \dots, m} |x_i|$
- $\|x\|_p = \sqrt[p]{\sum_{i=1}^m |x_i|^p}$

Inner Product

$$\langle x, y \rangle = x^T y$$

$$\langle x, x \rangle = \|x\|_2^2$$

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.
 $z^T(y_S - y) = 0 \quad \forall z \in S$

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
- $S \cap S_\perp = \{0\}$

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

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

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

$$\therefore \Pi_S(y) \in S$$

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

□

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

□

Orthogonal Set

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

1. $\|a_i\|_2 = 1 \quad \forall i = 1, \dots, n$
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 $\{a_1, a_2, \dots, a_n\}$. The procedure of finding an orthogonal set $\{q_1, q_2, \dots, q_n\}$ so that $\text{span}\{a_1, a_2, \dots, a_n\} = \text{span}\{q_1, q_2, \dots, q_n\}$ is called "Gram-Schmidt" procedure.

Orthogonal/Unitary Matrix

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

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

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

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

□

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

$$\dim(R(A)) + \dim(\text{Null}(A)) = n$$

Proof

$$\begin{aligned} \dim(R(A^T)) + \dim(R(A^T)_{\perp}) &= n \\ \dim(R(A^T)) &= \text{rank}(A^T) = \text{rank}(A) = \dim(R(A)) \\ \dim(R(A^T)_{\perp}) &= \text{Null}(A) = n \end{aligned}$$

□

Complexity of Matrix Computation

Flops

Arithmetic operations including addition, subtraction, 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 noise follow multi-variate Gaussian distribution

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

\Rightarrow 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 , $t = 0, 1, 2, \dots$,

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

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

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

Collect

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

\Rightarrow Solve LS problem to obtain α_i , $i = 1, \dots, p$

Polynomial model

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

Collect

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

\Rightarrow Solve LS problem to obtain α_i , $i = 1, \dots, p$

Basis Representation

Basis matrix

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

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

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 $\Rightarrow x$ is sparse

Fourier basis

Discrete Fourier Transform (DFT)

Inverse Discrete Fourier Transform (IDFT)

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

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

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

LTI System

Lienar time invariant system

$$\begin{aligned}
 & \text{-- Linearity} \quad \left\{ \begin{array}{l} x_t^1 \rightarrow \boxed{\text{LTI}} \rightarrow y_t^1 \\ x_t^2 \rightarrow \boxed{\text{LTI}} \rightarrow y_t^2 \\ \alpha x_t^1 + \beta x_t^2 \rightarrow \boxed{\text{LTI}} \rightarrow \alpha y_t^1 + \beta y_t^2 \end{array} \right. \\
 & \text{-- Time Invariant} \quad x_{t-d} \rightarrow \boxed{\text{LTI}} \rightarrow y_{t-d}
 \end{aligned}$$

Impulse $\delta(t)$

$$\begin{cases} \delta(t) = 0, t \neq 0 \\ \int_{-\infty}^{+\infty} \delta(t) dt = 1 \end{cases}$$

LTI system is characterized by an impulse response

$$h_t, \quad t = 0, \dots, p-1$$

$\Rightarrow x_t$ 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 $\{h_t\}_{t=0}^p$

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

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

Deconvolution

Given $\{y_t\}_{t=0}^{T-1}$ and $h_{t=0}^p$, find out $\{x_t\}_{t=0}^{T-1}$

Put them into a linear system of equations

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

\Rightarrow 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

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

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

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

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

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

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 .

\Rightarrow

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

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

For each λ_i find v_i such that

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

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

- Even if $A \in \mathbb{R}^{m \times m}$ is real-valued, λ_i and v_i could be complex. *e.g.*

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \Rightarrow \lambda = \pm j$$

- But if $\lambda_i \in \mathbb{R}$ is real-valued, $v_i \in \mathbb{R}^m$ should be in \mathbb{R}^m .

\Rightarrow (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.

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

eigenvalue decomposition (EVD)

\Rightarrow

$$AV = VD$$

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

\Leftrightarrow

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

\Rightarrow

A is diagonalizable *iff.* there is a set of linear indep. vectors which are eigenvectors 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. $\text{Tr}(A) = \sum_{i=1}^m \lambda_i$
3. $\text{rank}(A)$ = 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. $\text{Tr}(A) = \text{Tr}(VDV^{-1}) = \text{Tr}(DV^{-1}V) = \text{Tr}(D) = \sum_{i=1}^m \lambda_i$

□

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.

When Diagonalizable

Distinct eigenvalues

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

If A has $\lambda_1 \neq \lambda_2 \neq \dots \neq \lambda_k$, then the corresponding eigenvectors v_1, \dots, v_k are linearly independent.

Proof

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

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

Then

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

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

On the other hand,

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

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

(1) - (2):

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

Then

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

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

On the other hand

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

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

(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, \dots, v_k are linearly **independent** \square

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: $s = \dim(\text{Null}(A - \lambda I))$

Then

$$s \leq r$$

(1) *Proof*

Let $\dim(\text{Null}(A - \bar{\lambda}I)) = s$ and $\{q_1, \dots, q_s\}$ be an orthogonal basis for $\text{Null}(A - \bar{\lambda}I)$.

Then $Q_1 = [q_1, \dots, q_s] \in \mathbb{R}^{m \times s}$ 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

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

So

$$r \geq s$$

\square

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 $\|v\|_2 = 1$.

$$\implies Av = \lambda v$$

$$\implies v^H Av = \lambda v^H v = \lambda$$

$$\stackrel{\text{take } H}{\implies} v^H A^H v = \lambda^*$$

Since A is Hermitian, $A = A^H$ and hence $\lambda = \lambda^*$

So λ is real-valued \square

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

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

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

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

\square

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 $\lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{C}$ 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 $[T]_{ii} = \lambda_i, i = 1, 2, \dots, m$

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

$$U^H A^H U = T^H$$

\implies

$$T = T^H$$

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

$$U^H AU = T$$

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 eigenvector 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 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$

Construct a unitary matrix U_1

$$U_1 = [v_1 \quad V_1] \in \mathbb{C}^{m \times m}$$

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

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

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 $\lambda_2, \lambda_3, \dots, \lambda_m$

Let $v_2 \in \mathbb{C}^{m-1}$ be the eigenvector of A_1 associated with λ_2 .

Construct a unitary matrix U_2

$$U_2 = [v_2 \quad V_2] \in \mathbb{C}^{(m-1) \times (m-1)}$$

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

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

\implies

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

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

$$\begin{aligned} \tilde{U}_m^H \dots \tilde{U}_3^H \tilde{U}_2^H (U_1 A U_1) \tilde{U}_2 \tilde{U}_3 \dots \tilde{U}_m \\ \underbrace{\hspace{10em}}_{\triangleq U \text{ unitary}} \\ = \begin{bmatrix} \lambda_1 & \times & \times & \times \\ & \lambda_2 & \times & \times \\ & & \ddots & \times \\ & & & \lambda_m \end{bmatrix} \triangleq T \end{aligned}$$

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

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

and

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

Proof

According to the EVD of A ,

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

\Rightarrow

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

Since

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

So,

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

\Rightarrow

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

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 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ are ordered eigenvalues. Then

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

Proof ideas:

1.

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

where v_k is the eigenvector associated with λ_k of A .

2.

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

1+2 \Rightarrow

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

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

Proof

1.

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

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

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

Therefore,

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

2.

Let $y = V^H x$

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

□

Weyl's inequality

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

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

Proof 1.

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

where

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

So,

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

Therefore,

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

Proof 2.

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

Proof 3.

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

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

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

Repeat r times, we can get

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

where $\text{rank}(B) \leq r$ □

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

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

\Rightarrow

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

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

Proof 5.

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

\dagger : $e_i = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, only the i th entry equal to 1. □

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

Proof 6.

Let $\tilde{U} = \begin{bmatrix} U & \tilde{U} \end{bmatrix}$ be unitary. Then

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

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

Therefore,

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

□

□

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.

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

Assumption:

- A is diagonalizable (in particular, A is Hermitian (symmetric)). $A = V\Lambda V^H$.
- $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_m|$ 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 $V = [v_1, v_2, \dots, v_m]$

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

\Rightarrow

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

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

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

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

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

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

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

Deflation:

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

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

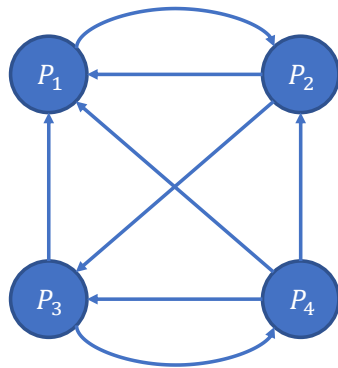
Importance score

- 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

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

The model

e.g.



$$\begin{aligned} v_1 &= 1 \cdot v_2 + \frac{1}{2}v_3 + \frac{1}{3}v_4 \\ v_2 &= 1 \cdot v_1 + \frac{1}{3}v_4 \end{aligned}$$

...

\Rightarrow

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

\Rightarrow

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

PageRank Problem

Find solution $v \geq 0$ satisfying $Av = v$

The four 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 equal to 1 (column-stochastic matrix).

$$\Rightarrow \mathbf{1}^T A = \mathbf{1}^T$$

$$\Rightarrow A^T \cdot \mathbf{1} = \mathbf{1}$$

$$\Rightarrow A^T \text{ has an eigenvalue equal to 1}$$

Since A and A^T have the same set of eigenvalues

$$\Rightarrow A \text{ has an eigenvalue equal to 1}$$

Theorem 6

Let A be non-negative and column(row)-stochastic, then

1. $\lambda = 1$ is an eigenvalue of A
2. $|\lambda| \leq 1$ for all eigenvalues of A

Proof Let $Av = \lambda v$ for some $|\lambda| \neq 1$.

Let $|v_k| = \max\{|v_1|, |v_2|, \dots, |v_m|\}$.

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

\Rightarrow

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

\Rightarrow

$$|\lambda| \leq 1$$

□

Disjoint page networks

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

$$A_1 v_1 = v_1$$

$$A_2 v_2 = v_2$$

$$A \left(\alpha_1 \begin{pmatrix} v_1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ v_2 \end{pmatrix} \right) = \left(\alpha_1 \begin{pmatrix} v_1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ v_2 \end{pmatrix} \right)$$

Modified model

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

This is positive column-stochastic

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

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

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

\implies

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

Contradict! So the above theorem is true. \square

Answer to Question 3