SI231b: Matrix Computations

Lecture 16: Eigenvalue Computations

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University

Nov. 02, 2022

Recap: (Inverse) Power Iteration

Power iteration

- compute the eigenvalue of the biggest magnitude and corresponding eigenvector
- convergence rate depends on $|\frac{\lambda_2}{\lambda_1}|$
  - λ_1 and λ_2 are the first and second biggest eigenvalue with magnitude

Inverse Power iteration

- compute the eigenvalue of the smallest magnitude and corresponding eigenvector
- convergence rate depends on $|\frac{\lambda_{n-1}}{\lambda_n}|$
  - λ_n and λ_{n-1} are the smallest and second to smallest eigenvalue with magnitude

Outline

- Inverse power iteration with shift
- (Inverse) Power iteration with deflation
- Block power iteration
- Subspace iteration

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
```

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

- 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.

- the power method finds the largest eigenvalue (in magnitude) only
- can we compute more eigenvalues and eigenvectors?
- there are many ways and let's consider a simple method called deflation
- consider a Hermitian $\mathbf{A}$ with $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n|$, we have

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

- Deflation: use the power method to obtain $\mathbf{v}_1, \lambda_1$, do the subtraction

$$\mathbf{A} := \mathbf{A} - \lambda_1 \mathbf{v}_1 \mathbf{v}_1^H = \sum_{i=2}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^H,$$

  and repeat until all the eigenvectors and eigenvalues are found

  - deflation can only be used for Hermitian/real symmetric matrices

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 magnitude.

Block Power Iteration: applying power iteration to several vectors at once. Sometimes it is called **Simultaneous iteration**.

Define $\mathbf{V}^{(0)}$ to be the $n \times r$ matrix,

$$\mathbf{V}^{(0)} = \begin{bmatrix} v_1^{(0)} & v_2^{(0)} & \cdots & v_r^{(0)} \end{bmatrix}.$$

After k steps of applying $\mathbf{A}$, we obtain

$$\mathbf{V}^{(k)} = \mathbf{A}^k \mathbf{V}^{(0)} = \begin{bmatrix} v_1^{(k)} & v_2^{(k)} & \cdots & v_r^{(k)} \end{bmatrix}.$$

Assume

1. The leading $r+1$ eigenvalues are distinct in absolute value;

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_r| > |\lambda_{r+1}| \geq |\lambda_{r+2}| \geq \cdots |\lambda_n|$$

2. All the leading principle sub-matrices $\mathbf{Q}^T \mathbf{V}^{(0)}$ are nonsingular.
   - $\mathbf{Q}$ is the matrix with $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_r$ as columns;
   - $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_r$ are eigenvectors associated with eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_r$.

```
choose V^(0) with r linear independent columns
for k = 1, 2, ···
      V^(k) = AV^(k-1)
      Q^(k)R^(k) = V^(k)   reduced QR factorization
end
```

Under the assumptions, we have as $k \to \infty$,

- For real symmetric matrix $\mathbf{A}$ ($\mathbf{Q}$ has orthonormal columns)

$$\|\mathbf{q}_j^{(k)} - (\pm q_j)\| = \mathcal{O}(C^k),$$

  for $1 \le j \le r$, where $C < 1$ is the constant

$$C = \max_{1 \le k \le r} \frac{|\lambda_{k+1}|}{|\lambda_k|}$$

- For unsymmetric matrix $\mathbf{A}$ ($\mathbf{Q}$ does not have orthonormal columns)

$$\mathcal{R}(\mathbf{Q}^{(k)}) \to \mathcal{R}(\mathbf{Q})$$

For Unnormalized Simultaneous Iteration, as $k \to \infty$, the vectors $\mathbf{q}^{(1)}$, $\mathbf{q}^{(2)}$, $\cdots$, $\mathbf{q}^{(r)}$ all converge to multiples of the same dominant eigenvector $\mathbf{q}_1$. Therefore, they form an ill-conditioned basis of span $\left\{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \cdots, \mathbf{q}^{(r)}\right\}$.

The remedy is simple, we should build orthonormal basis at each iteration $\rightsquigarrow$ **Simultaneous Iteration/Subspace Iteration**

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 $\mathbf{A}^k\mathbf{Q}^{(0)}$

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

Hermitian/real symmetric matrices:

- Simultaneous convergence of eigenvectors

$$\|\mathbf{q}_j^{(k)} - (\pm q_j)\| = \mathcal{O}(C^k),$$

for $1 \leq j \leq r,\ C = \frac{\lambda_{r+1}}{\lambda_r}$

QR Iteration:

$\mathbf{A}^{(0)} = \mathbf{A}$
for $k = 1,\ 2,\ \cdots$
    $\mathbf{Q}^{(k)}\mathbf{R}^{(k)} = \mathbf{A}^{(k-1)}$ QR factorization of $\mathbf{A}^{(k-1)}$
    $\mathbf{A}^{(k)} = \mathbf{R}^{(k)}\mathbf{Q}^{(k)}$
end

Facts:

- $\mathbf{A}^{(k)}$ is similar to $\mathbf{A}$
- Eigenvalues of $\mathbf{A}^{(k)}$ should be easier to compute than that of $\mathbf{A}$.
- $\mathbf{A}^{(k)}$ should converge fast (expected) to a form whose eigenvalues are easily computed.
    - upper triangular form

Readings

You are supposed to read

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

  Lecture 26, 28

- Gene H. Golub and Charles F. Van Loan. *Matrix Computations*, Johns Hopkins University Press, 2013.

  Chapter 7.3 – 7.4