# SI231b: Matrix Computations

## Lecture 18: Rayleigh Quotient for Eigenvalues of Hermitian Matrices

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University

Nov. 16, 2021

# 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 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**

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

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 $\lambda_k$?

### Least Square formulation of Eigenvalue Computation

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

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

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

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

### Rayleigh Quotient

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

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

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
```

$$\lambda_0 = r(\mathrm{v}_0) = \frac{\mathrm{v}_0^H \mathrm{A} \mathrm{v}_0}{\mathrm{v}_0^H \mathrm{v}_0}$$

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

$$\mathrm{v}_k = (\mathrm{A} - \lambda_{k-1}\mathrm{I})^{-1}\mathrm{v}_{k-1} \qquad \texttt{solve } (\mathrm{A} - \lambda_{k-1}\mathrm{I})\mathrm{v}_k = \mathrm{v}_{k-1}$$

$$\mathrm{v}_k = \frac{\mathrm{v}_k}{\|\mathrm{v}_k\|_2}$$

$$\lambda_k = (\mathrm{v}_k)^H \mathrm{A} \mathrm{v}_k$$

**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 $\lambda_1$ and $\lambda_n$ for A
- Proof:
  - by a change of variable $y = V^H x$, we have

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

  - 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 $\lambda_k$ 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 $\lambda_k$ and the maximum value over that subspace gives an upper bound on $\lambda_{n-k+1}$.
- 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 $\beta_j$, then

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

# Special Case of Cauchy Interlace Theorem

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 $m = n - 1$,

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

# Readings

You are supposed to read

- Gene H. Golub and Charles F. Van Loan. *Matrix Computations*, Johns Hopkins University Press, 2013. 1997.

  Chapter 8.1

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

  Lecture 27