Proof of the Property (cont'd)

Property 3: Let v_1 be an eigenvector associated with λ_1

Let $Z^{(n-1)} = \text{span}\{v_1\}^{\perp}$ $(n-1)$-dimensional subspace

We want to show $Z^{(n-1)}$ is an invariant subspace for A.

Pick any $z \in Z^{(n-1)}$. Then, $z^H v_1 = 0$

$$(Az)^H v_1 = z^H \underbrace{A^H}_{} v_1 = z^H (A v_1) = \lambda_1 \underbrace{z^H v_1}_{} = 0$$

$\Rightarrow Az \in Z^{(n-1)}$.

Using the Fact, we know that there is an eigenvector $v_2 \in Z^{(n-1)}$ of A.

Proof of the Property (cont'd)

Next, let $Z^{(n-2)} = \text{span}\{v_1, v_2\}^{\perp}$.

Likewise, we can show $Z^{(n-2)}$ is an invariant subspace for A and thus, there is an eigenvector $v_3 \in Z^{(n-2)}$ of A.

Finite induction completes the proof.

Eigendecomposition for Hermitian Matrices

Theorem

Every $\mathbf{A} \in \mathbb{H}^n$ *admits an eigendecomposition*

$V^{-1} = V^H$

$$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^H,$$

where $\mathbf{V} \in \mathbb{C}^{n\times n}$ *is unitary,* $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ *with* $\lambda_i \in \mathbb{R}$ *for all* i*. In addition, if* $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{V}$ *can be taken as a real orthogonal matrix.*

- A special case of Schur decomposition
- No need of assuming distinct eigenvalues

Corollary: If $\mathbf{A} \in \mathbb{H}^n$, $\mu_i = \gamma_i$ for all i

Interpretation of Eigendecomposition in $\mathbb{S}^n$

$A \in \mathbb{S}^n$

$A = V \Lambda V^T$

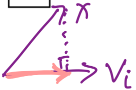
$Ax = V \Lambda V^T x$

$$\mathbf{x} \longrightarrow \boxed{\mathbf{A}} \xrightarrow{\ \mathbf{y} = \mathbf{A}\mathbf{x}\ }$$

$v_1, \cdots, v_n$ orthonormal

$$\mathbf{x} \longrightarrow \boxed{\mathbf{V}^{\mathbf{T}}} \xrightarrow{\ \mathbf{V}^{\mathbf{T}}\mathbf{x}\ } \boxed{\mathbf{\Lambda}} \xrightarrow{\ \mathbf{\Lambda}\mathbf{V}^{\mathbf{T}}\mathbf{x}\ } \boxed{\mathbf{V}} \xrightarrow{\ \mathbf{y} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathbf{T}}\mathbf{x} = \mathbf{A}\mathbf{x}\ }$$

$$V^T x = \begin{bmatrix} v_1^T x \\ \vdots \\ v_n^T x \end{bmatrix}$$



$(v_i^T x) v_i$

$$\Lambda (V^T x) = \begin{bmatrix} \lambda_1 v_1^T x \\ \vdots \\ \lambda_n v_n^T x \end{bmatrix}$$

1. $\mathbf{V}^T\mathbf{x}$: Let $\mathbf{x}$ resolve into $\mathbf{v}_1, \ldots, \mathbf{v}_n$
2. $\Lambda(\mathbf{V}^T\mathbf{x})$: Scale the ith coordinate of $(\mathbf{V}^T\mathbf{x})$ by λ_i
3. $\mathbf{V}(\Lambda\mathbf{V}^T\mathbf{x})$: Reconstitute $(\Lambda\mathbf{V}^T\mathbf{x})$ with basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$

$Vz = z_1 v_1 + \cdots + z_n v_n$, $\quad V(\Lambda V^T x) = (\lambda_1 v_1^T x) v_1 + \cdots + (\lambda_n v_n^T x) v_n$

Courant-Fischer Min-Max Theorem

For $\mathbf{A} \in \mathbb{H}^{n\times n}$, let $\lambda_k(\mathbf{A})$ denote the kth largest eigenvalue of $\mathbf{A}$, i.e.,

$$\lambda_n(\mathbf{A}) \leq \cdots \leq \lambda_1(\mathbf{A})$$

real eigenvalues

Theorem

For any $\mathbf{A} \in \mathbb{H}^{n\times n}$ and $k = 1, \ldots, n$,

$$\lambda_k(\mathbf{A}) = \max_{\substack{\mathcal{S}\subseteq\mathbb{C}^n: \\ \dim(\mathcal{S})=k}} \min_{\substack{\mathbf{y}\in\mathcal{S}, \\ \mathbf{y}\neq\mathbf{0}}} \frac{\mathbf{y}^H\mathbf{A}\mathbf{y}}{\mathbf{y}^H\mathbf{y}} = \max_{\substack{S\subseteq\mathbb{C}^n \\ \dim(S)=k}} \min_{\substack{y\in S \\ \|y\|_2=1}} y^H A y$$

all k-dim subspaces of $\mathbb{C}^n$

$$= \min_{\substack{\mathcal{S}\subseteq\mathbb{C}^n: \\ \dim(\mathcal{S})=n-k+1}} \max_{\substack{\mathbf{y}\in\mathcal{S}, \\ \mathbf{y}\neq\mathbf{0}}} \frac{\mathbf{y}^H\mathbf{A}\mathbf{y}}{\mathbf{y}^H\mathbf{y}}$$

$k=n$: $S=\mathbb{C}^n$

$$\lambda_n(A) = \min_{\substack{y\in\mathbb{C}^n \\ y\neq 0}} \frac{y^H A y}{y^H y}$$

$R_{\mathbf{A}}(\mathbf{y}) = \frac{\mathbf{y}^H\mathbf{A}\mathbf{y}}{\mathbf{y}^H\mathbf{y}}$, $\mathbf{y} \neq \mathbf{0}$ is called the Rayleigh–Ritz quotient

- $R_{\mathbf{A}}(\mathbf{y})$ can be replaced with $\mathbf{y}^H\mathbf{A}\mathbf{y}$, $\|\mathbf{y}\|_2 = 1$
- If $\mathbf{y}$ is an eigenvector of $\mathbf{A}$, $R_{\mathbf{A}}(\mathbf{y})$ is its associated eigenvalue
- **Consequence** of theorem: $\lambda_n(\mathbf{A}) \leq R_{\mathbf{A}}(\mathbf{y}) \leq \lambda_1(\mathbf{A})$

$$\underline{\lambda_n(A) \le R_A(y) \le \lambda_1(A)}$$

$A \in \mathbb{H}^n \Rightarrow \exists$ orthonormal eigenvectors $\underbrace{v_1, \cdots, v_n}_{\text{orthonormal basis of } \mathbb{C}^n} \in \mathbb{C}^n$

Let $y \in \mathbb{C}^n$. Then, $y = \alpha_1 v_1 + \cdots + \alpha_n v_n$; $\alpha_i \in \mathbb{C}$

$$y^H A y = y^H (\alpha_1 \lambda_1 v_1 + \cdots + \alpha_n \lambda_n v_n)$$

$$= |\alpha_1|^2 \lambda_1 \underbrace{\|v_1\|_2^2}_{=1} + \cdots + |\alpha_n|^2 \lambda_n \underbrace{\|v_n\|_2^2}_{=1}$$

$$= \lambda_1 |\alpha_1|^2 + \cdots + \lambda_n |\alpha_n|^2, \quad \lambda_1 \ge \cdots \ge \lambda_n$$

Let y be a unit vector.

$$\|y\|_2^2 = (\alpha_1 v_1 + \cdots + \alpha_n v_n)^H (\alpha_1 v_1 + \cdots + \alpha_n v_n)$$

$$= |\alpha_1|^2 + \cdots + |\alpha_n|^2 = 1$$

$y^H A y$ achieves maximum λ_1 when $|\alpha_1|^2 = 1$

minimum λ_n $|\alpha_n|^2 = 1$

Proof

Poincaré's Inequality: Let S be a subspace of $\mathbb{C}^n$ with $\dim(S) = k$. There exist unit vectors $\mathbf{x}, \mathbf{y} \in S$ s.t. $\mathbf{x}^H\mathbf{A}\mathbf{x} \leq \lambda_k(\mathbf{A})$ and $\mathbf{y}^H\mathbf{A}\mathbf{y} \geq \lambda_{n+1-k}(\mathbf{A})$.

Proof: Pick any $v_k, \cdots, v_n$ orthonormal eigenvectors in $\mathbb{C}^n$. $\quad v_i \to \lambda_i(A)$

Let $N = \text{span}\{v_k, \cdots, v_n\}$ $\quad \dim(N) = n-k+1$.

N must intersect S on at least a single line

(because $\underbrace{\dim(S+N)}_{\leq n} = \underbrace{\dim(S)}_{k} + \underbrace{\dim(N)}_{n-k+1} - \underbrace{\dim(S \cap N)}_{\geq 1}$)

Pick any $x \in S \cap N$ with $\|x\|_2 = 1$

$x \in N \Rightarrow x = \sum_{i=k}^{n} \alpha_i v_i$. $\quad \|x\|_2 = 1$, $v_k, \cdots, v_n$ orthonormal

$\Rightarrow \sum_{i=k}^{n} |\alpha_i|^2 = 1$

$$x^HAx = \left(\sum_{i=k}^{n} \alpha_i^* v_i^H\right)\left(\sum_{i=k}^{n} \alpha_i \lambda_i(A) v_i\right) = \sum_{i=k}^{n} |\alpha_i|^2 \lambda_i(A)$$

$\leq \lambda_k(A)$

The second part can be proved by setting A to $-A$. $\blacksquare$

Proof (cont'd)

Now let $S = \mathrm{span}\{v_1, \dots, v_k\}$, $v_i \to \lambda_i(A)$

$v_1, \dots, v_k$ orthonormal eigenvectors

$$\lambda_k(A) = v_k^H (\lambda_k(A) v_k) = v_k^H A v_k = \min_{\substack{x \in S \\ \|x\|_2 = 1}} x^H A x$$

$$\leq \max_{\substack{S' \subseteq \mathbb{C}^n \\ \dim(S') = k}} \left(\min_{\substack{x \in S' \\ \|x\|_2 = 1}} x^H A x \right). \quad (1)$$

On the other hand, from the Poincaré's Inequality,

$$\lambda_k(A) \geq x^H A x \text{ for some } x \in S' \cap N, \quad \|x\|_2 = 1$$

S' k-dim subspace of $\mathbb{C}^n$

$$\geq \min_{\substack{x \in S' \\ \|x\|_2 = 1}} x^H A x$$

N in the proof of Poincaré

Proof (cont'd)

Since S' can be any k-dim subspace of $\mathbb{C}^n$,

$$\lambda_k(A) \geq \max_{\substack{S' \subseteq \mathbb{C}^n \\ \dim(S')=k}} \min_{\substack{x \in S' \\ \|x\|_2=1}} x^H A x \qquad (2)$$

Combining (1) and (2) gives $\lambda_k(A) = \max\min \cdots$

The second equation can be proved similarly using the second part of Poincaré's Inequality.

Matrix Computations
Chapter 4: Eigenvalues, Eigenvectors, and Eigendecomposition

Section 4.4 Power Iteration and QR Iteration

Jie Lu
ShanghaiTech University

The Power Method

- A method for numerically computing an eigenvector of a given matrix
- Simple, though not the best in convergence speed
  - A comprehensive coverage of various computational methods for the eigenvalue problem can be found in Chapter 7 of textbook
- Suitable for large-scale sparse problems, e.g., PageRank

The Power Method/Power Iteration

Suppose $\mathbf{A} \in \mathbb{C}^{n\times n}$ admits an eigendecomposition $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1}$

The eigenvalues of $\mathbf{A}$ are ordered as $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$

eigenvector v_1

Algorithm: Power Method
input: $\mathbf{A} \in \mathbb{C}^{n\times n}$ and an initial guess $\mathbf{v}^{(0)} \in \mathbb{C}^n$
for $k = 1, 2, \ldots$ (until a termination criterion is satisfied)
$\quad \tilde{\mathbf{v}}^{(k)} = \mathbf{A}\mathbf{v}^{(k-1)}$
$\quad \mathbf{v}^{(k)} = \tilde{\mathbf{v}}^{(k)}/\|\tilde{\mathbf{v}}^{(k)}\|_2$
$\quad \lambda^{(k)} = [\mathbf{v}^{(k)}]^H\mathbf{A}\mathbf{v}^{(k)}$
end
output: $\mathbf{v}^{(k)}$, $\lambda^{(k)}$

Complexity per iteration: $\mathcal{O}(n^2)$, or $\mathcal{O}(nzz(\mathbf{A}))$ for sparse $\mathbf{A}$

Result: $\mathrm{dist}(\mathrm{span}\{\mathbf{v}^{(k)}\}, \mathrm{span}\{\mathbf{v}_1\}) \to 0$ and $\lambda^{(k)} \to \lambda_1$ as $k \to \infty$

The convergence rates depend on $|\lambda_2|/|\lambda_1|$

Analysis of The Power Method

A has an eigendecomposition $\Rightarrow \exists$ linearly independent eigenvectors $v_1, \ldots, v_n$ basis of $\mathbb{C}^n$

Let the initial guess

$V = [v_1 \cdots v_n]$

$$\mathbf{v}^{(0)} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n = \mathbf{V}\boldsymbol{\alpha}$$

We require $\alpha_1 \neq 0$ (random guess essentially works). Then,

$$\mathbf{A}^k \mathbf{v}^{(0)} = \underbrace{\mathbf{V}\Lambda^k \mathbf{V}^{-1}}_{\text{eigendecomposition of } A^k} \mathbf{v}^{(0)} = \sum_{i=1}^{n} \alpha_i \lambda_i^k \mathbf{v}_i = \alpha_1 \lambda_1^k \left(\mathbf{v}_1 + \underbrace{\sum_{i=2}^{n} \frac{\alpha_i}{\alpha_1} \left(\frac{\lambda_i}{\lambda_1} \right)^k \mathbf{v}_i}_{=\mathbf{r}^{(k)}} \right)$$

where $\mathbf{r}^{(k)}$ is a residual satisfying

$$\|\mathbf{r}^{(k)}\|_2 \leq \sum_{i=2}^{n} \left| \frac{\alpha_i}{\alpha_1} \right| \left| \frac{\lambda_i}{\lambda_1} \right|^k \|\mathbf{v}_i\|_2 \leq \left| \frac{\lambda_2}{\lambda_1} \right|^k \sum_{i=2}^{n} \left| \frac{\alpha_i}{\alpha_1} \right| \|\mathbf{v}_i\|_2 \to 0 \text{ as } k \to \infty$$

$\to [0, 1)$

Analysis of The Power Method (cont'd)

Note from the algorithm that $\mathbf{v}^{(k)} \in \mathrm{span}\{\mathbf{A}^k \mathbf{v}^{(0)}\}$

In fact, it can be verified that $\mathbf{v}^{(k)} = \dfrac{\mathbf{A}^k \mathbf{v}^{(0)}}{\|\mathbf{A}^k \mathbf{v}^{(0)}\|_2}$

Hence, $\mathbf{v}^{(k)}$ converges to an eigenvector associated with λ_1, i.e.,

$$\mathrm{dist}(\mathrm{span}\{\mathbf{v}^{(k)}\}, \mathrm{span}\{\mathbf{v}_1\}) = \mathcal{O}(|\tfrac{\lambda_2}{\lambda_1}|^k)$$

Accordingly,

$$|\lambda^{(k)} - \lambda_1| = \mathcal{O}(|\frac{\lambda_2}{\lambda_1}|^k)$$

The convergence is slow if $|\lambda_2|$ is closer to $|\lambda_1|$

The two conditions in red require that

- λ_1 is a dominant eigenvalue (i.e., > all the other eigenvalues in modulus)
- The initial guess has a component in the direction of the corresponding dominant eigenvector

Without these conditions, the power method does not necessarily converge

Deflation

- The power method only computes the dominant eigenvalue and eigenvector
- How can we compute all the eigenvalues with the corresponding eigenvectors?

Consider a Hermitian matrix **A** with $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|$
Express **A** using the outer-product representation

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^H = V \Lambda V^H$$

Deflation: Use the power method to obtain $\mathbf{v}_1$, λ_1. Then, do the subtraction

$$\mathbf{A} - \lambda_1 \mathbf{v}_1 \mathbf{v}_1^H = \sum_{i=2}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^H + 0 \cdot v_1 v_1^H$$

eigendecomposition

Apply the power method to the above matrix and obtain $\mathbf{v}_2$, λ_2
Repeat until all the eigenvalues and eigenvectors are found

- Stop when completing the kth iteration gives the first k eigen-pairs