Unit 4. Eigenvalues

Numerical Analysis

EE/NTHU

March 22, 2017

Eigenvalues and Eigenvectors

Definition 4.1.1.

Given a real $n \times n$ matrix $\mathbf{A}$, the number $\lambda \in \mathbb{C}$ is an eigenvalue of $\mathbf{A}$ if there is a nonnull vector $\mathbf{x} \in \mathbb{R}$ such that

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}. \tag{4.1.1}$$

The vector $\mathbf{x}$ is the eigenvector associated with the eigenvalue λ and the set of eigenvalues of $\mathbf{A}$ is the spectrum of $\mathbf{A}$, denoted by $\sigma(\mathbf{A})$.

Theorem 4.1.2.

The eigenvalue λ corresponding to the eigenvector $\mathbf{x}$ can be determined by computing the Rayleigh quotient

$$\lambda = \mathbf{x}^T\mathbf{A}\mathbf{x}/\mathbf{x}^T\mathbf{x}. \tag{4.1.2}$$

Definition 4.1.3.

The characteristic polynomial of the real $n \times n$ matrix $\mathbf{A}$ is

$$p_A(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}). \tag{4.1.3}$$

Eigenvalues and Eigenvectors, II

Theorem 4.1.4.

The eigenvalue λ is the solution of the characteristic equation

$$p_A(\lambda) = 0. \tag{4.1.4}$$

- Since the characteristic polynomial is of degree n with respect to λ, there exist n eigenvalues of $\mathbf{A}$. But these eigenvalues may not be distinct from each other.

Theorem 4.1.5.

Given real $n \times n$ matrix $\mathbf{A}$ with the eigenvalues λ_i, $i = 1, \ldots, n$ then

$$\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i, \tag{4.1.5}$$

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i. \tag{4.1.6}$$

Eigenvalues and Eigenvectors, III

Theorem 4.1.6.

Given a real $n \times n$ matrix $\mathbf{A}$ with the eigenvalues λ_i, $i = 1, \ldots, n$, then

1. $\mathbf{A}$ is singular if and only if there is a $\lambda_i = 0$, $1 \le i \le n$.
2. Complex eigenvalues of $\mathbf{A}$ occur in conjugate pairs.
3. The matrix polynomial $p_A(\mathbf{A})$ satisfies

$$p_A(\mathbf{A}) = \mathbf{0}. \tag{4.1.7}$$

Theorem 4.1.7.

The spectral radius of a real matrix $\mathbf{A}$ is defined as

$$\rho(\mathbf{A}) = \max_{\lambda \in \sigma(\mathbf{A})} |\lambda|. \tag{4.1.8}$$

Eigenvalues and Eigenvectors, IV

Theorem 4.1.8.

If the real matrix $\mathbf{A}$ has the following block triangular form

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1k} \\ 0 & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2k} \\ \vdots & \vdots & \ddots & \mathbf{A}_{2k} \\ 0 & \cdots & 0 & \mathbf{A}_{kk} \end{bmatrix} \tag{4.1.9}$$

Then

$$p_A(\lambda) = p_{A_{11}}(\lambda) \cdot p_{A_{22}}(\lambda) \cdots p_{A_{kk}}(\lambda), \tag{4.1.10}$$

$$\sigma(\mathbf{A}) = \bigcup_{j=1}^{k} \sigma(\mathbf{A}_{jj}). \tag{4.1.11}$$

Theorem 4.1.9.

If the real matrix $\mathbf{A}$ is triangular then

$$\sigma(\mathbf{A}) = \{a_{ii} | i = 1, \cdots, n\}.$$

Eigenvalues and Eigenvectors, V

Definition 4.1.10.

Given a real $n \times n$ matrix $\mathbf{A}$ with the characteristic polynomial

$$p_A(\lambda) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{\delta_i} \tag{4.1.12}$$

with $\lambda_i \neq \lambda_j$, $1 \leq i, j \leq k$, and $\sum_{i=1}^{k} \delta_i = n$. The algebraic multiplicity of λ_i is δ_i, $1 \leq i \leq n$.

Theorem 4.1.11.

The set of eigenvectors associated with a eigenvalue, λ, of a real $n \times n$ matrix $\mathbf{A}$ forms a subspace. The dimension of this subspace is called geometric multiplicity of the eigenvalue λ. For any λ the geometric multiplicity is less than or equal to the algebraic multiplicity.

The Power Method

- The power method can be used to estimate the largest and the smallest eigenvalues.
- These eigenvalues are needed for the condition number of a linear system solution, and other applications.
- In this section, we assume the real $n \times n$ matrix $\mathbf{A}$ has n distinct real eigenvalues and $\mathbf{x}_i$ is the eigenvector associated with eigenvalue λ_i, $i = 1, \cdots, n$.
- It is further assumed that the eigenvalues are ordered as

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \cdots \geq |\lambda_n|, \tag{4.1.13}$$

  where λ_1 has the algebraic multiplicity of 1.
- Under these assumptions, λ_1 is called the dominant eigenvalue of matrix $\mathbf{A}$.

The Power Method, II

Algorithm 4.1.12. Power Method

Given a diagnosable matrix $\mathbf{A}$ and an arbitrary initial vector $\mathbf{q}^{(0)}$, then

$$\mathbf{q}^{(k)} = \frac{\mathbf{A}^k\mathbf{q}^{(0)}}{\|\mathbf{A}^k\mathbf{q}^{(0)}\|_2}, \tag{4.1.14}$$

$$\nu^{(k)} = (\mathbf{q}^{(k)})^T\mathbf{A}\mathbf{q}^{(k)}. \tag{4.1.15}$$

- Since $\mathbf{A}$ is diagnosable, the eigenvectors form a basis of $\mathbb{R}^n$. Thus, $\mathbf{q}^{(0)}$ can be expressed as

$$\mathbf{q}^{(0)} = \sum_{i=1}^{n} \alpha_i\mathbf{x}_i = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_n\mathbf{x}_n$$

  Then

$$\mathbf{A}\mathbf{q}^{(0)} = \mathbf{A}\sum_{i=1}^{n} \alpha_i\mathbf{x}_i = \sum_{i=1}^{n} \alpha_i\lambda_i\mathbf{x}_i$$

The Power Method, III

- And

$$\mathbf{A}^k\mathbf{q}^{(0)} = \mathbf{A}^k\sum_{i=1}^{n}\alpha_i\mathbf{x}_i = \sum_{i=1}^{n}\alpha_i\mathbf{A}^k\mathbf{x}_i = \sum_{i=1}^{n}\alpha_i\lambda_i^k\mathbf{x}_i$$

$$= \alpha_1\lambda_1^k\Big(\mathbf{x}_1 + \sum_{i=2}^{n}\frac{\alpha_i}{\alpha_1}(\frac{\lambda_i}{\lambda_1})^k\mathbf{x}_i\Big)$$

When $k \to \infty$,

$$\mathbf{A}^k\mathbf{q}^{(0)} \to \alpha_1\lambda_1^k\mathbf{x}_1, \text{ if } \alpha_1 \neq 0.$$

Or, we can write

$$\mathbf{q}^{(k)} = \frac{\alpha_1\lambda_1^k(\mathbf{x}_1 + \mathbf{y}^{(k)})}{\|\alpha_1\lambda_1^k(\mathbf{x}_1 + \mathbf{y}^{(k)})\|_2}$$

where $\mathbf{y}^{(k)} = \sum_{i=2}^{n}\frac{\alpha_i}{\alpha_1}(\frac{\lambda_i}{\lambda_1})^k\mathbf{x}_i$, and $\mathbf{y}^{(k)} \to 0$ when $k \to \infty$.

Also, if $\alpha_1 \neq 0$, as $k \to \infty$

$$\mathbf{q}^{(k)} \to \mathbf{x}_1 \qquad (4.1.16)$$

$$\nu^{(k)} \to \lambda_1 \qquad (4.1.17)$$

The Power Method, IV

Theorem 4.1.13.

Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a diagnosable matrix whose eigenvalues satisfy equation (4.1.13). Assuming $\alpha_1 \neq 0$, there is a constant $C > 0$ such that

$$\|\tilde{\mathbf{q}}^{(k)} - \mathbf{x}_1\|_2 \leq C\left|\frac{\lambda_2}{\lambda_1}\right|^k, \qquad k \geq 1, \qquad (4.1.18)$$

where

$$\tilde{\mathbf{q}}^{(k)} = \frac{\mathbf{q}^{(k)}\|\mathbf{A}^k\mathbf{q}^{(0)}\|_2}{\alpha_1\lambda_1^k} = \mathbf{x}_1 + \sum_{i=2}^{n}\frac{\alpha_i}{\alpha_1}(\frac{\lambda_i}{\lambda_1})^k\mathbf{x}_i, \qquad k = 1, 2, \ldots. \qquad (4.1.19)$$

$$\|\tilde{\mathbf{q}}^{(k)} - \mathbf{x}_1\|_2 = \Big\|\sum_{i=2}^{n}\frac{\alpha_i}{\alpha_1}(\frac{\lambda_i}{\lambda_1})^k\mathbf{x}_i\Big\|_2 \leq \Big(\sum_{i=2}^{n}\Big[\frac{\alpha_i}{\alpha_1}(\frac{\lambda_i}{\lambda_1})^k\Big]^2\Big)^{1/2}$$

$$\leq \Big(\sum_{i=2}^{n}\Big[\frac{\alpha_i}{\alpha_1}(\frac{\lambda_2}{\lambda_1})^k\Big]^2\Big)^{1/2} = \left|\frac{\lambda_2}{\lambda_1}\right|^k\Big[\sum_{i=2}^{n}(\frac{\alpha_i}{\alpha_1})^2\Big]^{1/2}$$

Thus, $C = \Big[\sum_{i=2}^{n}(\frac{\alpha_i}{\alpha_1})^2\Big]^{1/2}$, and C is independent of k.

The Power Method, V

- The preceding theorem states that the power method converges with the rate $\left|\frac{\lambda_2}{\lambda_1}\right|$.
  - If $|\lambda_2| \ll |\lambda_1|$ then it converges quickly,
  - On the other hand, if $|\lambda_2| \approx |\lambda_1|$ then it converges slowly.
- $\tilde{\mathbf{q}}^{(k)}$ converges to $\mathbf{x}_1$
- Since $\tilde{\mathbf{q}}^{(k)} = \dfrac{\mathbf{q}^{(k)}\|\mathbf{A}^k\mathbf{q}^{(0)}\|_2}{\alpha_1\lambda_1^k}$, $\mathbf{q}^{(k)}$ converges, too.
  - See textbook.
- And $\nu^{(k)} = (\mathbf{q}^{(k)})^T\mathbf{A}\mathbf{q}^{(k)}$ converges to λ_1.
  - By the rate $\left|\frac{\lambda_2}{\lambda_1}\right|$.
- If $\mathbf{A}$ is real and symmetric and $\alpha_1 \neq 0$, then it can be shown that

$$|\lambda_1 - \nu^{(k)}| \leq |\lambda_1 - \lambda_n| \cdot \tan^2(\theta_0) \cdot \left|\frac{\lambda_2}{\lambda_1}\right|^{2k}. \tag{4.1.20}$$

  where $\cos(\theta_0) = |\mathbf{x}_1^T\mathbf{q}^{(0)}| \neq 0$.
  - In this case, the convergence rate is quadratic, $\left|\frac{\lambda_2}{\lambda_1}\right|^2$.

Stopping Criteria

- So far, we know that $\lim_{k\to\infty} \mathbf{q}^{(k)} = \mathbf{x}_1$, the eigenvector associated with λ_1 of matrix $\mathbf{A}$, and $\lim_{k\to\infty} \nu^{(k)} = \lambda_1$, which is the eigenvalue with the largest module.
- Define the residue at iteration k as

$$\mathbf{r}^{(k)} = \mathbf{A}\mathbf{q}^{(k)} - \nu^{(k)}\mathbf{q}^{(k)}, \qquad k \geq 1. \tag{4.1.21}$$

  Then, as $k \to \infty$, $\mathbf{r}^{(k)} \to \mathbf{0}$. One can use $\|\mathbf{r}^{(k)}\|_2$ as a stopping criterion.
- In fact, it has been shown that

$$|\lambda_1 - \nu^{(k)}| \simeq \frac{\|\mathbf{r}^{(k)}\|_2}{|(\mathbf{w}^{(k)})^T\mathbf{q}^{(k)}|}, \qquad k \geq 1, \tag{4.1.22}$$

  where $\mathbf{w}^{(k)}$ satisfies $(\mathbf{w}^{(k)})^T\mathbf{A} = \nu^{(k)}(\mathbf{w}^{(k)})^T$, and as $k \to \infty$, $\mathbf{w}^{(k)} \to \mathbf{w}$ and $\mathbf{w}^T\mathbf{A} = \lambda_1\mathbf{w}^T$ is the left eigenvector associated with λ_1.
  If $\mathbf{A}$ is symmetric the $\mathbf{w} = \mathbf{q}$.
- One approach is to use Eq. (4.1.22) as the stopping criterion.

The Algorithm

Algorithm 4.1.14. The Power Method

Given a diagnosable matrix $\mathbf{A}$, an initial vector $\mathbf{q}^{(0)}$, a small number ϵ and a large integer $maxiter$, let

$tol = 1 + \epsilon$, $\mathbf{q}^{(0)} = \dfrac{\mathbf{q}^{(0)}}{\|\mathbf{q}^{(0)}\|_2}$, and $k = 0$,

while $(tol \geq \epsilon$ and $k \leq maxiter)$ {

$$\mathbf{z} = \mathbf{A}\mathbf{q}^{(k)} ,$$

$$k = k + 1,$$

$$\mathbf{q}^{(k)} = \frac{\mathbf{z}}{\|\mathbf{z}\|_2} ,$$

$$\nu^{(k)} = (\mathbf{q}^{(k)})^T \mathbf{A}\mathbf{q}^{(k)} ,$$

$$\mathbf{r}^{(k)} = \mathbf{A}\mathbf{q}^{(k)} - \nu^{(k)}\mathbf{q}^{(k)} ,$$

$$(\mathbf{u}^{(k)})^T = (\mathbf{q}^{(k)})^T \mathbf{A} ,$$

$$\mathbf{w}^{(k)} = \frac{\mathbf{u}}{\|\mathbf{u}\|_2} ,$$

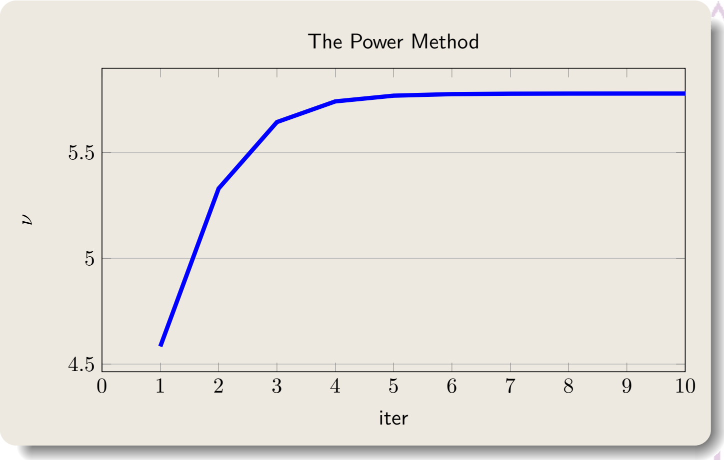
$$tol = \frac{\|\mathbf{r}^{(k)}\|_2}{|(\mathbf{w}^{(k)})^T \mathbf{q}^{(k)}|} ,$$

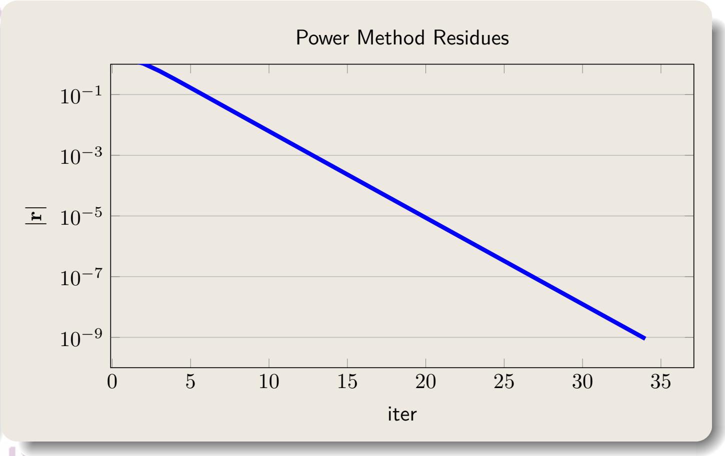
}

The Algorithm, II

- Note that it can also check for $|(\mathbf{w}^{(k)})^T \mathbf{q}^{(k)}|$. If this number is 0, then λ_1 does not have the algebraic multiplicity of 1.
- If the algebraic multiplicity of λ_1 is greater than 1, then $\mathbf{q}^{(k)}$ may not converge though $\nu^{(k)}$ is convergent.
  - $\mathbf{r}^{(k)}$ may not be convergent either.
  - Thus, using $\|\mathbf{r}^{(k)}\|_2$ as the stopping criterion may not work.
- If $\mathbf{A}$ is symmetric then the left eigenvector $\mathbf{q}_i$ of a eigenvalue λ_i is always the same as the right eigenvector $\mathbf{w}_i$ associated with λ_i, and $|(\mathbf{w}^{(k)})^T \mathbf{q}^{(k)}| = 1$.
- For asymmetric $\mathbf{A}$, this property does not hold but $|(\mathbf{w}^{(k)})^T \mathbf{q}^{(k)}|$ is convergent to a single number as both eigenvectors converges.

The Power Method, Example





- Using the resistor network example of Unit 3
- The power method is shown to be convergent with a constant rate
  - $\lambda_1 = 5.77846.$

The Power Method, Complexity

- The first power method algorithm, Algorithm (4.1.12), is usually formulated as

$$\mathbf{q}^{(k+1)} = \frac{\mathbf{A}\mathbf{q}^{(k)}}{\|\mathbf{A}\mathbf{q}^{(k)}\|_2}, \tag{4.1.23}$$

$$\nu^{(k+1)} = (\mathbf{q}^{(k+1)})^T \mathbf{A}\mathbf{q}^{(k+1)}. \tag{4.1.24}$$

- The computation is dominated by $\mathbf{A}\mathbf{q}^{(k)}$.
  - Matrix-vector multiplication.
  - Computational complexity is $\mathcal{O}(n^2)$ per iteration.
  - Overall computational complexity is $\mathcal{O}(N_{iter} \times n^2)$.
    - N_{iter} is the number of iterations needed to reach to a converged solution.
    - A function of $\left|\dfrac{\lambda_2}{\lambda_1}\right|$.
- The computation of the second form of power method, Algorithm (4.1.14), is also dominated by $\mathbf{A}\mathbf{q}^{(k)}$ and $(\mathbf{q}^{(k)})^T\mathbf{A}$.
  - Matrix-vector multiplications.
  - Overall computational complexity remains $\mathcal{O}(N_{iter} \times n^2)$.
  - But with a larger coefficient.

Inverse Power Method

- The power method can be modified to find the eigenvalue with the smallest module, and the eigenvector associated.

Algorithm 4.1.15. Inverse Power Method

Given a diagnosable matrix $\mathbf{A}$ and an arbitrary initial guess $\mathbf{q}^{(0)}$ with unit Euclidean norm, iterate for $k = 1, \ldots$.

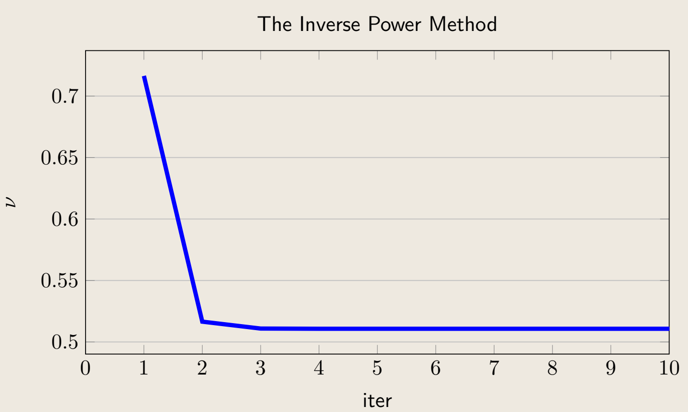
$$\mathbf{A}\mathbf{z}^{(k)} = \mathbf{q}^{(k-1)}, \tag{4.1.25}$$

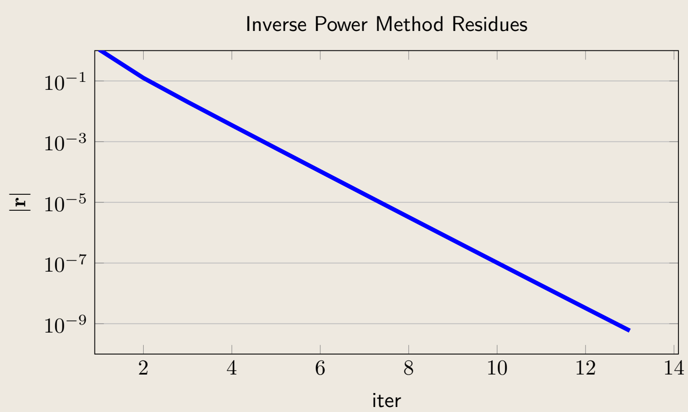
$$\mathbf{q}^{(k)} = \frac{\mathbf{z}^{(k)}}{\|\mathbf{z}^{(k)}\|_2}, \tag{4.1.26}$$

$$\mu^{(k)} = (\mathbf{q}^{(k)})^T \mathbf{A}\mathbf{q}^{(k)}. \tag{4.1.27}$$

- Note that the vector $\mathbf{z}^{(k)}$ can be found using LU decomposition or any linear system solution method.
- Other than that this is the power method and its convergence rate is determined by $\left|\frac{\lambda_n}{\lambda_{n-1}}\right|$.

Inverse Iteration Example





- Using the resistor network example of Unit 3
- The power method is shown to be convergent with a constant rate
  - $\lambda_7 = 0.510711$.

The Inverse Power Method with Shifting

- The inverse power method can be generalized to find eigenvalue that is closest to a specific number, ω, and the eigenvector associated.

Algorithm 4.1.16. Inverse Power Method with Shifting

Given a diagnosable matrix $\mathbf{A}$ and an arbitrary initial guess $\mathbf{q}^{(0)}$ with unit Euclidean norm and $\omega \in \mathbb{R}$, iterate for $k = 1, \ldots$.

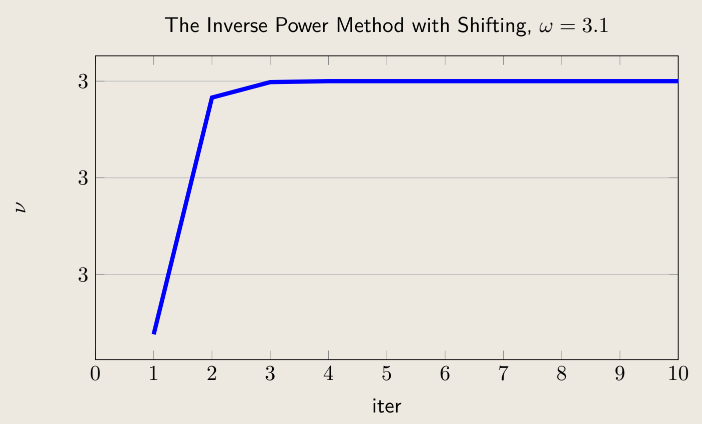
$$(\mathbf{A} - \omega\mathbf{I})\mathbf{z}^{(k)} = \mathbf{q}^{(k-1)}, \tag{4.1.28}$$

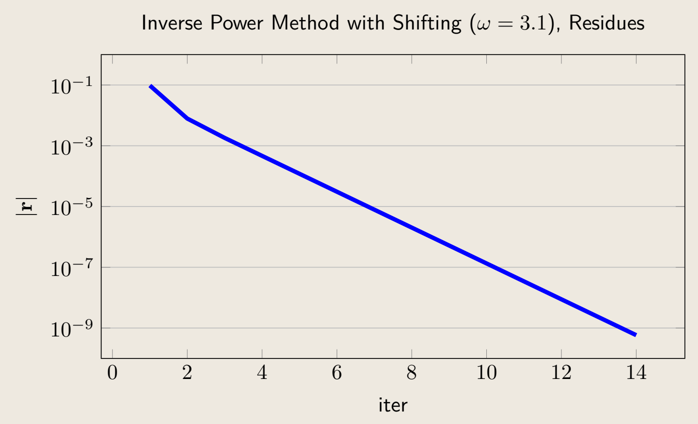
$$\mathbf{q}^{(k)} = \frac{\mathbf{z}^{(k)}}{\|\mathbf{z}^{(k)}\|_2}, \tag{4.1.29}$$

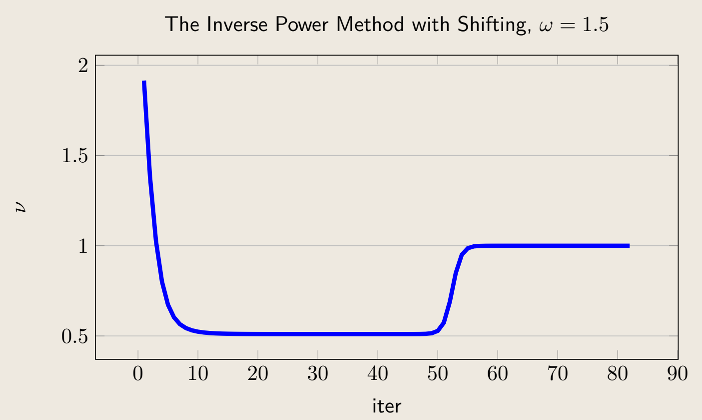
$$\mu^{(k)} = (\mathbf{q}^{(k)})^T\mathbf{A}\mathbf{q}^{(k)}. \tag{4.1.30}$$

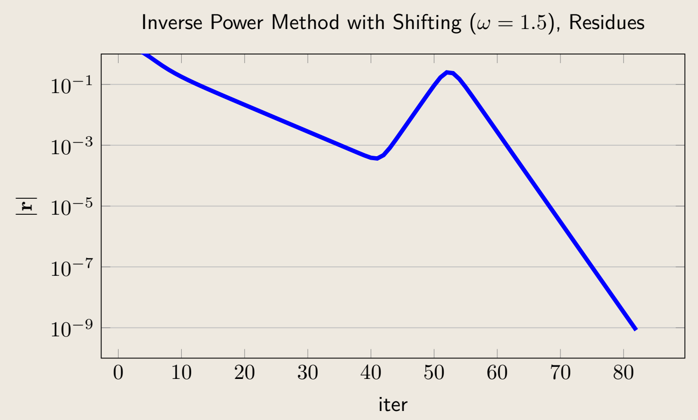
- Except Eq. (4.1.28), this algorithm is identical to the inverse power method.
- And the convergence rate is determined by the ratio of the two eigenvalues that are closest to ω.

Inverse Power Method with Shifting Examples









Power Methods

- In our development we have assumed that $\lambda_1 > \lambda_2$, or the algebraic multiplicity of λ_1 is 1. But, the power method is convergent when $\lambda_1 = \lambda_2$, since the vectors generated by Eq. (4.1.14) converge to the subspace spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$, and ν converges to λ_1. (The original form of power method, Algorithm (4.1.12).)
- When $\lambda_1 = -\lambda_2$ then the power method would oscillate and not converge.
- If $\lambda_1 = \overline{\lambda_2}$, it would also oscillate and not converge.
- The inverse power method with shifting is very effective in find eigenvalues and eigenvectors for diagnosable matrices.
- With ω is properly positioned, the eigenvalue closest to ω can be quickly found. The convergence rate can be improved greatly if ω is close to λ_i.
- Initial guess $\mathbf{q}^{(0)}$ can affect the convergence rate as shown in the last example. But, due to the computer round off, the eigenvalue closest to ω is always found, even $\mathbf{q}^{(0)}$ is an eigenvector of a different eigenvalue.
- The inverse power method with shifting is more effective even though it needs more operations for each iteration: matrix addition and forward and backward substitutions. (note that LU decomposition needs to be done only once.)

Summary

- The power method.
  - Simple form.
  - More elaborated form.
- The inverse power method.
- Power method with shifting.