CS 3331 Numerical Methods

Lecture 5: Eigenvalue Problem

Cherung Lee

Outline

- Linear algebra.
- Upper triangular matrix
- The power method.
  - Speedup methods
- The orthogonal iteration.
- The QR method.
- Singular value decomposition

Linear Algebra

Definition

- For a given $n \times n$ matrix $\mathbf{A}$, if a scalar λ and a nonzero vector $\mathbf{x}$ satisfies $\mathbf{Ax} = \lambda\mathbf{x}$, we say $(\lambda, \mathbf{x})$ is an eigenpair of $\mathbf{A}$.

- If there are n linearly independent eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, then $\mathbf{A}$ has the eigen-decomposition:

$$\mathbf{A} = \mathbf{X\Lambda X}^{-1}$$

where $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots \mathbf{x}_n]$ and $\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$.

Gerschgorin circle theorem LVF pp.12

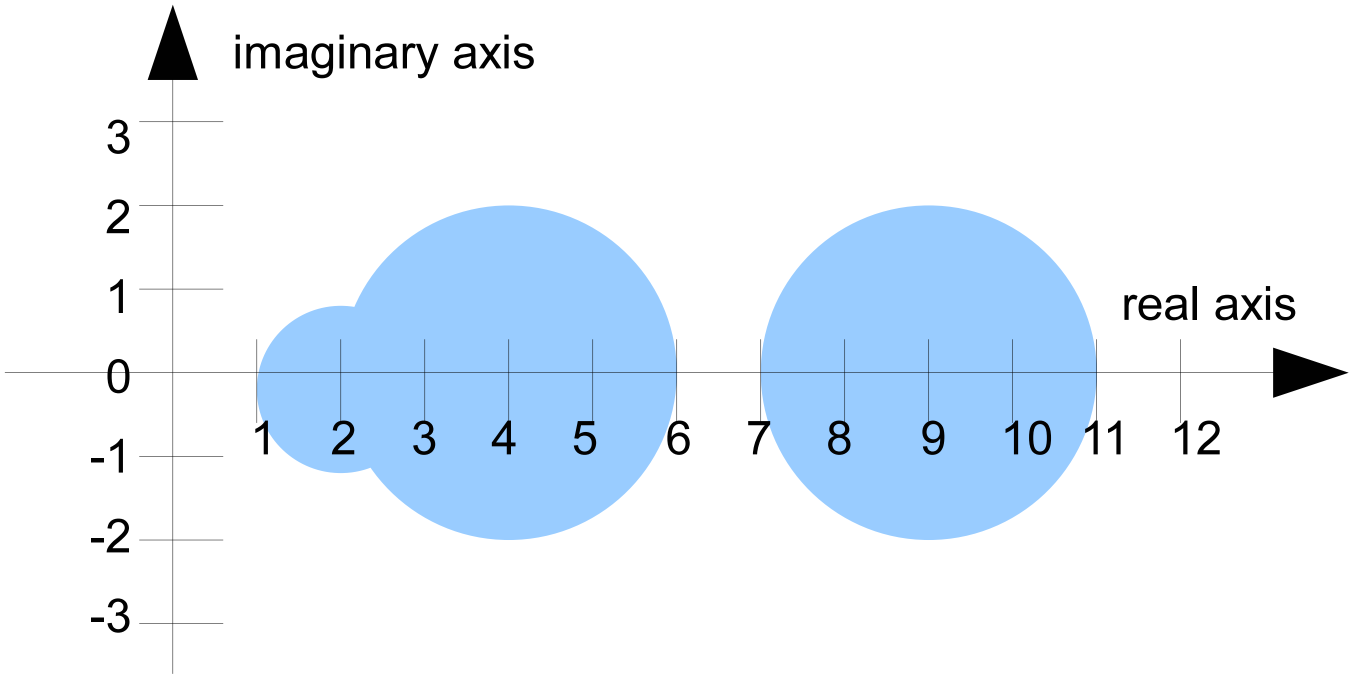
- For a given $n \times n$ matrix $\mathbf{A}$, define
$$r_i = \sum_{j=1, j \neq i}^{n} |a_{ij}|.$$
Then each eigenvalue of $\mathbf{A}$ is in at least one of the disk $\{z : |z - a_{ii}| < r_i\}$.

- If there is a union of k disks, disjoint from the other disks, then exact k eigenvalues lie within the union.

- Example:

$$\mathbf{A} = \begin{pmatrix} 4 & 1 & 1 \\ 0 & 2 & 1 \\ -2 & 0 & 9 \end{pmatrix}$$

$C_1 = \{z : |z-4| \leq 2\}$, $C_2 = \{z : |z-2| \leq 1\}$, $C_3 = \{z : |z-9| \leq 2\}$



Residual and Rayleigh quotient LVF pp.194

- Let $(\mu, \mathbf{z})$ be an approximation to an eigenpair of $\mathbf{A}$. Its residual is

$$\mathbf{r} = \mathbf{A}\mathbf{z} - \mu\mathbf{z}.$$

  - Small residual implies small backward error.

- If $\mathbf{z}$ is an approximation to an eigenvector of $\mathbf{A}$, then the Rayleigh quotient

$$\mu = \frac{\mathbf{z}^T\mathbf{A}\mathbf{z}}{\mathbf{z}^T\mathbf{z}}$$

  is an approximation to the corresponding eigenvalue.

  - Rayleigh quotient is the one minimizing $\|\mathbf{r}\|$.

Upper Triangular Matrix

Eigenvalues and characteristic polynomial

- The function $p(x) = \det(\mathbf{A} - x\mathbf{I})$ is called the characteristic polynomial of $\mathbf{A}$.
- The roots of $p(x) = 0$ are eigenvalues of $\mathbf{A}$.
- If $\mathbf{A}$ is triangular, $\det(\mathbf{A}) = \prod_{i=1}^{n} a_{ii}$.
  - The characteristic function of a triangular matrix is

$$f(x) = \det(\mathbf{A} - x\mathbf{I}) = \prod_{i=1}^{n} (a_{ii} - x)$$

  - The eigenvalues of $\mathbf{A}$ are the diagonal elements, $a_{11}, a_{22}, \ldots, a_{nn}$.

Eigenvectors of upper triangular matrices

- If $\mathbf{A}$ is upper triangular, the kth eigenvector has the form

$$\mathbf{x}_k = \begin{pmatrix} \mathbf{z}_1 \\ 1 \\ \mathbf{0} \end{pmatrix} \begin{matrix} \text{length } k-1 \\ \text{length } 1 \\ \text{length } n-k \end{matrix}$$

- Decompose $\mathbf{A}$ as $\begin{pmatrix} \mathbf{A}_1 & \mathbf{a}_{:,k} & \mathbf{A}_2 \\ & a_{kk} & \mathbf{a}_{k,:} \\ & & \mathbf{A}_3 \end{pmatrix}$. Using the equation $\mathbf{A}\mathbf{x}_k = \lambda_k \mathbf{x}_k$ one can obtain

$$\mathbf{A}_1 \mathbf{z}_1 + \mathbf{a}_{:,k} = \lambda_k \mathbf{z}_1$$

Suppose λ_k is not an eigenvalue of $\mathbf{A}_1$. $\mathbf{z}_1$ can be obtained by solving $(\mathbf{A}_1 - \lambda_k \mathbf{I})\mathbf{z}_1 = -\mathbf{a}_{:,k}$

Power Method

Power method LVF pp.190-193

- Algorithm

  1. Given an initial vector $\mathbf{p}_0$, $\|\mathbf{p}_0\| = 1$.

  2. For $i = 1, 2, \ldots$ until converged

     (a) $\mathbf{p}_i = \mathbf{A}\mathbf{p}_{i-1}$

     (b) $\mathbf{p}_i = \mathbf{p}_i / \|\mathbf{p}_i\|$ //normalization

Why it works?

- Suppose $\mathbf{A}$ has eigendecomposition $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}$.
  - $\mathbf{X} = [\mathbf{x}_1 \; \mathbf{x}_2 \; \cdots \mathbf{x}_n]$.
  - Suppose $1 = \lambda_1 > \lambda_2 > \cdots > \lambda_n > 0$
- Initial vector $\mathbf{p}_0 = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots a_n\mathbf{x}_n$.
  - $\mathbf{p}_1 = \mathbf{A}\mathbf{p}_0 = a_1\lambda_1\mathbf{x}_1 + a_2\lambda_2\mathbf{x}_2 + \cdots a_n\lambda_n\mathbf{x}_n$
- Without normalization,

$$\begin{aligned}\mathbf{p}_k &= \mathbf{A}\mathbf{p}_{k-1} = \mathbf{A}^2\mathbf{p}_{k-2} = \cdots = \mathbf{A}^k\mathbf{p}_0 \\ &= a_1\lambda_1^k\mathbf{x}_1 + a_2\lambda_2^k\mathbf{x}_2 + \cdots a_n\lambda_n^k\mathbf{x}_n \\ &\to a_1\mathbf{x}_1\end{aligned}$$

Convergence LVF pp.201

- Eigenvector: linear convergence with rate $|\lambda_2/\lambda_1|$.
  - Let $\mathbf{z}_k = \frac{1}{a_1\lambda_1^k}\mathbf{p}_k$. Then

$$\mathbf{z}_k - x_1 = \sum_{i=2}^{n} \frac{a_i}{a_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k x_i \longrightarrow \frac{a_2}{a_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k x_2.$$

- Eigenvalue: linear convergence with better convergence rate
  - Convergent rate is $|\lambda_2/\lambda_1|^2$ for symmetric matrices.

Speedup by shift LVF pp.196

- Matrix $\mathbf{B} = \mathbf{A} - b\mathbf{I}$ has the same eigenvectors as $\mathbf{A}$.

- Eigenvalues of $\mathbf{B}$ are $\lambda_1 - b, \lambda_2 - b, \cdots \lambda_n - b$.

- Suppose $|\lambda_1 - b|$ is still the largest number among $|\lambda_i - b|$
  - The convergent rate becomes $\rho = \max\left\{\frac{|\lambda_2 - b|}{|\lambda_1 - b|}, \frac{|\lambda_n - b|}{|\lambda_1 - b|}\right\}$.
  - The b that minimizes ρ is $b^* = (\lambda_2 + \lambda_n)/2$.

Speedup by shift-invert LVF pp.198-200

- Matrix $\mathbf{B} = \mathbf{A}^{-1}$ has the same eigenvectors as $\mathbf{A}$.
- Eigenvalues of $\mathbf{B}$ are $1/\lambda_1, 1/\lambda_2, \cdots 1/\lambda_n$.
  - The smallest eigenvalue becomes the largest one.
  - The convergent rate is $|\lambda_n/\lambda_{n-1}|$.
- Combining shift: $\mathbf{B} = (\mathbf{A} - b\mathbf{I})^{-1}$
  - $\mathbf{B}$ has the same eigenvectors as $\mathbf{A}$.
  - Eigenvalues of $\mathbf{B}$ are $1/(\lambda_1 - b), 1/(\lambda_2 - b), \cdots 1/(\lambda_n - b)$.
  - If $|\lambda_1 - b| < |\lambda_2 - b| \cdots < |\lambda_n - b|$, convergence rate is $\frac{|\lambda_1 - b|}{|\lambda_2 - b|}$.

Orthogonal Iteration

Orthogonal iteration JWD 156-159

- Can we compute more than one eigenvectors simultaneously?
- Problem: what is the normalization step?
- Algorithm for two eigenvectors

  1. Let $\mathbf{Z}_{(0)}$ be an $n \times 2$ orthogonal matrix.

  2. For $i = 1, 2, \ldots$ until converged

     (a) $\mathbf{Y}_{(i)} = \mathbf{A}\mathbf{Z}_{(i-1)}$.

     (b) Compute the QR decomposition of $\mathbf{Y}_{(i)}$,

$$\mathbf{Y}_{(i)} = \mathbf{Z}_{(i)}\mathbf{R}_{(i)}.$$

Why it works?

- span $\left\{\mathbf{Z}_{(i)}\right\} = $ span $\left\{\mathbf{Y}_{(i)}\right\} = $ span $\left\{\mathbf{A}\mathbf{Z}_{(i-1)}\right\} = \cdots = $ span $\left\{\mathbf{A}^i\mathbf{Z}_{(0)}\right\}$.
- Let $\mathbf{Z}_{(0)} = [\mathbf{q}_1, \mathbf{q}_2]$. Without orthogonalization, both $\mathbf{A}^i\mathbf{q}_1, \mathbf{A}^i\mathbf{q}_2$ converge to $\mathbf{x}_1$.

$$\begin{cases} \mathbf{A}^i\mathbf{q}_1 = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_n\mathbf{x}_n, & |\alpha_1| > |\alpha_2| > \cdots > |\alpha_n|; \\ \mathbf{A}^i\mathbf{q}_2 = \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \cdots + \beta_n\mathbf{x}_n, & |\beta_1| > |\beta_2| > \cdots > |\beta_n|. \end{cases}$$

  - However, with orthogonalization, $\mathbf{A}^i\mathbf{q}_2$ can get rid off the influence from $\mathbf{x}_1$.
  - Together, span$\{\mathbf{q}_1, \mathbf{q}_2\}$ converges to span$\{\mathbf{x}_1, \mathbf{x}_2\}$.

Where are eigenvalues and eigenvectors?

- In the power method, $\mathbf{p}_i$ converges to an eigenvector, and the Rayleigh quotient $\mathbf{p}_i^T \mathbf{A} \mathbf{p}_i / \mathbf{p}_i^T \mathbf{p}_i$ converges to the corresponding eigenvalue.
- When the orthogonal iteration is converging, $\mathbf{Z}_{(i)} \approx \mathbf{Z}_{(i-1)}$. The generalized Rayleigh quotient

$$\mathbf{Z}_{(i-1)}^T \mathbf{A} \mathbf{Z}_{(i-1)} \approx \mathbf{Z}_{(i)}^T \mathbf{Y}_{(i)} = \mathbf{R}_{(i)}$$

  converges to an upper triangular matrix $\mathbf{R}_{(i)}$

  - Eigenvalues approximations are on the diagonal of $\mathbf{R}_{(i)}$.
  - Eigenvector approximations can be solved by inverse power method.(LVF pp.206) or the methods for triangular matrices. (slide 8).

QR Method

QR method LVF pp.202-203

- How about computing all the eigenpairs?
- Algorithm

1. Let $\mathbf{A}_{(0)} = \mathbf{A}$.
2. For $i = 1, 2, \ldots$ until converged
   (a) Compute the QR decomposition of $\mathbf{A}_{(i-1)}$,
   $$\mathbf{A}_{(i-1)} = \mathbf{Q}_{(i)}\mathbf{R}_{(i)}.$$
   (b) Compute $\mathbf{A}_{(i)} = \mathbf{R}_{(i)}\mathbf{Q}_{(i)}$

Why it works?

$$\begin{aligned}\mathbf{A}_{(i)} &= \mathbf{R}_{(i)}\mathbf{Q}_{(i)} = \mathbf{Q}_{(i)}^T\mathbf{A}_{(i-1)}\mathbf{Q}_{(i)} \\ &= \mathbf{Q}_{(i)}^T\mathbf{R}_{(i-1)}\mathbf{Q}_{(i-1)}\mathbf{Q}_{(i)} \\ &= \mathbf{Q}_{(i)}^T\mathbf{Q}_{(i-1)}^T\mathbf{A}_{(i-2)}\mathbf{Q}_{(i-1)}\mathbf{Q}_{(i)} \\ &= \cdots \\ &= \mathbf{Q}_{(i)}^T\cdots\mathbf{Q}_{(1)}^T\mathbf{A}\mathbf{Q}_{(1)}\cdots\mathbf{Q}_{(i)}\end{aligned}$$

- If $\mathbf{Z}_{(0)} = \mathbf{I}$, $\mathbf{Z}_{(i)} = \mathbf{Q}_{(1)}\cdots\mathbf{Q}_{(i)}$.
  - Can be proved by induction. The base case is trivial.
  - Orthogonal iteration can be expressed as $\mathbf{A}\mathbf{Z}_{(i)} = \mathbf{Z}_{(i+1)}\mathbf{R}_{(i+1)}$.
  - $\mathbf{A}_{(i)} = \mathbf{Z}_{(i)}^T\mathbf{A}\mathbf{Z}_{(i)} = \mathbf{Z}_{(i)}^T\mathbf{Z}_{(i+1)}\mathbf{R}_{(i+1)} = \mathbf{Q}_{(i+1)}\mathbf{R}_{(i+1)}$
  - Since $\mathbf{Z}_{(i)}^T\mathbf{Z}_{(i+1)} = \mathbf{Q}_{(i+1)}$, $\mathbf{Z}_{(i+1)} = \mathbf{Z}_{(i)}\mathbf{Q}_{(i+1)}$.

But why this formulation?

- The QR decomposition costs $O(n^4)$ for n eigenpairs.
- An elegant algorithm. (LVF pp.204)

  1. Reduce $\mathbf{A}$ to upper Hessenberg, $\mathbf{H}_{(0)} = \mathbf{W}\mathbf{A}\mathbf{W}^T$ //$O(n^3)$
  2. For $i = 1, 2, \ldots$ until converged.
     (a) QR decomposed $\mathbf{H}_{(i-1)} = \mathbf{Q}_{(i)}\mathbf{R}_{(i)}$
     //$O(n^2)$ using Givens rotation
     (b) $\mathbf{H}_{(i)} = \mathbf{R}_{(i)}\mathbf{Q}_{(i)}$
     //$O(n^2)$. $\mathbf{H}_{(i)}$ is still upper Hessenberg

- Total time complexity is $O(n^3)$.

- Prove $\mathbf{H}_{(i)} = \mathbf{R}_{(i)}\mathbf{Q}_{(i)} = \mathbf{Q}_{(i)}^T\mathbf{H}_{(i-1)}\mathbf{Q}_{(i)}$ is upper Hessenberg.

  1. In 2(a), $\mathbf{H}_{(i-1)} = \mathbf{Q}_{(i)}\mathbf{R}_{(i)}$, matrix $\mathbf{Q}_{(i)}$ is upper Hessenberg. (Think about Gram-Schmidt process.)
  2. In 2(b), $\mathbf{R}_{(i)}\mathbf{Q}_{(i)}$ generates an upper Hessenberg matrix, since $\mathbf{R}_{(i)}$ is upper triangular.

- Adding shift: ρ_i (LVF pp.207,208)

  2(a) QR decomposed $\mathbf{H}_{(i-1)} - \rho_i\mathbf{I} = \mathbf{Q}_{(i)}\mathbf{R}_{(i)}$
  2(b) $\mathbf{H}_{(i)} = \mathbf{R}_{(i)}\mathbf{Q}_{(i)} + \rho_i\mathbf{I}$

  - $\mathbf{H}_{(i)} = \mathbf{R}_{(i)}\mathbf{Q}_{(i)} + \rho_i\mathbf{I} = \mathbf{Q}_{(i)}^T(\mathbf{H}_{(i-1)} - \rho_i\mathbf{I})\mathbf{Q}_{(i)} + \rho_i\mathbf{I} = \mathbf{Q}_{(i)}^T\mathbf{H}_{(i-1)}\mathbf{Q}_{(i)} - \rho_i\mathbf{Q}_{(i)}^T\mathbf{Q}_{(i)} + \rho_i\mathbf{I} = \mathbf{Q}_{(i)}^T\mathbf{H}_{(i-1)}\mathbf{Q}_{(i)}$.
  - If $\rho_i = r_{nn}$, the algorithm converges quadratically.

Singular Value Decomposition

Definition

- For an $m \times n$ matrix $\mathbf{A}$, there always exists an decomposition

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

  - $\mathbf{\Sigma}$ is diagonal. The diagonal elements are *singular values*.
  - $\mathbf{U}$ and $\mathbf{V}$ are orthogonal.

	Eigenvalue decomp $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}$	Singular value decomp $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$
matrix shape	square	any shape
Existence	not always	always
Values	no restricted	always ≥ 0
Relation	The eigenvalue decomposition of $\mathbf{A}^T\mathbf{A}$ is $\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T$; The eigenvalue decomposition of $\mathbf{A}\mathbf{A}^T$ is $\mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T$	

Matrix 2-norm

- Recall that $\|\mathbf{A}\|_2 = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 = \max_{\|\mathbf{x}\|_2=1} \sqrt{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}}$
- Use the relation, $\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T$.

$$\|\mathbf{A}\|_2 = \max_{\|\mathbf{x}\|_2=1} \sqrt{\mathbf{x}^T\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{x}}.$$

  - Matrix $\mathbf{A}^T\mathbf{A}$ is symmetric semipositive definite.
  - Suppose the singular values in $\mathbf{\Sigma}$ are sorted in descending order. $\mathbf{V}^T\mathbf{x} = \mathbf{e}_1$ gives the maximum value, which is the largest singular value.
  - $\mathbf{V}^T\mathbf{x} = \mathbf{e}_1$ means $\mathbf{x}$ is the first column of $\mathbf{V}$, the left singular vector corresponding to the largest singular value.

How to compute SVD? LVF pp.210

- Algorithm

  1. Compute the eigenvalue decomposition of $\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$.
  2. Compute the QR decomposition of $\mathbf{AV} = \mathbf{UR}$.
  3. Let $\mathbf{\Sigma} = \sqrt{\mathbf{\Lambda}}$.
  4. $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is the SVD of $\mathbf{A}$.

- A more cheap and numerically stable algorithm is like QR method, but it is rather complicated.