

Eigenvalues and eigenvectors

Let $A \in \mathbb{R}^{n \times n}$. If $0 \neq v \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$ satisfy

$$Av = \lambda v$$

then λ is called **eigenvalue**, and v is called **eigenvector**.

Given a matrix, we want to approximate its eigenvalues and eigenvectors.
Some applications:

- Structural engineering (natural frequency, earthquakes)
- Electromagnetics (resonance cavity)
- Google's PageRank algorithm
- ...

The characteristic polynomial

The eigenvalues of a matrix are the roots of **the characteristic polynomial**

$$p(\lambda) := \det(\lambda I - A) = 0$$

However, computing the roots of a polynomial is a very ill-conditioned problem! We cannot use this approach to compute the eigenvalues.

Eigenvalues and eigenvectors

Algorithms that compute the eigenvalues/eigenvectors of a matrix are divided into two categories:

- ① Methods that compute all the eigenvalues/eigenvectors at once.
- ② Methods that compute only a few (possibly one) eigenvalues/eigenvectors.

The methods are also different whether the matrix is symmetric or not. In this lesson we will discuss methods of type 2.

Diagonalizable matrices

Definition

We say that a matrix $A \in \mathbb{C}^{n \times n}$ is diagonalizable if there exists a non singular matrix U and a diagonal matrix D such that $U^{-1}AU = D$.

The diagonal elements of D are the eigenvalues of A and the column u_i of U is an eigenvector of A relative to the eigenvalue $D_{i,i}$.

Since a nonzero scalar multiple of an eigenvector is still an eigenvector, we can choose U such that $\|u_i\|_2 = 1$ for $i = 1, \dots, n$.

Finally, we observe that if A is diagonalizable, since U is non singular, then the vectors $\{u_1, \dots, u_n\}$ form a basis of $\mathbb{C}^n$.

From now on, we assume that the eigenvalues are numbered in decreasing order (in modulus), i.e.

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$$

Eigenvalues/eigenvectors of a symmetric matrix (refresh)

Theorem

All the eigenvalues of a real symmetric matrix are **real**. Moreover, there exists a basis of eigenvectors $u_1, \dots, u_n$, i.e.

$$Au_i = \lambda_i u_i$$

that are orthonormal, i.e.

$$(u_i, u_j) = \delta_{ij}$$

The power method

To approximate the eigenvalue of A largest in modulus, and its eigenvector.

Power method

v_0 = some vector with $\|v_0\|_2 = 1$.

for $k = 1, 2, \dots$ (requires a stopping criterion)

$$w = Av_{k-1}$$

apply A

$$v_k = w / \|w\|_2$$

normalize

$$\mu_k = (v_k)^H Av_k$$

Rayleigh quotient

end

Theorem

There is a constant $C > 0$ such that given a diagonalizable matrix $A \in \mathbb{C}^{n \times n}$, assuming $|\lambda_1| > |\lambda_2|$ and $v_0 = \sum_{i=1}^n \alpha_i u_i$, with $\alpha_1 \neq 0$, then there is a sequence $c_k \in \mathbb{C}$ such that

$$\|c_k v_k - u_1\|_2 \leq C \left| \frac{\lambda_2}{\lambda_1} \right|^k. \quad (1)$$

Proof

We expand v_0 on the eigenvector basis $\{u_1, \dots, u_n\}$ chosen s.t. $\|u_i\|_2 = 1$ for $i = 1, \dots, n$:

$$v_0 = \sum_{i=1}^n \alpha_i u_i, \quad \text{with } \alpha_1 \neq 0$$

It holds

$$A^k v_0 = \sum_{i=1}^n \alpha_i \lambda_i^k u_i \quad \text{and} \quad v_k = \frac{A^k v_0}{\|A^k v_0\|_2}$$

We then define

$$\tilde{v}_k = \frac{A^k v_0}{\alpha_1 \lambda_1^k} = u_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1} \left(\frac{\lambda_i}{\lambda_1} \right)^k u_i$$

At this point, it holds

$$\|\tilde{v}_k - u_1\|_2 = \left\| \sum_{i=2}^n \frac{\alpha_i}{\alpha_1} \left(\frac{\lambda_i}{\lambda_1} \right)^k u_i \right\|_2 \leq \sum_{i=2}^n \left\| \frac{\alpha_i}{\alpha_1} \left(\frac{\lambda_i}{\lambda_1} \right)^k u_i \right\|_2 = \sum_{i=2}^n \left| \frac{\alpha_i}{\alpha_1} \right| \left| \frac{\lambda_i}{\lambda_1} \right|^k$$

So, we obtain

$$\|\tilde{v}_k - u_1\|_2 \leq \sum_{i=2}^n \left| \frac{\alpha_i}{\alpha_1} \right| \left| \frac{\lambda_i}{\lambda_1} \right|^k \leq (n-1) \cdot \max_{i=2,\dots,n} \left(\left| \frac{\alpha_i}{\alpha_1} \right| \right) \left| \frac{\lambda_2}{\lambda_1} \right|^k = C \left| \frac{\lambda_2}{\lambda_1} \right|^k,$$

where we have defined $C = (n-1) \cdot \max_{i=2,\dots,n} \left(\left| \frac{\alpha_i}{\alpha_1} \right| \right)$. Observing that C does not depend on k , and that $\tilde{v}_k = \frac{\|A^k v_0\|_2}{\alpha_1 \lambda_1^k} v_k$ the estimate (1) follows.

The previous theorem implies that the $c_k v_k$ converges to u_1 when $k \rightarrow \infty$. Since $c_k v_k$ is just a scalar multiple of v_k , it means that the linear space spanned by v_k (the so-called autospace) “converges” to the linear space spanned by u_1 . Thus v_k tends to be an eigenvector for the eigenvalue λ_1 .

In order to better understand (1), assume $A \in \mathbb{R}^{n \times n}$ is a (real) symmetric matrix and $v_0 \in \mathbb{R}^n$. In this case, choose real eigenvectors; the c_k in the previous proof are real valued. Further inspecting the proof, $|c_k|$ tends towards 1 as $k \rightarrow +\infty$. In particular $c_k = \frac{\|A^k v_0\|_2}{\alpha_1 \lambda_1^k}$ either tends towards 1, or towards -1 , or when $\lambda_1 < 0$ oscillates between $+1$ and -1 . Then one of the following four statements holds true:

- $v_k \xrightarrow{k \rightarrow +\infty} u_1$
- $v_k \xrightarrow{k \rightarrow +\infty} -u_1$
- $(-1)^k v_k \xrightarrow{k \rightarrow +\infty} u_1$
- $(-1)^k v_k \xrightarrow{k \rightarrow +\infty} -u_1$

We also have a convergence result for the approximation of λ_1 .

Theorem

There is a constant $C > 0$ such that given a diagonalizable matrix $A \in \mathbb{C}^{n \times n}$, if $|\lambda_1| > |\lambda_2|$ and $v_0 = \sum_{i=1}^n \alpha_i u_i$, with $\alpha_1 \neq 0$, then

$$|\mu_k - \lambda_1| \leq C \left(\left| \frac{\lambda_2}{\lambda_1} \right|^k \right), \quad \text{for } k \rightarrow +\infty.$$

Furthermore, if $A \in \mathbb{R}^{n \times n}$ is a (real) symmetric matrix, then it holds

$$|\mu_k - \lambda_1| \leq C \left(\left| \frac{\lambda_2}{\lambda_1} \right|^{2k} \right), \quad \text{for } k \rightarrow +\infty.$$

Remark

if $|\lambda_2| \ll |\lambda_1|$ the convergence will be fast. On the other hand, if $|\lambda_2| \approx |\lambda_1|$ the convergence will be slow.

Some observations

One of the hypotheses of the previous results is $\alpha_1 \neq 0$, where α_i are defined such that $v_0 = \sum_{i=1}^n \alpha_i u_i$. Clearly, $u_1, \dots, u_n$ are unknown and we cannot check if v_0 satisfies this hypothesis.

Practically this is not a real obstacle. If unfortunately we choose v_0 s.t $\alpha_1 = 0$, there are two possible cases:

- in *exact arithmetic*, if $\alpha_2 \neq 0$ and $|\lambda_2| > |\lambda_3|$, define $\hat{v}_k = A^k v_0 / (\alpha_2 \lambda_2^k)$. Then $\hat{v}_k \rightarrow u_2$ and $\mu_k \rightarrow \lambda_2$. Otherwise, convergence depends on the first nonzero coefficient and the corresponding spectral gap.
- in *finite arithmetic*, during the iterations of the Power Method, round-off errors can cause the appearance of a non-zero component in the direction of u_1 in a certain v_k . If this happens, the dominant component may eventually grow and lead to convergence towards the dominant eigenvalue λ_1 and its corresponding eigenvector u_1 .

Stopping criterion

A simple stopping criterion for the power method is based on the residual:

$$\text{Stop when } \|Av_k - \mu_k v_k\| \leq \text{tol}$$

How can we compute other eigenvalues and eigenvectors?

Choose a shift $\mu \in \mathbb{C}$ outside the spectrum of A . We want the eigenvalue closest to μ :

$$J \in \operatorname{argmin}_i |\mu - \lambda_i|, \quad \text{target eigenvalue } \lambda_J.$$

If $Au_i = \lambda_i u_i$, then

$$(A - \mu I)^{-1} u_i = \frac{1}{\lambda_i - \mu} u_i.$$

Thus the closest eigenvalue becomes the largest in modulus for the inverse shifted matrix. Apply the power method through linear solves with $A - \mu I$. Convergence to a single direction requires a uniquely closest eigenvalue and a nonzero initial component in its eigendirection, as specified in the theorem below.

Inverse power method with shift μ

Choose $\mu \notin \sigma(A)$, and v_0 with $\|v_0\|_2 = 1$. Factor $A - \mu I$ once and reuse the factors.

Inverse power method

Input: A , μ , v_0 , $\text{tol} > 0$, $\text{MAXITER} \geq 1$.

Set $\mu_0 = v_0^H A v_0$.

If $\|A v_0 - \mu_0 v_0\|_2 \leq \text{tol}$, return (μ_0, v_0) .

for $k = 1, \dots, \text{MAXITER}$

 Solve $(A - \mu I)w = v_{k-1}$.

$v_k = w / \|w\|_2$, $\mu_k = v_k^H A v_k$.

 If $\|A v_k - \mu_k v_k\|_2 \leq \text{tol}$, return (μ_k, v_k) .

end

Return (μ_k, v_k) and report that the tolerance was not met.

The residual tests whether an eigenpair has been approximated; it does not ensure that the shift is outside the spectrum or that the desired eigenvalue was selected.

Theorem

Assume A is diagonalizable, $\mu \notin \sigma(A)$, and $|\mu - \lambda_J| < |\mu - \lambda_i| \forall i = 1, \dots, n, i \neq J$, and $v_0 = \sum_{i=1}^n \alpha_i u_i$, with $\alpha_J \neq 0$. Then

$$\lim_{k \rightarrow +\infty} \mu_k = \lambda_J$$

and there is a sequence $c_k \in \mathbb{C}$ such that

$$\lim_{k \rightarrow +\infty} \|c_k v_k - u_J\|_2 = 0,$$

proof

The inverse power method is just a power method applied to $(A - \mu I)^{-1}$, and the previous results apply. Indeed, given μ and since $Au_i = \lambda_i u_i$, then $(A - \mu I)u_i = (\lambda_i - \mu)u_i$, and then $\frac{1}{\lambda_i - \mu} u_i = (A - \mu I)^{-1} u_i$. For the assumption, the largest (in modulus) eigenvalue of $(A - \mu I)^{-1}$ is $\frac{1}{\lambda_J - \mu}$, and the relative eigenvector is u_J .

Note that if $\mu = 0$, the method approximates the eigenvalue of A that is smallest in modulus.