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 an **eigenvalue** and v is an associated **eigenvector**.

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

- Structural engineering (natural frequency, earthquakes)
- Electromagnetics (resonance cavity)
- ...

The characteristic polynomial

The characteristic polynomial of A is

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

An eigenvalue is a root of this polynomial:

$$\lambda \text{ is an eigenvalue of } A \quad \Longleftrightarrow \quad p(\lambda) = 0.$$

Computing the polynomial explicitly and then finding its roots can be numerically unreliable: roots may be very sensitive to perturbations of the coefficients, especially near multiple roots, and forming the coefficients introduces additional rounding errors.

This approach is useful for small hand calculations. Practical numerical algorithms generally work directly with A, without forming p.

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 over $\mathbb{C}$

Definition

A matrix $A \in \mathbb{C}^{n\times n}$ is **diagonalizable over** $\mathbb{C}$ if there are a nonsingular $U \in \mathbb{C}^{n\times n}$ and a diagonal matrix D such that

$$U^{-1}AU = D, \qquad U = (u_1, \ldots, u_n), \qquad D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

The columns satisfy $Au_i = \lambda_i u_i$ and form a basis of $\mathbb{C}^n$. We choose $\|u_i\|_2 = 1$; the basis need not be orthogonal.

We number the eigenvalues in decreasing order of absolute value:

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

Spectral theorem: the real symmetric case

Theorem

Every real symmetric matrix $A \in \mathbb{R}^{n\times n}$ has real eigenvalues and an orthonormal basis of real eigenvectors:

$$Au_i = \lambda_i u_i, \qquad (u_i, u_j) = \delta_{ij}.$$

Equivalently, $A = UDU^T$, where U is real orthogonal and D is real diagonal.

In particular, every real symmetric matrix is diagonalizable over $\mathbb{R}$.

From now on, except in the slide on the complex extension, we work with $A \in \mathbb{R}^{n\times n}$ diagonalizable over $\mathbb{R}$ and real starting vectors. Symmetry is only required for the sharper eigenvalue estimate.

The power method

To approximate the eigenvalue of A largest in modulus and an associated eigenvector. All vectors are real. Use the Euclidean norm and the scalar product $(x, y) = y^T x$.

Power method

Choose $v^{(0)} \in \mathbb{R}^n$ with $\|v^{(0)}\|_2 = 1$.
for $k = 1, \ldots, \mathrm{MAXITER}$
  $w = Av^{(k-1)}$ — apply A
  If $w = 0$, print "the iterate is in $\ker(A)$" and exit.
  $v^{(k)} = w/\|w\|_2$ — normalize
  $\mu_k = (Av^{(k)}, v^{(k)})$ — Rayleigh quotient
  Stop if the eigenpair residual meets the tolerance (see below).
end

The Rayleigh formula above uses $\|v^{(k)}\|_2 = 1$; for an arbitrary nonzero vector v, the Rayleigh quotient is $(Av, v)/(v, v)$.

From repeated products to powers of A

Why is this called the *power* method, even though we never form A^k?

Without normalization, $z^{(0)} = v^{(0)}$ and $z^{(k)} = Az^{(k-1)}$ give

$$z^{(1)} = Av^{(0)}, \qquad z^{(2)} = A^2 v^{(0)}, \qquad z^{(k)} = A^k v^{(0)}.$$

The power method normalizes at every step. For $k = 1$,

$$v^{(1)} = \frac{Av^{(0)}}{\|Av^{(0)}\|_2}.$$

For $k = 2$, put $s_1 = \|Av^{(0)}\|_2 > 0$. Then

$$Av^{(1)} = \frac{A^2 v^{(0)}}{s_1}, \qquad \|Av^{(1)}\|_2 = \frac{\|A^2 v^{(0)}\|_2}{s_1}.$$

Therefore the normalization factor from the first step cancels:

$$v^{(2)} = \frac{Av^{(1)}}{\|Av^{(1)}\|_2} = \frac{A^2 v^{(0)}/s_1}{\|A^2 v^{(0)}\|_2/s_1} = \frac{A^2 v^{(0)}}{\|A^2 v^{(0)}\|_2}.$$

The normalized-power identity: induction

Suppose $A^j v^{(0)} \neq 0$ for the steps considered. We prove

$$\boxed{v^{(k)} = \frac{A^k v^{(0)}}{\|A^k v^{(0)}\|_2}}.$$

The first step was checked above. If the identity holds at step $k-1$, set $s_{k-1} = \|A^{k-1} v^{(0)}\|_2 > 0$. Linearity and homogeneity of the norm give

$$Av^{(k-1)} = \frac{A^k v^{(0)}}{s_{k-1}}, \qquad \|Av^{(k-1)}\|_2 = \frac{\|A^k v^{(0)}\|_2}{s_{k-1}}.$$

Normalizing cancels s_{k-1}, proving the identity at step k.

Thus $v^{(k)}$ has the direction of $A^k v^{(0)}$ and unit norm. Computing the next vector requires one matrix–vector product; no power A^k is formed. Normalization keeps the vectors at a controlled magnitude.

This identity holds for **any matrix** A while the products are nonzero.

Convergence of the power method

Theorem

Let $A \in \mathbb{R}^{n \times n}$ be diagonalizable over $\mathbb{R}$, with real normalized eigenvectors u_i and eigenvalues ordered by absolute value. Assume

$$|\lambda_1| > |\lambda_2|, \qquad v^{(0)} = \sum_{i=1}^{n} \alpha_i u_i, \qquad \alpha_1 \neq 0.$$

Then there are a constant $C > 0$ and real scalars c_k such that

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

Proof

We expand $v^{(0)}$ on the eigenvector basis $\{u_1, \ldots, u_n\}$ chosen s.t. $\|u_i\|_2 = 1$ for $i = 1, \ldots, n$:

$$v^{(0)} = \sum_{i=1}^{n} \alpha_i u_i, \qquad \text{with } \alpha_1 \neq 0$$

Using the normalized-power identity just proved, we have

$$A^k v^{(0)} = \sum_{i=1}^{n} \alpha_i \lambda_i^k u_i \qquad \text{and} \qquad v^{(k)} = \frac{A^k v^{(0)}}{\|A^k v^{(0)}\|_2}$$

We then define

$$\widetilde{v}^{(k)} = \frac{A^k v^{(0)}}{\alpha_1 \lambda_1^k} = u_1 + \sum_{i=2}^{n} \frac{\alpha_i}{\alpha_1} \left(\frac{\lambda_i}{\lambda_1}\right)^k u_i$$

At this point, it holds

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

So, we obtain

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

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 $\widetilde{v}^{(k)} = \frac{\|A^k v^{(0)}\|_2}{\alpha_1 \lambda_1^k} v^{(k)}$ the estimate (1) follows.

Hence $c_k v^{(k)} \to u_1$: the line spanned by $v^{(k)}$ approaches the eigenspace $\mathrm{span}\{u_1\}$. The vectors $v^{(k)}$ themselves need not have a limit, because their sign may vary with k.

Convergence up to a sign

Remark

We proved that

$$\widetilde{v}^{(k)} = \frac{A^k v^{(0)}}{\alpha_1 \lambda_1^k} \longrightarrow u_1, \qquad \|u_1\|_2 = 1.$$

Therefore

$$|c_k| = \frac{\|A^k v^{(0)}\|_2}{|\alpha_1|\,|\lambda_1|^k} = \|\widetilde{v}^{(k)}\|_2 \longrightarrow 1.$$

The sign of c_k is the sign of $\alpha_1 \lambda_1^k$. Consequently,

$$\left\| v^{(k)} - \text{sign}(\alpha_1 \lambda_1^k) u_1 \right\|_2 \longrightarrow 0.$$

- If $\lambda_1 > 0$, then $v^{(k)} \to \text{sign}(\alpha_1) u_1$.
- If $\lambda_1 < 0$, the iterates alternate signs, while $(-1)^k v^{(k)} \to \text{sign}(\alpha_1) u_1$.

Convergence of the eigenvalue approximation

Theorem

Let $A \in \mathbb{R}^{n \times n}$ be diagonalizable over $\mathbb{R}$, with $|\lambda_1| > |\lambda_2|$, and let $v^{(0)} = \sum_{i=1}^{n} \alpha_i u_i$ with $\alpha_1 \neq 0$. Then the Rayleigh estimates $\mu_k = (Av^{(k)}, v^{(k)})$ satisfy

$$|\mu_k - \lambda_1| \leq C \left| \frac{\lambda_2}{\lambda_1} \right|^k \quad \text{for all sufficiently large } k,$$

where $C > 0$ is independent of k.

For a real symmetric A, we also state without proof the sharper estimate

$$|\mu_k - \lambda_1| \leq C \left| \frac{\lambda_2}{\lambda_1} \right|^{2k} \quad \text{for all sufficiently large } k.$$

Remark

If $|\lambda_2| \ll |\lambda_1|$, convergence is fast. If $|\lambda_2| \approx |\lambda_1|$, convergence is 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, $u_1, \ldots, u_n$ are unknown and we cannot check if $v^{(0)}$ satisfies this hypothesis.
A random starting vector is a usual practical choice. If we choose $v^{(0)}$ s.t $\alpha_1 = 0$, there are two possible cases:

- In *exact arithmetic*, $A^k v^{(0)}$ stays in the span of the eigenvectors with $\alpha_i \neq 0$. If one of their nonzero eigenvalues has uniquely largest modulus, the method targets that eigenpair instead. Equal moduli may prevent convergence to one direction. If $Av^{(0)} = 0$, the initial vector is already a zero-eigenvalue eigenvector.
- In *floating-point 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:

$$\text{Stop when} \quad \left\| A v^{(k)} - \mu_k v^{(k)} \right\|_2 \leq \texttt{tol}$$

Here $\texttt{tol} > 0$ is a prescribed residual threshold. A small residual means that the eigenvalue equation is nearly satisfied; it does not by itself certify that the eigenvalue is the dominant one.

Always impose an iteration limit and report if the residual test is not met.

The power method in complex arithmetic

Take $A \in \mathbb{C}^{n\times n}$ and $v^{(0)} \in \mathbb{C}^n$ with unit norm. Use $(x, y) = y^H x = \overline{y}^T x$ and $\|x\|_2 = \sqrt{(x, x)}$, where H denotes conjugate transpose.

Same iteration, Hermitian Rayleigh quotient:

$$w = Av^{(k-1)}, \qquad v^{(k)} = \frac{w}{\|w\|_2}, \qquad \mu_k = (Av^{(k)}, v^{(k)}) = (v^{(k)})^H Av^{(k)}.$$

The convergence theory still applies

Assume A is diagonalizable over $\mathbb{C}$, $|\lambda_1| > |\lambda_2|$, and $v^{(0)} = \sum_i \alpha_i u_i$ with $\alpha_1 \neq 0$. The preceding general vector and eigenvalue estimates remain valid:

$$c_k v^{(k)} \to u_1, \qquad c_k \in \mathbb{C}, \qquad |c_k| \to 1, \qquad \mu_k \to \lambda_1.$$

Convergence of the vectors is up to a complex phase.

How can we compute other eigenvalues and eigenvectors?

Choose a real shift μ and assume it is not an eigenvalue of A. We want the eigenvalue closest to μ:

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

If $Au_i = \lambda_i u_i$, then

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

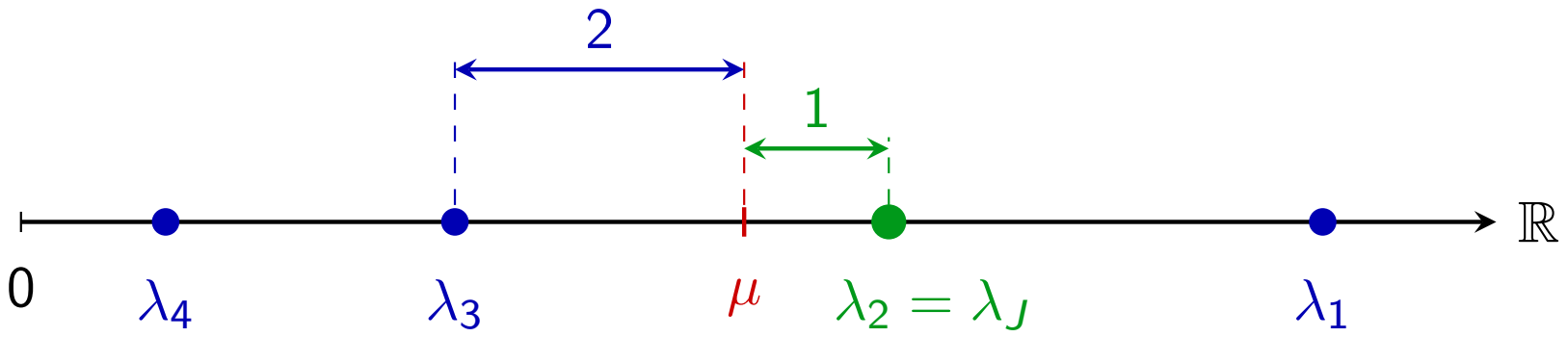
Thus the closest eigenvalue becomes the largest in modulus for the inverse shifted matrix. Apply the power method through linear solves with $A - \mu I$. Convergence to a single eigendirection requires a simple eigenvalue

uniquely closest to the shift, and a nonzero initial component in its eigendirection, as specified below.

Identifying the target eigenvalue λ_J

For a real symmetric positive definite matrix, all eigenvalues are positive. Consider the example

$$\lambda_1 = 9, \quad \lambda_2 = 6, \quad \lambda_3 = 3, \quad \lambda_4 = 1, \qquad \mu = 5.$$



$$\big(|\mu - \lambda_1|, |\mu - \lambda_2|, |\mu - \lambda_3|, |\mu - \lambda_4|\big) = (4, 1, 2, 4).$$

Thus $\operatorname{argmin}_i |\mu - \lambda_i| = \{2\}$: the minimizing index is $J = 2$, and the target eigenvalue is $\lambda_J = 6$.

For $(A - \mu I)^{-1}$, the eigenvalues are $\frac{1}{4}$, 1, $-\frac{1}{2}$, $-\frac{1}{4}$: the one corresponding to λ_J has the largest modulus.

Inverse power method with shift μ

Assume $\mu \in \mathbb{R}$ is not an eigenvalue of A. Keep this shift fixed and choose $v^{(0)} \in \mathbb{R}^n$ with $\|v^{(0)}\|_2 = 1$.

Inverse power method

Input: A, μ, $v^{(0)}$, $\mathrm{tol} > 0$, $\mathrm{MAXITER} \geq 1$.
Set $\mu_0 = (Av^{(0)}, v^{(0)})$.
If $\|Av^{(0)} - \mu_0 v^{(0)}\|_2 \leq \mathrm{tol}$, return $(\mu_0, v^{(0)})$.
for $k = 1, \ldots, \mathrm{MAXITER}$
  Solve $(A - \mu I)w = v^{(k-1)}$.
  $v^{(k)} = w/\|w\|_2$, $\mu_k = (Av^{(k)}, v^{(k)})$.
  If $\|Av^{(k)} - \mu_k v^{(k)}\|_2 \leq \mathrm{tol}$, return $(\mu_k, v^{(k)})$.
end
Return $(\mu_k, v^{(k)})$ and report that the tolerance was not met.

The shift μ stays fixed; μ_k estimates an eigenvalue of A.

Implementation note: reuse the factorization

At every iteration, the inverse power method solves

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

The right-hand side changes, but the matrix is the same because the shift μ is fixed.

Factor once, solve repeatedly

Before the iterations, compute an LU factorization with pivoting:

$$P(A - \mu I) = LU.$$

At iteration k, reuse these same factors and solve

$$Lz = Pv^{(k-1)} \quad \text{(forward substitution)}, \qquad Uw = z \quad \text{(back substitution)}.$$

Then normalize w to obtain $v^{(k)}$.

For a dense matrix, the factorization costs $O(n^3)$ once; each pair of triangular solves costs $O(n^2)$. There is no need to form $(A - \mu I)^{-1}$.

Convergence of the inverse power method

Theorem

Let $A \in \mathbb{R}^{n\times n}$ be diagonalizable over $\mathbb{R}$. Assume the real shift μ is not an eigenvalue of A, and

$$|\lambda_J - \mu| < |\lambda_i - \mu| \quad (i \neq J), \qquad v^{(0)} = \sum_{i=1}^{n} \alpha_i u_i, \quad \alpha_J \neq 0.$$

Then $\mu_k \to \lambda_J$, and there are real scalars c_k such that

$$c_k v^{(k)} \longrightarrow u_J, \qquad |c_k| \longrightarrow 1.$$

If $\lambda_J > \mu$, the vectors converge up to a fixed sign; if $\lambda_J < \mu$, their signs alternate. In either case they approach the eigendirection $\operatorname{span}\{u_J\}$.

Inverse power method: proof of convergence

Proof

The matrix $B = (A - \mu I)^{-1}$ is diagonalizable over $\mathbb{R}$, with

$$Bu_i = \frac{1}{\lambda_i - \mu} u_i.$$

Its eigenvalue $1/(\lambda_J - \mu)$ has uniquely largest absolute value. Apply the power-method eigenvector result to B: there are real scalars c_k such that $z^{(k)} := c_k v^{(k)} \to u_J$, with $|c_k| \to 1$. By invariance under scaling and continuity,

$$\mu_k = \frac{(Az^{(k)}, z^{(k)})}{(z^{(k)}, z^{(k)})} \longrightarrow \frac{(Au_J, u_J)}{(u_J, u_J)} = \lambda_J.$$

The sign behavior follows from the power-method result applied to the dominant eigenvalue $1/(\lambda_J - \mu)$ of B.

Inverse power method without shift ($\mu = 0$)

Assume A is nonsingular and diagonalizable over $\mathbb{R}$, with $0 < |\lambda_n| < |\lambda_{n-1}|$. Since $A^{-1}u_i = (1/\lambda_i)u_i$, the eigenvalue of A^{-1} largest in modulus is $1/\lambda_n$.

Inverse power method

Choose $v^{(0)} \in \mathbb{R}^n$ with $\|v^{(0)}\|_2 = 1$ and a nonzero component along u_n in the eigenvector basis.
Input: A, $v^{(0)}$, $\mathrm{tol} > 0$, $\mathrm{MAXITER} \geq 1$.
Set $\mu_0 = (Av^{(0)}, v^{(0)})$.
If $\|Av^{(0)} - \mu_0 v^{(0)}\|_2 \leq \mathrm{tol}$, return $(\mu_0, v^{(0)})$.
for $k = 1, \ldots, \mathrm{MAXITER}$
  Solve $Aw = v^{(k-1)}$.
  $v^{(k)} = w/\|w\|_2$, $\mu_k = (Av^{(k)}, v^{(k)})$.
  If $\|Av^{(k)} - \mu_k v^{(k)}\|_2 \leq \mathrm{tol}$, return $(\mu_k, v^{(k)})$.
end
Return $(\mu_k, v^{(k)})$ and report that the tolerance was not met.

Thus $\mu_k \to \lambda_n$, the eigenvalue of A smallest in modulus. For an SPD matrix, this is simply the smallest eigenvalue.

Factor $PA = LU$ once and reuse the factors in every solve; do not form A^{-1}.