SI231b: Matrix Computations

Lecture 15: Eigenvalue Computations

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University

Oct. 31, 2022

Recap: Eigenvalue Revealing Decomposition

Factorize a matrix to a form in which eigenvalues are explicitly displayed

- Diagonalization, $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1}$, exists if and only if $\mathbf{A}$ is nondefective.
- Schur decompositon, $\mathbf{A} = \mathbf{Q}\mathbf{T}\mathbf{Q}^H$ always exists.
- Jordan canonical form, $\mathbf{A} = \mathbf{S}\mathbf{J}\mathbf{S}^{-1}$ always exists (will not be introduced in our lecture), where

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & & & \\ & \mathbf{J}_2 & & \\ & & \ddots & \\ & & & \mathbf{J}_k \end{bmatrix}$$

with

$$\mathbf{J}_i = \begin{bmatrix} \lambda_i & & & \\ & \lambda_i & & \\ & & \ddots & \\ & & & \lambda_i \end{bmatrix}, \quad \text{or} \quad \mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

Outline

- Facts About Eigenvalues
- Power Iteration
- Inverse Iteration
- Subspace Iteration

- Eigenvalues of Hermitian matrices are real

$$\lambda(\mathbf{A}) \in \mathbb{R}, \quad \text{for } \mathbf{A} \in \mathbb{C}^{n\times n}, \mathbf{A} = \mathbf{A}^H$$

- Eigenvalues of real symmetric matrices are real

$$\lambda(\mathbf{A}) \in \mathbb{R}, \quad \text{for } \mathbf{A} \in \mathbb{R}^{n\times n},\ \mathbf{A} = \mathbf{A}^T$$

- Eigenvectors of real symmetric matrices are also real
- Complex eigenvalues of real matrices appear in conjugate pair.
  - For $\mathbf{A} \in \mathbb{R}^{n\times n}$, if $(\lambda, \mathbf{v})$ is an eigenpair, then also $(\lambda^*, \mathbf{v}^*)$
- Skew-Hermitian matrices ($\mathbf{A} = -\mathbf{A}^H$) have only pure imaginary eigenvalues
- Hermitian/real symmetric matricres are diagonalizable.

The Largest Eigenvalue and Associated Eigenvector

Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be diagonalizable, i.e., $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1}$ with $\mathbf{V} = [\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots\ \mathbf{v}_n]$, and $\Lambda = \mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)$. Assume that

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|.$$

The following iteration generates a sequence of $(\lambda^{(k)},\ \mathbf{q}^{(k)})$ that converges to $(\lambda_1,\ \mathbf{v}_1)$.

Power Iteration:

random selection $\mathbf{q}^{(0)} \in \mathbb{C}^n$

for $k = 1,\ 2,\ \cdots$

$$\mathbf{z}^{(k)} = \mathbf{A}\mathbf{q}^{(k-1)}$$

$$\mathbf{q}^{(k)} = \frac{\mathbf{z}^{(k)}}{\|\mathbf{z}^{(k)}\|_2}$$

$$\lambda^{(k)} = (\mathbf{q}^{(k)})^H\mathbf{A}\mathbf{q}^{(k)}$$

end

Convergence of Power Iteration

The Power Iteration can only compute the largest eigenvalue and associated eigenvector with convergence rate

- $|\lambda^{(k)} - \lambda_1| = \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$
- $\|\mathbf{q}^{(k)} - \mathbf{v}_1\| = \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$
- can have slow convergence when λ_2 is close to λ_1 in magnititude, i.e., $\left|\frac{\lambda_2}{\lambda_1}\right|$ is close to 1.
- The Raleigh Quotient

$$(\mathbf{q}^{(k)})^H \mathbf{A} \mathbf{q}^{(k)} \quad \text{or} \quad \frac{\mathbf{v}^H \mathbf{A} \mathbf{v}}{\mathbf{v}^H \mathbf{v}} \text{in general}$$

  is an approximation of corresponding eigenvalues.

The Smallest Eigenvalue in Magnitude and Associated Eigenvector

```
random selection q^(0) ∈ ℂ^n
for k = 1, 2, ···
        z^(k) = A^{-1} q^(k-1)
        q^(k) = z^(k) / ||z^(k)||_2
        λ^(k) = (q^(k))^H A q^(k)
end
```

$$\texttt{random selection } \mathbf{q}^{(0)} \in \mathbb{C}^n$$
$$\texttt{for } k = 1,\ 2,\ \cdots$$
$$\mathbf{z}^{(k)} = \mathbf{A}^{-1}\mathbf{q}^{(k-1)}$$
$$\mathbf{q}^{(k)} = \frac{\mathbf{z}^{(k)}}{\|\mathbf{z}^{(k)}\|_2}$$
$$\lambda^{(k)} = (\mathbf{q}^{(k)})^H \mathbf{A}\mathbf{q}^{(k)}$$
$$\texttt{end}$$

Facts:

- $(\lambda,\ \mathbf{v})$ is eigenpair of $\mathbf{A}$, so $(\lambda^{-1},\ \mathbf{v})$ is eigenpair of $\mathbf{A}^{-1}$
- Therefore, for the inverse power iteration,

$$\lambda^{(k)} \to \lambda_n, \quad \mathbf{q}^{(k)} \to \mathbf{v}_n$$

where λ_n is the eigenvalue of $\mathbf{A}$ with the smallest magnitude, associated with eigenvector $\mathbf{v}_n$.

Suppose μ is not an eigenvalue of $\mathbf{A}$, the inverse iteration is given by

Inverse Iteration with Shift:

```
random selection q^(0) ∈ C^n
for k = 1, 2, ...
    z = (A − μI)^{-1} q^(k−1)      solve (A − μI)z = q^(k−1)
    q^(k) = z / ||z||_2
    λ^(k) = (q^(k))^H A q^(k)
end
```

- compute the eigenvalue closest to μ
- convergence rate

$$\left|\frac{\mu-\lambda_j}{\mu-\lambda_k}\right|$$

where λ_j and λ_k are the closest and second closest eigenvalues to μ.

Efficiency per iteration vs Number of iterations?

Power Iterations for a Set of Vectors

From the Power Iteration, we know that

- $\mathbf{A}^k\mathbf{q}_0$ converges to the eigenvector associated with the largest eigenvalue in magnititude.
- if we start with a set of linearly independent vectors $\{\mathbf{q}_1,\ \mathbf{q}_2,\ \cdots, \mathbf{q}_r\}$, then $\mathbf{A}^k\{\mathbf{q}_1,\ \mathbf{q}_2,\ \cdots, \mathbf{q}_r\}$ should converge (under suitable assumptions) to a subspace spanned by eigenvectors of $\mathbf{A}$ associated with r largest eigenvalues in magnititude.

Suppose there is a gap between the r largest eigenvalues in magnititude and λ_{r+1}, i.e, $|\lambda_1| \geq |\lambda_2| \geq \cdots |\lambda_r| > |\lambda_{r+1}|$

Subspace Iteration:

```
random selection Q^(0) with orthonormal columns
for k = 1, 2, ···
      Z_k = A Q^(k-1)
      Z_k = Q^(k) R^(k)   reduced QR factorization
end
```

- $\mathbf{Z}_k$ and $\mathbf{Q}^{(k)}$ has the same column space
- equal to the column space of $\mathbf{A}^k\mathbf{Q}^{(0)}$

- $\mathbf{Q}^{(k)}$ converge to subspace associated with r largest eigenvalues in magnititude (dominant invariant subspace).
- $\mathrm{diag}\left(\left(\mathbf{Q}^{(k)}\right)^H \mathbf{A}\mathbf{Q}^{(k)}\right) \to \{\lambda_1,\ \lambda_2,\ \cdots,\lambda_r\}$
- $\|\mathbf{q}_i^{(k)} - \mathbf{v}_i\| = \mathcal{O}\left(\left|\dfrac{\lambda_{r+1}}{\lambda_i}\right|^k\right),\ i = 1,\ 2,\ \cdots,\ r$
- $\left|\lambda_i^{(k)} - \lambda_i\right| = \mathcal{O}\left(\left|\dfrac{\lambda_{r+1}}{\lambda_i}\right|^k\right),\ i = 1,\ 2,\ \cdots,\ r$
- also called simultaneously iteration or orthogonal iteration
- when $r = n$, it coincides with QR iteration

Readings

You are supposed to read

- Lloyd N. Trefethen and David Bau III. *Numerical Linear Algebra*, SIAM, 1997.

  Lecture 25, 26