Lecture 19:

Inverse power method with shift

Goal: Take $\mu \in \mathbb{R}$. Find the eigenvalue of A closest to μ.

Observation: Consider $B = A - \mu I$. Then B has eigenvalues:

$$\{\lambda_1 - \mu, \lambda_2 - \mu, \ldots, \lambda_n - \mu\} \leftarrow$$

Inverse Power method find eigenvalues such that $|\lambda_j - \mu|$ is the smallest.

$\therefore$ λ_j closest to μ can be found.

Algorithm: (Inverse power method with shift)

Step 1: Take $\mu \in \mathbb{R}$. Pick $\vec{x}^{(0)}$ such that $\|\vec{x}^{(0)}\|_\infty = 1$.

Step 2: For $k = 1, 2, \ldots$

Solve: $(A - \mu I)\vec{w} = \vec{x}^{(k-1)}$ for $\vec{w}$.

Let: $\vec{x}^{(k)} = \dfrac{\vec{w}}{\|\vec{w}\|_\infty}$.

Let $\rho_k = \|A\vec{x}^{(k)}\|_\infty$ ($\rho_k \to |\lambda_j|$ as $k \to \infty$)

Recall:

Power's method reads: $\vec{x}^{(k+1)} = \dfrac{A\vec{x}^{(k)}}{\|A\vec{x}^{(k)}\|_\infty}$ for $k=0,1,\dots$

$$\Rightarrow \quad \vec{x}^{(k)} = \frac{A^k \vec{x}^{(0)}}{\|A^k \vec{x}^{(0)}\|_\infty}$$

Suppose A is diagonalizable. That's, we can assume $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ form a basis for $\mathbb{C}^n$.

Take $\vec{x}^{(0)} = a_1\vec{x}_1 + a_2\vec{x}_2 + \dots + a_n\vec{x}_n$ (assuming $a_1 \neq 0$)

$$\vec{x}^{(k)} = \frac{a_1 \lambda_1^k \left[\vec{x}_1 + \sum_{j=2}^{n} \frac{a_j}{a_1} \left(\frac{\lambda_j}{\lambda_1}\right)^k \vec{x}_j \right]}{\left\| a_1 \lambda_1^k \left[\vec{x}_1 + \sum_{j=2}^{n} \frac{a_j}{a_1} \left(\frac{\lambda_j}{\lambda_1}\right)^k \vec{x}_j \right] \right\|_\infty} \approx \underbrace{\frac{a_1}{|a_1|} \frac{\vec{x}_1}{\|\vec{x}_1\|_\infty} \frac{\lambda_1^k}{|\lambda_1|^k}}_{\vec{v}}$$

when k is big

(Because $1 > \left|\frac{\lambda_2}{\lambda_1}\right| \geq \left|\frac{\lambda_3}{\lambda_1}\right| \geq \dots \geq \left|\frac{\lambda_n}{\lambda_1}\right|$)

Rate of convergence of Power's method depends on $\left|\frac{\lambda_2}{\lambda_1}\right|$

- λ_2 ← second largest eigenvalue
- λ_1 ← largest eigenvalue

In fact,

$$\|A\vec{x}^{(k)}\|_\infty \to \left\| A\left(\frac{a_1}{|a_1|} \frac{\vec{x}_1}{\|\vec{x}_1\|_\infty}\right)\right\|_\infty = \left\| \frac{a_1}{|a_1|} \lambda_1 \frac{\vec{x}_1}{\|\vec{x}_1\|_\infty} \right\|_\infty = |\lambda_1|$$

$$\|A\vec{x}^{(k)}\|_\infty = |\lambda_1| + O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$$

Convergence rate:

$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|$

1. Power method:

Converges if $\eta \stackrel{\text{def}}{=} \left|\frac{\lambda_2}{\lambda_1}\right| < 1$ and $\langle \vec{v}_1, \vec{x}^{(0)} \rangle \neq 0$ ($\vec{v}_1$ = eigenvector of λ_1)

Also, $\rho_k = \|A\vec{x}^{(k)}\|_\infty = |\lambda_1| + O(\eta^k)$ (Slow convergence if $\eta \approx 1$!)

2. Inverse Power method:

Converges if $\left|\frac{1/\lambda_{n-1}}{1/\lambda_n}\right| = \left|\frac{\lambda_n}{\lambda_{n-1}}\right| < 1$ and $\langle \vec{v}_n, \vec{x}^{(0)} \rangle \neq 0$ ($\vec{v}_n$ = eigenvector of λ_n)

Also, $\rho_k = \|A\vec{x}^{(k)}\|_\infty = |\lambda_n| + O(\eta^k)$ (Slow convergence if $\eta \approx 1$!)

3. Inverse Power method with shift, let λ_j be closest to μ.

Converges if: $\eta = \max_{m \neq j} \left|\frac{\lambda_j - \mu}{\lambda_m - \mu}\right| < 1$ and $\langle \vec{v}_j, \vec{x}^{(0)} \rangle \neq 0$ ($\vec{v}_j$ = eigenvector of λ_j.)

$\rho_k = \|A\vec{x}^{(k)}\|_\infty = |\lambda_j| + O(\eta^k)$ (Slow convergence if $\eta \approx 1$!)

How to speed up convergence? Let $A \in M_{n\times n}(\mathbb{R})$

Idea: Use Inverse Power method with shift, update μ in each iteration (such that μ is closer to a real eigenvalue in each iteration)

Then: $\eta \overset{def}{:=} \max_{m\neq j} \left| \frac{\lambda_j - \mu}{\lambda_m - \mu} \right|$ becomes smaller and smaller $\Rightarrow$ Converges faster and faster!

Definition: (Rayleigh quotient) Let $\vec{v} \neq \vec{0} \in \mathbb{R}^n$, $A \in M_{n\times n}$. Then, the Rayleigh quotient is defined as:

$$R(\vec{v}, A) = \frac{\vec{v}^* A \vec{v}}{\vec{v}^* \vec{v}}.$$

Remark: Let A be symmetric positive definite. Then: all eigenvalues:

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are real.

Then: $\lambda_n \leq R(\vec{v}, A) \leq \lambda_1$ and

$R(\vec{v}, A) = \lambda_1$ when $\vec{v} = \vec{v}_1$ = eigenvector of λ_1.

$R(\vec{v}, A) = \lambda_n$ when $\vec{v} = \vec{v}_n$ = eigenvector of λ_n

$R(\vec{v}, A)$ can be regarded as the approximation of eigenvalue λ_j, given that $\vec{v}$ is closed to $\vec{v}_j$.

Rayleigh Quotient Iteration

Let $A \in M_{n\times n}(\mathbb{R})$

Initiate $\vec{x}^{(0)}$ such that $\vec{x}^{(0)T}\vec{x}^{(0)} = 1$

Initiate μ_0 = initial guess of desired eigenvalue.

Solve: $(A - \mu_0 I)\vec{z}_1 = \vec{x}^{(0)}$

Let $\vec{x}^{(1)} = \dfrac{\vec{z}_1}{\|\vec{z}_1\|_2}$ $\left(\|\vec{x}\|_2 \overset{def}{=} \sqrt{\vec{x}^*\vec{x}} \right)$

Let $\mu_1 = R(\vec{x}^{(1)}, A) = \vec{x}^{(1)*} A \vec{x}^{(1)}$ (Improve μ_0 such that it is closer to an actual eigenvalue)

Keep iteration going!

Algorithm: (Rayleigh Quotient Iteration)

Input: $\vec{x}^{(0)}$ s.t. $\|\vec{x}^{(0)}\|_2 = 1$ and μ_0

Output: μ_k = eigenvalue

For $k = 0, 1, 2, \ldots$

Step 1: Solve $(A - \mu_k I)\vec{z}_{k+1} = \vec{x}^{(k)}$

Step 2: Let $\vec{x}^{(k+1)} = \dfrac{\vec{z}_{k+1}}{\|\vec{z}_{k+1}\|_2}$.

Step 3: $\mu_{k+1} = R(\vec{x}^{(k+1)}, A)$

Example: Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 1 \\ 3 & 2 & 1 \end{pmatrix}$.

Eigenvalues: $\lambda_1 = 3+\sqrt{5}$, $\lambda_2 = 3-\sqrt{5}$, $\lambda_3 = -2$.

Want to find $3+\sqrt{5}$.

Let $\vec{x}^{(0)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $\mu_0 = 200$

Then: $\vec{x}^{(1)} \approx \begin{pmatrix} -0.57927 \\ -0.57348 \\ -0.57927 \end{pmatrix}$ with $\mu_1 = 5.3355$

$\vdots$

$\mu_3 = 5.2361 \approx 3+\sqrt{5}$!

Converges very fast!

Remark:

- RQI works for SPD A
- May or may not work for other A.

Lecture 19:

What if λ_1 has multiplicity > 1?

Consider the case when A is diagonalizable.

Let $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$ be the basis of eigenvectors with eigenvalues equal to $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Assume that: $\lambda_1 = \lambda_2 = \cdots = \lambda_i > |\lambda_{i+1}| \geq \cdots \geq |\lambda_n|$.

Let $\vec{x}^{(0)} = a_1\vec{x}_1 + a_2\vec{x}_2 + \cdots + a_n\vec{x}_n$ (with $a_1 \neq 0$)

Easy to check:

$$\vec{x}^{(k)} = \frac{\lambda_1^k\left(a_1\vec{x}_1 + \cdots + a_i\vec{x}_i + \left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k\vec{x}_{i+1} + \cdots + \left(\frac{\lambda_n}{\lambda_i}\right)^k\vec{x}_n\right)}{\left\|\lambda_1^k\left(a_1\vec{x}_1 + \cdots + a_i\vec{x}_i + \left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k\vec{x}_{i+1} + \cdots + \left(\frac{\lambda_n}{\lambda_i}\right)^k\vec{x}_n\right)\right\|_\infty}$$

$$\rightarrow \frac{a_1\vec{x}_1 + \cdots + a_i\vec{x}_i}{\|a_1\vec{x}_1 + \cdots + a_i\vec{x}_i\|_\infty} \text{ as } k \to \infty.$$

↑ Eigenvector of λ_1.

Also, $\|A\vec{x}^{(k)}\|_\infty \longrightarrow \left\|\frac{A(a_1\vec{x}_1 + \cdots + a_i\vec{x}_i)}{\|a_1\vec{x}_1 + \cdots + a_i\vec{x}_i\|_\infty}\right\|_\infty = |\lambda_1|$ as $k \to \infty$

Remark: The condition on multiplicity ($= 1$) can be relaxed.

Method 2: QR method

Preliminary: QR factorization

Definition: $Q \in M_{n\times n}(\mathbb{R})$ is orthogonal if $Q^TQ = I_n$

Remark: - $Q^{-1} = Q^T$

- Columns of Q forms orthonormal set.

Revisit: Gram- Schmidt orthogonalization

Let $A = (\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n)$ are linearly independent.

G-S process converts $\{\vec{a}_1, \dots, \vec{a}_n\}$ to orthonormal set $\{\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n\}$

G-S process:

Step 1: Let $\tilde{q}_1 = \vec{a}_1$. Normalize: $\vec{q}_1 = \dfrac{\tilde{q}_1}{\|\tilde{q}_1\|_2}$. Let $\alpha_{11} = \|\tilde{q}_1\|_2$

Step 2: Define: $\tilde{q}_2 = \vec{a}_2 - \alpha_{12}\vec{q}_1$. Choose α_{12} such that:

$\tilde{q}_2^T \vec{q}_1 = 0$. Then: $\alpha_{12} = \vec{q}_1^T \vec{a}_2$.

Normalize: $\vec{q}_2 = \dfrac{\tilde{q}_2}{\|\tilde{q}_2\|_2}$. Let $\alpha_{22} = \|\tilde{q}_2\|_2$.

Step 3: Suppose $\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_{k-1}$ are constructed.

Let $\tilde{q}_k = \vec{a}_k - (\alpha_{1k}\vec{q}_1 + \alpha_{2k}\vec{q}_2 + \ldots + \alpha_{k-1,k}\vec{q}_{k-1})$ where $\alpha_{jk} = \vec{q}_j^{\,T}\vec{a}_k$.

Then: $\tilde{q}_k^T \vec{q}_i = 0$ for $i = 1, 2, \ldots, k-1$. Let $\alpha_{kk} = \|\tilde{q}_k\|_2$

Normalize: $\vec{q}_k = \tilde{q}_k / \|\tilde{q}_k\|_2$.

In summary,

$$\begin{cases} \vec{a}_1 = \alpha_{11}\vec{q}_1 \\ \vec{a}_2 = \alpha_{12}\vec{q}_1 + \alpha_{22}\vec{q}_2 \\ \vec{a}_3 = \alpha_{13}\vec{q}_1 + \alpha_{23}\vec{q}_2 + \alpha_{33}\vec{q}_3 \\ \quad\vdots \\ \vec{a}_k = \alpha_{1k}\vec{q}_1 + \alpha_{2k}\vec{q}_2 + \dots + \alpha_{kk}\vec{q}_k \end{cases}$$

This is equivalent to:

$$A = \begin{pmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \\ | & | & & | \end{pmatrix} = \underbrace{\begin{pmatrix} | & & | \\ \vec{q}_1 & \cdots & \vec{q}_n \\ | & & | \end{pmatrix}}_{Q} \underbrace{\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ & \alpha_{22} & \cdots & \alpha_{2n} \\ & & \ddots & \vdots \\ & & & \alpha_{nn} \end{pmatrix}}_{R}$$

<u>Remark:</u>

- Q = orthonormal; R = upper triangular
- Q may NOT be a square matrix ($Q \in M_{m \times n}$)
  R is a square matrix ($R \in M_{n \times n}$)
- Factorization of $A = QR$ is called the QR factorization.