Lecture 8: Power/Inverse iteration, Rayleigh quotient iteration

School of Mathematical Sciences, Xiamen University

1. Eigenvalue problem and polynomial rootfinding problem

- The eigenvalues of a matrix $\mathbf{A} \in \mathbb{C}^{m\times m}$ are the m roots of its characteristic polynomial $p(z) = \det(z\mathbf{I} - \mathbf{A})$.
- Suppose we have the monic polynomial

$$p(z) = z^m + a_{m-1}z^{m-1} + \cdots + a_1 z + a_0.$$

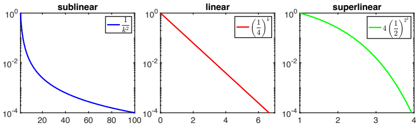
It is not hard to verify that $p(z) = \det(z\mathbf{I} - \mathbf{A})$, where the $m \times m$ matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{bmatrix} 0 & & & & & -a_0 \\ 1 & 0 & & & & -a_1 \\ & 1 & 0 & & & -a_2 \\ & & 1 & \ddots & & \vdots \\ & & & \ddots & 0 & -a_{m-2} \\ & & & & 1 & -a_{m-1} \end{bmatrix}.$$

The matrix $\mathbf{A}$ is called a *companion matrix* corresponding to $p(z)$.

- Any eigenvalue solver must be iterative because no explicit root expressing formula exists for polynomial of degree ≥ 5. The goal of an eigenvalue solver is to produce sequences of numbers that converge rapidly towards eigenvalues.
- Convergence rate

  Let $e_1, e_2, \ldots$ be a sequence of nonnegative numbers representing errors in some iterative process that converge to zero, and suppose there are a positive constant c and an exponent α such that for all sufficiently large k, $e_{k+1} \leq c(e_k)^\alpha$. Then,
  (1) $\alpha = 1$ and $c < 1$, *linear convergence* or *geometric convergence*;
  (2) $\alpha = 2$, *quadratic convergence*;
  (3) $\alpha = 3$, *cubic convergence*; ...



2. Rayleigh quotient

- The *Rayleigh quotient* of a nonzero vector $\mathbf{x} \in \mathbb{C}^m$ with respect to $\mathbf{A}$ is the scalar

$$r(\mathbf{x}) = \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}}.$$

Theorem 1

Let $\{\lambda, \mathbf{v}\}$ *be an eigenpair of the matrix* $\mathbf{A}$, *i.e.*, $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ *and* $\mathbf{v} \neq 0$.

(i) *If* $\mathbf{A}$ *is non-normal, then the Rayleigh quotient* $r(\mathbf{x})$ *is generally a linearly accurate estimate of the eigenvalue* λ, *i.e.*,

$$|r(\mathbf{x}) - \lambda| = \mathcal{O}(\|\mathbf{x} - \mathbf{v}\|_2), \quad \text{as} \quad \mathbf{x} \to \mathbf{v}.$$

(ii) *If* $\mathbf{A}$ *is normal, then the Rayleigh quotient* $r(\mathbf{x})$ *is a quadratically accurate estimate of the eigenvalue* λ, *i.e.*,

$$|r(\mathbf{x}) - \lambda| = \mathcal{O}(\|\mathbf{x} - \mathbf{v}\|_2^2), \quad \text{as} \quad \mathbf{x} \to \mathbf{v}.$$

Hint:

For simplicity, we assume that $\|\mathbf{v}\|_2 = 1$ and consider a Schur form

$$\mathbf{T} = \mathbf{Q}^*\mathbf{A}\mathbf{Q}$$

with $t_{11} = \lambda$ and $\mathbf{Q}\mathbf{e}_1 = \mathbf{v}$. Let $\mathbf{y} = \mathbf{Q}^*\mathbf{x}$. Then $\mathbf{y} \to \mathbf{e}_1$ as $\mathbf{x} \to \mathbf{v}$. $\square$

3. Power iteration

Algorithm 1: Power iteration

$\mathbf{v}^{(0)} =$ some vector with $\|\mathbf{v}^{(0)}\|_2 = 1$
for $k = 1, 2, 3, \ldots,$
  $\mathbf{w} = \mathbf{A}\mathbf{v}^{(k-1)}$
  $\mathbf{v}^{(k)} = \mathbf{w}/\|\mathbf{w}\|_2$
  $\lambda^{(k)} = (\mathbf{v}^{(k)})^* \mathbf{A}\mathbf{v}^{(k)}$
end

- Termination conditions.
- One application: Google's Pagerank.
- Power iteration can find only an approximate eigenpair corresponding to the eigenvalue with the largest magnitude. The convergence is *linear*, which is very slow if the largest two eigenvalues are close in magnitude.

- Assume that $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ is diagonalizable with $\|\mathbf{v}_1\|_2 = 1$ and

$$\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_m\}, \quad |\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_m|.$$

Let $\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix} := \mathbf{V}^{-1}\mathbf{v}^{(0)}$. If $\alpha_1 \neq 0$, then we have

$$\mathbf{A}^k\mathbf{v}^{(0)} = \mathbf{V}\mathbf{\Lambda}^k\mathbf{V}^{-1}\mathbf{v}^{(0)} = \mathbf{V}\begin{bmatrix} \alpha_1\lambda_1^k \\ \alpha_2\lambda_2^k \\ \vdots \\ \alpha_m\lambda_m^k \end{bmatrix} = \alpha_1\lambda_1^k\mathbf{V}\begin{bmatrix} 1 \\ \dfrac{\alpha_2}{\alpha_1}\dfrac{\lambda_2^k}{\lambda_1^k} \\ \vdots \\ \dfrac{\alpha_m}{\alpha_1}\dfrac{\lambda_m^k}{\lambda_1^k} \end{bmatrix}.$$

By $\mathbf{v}^{(k)} = \dfrac{\mathbf{A}^k\mathbf{v}^{(0)}}{\|\mathbf{A}^k\mathbf{v}^{(0)}\|_2}$, we have $\mathrm{e}^{-\mathrm{i}\theta_k}\mathbf{v}^{(k)} \to \mathbf{v}_1$ and $\lambda^{(k)} \to \lambda_1$, where $\theta_k = k\theta + \theta_0$ with $\mathrm{e}^{\mathrm{i}\theta} = \lambda_1/|\lambda_1|$ and $\mathrm{e}^{\mathrm{i}\theta_0} = \alpha_1/|\alpha_1|$.

Theorem 2

Suppose that $\mathbf{A}$ *is diagonalizable, i.e.,* $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ *with*

$$\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_m\}.$$

Furthermore, suppose $\mathbf{e}_1^*\mathbf{V}^{-1}\mathbf{v}^{(0)} \neq 0$, $\mathbf{e}_2^*\mathbf{V}^{-1}\mathbf{v}^{(0)} \neq 0$, $\|\mathbf{v}_1\|_2 = 1$, *and*

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_m|.$$

Then the iterates of power iteration satisfy, as $k \to \infty$,

$$\|\mathrm{e}^{-\mathrm{i}\theta_k}\mathbf{v}^{(k)} - \mathbf{v}_1\|_2 = \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right),$$

and

$$|\lambda^{(k)} - \lambda_1| = \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right) \quad \textit{or} \quad \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right).$$

4. Inverse iteration

Proposition 3

For any μ that is not an eigenvalue, the eigenvectors of $(\mathbf{A} - \mu\mathbf{I})^{-1}$ are the same as the eigenvectors of $\mathbf{A}$, and the corresponding eigenvalues are $\{(\lambda_j - \mu)^{-1}\}$, where $\{\lambda_j\}$ are the eigenvalues of $\mathbf{A}$.

Algorithm 2: Inverse iteration

$\mathbf{v}^{(0)}$ = some vector with $\|\mathbf{v}^{(0)}\|_2 = 1$

for $k = 1, 2, 3, \ldots,$

  Solve $(\mathbf{A} - \mu\mathbf{I})\mathbf{w} = \mathbf{v}^{(k-1)}$ for $\mathbf{w}$

  $\mathbf{v}^{(k)} = \mathbf{w}/\|\mathbf{w}\|_2$

  $\lambda^{(k)} = (\mathbf{v}^{(k)})^* \mathbf{A} \mathbf{v}^{(k)}$

end

- We call μ the *shift* of inverse iteration. Like power iteration, inverse iteration exhibits only *linear* convergence.
- Other important issue about stability: TreBau Exercise 27.5

Theorem 4

Suppose that $\mathbf{A}$ *is diagonalizable, i.e.,* $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ *with*

$$\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_m\}.$$

Suppose λ_j *is the closest eigenvalue to* μ, λ_l *is the second closest, and they satisfy*

$$|\lambda_j - \mu| < |\lambda_l - \mu| \leq |\lambda_i - \mu|$$

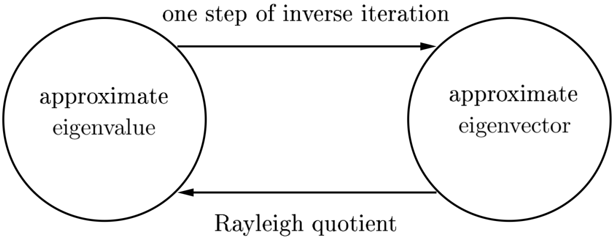
for each $i \neq j$. *Furthermore, suppose* $\mathbf{e}_j^* \mathbf{V}^{-1} \mathbf{v}^{(0)} \neq 0$, $\mathbf{e}_l^* \mathbf{V}^{-1} \mathbf{v}^{(0)} \neq 0$, *and* $\|\mathbf{v}_j\|_2 = 1$. *Then the iterates of inverse iteration satisfy, as* $k \to \infty$,

$$\|\mathrm{e}^{-\mathrm{i}\theta_k} \mathbf{v}^{(k)} - \mathbf{v}_j\|_2 = \mathcal{O}\left(\left| \frac{\lambda_j - \mu}{\lambda_l - \mu} \right|^k \right), \quad \text{(Exercise : } \theta_k = ?\text{)}$$

and

$$|\lambda^{(k)} - \lambda_j| = \mathcal{O}\left(\left| \frac{\lambda_j - \mu}{\lambda_l - \mu} \right|^k \right) \quad \textit{or} \quad \mathcal{O}\left(\left| \frac{\lambda_j - \mu}{\lambda_l - \mu} \right|^{2k} \right).$$

5. Rayleigh quotient iteration



Algorithm 3: Rayleigh quotient iteration

$\mathbf{v}^{(0)} =$ some vector with $\|\mathbf{v}^{(0)}\|_2 = 1$
$\lambda^{(0)} = (\mathbf{v}^{(0)})^* \mathbf{A} \mathbf{v}^{(0)}$
for $k = 1, 2, 3, \ldots,$
  Solve $(\mathbf{A} - \lambda^{(k-1)} \mathbf{I})\mathbf{w} = \mathbf{v}^{(k-1)}$ for $\mathbf{w}$
  $\mathbf{v}^{(k)} = \mathbf{w}/\|\mathbf{w}\|_2$
  $\lambda^{(k)} = (\mathbf{v}^{(k)})^* \mathbf{A} \mathbf{v}^{(k)}$
end

Theorem 5

Rayleigh quotient iteration converges to an eigenpair

$$\{\lambda, \mathbf{v}\}, \qquad \|\mathbf{v}\|_2 = 1,$$

for all except a set of measure zero of starting vectors $\mathbf{v}^{(0)}$. *When it converges, the convergence is ultimately quadratic* ($\alpha = 2$) *for non-normal case or cubic* ($\alpha = 3$) *for normal case in the sense that if* $\mathrm{e}^{-\mathrm{i}\theta_k}\mathbf{v}^{(k)}$ *is sufficiently close to the eigenvector* $\mathbf{v}$, *then*

$$\|\mathrm{e}^{-\mathrm{i}\theta_{k+1}}\mathbf{v}^{(k+1)} - \mathbf{v}\|_2 = \mathcal{O}(\|\mathrm{e}^{-\mathrm{i}\theta_k}\mathbf{v}^{(k)} - \mathbf{v}\|_2^\alpha)$$

and

$$|\lambda^{(k+1)} - \lambda| = \mathcal{O}(|\lambda^{(k)} - \lambda|^\alpha)$$

as $k \to \infty$.

Example:

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}, \quad \mathbf{v}^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Big/ \sqrt{3}.$$

The eigenvalue $\lambda = 5.214319743377$

Power iteration:

$$\lambda^{(0)} = 5$$
$$\lambda^{(1)} = 5.1818\ldots$$
$$\lambda^{(2)} = 5.2081\ldots$$
$$\lambda^{(3)} = 5.2130\ldots$$

Rayleigh quotient iteration:

$$\lambda^{(0)} = 5$$
$$\lambda^{(1)} = 5.2131\ldots$$
$$\lambda^{(2)} = 5.214319743184\ldots$$

The convergence of Rayleigh quotient iteration is spectacular: each iteration triples the number of digits of accuracy.