SI231b: Matrix Computations

Lecture 15: Rayleigh Quotient for Eigenvalues of Hermitian Matrices

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University

Nov. 3, 2020

Eigenvalues of Hermitian/real symmetric matrices

In this lecture, we focus on eigenvalues of Hermitian/real symmetric matrices whose eigenvalues are real, then we have simplified results.

- power iteration + deflation can be used to compute more eigenpairs, i.e., once $(\lambda_1,\ v_1)$ is computed, applying power iteration to $A = A - \lambda_1 v_1 v_1^H$ gives $(\lambda_2,\ v_2)$.
- QR iteration + Hessenberg reduction:
  - The Hessebberg reduction reduces A to a tridiagonal matrix
  - QR iteration (with shifts) forces to converge to a bi-diagonal matrix
- Subspace iteration: the orthogonal basis converge to the dominant invariant subspace (for general matrices)
  - the orthogonal basis converge to the associated eigenvectors of dominant eigenvalues

Eigenvalue equation

$$f(\lambda, \mathrm{v}) = \mathrm{Av} - \lambda \mathrm{v}.$$

To differentiate, we obtain

$$\delta f = (\mathrm{A} - \lambda \mathrm{I})\delta \mathrm{v} - (\delta\lambda)\mathrm{v}.$$

Newton's method gives

$$f(\lambda, \mathrm{v}) + \delta f = 0,$$

i.e., at the k-th Newton step,

$$\begin{aligned} 0 &= f(\lambda_k, \mathrm{v}_k) + \delta f(\lambda_k, \mathrm{v}_k) \\ &= (\mathrm{A} - \lambda_k \mathrm{I})(\mathrm{v}_k + \delta \mathrm{v}) - (\delta\lambda_k)\mathrm{v}_k \\ &= (\mathrm{A} - \lambda_k \mathrm{I})\mathrm{v}_{k+1} - (\delta\lambda_k)\mathrm{v}_k \end{aligned}$$

This gives $\mathrm{v}_{k+1} = (\delta\lambda_k)(\mathrm{A} - \lambda_k \mathrm{I})^{-1}\mathrm{v}_k$, where $\delta\lambda_k$ is some normalizing constant.

How to update λ_k?

Least Square formulation of Eigenvalue Computation

Suppose $\tilde{v}$ is an approximate eigenvector, we want to find the corresponding best approximate eigenvalue $\tilde{\lambda}$. This can be achieved by solving

$$\min_{\mu} \|A\tilde{v} - \mu\tilde{v}\|_2^2.$$

The best approximate $\tilde{\lambda}$ is given by

$$\begin{aligned}\tilde{\lambda} &= \arg\min_{\mu} \|A\tilde{v} - \mu\tilde{v}\|_2^2 \\ &= \frac{\tilde{v}^H A\tilde{v}}{\tilde{v}^H\tilde{v}}\end{aligned}$$

Rayleigh Quotient

For any $x \in \mathbb{C}^n$ with $x \neq 0$, the Rayleigh Quotient is given by

$$r(x) = \frac{x^H A x}{x^H x}$$

Rayleigh Quotient

The Rayleigh Quotient is a continuous function except at $x = 0$, and its gradient denoted by $\nabla r(x)$ is given by

$$\nabla r(x) = \frac{2}{x^H x}(Ax - r(x)x)$$

- at an eigenvector of A, the gradient is a zero vector
- if $\nabla r(x) = 0$, x is an eigenvector and $r(x)$ is the corresponding eigenvalue.
- eigenctors of A are the stationary points of the function $r(x)$

Together with the Newton iteration for computing v_{k+1} and the Rayleigh Quotient, we obtain the Rayleigh Quotient iteration for computing the eigenpair.

Rayleigh Quotient Iteration:

```
random selection of v_0 ∈ C^n
```

$$\lambda_0 = r(v_0) = \frac{v_0^H A v_0}{v_0^H v_0}$$

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

$$v_k = (A - \lambda_{k-1} I)^{-1} v_{k-1} \qquad \texttt{solve } (A - \lambda_{k-1} I) v_k = v_{k-1}$$

$$v_k = \frac{v_k}{\|v_k\|_2}$$

$$\lambda_k = (v_k)^H A v_k$$

end

- inverse iteration with shift, and shift varies per iteration
- cubic convergence for Hermitian/real symmetric matrices

Theorem[Rayleigh-Ritz]. Let A be a Hermitian matrix. It holds that

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

$$\lambda_{\min} = \min_{x\in\mathbb{C}^n, \|x\|_2=1} x^H A x, \qquad \lambda_{\max} = \max_{x\in\mathbb{C}^n, \|x\|_2=1} x^H A x$$

- provides information about λ_1 and λ_n for A
- Proof:
  - by a change of variable $y = V^H x$, we have

    $$x^H A x = y^H \mathbf{\Lambda} y = \sum_{i=1}^n \lambda_i |y_i|^2 \leq \lambda_1 \sum_{i=1}^n |y_i|^2 = \lambda_1 \|V^H x\|_2^2 = \lambda_1 \|x\|_2^2$$

  - we thus have $\max_{\|x\|_2=1} x^H A x \leq \lambda_1$
  - since $v_1^H A v_1 = \lambda_1$, the above equality is attained
  - the results $x^H A x \geq \lambda_n \|x\|_2^2$ and $\min_{\|x\|_2=1} x^H A x = \lambda_n$ are proven in the same way

Question: how about λ_k for any $k \in \{1, \ldots, n\}$? Do we have a similar variational characterization as that in the Rayleigh-Ritz theorem?

Theorem[Courant-Fischer]. Let A be a Hermitian matrix, and let $\mathcal{S}_k$ denote any subspace of $\mathbb{C}^n$ and of dimension k. For any $k \in \{1, \ldots, n\}$, it holds that

$$\begin{aligned} \lambda_k &= \min_{\mathcal{S}_{n-k+1} \subseteq \mathbb{C}^n} \max_{\mathrm{x} \in \mathcal{S}_{n-k+1}, \|\mathrm{x}\|_2 = 1} \mathrm{x}^H \mathrm{A} \mathrm{x} \\ &= \max_{\mathcal{S}_k \subseteq \mathbb{C}^n} \min_{\mathrm{x} \in \mathcal{S}_k, \|\mathrm{x}\|_2 = 1} \mathrm{x}^H \mathrm{A} \mathrm{x} \end{aligned}$$

(requires a proof)

- optima is achieved when $\mathcal{S}_k = \mathrm{span}\{\mathrm{v}_1, \mathrm{v}_2, \cdots, \mathrm{v}_k\}$ and $\mathcal{S}_{n-k+1} = \mathrm{span}\{\mathrm{v}_k, \mathrm{v}_{k+1}, \cdots, \mathrm{v}_n\}$
- The Rayleigh-Ritz Theorem is a special case of the Courant-Fischer minimax theorem when $k = 1$ and $k = n$

From the Courant-Fischer minimax theorem, we know that

- given any k-dimensional subspace $\mathcal{S}$, the smallest value of the Rayleigh quotient over that subspace is a lower bound on λ_k and an upper bound on λ_{n-k+1}.
- practically used to bound eigenvalues (Hermitian/real symmetric matrices)
- numerical range is often used to bound eigenvalues of non-Hermitian matrices (self-study for your own interest)

One step further, we have the Cauchy interlace theorem, which relates the eigenvalues of a block Rayleigh quotient to the eigenvalues of the corresponding matrix.

Theorem[Cauchy interlace]. Suppose A is Hermitian/real symmetric, and let V be a matrix with m orthonormal columns. Then the eigenvalues of $\mathrm{V}^H\mathrm{AV}$ interlace the eigenvalues of A. That is, if A has eigenvalues $\alpha_1, \alpha_2, \cdots, \alpha_n$ and $\mathrm{V}^H\mathrm{AV}$ has eigenvalues β_j, then

$$\beta_j \in [\alpha_{n-m+j}, \ \alpha_j]$$

If B is a principle submatrix of $m \times m$ for an $n \times n$ Hermitian/real symmetric matrix A. Suppose A has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and B has eigenvalues $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_m$, then

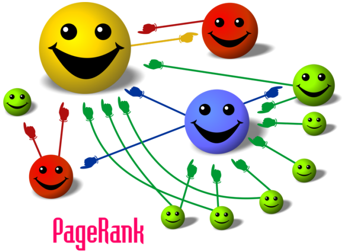
$$\lambda_{n-m+j} \leq \beta_j \leq \lambda_j.$$

Specifically, when $m = n - 1$,

$$\lambda_n \leq \beta_{n-1} \leq \lambda_{n-1} \leq \cdots \leq \lambda_2 \leq \beta_1 \leq \lambda_1$$

Case Study: PageRank Problem

- PageRank is an algorithm used by Google to rank the pages of a search result.
- the idea is to use counts of links of various pages to determine pages' importance.



Source: Wiki.

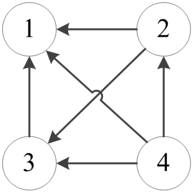
- further reading: **[Bryan-Tanya2006]**
  - K. Bryan and L. Tanya. "The 25, 000, 000, 000 eigenvector: The linear algebra behind Google," *SIAM Review*, vol. 48, no. 3, pp. 569–581, 2006.

- Model:

$$\sum_{j \in \mathcal{L}_i} \frac{v_j}{c_j} = v_i, \quad i = 1, \ldots, n,$$

where c_j is the number of outgoing links from page j; $\mathcal{L}_i$ is the set of pages with a link to page i; v_i is the importance score of page i.

- example:



$$\overbrace{\begin{bmatrix} 0 & \frac{1}{2} & 1 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}}^{\text{A}} \overbrace{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}}^{\text{v}} = \overbrace{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}}^{\text{v}}.$$

PageRank Problem

- let $A \in \mathbb{R}^{n \times n}$ be a matrix such that $a_{ij} = 1/c_j$ if $j \in \mathcal{L}_i$ and $a_{ij} = 0$ if $j \notin \mathcal{L}_i$
- Problem: find a non-negative v such that $Av = v$
  - A is extremely large and sparse, and we want to use the power method
- Questions:
  - does a solution to $Av = v$ exist? Or, is $\lambda = 1$ an eigenvalue of A?
  - does $Av = v$ have a non-negative solution? Or, does a non-negative eigenvector associated with $\lambda = 1$ exist?
  - is the solution to $Av = v$ unique? Or, would there exist more than one eigenvector associated with $\lambda = 1$?
    - a unique solution is desired for this problem
  - is $\lambda = 1$ the only eigenvalue that is the largest in modulus?
    - this is required for the power method

Some Notation and Conventions

- notation:
  - $x \geq y$ means that $x_i \geq y_i$ for all i
  - $x > y$ means that $x_i > y_i$ for all i
  - $x \ngeq y$ means that $x \geq y$ does not hold
  - the same notations apply to matrices
- conventions:
  - x is said to be non-negative if $x \geq 0$, and non-positive if $-x \geq 0$
  - x is said to be positive if $x > 0$, and negative if $-x > 0$
  - the same conventions apply to matrices
  - a square A is said to be column-stochastic if $A \geq 0$ and $A^T 1 = 1$
    - a column-stochastic A has every column a_i satisfying $a_i^T 1 = \sum_{j=1}^{n} a_{ji} = 1$

PageRank Matrix Properties

- in PageRank, A is column-stochastic if all pages have outgoing links
  - see the literature to see how to deal with cases where some pages do not have outgoing links (dangling nodes)

Property: Let A be column-stochastic. Then,

1. $\lambda = 1$ is an eigenvalue of A
2. $|\lambda| \leq 1$ for any eigenvalue λ of A

- Implications:
  - a solution to $Av = v$ does exist, though it doesn't say if $v \geq 0$ or not
  - $\lambda = 1$ is an eigenvalue that has the largest modulus, but we don't know if it is the *only* eigenvalue that has the largest modulus
- we resort to non-negative matrix theory to answer the rest of the questions

Theorem[Perron-Frobenius] Let A be square positive. There exists an eigenvalue ρ of A such that

1. ρ is real and $\rho > 0$
2. $|\lambda| < \rho$ for any eigenvalue λ of A with $\lambda \neq \rho$
3. there exists a positive eigenvector associated with ρ
4. the algebraic multiplicity of ρ is 1 (so the geometric multiplicity of ρ is also 1)

A weaker result for general non-negative matrices:

Theorem. Let A be square non-negative. There exists an eigenvalue ρ of A such that

1. ρ is real and $\rho \geq 0$
2. $|\lambda| \leq \rho$ for any eigenvalue λ of A
3. there exists a non-negative eigenvector associated with ρ

PageRank Matrix Properties

- further implication by the above theorem:
  - a non-negative solution to $\mathrm{A}\mathrm{v} = \mathrm{v}$ exists, though it doesn't say if there exists another solution
  - even worse, it is not known if there exists another solution v such that $\mathrm{v} \ngeq 0$
- PageRank actually considers a modified version of A

$$\tilde{\mathrm{A}} = (1-\beta)\mathrm{A} + \beta \begin{bmatrix} 1/n & \dots & 1/n \\ \vdots & & \vdots \\ 1/n & \dots & 1/n \end{bmatrix}$$

where $0 < \beta < 1$ (typical value is $\beta = 0.15$)

- $\tilde{\mathrm{A}}$ is positive
- further implications
  - $\lambda = 1$ is the *only* eigenvalue that has the largest modulus
  - there exists *only* one eigenvector associated with $\lambda = 1$; that eigenvector is either positive or negative